

UNIVERSITÉ DE NANTES
FACULTÉ DE MÉDECINE

Étude d'association cas-témoins
pour l'identification de gènes de prédisposition
au cancer colorectal sporadique

THÈSE DE DOCTORAT

Ecole Doctorale : BIOLOGIE-SANTÉ
Discipline : Sciences de la vie et de la santé
Spécialité : Génétique Moléculaire

*Présentée
et soutenue publiquement par*

Sébastien ROBIOU du PONT

Le 07 juillet 2009, devant le jury ci-dessous

Président	M. S. BRULEY des VARANNES, PU-PH, NANTES
Rapporteurs	Mme. S. OLSCHWANG, CR1-PH, MARSEILLE M. J. MOSSER, PU-PH, RENNES
Examineurs	Mme. M. BLAYAU, PH, RENNES Mme. T. MATYSIAK-BUDNIK, MCU-PH, NANTES M. J-F. MOSNIER, PU-PH, NANTES
Directeur de thèse :	M. S. BÉZIEAU, MCU-PH, NANTES
Co-encadrant :	M. S. KÜRY, IH, NANTES

LISTE DES FIGURES, TABLEAUX, ENCADRE ET ANNEXES.....	VI
LISTE DES ABREVIATIONS.....	X
INTRODUCTION.....	1
CHAPITRE I : ÉTUDE BIBLIOGRAPHIQUE.....	3
I. LE CANCER	4
I.1. La tumeur	4
I.2. La croissance cellulaire.....	5
I.2.1. La croissance des cellules « normales »	5
I.2.2. Le cycle cellulaire	5
I.2.2.1. Les différentes phases	5
I.2.2.2. La régulation du cycle cellulaire	6
I.2.3. La croissance des cellules cancéreuses	7
I.2.3.1. Les gènes de dépendance aux facteurs de croissance.....	7
I.2.3.2. Les gènes de contrôle du cycle cellulaire	7
I.2.3.3. Les gènes de l'immortalisation cellulaire	7
I.2.3.4. Les gènes de l'apoptose.....	8
I.2.3.5. Les gènes de néo-vascularisation	8
I.2.3.6. Les gènes de la mobilité cellulaire	8
I.3. Les étapes de la cancérisation	8
I.3.1. Les trois étapes de la cancérisation	8
I.3.1.1. L'initiation	8
I.3.1.2. La promotion.....	9
I.3.1.3. L'invasion tumorale	9
I.4. Illustration par le cancer colorectal.....	11
I.4.1. Caractéristiques anatomiques et fonctionnelles du gros intestin.....	11
I.4.1.1. Anatomie.....	11
I.4.1.2. Fonctions.....	12
I.4.2. Épidémiologie des cancers du côlon	12
I.4.3. Modèle de la carcinogenèse colorectale	12
II. GENETIQUE DES CANCERS	14
II.1. Les différentes formes de cancers	14
II.1.1. Les cancers héréditaires.....	14
II.1.2. Les cancers non mendéliens.....	15
II.1.2.1. Les formes familiales non mendéliennes	15
II.1.2.2. Les formes sporadiques	15
II.2. Les gènes impliqués dans les cancers.....	15
II.2.1. Les oncogènes	15
II.2.1.1. Définition	15
II.2.1.2. Fonctions et mécanismes d'action des proto-oncogènes.....	16
II.2.1.2.1. Mécanismes d'action d'un proto-oncogène	16
II.2.1.2.2. Les familles d'oncoprotéines impliquées dans la transduction des signaux de prolifération.....	17
II.2.1.3. Les oncogènes dans les cancers héréditaires	18
II.2.2. Les gènes suppresseurs de tumeur.....	19
II.2.2.1. Définition	19
II.2.2.2. Le modèle de Knudson et Comings.....	19
II.2.2.3. Exemples de gènes suppresseurs de tumeur impliqués dans des formes familiales de cancer .	20
II.2.2.3.1. Le gène <i>Rb</i>	21
II.2.2.3.2. Les gènes <i>BRCA1</i> et <i>BRCA2</i>	21
II.2.2.3.3. Autres gènes suppresseurs	21
II.2.3. Les gènes impliqués dans la réparation de l'ADN	23

II.2.4.	Les télomérases.....	23
II.2.5.	Les petits fragments d'ARN.....	24
II.3.	Génétique des cancers colorectaux.....	25
II.3.1.	Les différentes voies altérées dans les cancers colorectaux.....	25
II.3.1.1.	La voie WNT/ β -caténine.....	25
II.3.1.2.	La voie de RAS/RAF/MAPK.....	26
II.3.1.3.	La voie TGF β	27
II.3.1.4.	La voie p53.....	28
II.3.2.	Les différentes formes d'instabilités rencontrées.....	29
II.3.2.1.	Les instabilités génomiques.....	29
II.3.2.1.1.	Les gènes cibles des instabilités microsatellitaires.....	31
II.3.2.2.	Les instabilités chromosomiques.....	32
II.3.2.2.1.	Les gènes atteints dans les instabilités chromosomiques.....	33
II.3.2.2.2.	Les mécanismes impliqués dans les instabilités chromosomiques.....	35
II.3.2.3.	Les altérations épigénétiques et les instabilités épigénétiques.....	38
II.3.3.	Prédisposition génétique aux cancers colorectaux.....	39
II.4.	Les formes familiales.....	42
II.4.1.	La polypose adénomateuse familiale (PAF).....	42
II.4.1.1.	Contexte historique.....	42
II.4.1.2.	Polypose adénomateuse familiale atténuée.....	43
II.4.1.3.	Les variants génétiques d'APC associés à la PAF.....	44
II.4.1.4.	Corrélation génotype-phénotype dans les PAF.....	46
II.4.2.	Polypose associée à <i>MUTYH</i> (MAP).....	46
II.4.2.1.	Histoire et caractérisation.....	46
II.4.2.2.	Le gène <i>MUTYH</i> et les mutations associées à la MAP.....	47
II.4.2.2.1.	Mécanisme d'action du système BER.....	47
II.4.2.2.2.	Mutations du gène <i>MUTYH</i>	49
II.4.3.	Le syndrome de Lynch.....	51
II.4.3.1.	Historique et caractérisation.....	51
II.4.3.2.	Mutations et polymorphismes mis en cause dans le syndrome de Lynch.....	54
II.4.3.3.	Corrélation phénotype/génotype.....	56
II.4.3.4.	Instabilité microsatellitaire (MSI).....	56
II.5.	Les formes sporadiques de cancers colorectaux : des maladies complexes.....	58
II.5.1.	Définition d'une maladie complexe ou multifactorielle.....	58
II.5.2.	Les facteurs environnementaux dans les cancers colorectaux.....	58
II.5.2.1.	Les macro-constituants alimentaires qui protègent du CCR.....	58
II.5.2.1.1.	Les fruits, les légumes et les autres micronutriments.....	58
II.5.2.1.2.	Les fibres alimentaires.....	59
II.5.2.1.3.	Les acides gras polyinsaturés n-3.....	60
II.5.2.2.	Les macro-constituants alimentaires qui prédisposent au CCR.....	61
II.5.2.2.1.	Les graisses animales et la viande rouge.....	61
II.5.2.2.2.	Les glucides hautement assimilables et la charge glycémique importante.....	63
II.5.2.3.	Les micronutriments.....	63
II.5.2.3.1.	Le calcium.....	63
II.5.2.3.2.	La vitamine D.....	63
II.5.2.3.3.	L'acide folique.....	64
II.5.2.3.4.	Le sélénium.....	64
II.5.2.3.5.	Les antioxydants.....	65
II.5.2.3.6.	Les métabolites secondaires des plantes.....	65
II.5.2.4.	Les facteurs non-alimentaires.....	66
II.5.2.4.1.	Les anti-inflammatoires non stéroïdiens.....	66
II.5.2.4.2.	L'activité physique et l'indice de masse corporelle.....	67
II.5.2.4.3.	Le tabac.....	69
II.5.2.4.4.	L'alcool.....	70
II.5.2.4.5.	L'hormonothérapie substitutive.....	70
II.5.3.	Les facteurs génétiques de prédisposition.....	71
II.5.3.1.	Définition d'un variant ou polymorphisme de prédisposition.....	71

II.5.3.2.	Les polymorphismes de prédisposition impliqués dans les cancers colorectaux	71
II.5.3.2.1.	La voie de l'inflammation.....	72
II.5.3.2.2.	La voie de détoxification	77
II.5.3.2.3.	La voie des radicaux mono-carbonés.....	80
II.5.3.2.4.	Voie de l'insuline	83
II.5.3.2.5.	Voie de réparation de l'ADN	86
III.	METHODES D'IDENTIFICATION DES GENES ET DES POLYMORPHISMES DE PREDISPOSITION DANS LES MALADIES COMPLEXES ..	96
III.1.	Une méthode de choix : L'étude d'association cas-témoins.....	96
III.1.1.	Une étape clé : la constitution des cohortes	96
III.1.2.	Précautions d'analyse	96
III.1.2.1.	Défaillance du recrutement	97
III.1.2.1.1.	L'homogénéité des cohortes.....	97
III.1.2.1.2.	La taille des cohortes	97
III.1.2.2.	Notion de SNP causal ou de marqueur de la pathologie	97
III.1.3.	Les paramètres à analyser.....	98
III.1.3.1.	Le seuil critique de significativité.....	98
III.1.3.2.	L'estimation du risque relatif	98
III.2.	Alternative d'étude d'association	99
III.3.	Les différentes approches permettant l'identification des gènes de prédisposition	99
III.3.1.	L'approche « gènes candidats »	99
III.3.2.	L'approche « génome complet »	100
III.3.2.1.	Méthodes pour augmenter l'efficacité des études d'association pangénomiques	100
III.3.2.1.1.	L'approche multi-étapes	100
III.3.2.1.2.	L'approche requérant des populations avec un effet fondateur.....	101
III.3.2.1.3.	L'approche par mélange d'ADN.....	102
III.4.	Impact de la fonction d'un SNP lié à une prédisposition.....	102
III.4.1.	Corrélation entre la localisation d'un SNP et son effet prédisposant	102
III.4.2.	Analyse du caractère fonctionnel d'un SNP	103
III.4.2.1.	Analyse bioinformatique.....	103
III.4.2.2.	Analyse de l'expression de l'ARN messenger	103
III.4.2.3.	Analyse par minigène	104
III.4.2.4.	Analyses par cassette d'expression	104
IV.	CONCLUSION DE LA PARTIE BIBLIOGRAPHIQUE	105
	CHAPITRE II : ÉTUDE EXPERIMENTALE	106
I.	INTRODUCTION	107
II.	APPROCHE « GENES CANDIDATS »	109
II.1.	Données de départ.....	109
II.1.1.	Les polymorphismes choisis	109
II.1.1.1.	Les polymorphismes des gènes de la voie de l'inflammation	109
II.1.1.2.	Les polymorphismes des gènes de la voie de détoxification des xénobiotiques.....	110
II.1.1.3.	Les polymorphismes des gènes de la voie des radicaux monocarbonés	111
II.1.1.4.	Les polymorphismes des gènes de la voie de l'insuline	112
II.1.1.5.	Les polymorphismes des gènes de la voie de réparation de l'ADN	112
II.1.2.	Méthode d'analyse	113
II.1.2.1.	Analyse des SNP	113
II.1.2.2.	Analyse des polymorphismes complexes.....	114
II.2.	Etude de 6 polymorphismes des cytochromes P450 (CYP) (Article 1).....	115
II.2.1.	Supplementary Data	127
II.2.1.1.	Supplementary Table S1 : TaqMan-based genotyping analysis of the selected SNPs	127
II.2.1.2.	Supplementary data S2 : Analysis of interaction between allelic variants of CYP genes and environmental covariates.....	128

II.2.1.3.	Supplementary Table S3 : Analysis of association to CRC risk of multiple CYP SNPs combinations interacting with the body mass.....	130
II.3.	Etude de 52 SNP de faible prédisposition au cancer colorectal sporadique (Article 2).....	132
II.3.1.	Supplementary data.....	149
II.3.1.1.	Additional file 1: Supplemental materials and methods.....	149
II.3.1.2.	Additional file 2: List of primers and probes used for the study.....	153
II.3.1.3.	Additional file 3: Statistical analysis of association between CRC risk and polymorphisms selected for the study using conditional logistic regression.	156
II.3.1.4.	Additional file 4: Test of the different possible models of genotypic combinations related to a given set of five polymorphisms picked among the 52 polymorphisms analyzed in the study.....	159
II.3.1.5.	Additional file 5: Estimation by use of bootstrapping of the external validity of the model showing association between CRC risk and genotypes combinations	159
II.4.	Etude du gène <i>MUTYH</i> (Article 3)	161
III.	APPROCHE GENOME-WIDE ASSOCIATION	172
III.1.	Introduction.....	172
III.2.	Matériels et Méthodes.....	174
III.2.1.	Constitution des cohortes et préparation des ADN.....	174
III.2.2.	La méthode SNP-MaP	174
III.2.2.1.	Sélection des individus intégrés dans les pools d'ADN	175
III.2.2.1.1.	Groupe d'individus classés par catégories d'âge (projet 1) : Groupe « catégorie d'âge » ..	175
III.2.2.1.2.	Groupe d'individus classés par localisation géographique (projet 2) : Groupe « Vendéens »	176
III.2.2.2.	Construction des pools d'ADN pour l'analyse SNP-MaP.....	178
III.2.2.3.	Hybridation sur puce	178
III.2.3.	Analyses statistiques.....	178
III.2.3.1.	Calculs des fréquences alléliques	178
III.2.3.2.	Calculs des OR et p-value	182
III.2.4.	Choix des SNP d'intérêt.....	183
III.2.4.1.	Graphique : $-\log_{10}(p\text{-value}) = f(\text{position chromosomique})$	183
III.2.4.2.	Analyses bioinformatiques.....	183
III.2.4.3.	Critères de sélection d'un SNP à analyser en détail.....	185
III.2.4.3.1.	Nombre de fois où le SNP apparait dans les différentes analyses des projets 1 et 2 ...	186
III.2.4.3.2.	Concordance des OR calculés pour le SNP entre les projets 1 et 2	186
III.2.4.3.3.	Pertinence du gène candidat proche du SNP avec le cancer	186
III.2.4.3.4.	Pertinence bibliographique avec le cancer colorectal du gène proche du SNP	186
III.2.4.3.5.	Localisation précise du SNP par rapport au gène candidat	187
III.2.5.	<i>5'-nuclease assay</i>	187
III.2.6.	Etude de réplication.....	188
III.2.6.1.	Population de réplication : Zanke <i>et al.</i> (population Arctic).....	188
III.2.6.1.1.	Recrutement	188
III.2.6.1.2.	Génotypage.....	188
III.2.6.1.3.	Analyse statistique.....	189
III.2.7.	Etude comparative de l'expression des gènes d'intérêt entre les tissus sains et tumoraux	189
III.2.7.1.	Principe.....	189
III.2.7.1.1.	Recrutement des échantillons et extraction de l'ARN et de l'ADN.....	189
III.2.7.1.2.	<i>5'-nuclease Assay</i>	189
III.2.7.1.3.	PCR quantitative en temps réel	189
III.2.7.1.4.	Analyses des données d'expression.....	190
III.2.7.2.	Etude des transcrits du gène d'intérêt.....	191
III.2.8.	Analyse par le logiciel GenePool.....	192
III.3.	Résultats.....	194
III.3.1.	Méthode de validation des résultats	194
III.3.1.1.	Validation de l'approche SNP-MaP par comparaison avec les données de la littérature	194
III.3.1.2.	Validation par génotypage individuel des associations mises en évidence par la méthode SNP-MaP	195

III.3.2.	Résultats de génotypage par SNP-MaP.....	197
III.3.3.	Résultats de génotypage individuel.....	205
III.3.4.	Réplication sur d'autres populations du test d'association des SNP rs2306907 et rs3760454 associés dans notre population.....	206
III.3.5.	Expression dans les tissus sain et tumoral des gènes candidats <i>CDH13</i> et <i>UTP6 (HCA66)</i>	207
III.3.6.	Recherche par RT-PCR d'épissage alternatif de <i>CDH13</i> induit par le SNP rs2306907.....	208
III.3.7.	Résultat des analyses de GenePool.....	210
IV.	DISCUSSION.....	217
IV.1.	Les différentes stratégies employées.....	217
IV.1.1.	Choix de l'étude d'association cas-témoins.....	217
IV.1.2.	L'approche « gènes candidats ».....	217
IV.1.3.	L'approche pangénomique.....	219
IV.2.	Approche « gènes candidats ».....	222
IV.2.1.	Etude de six polymorphismes des cytochromes P450 (Article 1).....	222
IV.2.1.1.	Association entre les facteurs environnementaux et le risque de CCR sporadique.....	222
IV.2.1.2.	Association entre les facteurs environnementaux et les combinaisons génétiques des cytochromes P450.....	223
IV.2.2.	Etude de 52 SNP de faible prédisposition au cancer colorectal sporadique (Article 2).....	225
IV.2.3.	Etude du gène <i>MUTYH</i> (Article 3).....	229
IV.3.	L'approche <i>Genome-Wide Association</i>	232
IV.3.1.	Approche par pool d'ADN.....	232
IV.3.1.1.	Notion d'allélotype.....	232
IV.3.1.2.	Constitution des pools.....	233
IV.3.2.	Analyse informatique.....	234
IV.3.3.	Validation de la stratégie.....	235
IV.3.4.	Résultats.....	237
IV.3.5.	Réplication des associations au CCR pour le SNP rs2306907 du gène <i>CDH13</i> et rs3760454 du gène <i>UTP6</i>	238
IV.3.6.	Corrélation entre le génotype et l'expression de l'ARNm des gènes d'intérêt, <i>CDH13</i> et <i>UTP6</i>	239
IV.3.7.	Analyse par le logiciel GenePool.....	240
IV.4.	Les cadhérines et le rôle de la cadhérine-13 dans la prédisposition au cancer colorectal.....	241
V.	CONCLUSION ET PERSPECTIVES.....	244
	ANNEXES.....	246
	RÉFÉRENCES.....	250

LISTE DES FIGURES, TABLEAUX, ENCADRE ET ANNEXES

Figures

Figure 1 : Les différentes phases du cycle cellulaire.....	5
Figure 2 : Régulation du cycle cellulaire par les complexes kinases cyclines dépendantes (cdk)/cycline.....	6
Figure 3 : Processus multi-étapes de la carcinogenèse colorectale en relation avec les changements morphologiques rencontrés.....	10
Figure 4 : Anatomie macroscopique du gros intestin.....	11
Figure 5 : Modèle de carcinogenèse colorectale.....	13
Figure 6 : Les différentes voies de transduction.....	17
Figure 7 : La théorie de Knudson.....	20
Figure 8 : Mécanisme d'action antagoniste d'un gène suppresseur de tumeur.....	20
Figure 9 : Régulation de la β -caténine dans la voie Wnt.....	25
Figure 10 : Voie de signalisation des MAP-kinases.....	27
Figure 11 : Représentation simplifiée de la voie TGF- β	28
Figure 12 : Voie de p53.....	29
Figure 13 : Progression des tumeurs colorectales avec des instabilités microsatellites (MSI).....	30
Figure 14 : Progression des tumeurs colorectales avec des instabilités chromosomiques (CIN).....	32
Figure 15 : p53, une molécule carrefour dans la réponse aux dommages à l'ADN.....	34
Figure 16 : Contrôle du cycle cellulaire.....	36
Figure 17 : Progression des tumeurs colorectales possédant un phénotype méthylé des îlots CpG (CIMP).....	38
Figure 18 : Vue d'ensemble de la contribution génétique et environnementale dans les cancers colorectaux.....	40
Figure 19 : Répartition des différentes formes de cancers colorectaux.....	41
Figure 20 : Polypose adénomateuse classique d'un patient de 20 ans.....	43
Figure 21 : Représentation schématique de la protéine APC.....	45
Figure 22 : Voie de réparation de l'ADN par le système BER (<i>Base Excision Repair</i>).....	48
Figure 23 : Voie « <i>short-patch</i> » délimination des 8-oxoG par le système BER.....	49
Figure 24 : Distribution et localisation des mutations identifiées à ce jour dans le gène <i>MUTYH</i>	50
Figure 25 : Comparaison entre le développement de la polypose adénomateuse familiale (PAF) et celui du syndrome de Lynch.....	53
Figure 26 : Distribution des différentes mutations de <i>MSH2</i> , <i>MLH1</i> et <i>MSH6</i>	55
Figure 27 : Mécanismes participant à l'inflammation dans le développement du cancer.....	76
Figure 28 : Illustration des différentes phases du métabolisme des xénobiotiques.....	78
Figure 29 : Équilibre entre l'activité toxique des métabolites et la détoxification par les enzymes.....	79
Figure 30 : Les différentes voies de détoxification qu'utilise l'organisme pour éliminer les métabolites carcinogènes.....	79
Figure 31 : Représentation schématique du métabolisme des folates.....	82
Figure 32 : Principales voies de signalisation de l'insuline.....	85
Figure 33 : Réparation d'une lésion de l'ADN par le système NER.....	89
Figure 34 : Réparation des cassures doubles-brins d'ADN.....	90
Figure 35 : Réparation des mésappariements de l'ADN.....	92
Figure 36 : Mécanisme d'action de l'alkyltransférase O ⁶ -alkylguanine (AGT).....	94
Figure 37 : Association directe ou indirecte d'un polymorphisme à un gène.....	98
Figure 38 : Illustration de l'approche multi-étapes permettant de génotyper par GWA un nombre moins important de SNP.....	101
Figure 39 : Principe du système minigène.....	104
Figure 40 : Principe et résultats de la technique de génotypage par <i>5'-nuclease assay</i>	113

Figure 41 : Résultat de QMPSF pour deux individus après ajustement des pics <i>MLH1</i>	114
Figure 42 : Les différentes sondes de la puce permettant de calculer la fréquence allélique.....	179
Figure 43 : Exemple de calcul de l'OR d'un SNP à partir des fréquences alléliques obtenues par l'analyse des puces	182
Figure 44 : (a) Mode de sélection des gènes dans le cas d'un SNP intergénique. (b) Mode de sélection des gènes dans le cas de deux SNP distants de moins de 1Mb.....	184
Figure 45 : Représentation du graphique $-\log_{10}(\text{p-value})=f(\text{position chromosomique})$	185
Figure 46 : Exemple de calcul du $2^{-\Delta\Delta\text{CT}}$ pour un individu	190
Figure 47 : Représentation simplifiée des résultats de l'analyse GenePool	192
Figure 48 : Représentation des clusters de SNP pour trois réplicats.....	193
Figure 49 : OR associés aux allèles minoritaires des SNP (rs2306907, rs3760454) pour notre population (NANTES) et pour la population caucasienne canadienne ARTIC de l'équipe de B. ZANKE.	206
Figure 50 : Résultat de la PCR exons sur gel d'agarose à 1,2%.....	209
Figure 51 : Résultats de la PCR amplifiant la totalité de la séquence codante sur gel d'agarose à 1,2%	209
Figure 52 : Signalisation et régulation de la β -caténine impliquant l'E-cadhérine.....	242

Tableaux

Tableau 1 : Exemples de gènes suppresseurs de tumeur	22
Tableau 2 : Principaux gènes intervenant dans les différents systèmes de réparation de l'ADN	23
Tableau 3 : Causes génétiques des polyposes familiales	42
Tableau 4 : Critères d'Amsterdam I	51
Tableau 5 : Critères de Bethesda	57
Tableau 6 : Polymorphismes des gènes de la voie l'inflammation	77
Tableau 7 : Polymorphismes des gènes de la voie de détoxication des xénobiotiques.	80
Tableau 8 : Polymorphismes des gènes de la voie des radicaux monocarbonés	83
Tableau 9 : Polymorphismes des gènes de la voie de l'insuline	86
Tableau 10 : Polymorphismes du système de réparation BER	88
Tableau 11 : Polymorphismes du système de réparation NER.	90
Tableau 12 : Polymorphismes du système de réparation DSBR.....	91
Tableau 13 : Polymorphismes du système de réparation MMR	93
Tableau 14 : Polymorphismes du système de réparation directe	94
Tableau 15 : Taille approximative de l'effectif des patients dans une étude d'association cas-témoins pour différentes fréquences alléliques et en fonction des OR attendus	97
Tableau 16 : Distribution des effectifs des génotypes AA, AB et BB pour les patients et les témoins ..	98
Tableau 17 : Les différents polymorphismes de la voie de l'inflammation retenus dans notre approche « gènes candidats »	110
Tableau 18 : Les différents polymorphismes de la voie de détoxication des xénobiotiques retenus dans notre approche « gènes candidats ».	111
Tableau 19 : Les différents polymorphismes de la voie des radicaux monocarbonés retenus dans notre approche « gènes candidats ».	112
Tableau 20 : Les différents polymorphismes de la voie de l'insuline retenus dans notre approche « gènes candidats ».	112
Tableau 21 : Les différents polymorphismes de la voie de réparation de l'ADN choisis pour notre approche « gènes candidats »	112
Tableau 22 : Isoformes des cytochromes P450	116
Tableau 23 : Composition des sous-groupes « catégorie d'âge »	176
Tableau 24 : Composition des sous-groupes « Vendéens »	177
Tableau 25 : Composition des pools vendéens avec les effectifs de provenance des ADN.....	177
Tableau 26 : Fréquence moyenne élaguée pour l'allèle minoritaire est calculée à partir de celle obtenue pour la catégorie « patients » et comparée à la fréquence moyenne calculée à partir des puces témoins	180
Tableau 27 : Fréquence moyenne élaguée pour l'allèle minoritaire est calculée à partir de celles obtenues pour les catégories « patients » de moins et de plus de 60 ans et comparée à la fréquence moyenne calculée à partir des puces témoins	181
Tableau 28 : Fréquence moyenne élaguée pour l'allèle minoritaire est calculée à partir de celle obtenue pour le pool « patient » correspondant et comparée à la fréquence moyenne calculée à partir des puces témoins.....	181
Tableau 29 : Fréquence moyenne élaguée pour l'allèle minoritaire est calculée à partir de celles obtenues pour les deux pools vendéens et comparée à la fréquence moyenne calculée à partir des puces témoins	181
Tableau 30 : Séquences de référence pour l'analyse des polymorphismes d'intérêt.	187
Tableau 31 : Séquence des couples d'amorces permettant de vérifier la présence de transcrit alternatif.....	191
Tableau 32 : Comparaison entre les OR publiés et les OR calculés dans notre étude par la méthode SNP-MaP à partir des puces pangénomiques	194
Tableau 33 : Comparaison des résultats obtenus par le génotypage individuel et par la méthode SNP-MaP pour les polymorphismes rs1801282, rs1800566 et rs10505477	196

Tableau 34 : Tableau simplifié représentant tous les polymorphismes ayant permis de faire ressortir les gènes ayant un lien avec le cancer	201
Tableau 35 : Représentation des sept gènes mis en évidence à la fois par l'analyse portant sur le groupe d'individus classés par âge et par celle portant sur le groupe d'individus vendéens.....	204
Tableau 36 : SNP trouvés associés au cancer colorectal dans notre étude par la méthode SNP-MaP.	205
Tableau 37 : OR et p-value obtenus par génotypage individuel par sonde Taqman et par génotypage selon la méthode SNP-MaP pour les SNP d'intérêt.	205
Tableau 38 : OR calculés à partir du génotypage des rs2306907 et rs3760454 pour notre population d'étude (NANTES) et pour la population caucasienne canadienne (ARCTIC) de B. ZANKE.....	206
Tableau 39 : Rapports $2^{-\Delta\Delta CT}$ Cancer/ $2^{-\Delta\Delta CT}$ Normal moyens pour chaque catégorie.	207
Tableau 40 : Résultats des 200 meilleurs SNP obtenus pour cinq analyses différentes réalisées par le logiciel GenePool	216
Tableau 41 : Avantages et inconvénients des approches « gènes candidats » et GWA.....	221

Encadré

Encadré 1 : Méthode de Liu <i>et al.</i> (2003) permettant le calcul de la fréquence allélique à partir de la fluorescence des puces à ADN.	180
--	-----

Annexes

Annexe 1 : Questionnaire d'habitude de vie.....	247
Annexe 2 : Feuille de recueil clinique.	248
Annexe 3 : Organigramme représentant les différentes opérations réalisées dans l'approche SNP-MaP.....	249

LISTE DES ABREVIATIONS

8-oxoG	7,8-dihydro-2'-deoxyguanosine
A	Adénosine
ADN	Acide désoxyribonucléique
AINS	Anti-Inflammatoire Non Stéroïdien
Ala	Alanine
AP	Site apurinique/apyrimidinique
Arg	Arginine
ARN	Acide ribonucléique
ARNm	ARN messenger
Asn	Asparagine
Asp	Acide aspartique
ATP	Adénosine triphosphate
BER	<i>Base Excision Repair</i>
BRRS	<i>Bannayan-Ruvalcaba-Riley Syndrome</i>
C	Cytosine
CCR	Cancer Colorectal
CD	<i>Cowden Disease</i>
cdk	Kinase dépendante des cyclines
CHRPE	<i>Congenital Hypertrophy of Retinal Pigment Epithelium</i>
CIMP	<i>CpG Island Methylator Phenotype</i>
CIN	<i>Chromosomal Instability</i>
COX	Cyclooxygénase
CRP	Protéine C-réactive
C _T	<i>Cycle Threshold</i>
CYP	Cytochrome P450
Cys	Cystéine
DDCP	<i>DNA Damage Checkpoint</i>
DHA	Acide docosahexaénoïque
dHPLC	<i>Denaturing High Performance Liquid Chromatography</i>
DSBR	<i>Double Strand Break Repair</i>
dTMP	Désoxythymidine monophosphate
dUMP	Désoxyuridine monophosphate
EMX	Enzyme du Métabolisme des Xénobiotiques
EPA	Acide eicosapentaénoïque
FAM	5-Carboxyfluorescéine
FDR	<i>False Discovery Rate</i>
FJP	<i>Familial Juvenile Polyposis</i>
G	Guanosine
GDP/GTP	Guanosine diphosphate/Guanosine triphosphate
Gln	Glutamine
Glu	Acide glutamique
Gly	Glycine

GPI	Glycosyl-Phosphatidylinositol
GWA	<i>Genome-Wide Association</i>
GWAS	<i>Genome-Wide Association Study</i>
HEX	Hexachloro carboxyfluorescéine
His	Histidine
HNPCC	<i>Hereditary Non Polyposis Colorectal Cancer</i>
HR	Recombinaison Homologue
IC	Intervalle de Confiance
IHC	Immunohistochimie
IL	Interleukine
Ile	Isoleucine
IMC (<i>BMI</i>)	Indice de Masse Corporelle (<i>Body Mass Index</i>)
iNOS	Oxyde nitrique synthétase inductible
Leu	Leucine
LOH	<i>Loss of Heterozygoty</i> (perte d'hétérozygotie)
LOX	Lipoxygénase
Lys	Lysine
MAF	<i>Minor Allele Frequency</i>
MAP	<i>MUTYH Associated Polyposis</i>
Mb	Méga base
Met	Méthionine
méthylTHF	Méthyltétrahydrofolate
miARN	micro ARN
ml, µl	millilitre, microlitre
Mm	<i>Mismatch</i>
mM, µM	millimolaire, micromolaire
MMR	<i>MisMatch Repair</i>
MNU	N-Nitroso-N-Méthylurée
MPC	<i>Mitotic Checkpoint</i>
MRP	<i>Multidrug Related Protein</i>
MSI / MSI-H /	Microsatellite instable / Microsatellite instable High /
MSI-L	Microsatellite instable Low
NER	<i>Nucleotide Excision Repair</i>
NHEJ	Réparation par jointure non homologue
OR	Odds Ratio
PAF (<i>FAP</i>)	Polypose Adénomateuse Familiale (<i>Familial Adenomatous Polyposis</i>)
PAFA (<i>AFAP</i>)	Polypose Adénomateuse Familiale Atténuée (<i>Attenuated Familial Adenomatous Polyposis</i>)
pb	Paire de base
PCR	<i>Polymerase Chain Reaction</i>
PgP	P-glycoprotéine
Phe	Phénylalanine
PHRC	Programme Hospitalier de Recherche Clinique
PI3K	Phosphatidylinositol 3 kinase
PJS	<i>Peutz-Jeghers Syndrome</i>

Pm	<i>Perfect match</i>
Pro	Proline
RAS	<i>Relative Allele Signal</i> ou <i>Relative Allele Strength</i>
RB	Protéine du rétinoblastome
RCP	<i>Replication Checkpoint</i>
RER	<i>Replication error</i> (erreur de réplication)
ROS	Espèce réactive à l'oxygène
RR	Risque Relatif
RTK	Récepteurs des tyrosines kinases
RT-PCR	<i>Reverse transcription Polymerase Chain Reaction</i>
SAM	S-adénosylméthionine
Ser	Sérine
siARN	<i>small interfering</i> ARN
SNP	<i>Single Nucleotide Polymorphism</i>
SNP-MaP	<i>SNP Microarray and Pooling</i>
T	Thymidine
TDT	<i>Transmission Disequilibrium Test</i>
Thr	Thréonine
Trp	Tryptophane
Tyr	Tyrosine
Val	Valine
WAGR	<i>Wilms tumor-aniridia-genitourinary anomalies-mental retardation</i>

INTRODUCTION

En France, le cancer colorectal présente une incidence élevée et se place au troisième rang de la pathologie cancéreuse pour les deux sexes confondus, après le cancer du sein et de la prostate et avant le cancer du poumon. Du fait de ses caractéristiques épidémiologiques et de son pronostic sévère, le cancer colorectal constitue un réel enjeu de santé publique dans les pays industrialisés et notamment en France, où il représente un poids économique substantiel pour le système de protection sociale et pour la société dans son ensemble (SELKE *et al.* 2003).

Environ trois-quarts des cancers colorectaux (CCR) sont sporadiques et surviennent dans la population dite « à risque moyen », c'est-à-dire chez les individus sans antécédent personnel ou familial au premier degré de cancer ou de polype(s) adénomateux colorectal(aux). Les CCR sporadiques s'opposent aux cancers colorectaux « héréditaires » ou « familiaux » qui surviennent principalement dans le cadre de deux affections rares, à transmission autosomique dominante : la polypose adénomateuse familiale et le syndrome de Lynch ou HNPCC (acronyme pour *Hereditary Non Polyposis Colorectal Cancer*), pourvoyeuses respectivement de 1% et de 3 à 5% de l'ensemble des cancers colorectaux et associées à un niveau de risque de cancer colorectal considérable.

Contrairement aux CCR familiaux, affections mendéliennes monogéniques classiques transmises sur le mode autosomique dominant, la genèse des CCR sporadiques est méconnue et implique à la fois des facteurs génétiques et des facteurs environnementaux et nutritionnels. Ces derniers interviendraient avec le fond génétique des individus, notamment avec des gènes de susceptibilité mineure, dont les allèles variants de faible pénétrance agissent de manière cumulative et constituent la base de la prédisposition génétique.

Cette absence de données claires sur l'épidémiologie des cancers colorectaux sporadiques nous a motivés à mener une étude d'association cas-témoins de plus de 1000 patients caucasiens atteints de CCR sporadique et 1000 témoins caucasiens indemnes (les témoins étant appariés par âge, sexe et région géographique d'origine aux patients). Cette étude a pour but de rechercher et d'identifier des polymorphismes qui prédisposeraient aux cancers colorectaux sporadiques, pour en comprendre les facteurs de risque aussi bien génétiques qu'environnementaux, ainsi que le ou les mécanisme(s) de la carcinogenèse colorectale. Le but ultime serait, bien évidemment, la mise en évidence de marqueur(s) fiable(s) de prédisposition permettant le dépistage et la surveillance préventive de patients à risque.

Pour cela, deux approches différentes ont été réalisées : une approche « gènes candidats » et une approche pangénomique par la méthode SNP-MaP (*SNP Microarray and Pooling*). La démarche adoptée pour ce travail, ainsi que les résultats obtenus, seront, pour l'approche « gènes candidats »,

retracés au travers des publications. La première publication rapporte l'identification d'une combinaison d'allèles de six SNP des cytochromes P450 qui, en relation avec la consommation de viande rouge, augmente le risque de CCR (Article 1). La seconde publication présente l'analyse de 52 SNP de prédisposition au CCR (Article 2) et la troisième correspond à une étude du gène *MUTYH* (Article 3). Les résultats de l'approche pangénomique n'ayant, quant à eux, pas encore été valorisés par la publication d'article, les descriptions des méthodes et des résultats seront détaillées de manière exhaustive dans le présent document. Une discussion sera ensuite proposée sur les différents résultats obtenus par les deux approches réalisées, mais également, d'un point de vue plus général, sur les différentes approches utilisées.

Avant d'aborder la partie expérimentale proprement dite, il convient, dans un premier temps, d'établir un état des lieux des connaissances scientifiques sur le cancer et plus particulièrement sur la génétique des cancers colorectaux.

CHAPITRE I :

ÉTUDE

BIBLIOGRAPHIQUE

I. Le cancer

I.1. La tumeur

La tumeur est définie comme étant « un processus pathologique où la prolifération exagérée des cellules aboutit à une surproduction tissulaire qui persiste et a tendance à s'accroître (Synonymes : néoplasie, néoplasme) » (www.Larousse.fr).

Une tumeur est avant tout un amas de cellules. Cet amas a pour origine une cellule unique qui a acquis une capacité de prolifération exagérée par perte des phénomènes qui régulent sa croissance.

Une tumeur possède quatre grandes caractéristiques :

- elle se développe de manière anormale, tant quantitativement que qualitativement à partir de cellules d'un tissu organisé ;
- elle croît différemment des tissus voisins, c'est-à-dire que sa croissance est excessive et non coordonnée ;
- elle continue de proliférer, même après l'arrêt du stimulus qui lui a donné naissance ;
- elle a une prolifération biologiquement autonome, car elle est insensible aux signaux de régulation cellulaire.

Dans l'inconscient collectif, le terme « tumeur » est systématiquement associé au cancer, alors qu'il existe différents types de tumeurs : les tumeurs malignes, qui aboutiront aux cancers et les tumeurs bénignes, dont l'évolution, bien différente, est très variable.

Les tumeurs bénignes ne vont évoluer que localement, c'est-à-dire qu'elles ne donneront jamais de métastases, et qu'elles n'entraîneront pas la mort du sujet porteur, sauf dans le cas de complications mécaniques et/ou métaboliques (ex : tumeurs cérébrales et endocriniennes). La transformation en tumeur maligne est cependant possible. Par opposition, les tumeurs malignes, ou cancéreuses, se présentent sous la forme d'une masse délimitée ayant des prolongements, ce qui n'est pas le cas des formes bénignes. C'est de cette dernière caractéristique qu'est tiré le nom « cancer » signifiant crabe en latin. De plus, la capacité de croissance des tumeurs malignes à se développer est plus importante que celle des formes bénignes. Enfin, contrairement aux formes bénignes, les tumeurs cancéreuses ont la capacité d'envahir les tissus, organes et ganglions voisins.

Connaissant maintenant les différentes caractéristiques des tumeurs, nous nous intéresserons plus particulièrement à la cellule cancéreuse et aux aspects moléculaires de la carcinogénèse.

I.2. La croissance cellulaire

I.2.1. La croissance des cellules « normales »

Dans un tissu sain, les cellules, dites somatiques, remplissent différentes fonctions déterminées génétiquement. Une cellule somatique (de l'organisme) n'a que deux devenir possibles : soit elle se divise, soit elle meurt par apoptose (mort cellulaire programmée). En fonction de signaux extracellulaires variés, les cellules en division peuvent se reproduire soit :

- à l'identique, c'est l'autoreproduction, dans laquelle une cellule-mère va donner deux cellules-filles,
- en acquérant de nouvelles fonctions par différenciation.

Un tissu sain doit connaître un équilibre entre le nombre de divisions cellulaires et le nombre de cellules qui meurent par apoptose. Si un de ces phénomènes est perturbé, il évoluera inexorablement vers une tumeur. Pour maintenir cet équilibre précaire, la cellule possède différents systèmes de régulation qui interviennent en fonction des signaux qu'elle capte de son environnement et des cellules voisines. Ces signaux vont permettre à la cellule de jouer un rôle précis dans le tissu et vont ainsi l'engager dans un phénomène de prolifération, de différenciation, de morphogénèse, ou de mobilité.

I.2.2. Le cycle cellulaire

I.2.2.1. Les différentes phases

Le cycle cellulaire des cellules eucaryotes comprend quatre phases plus un stade nommé G₀, durant lequel les cellules ne se divisent pas (cellules quiescentes).

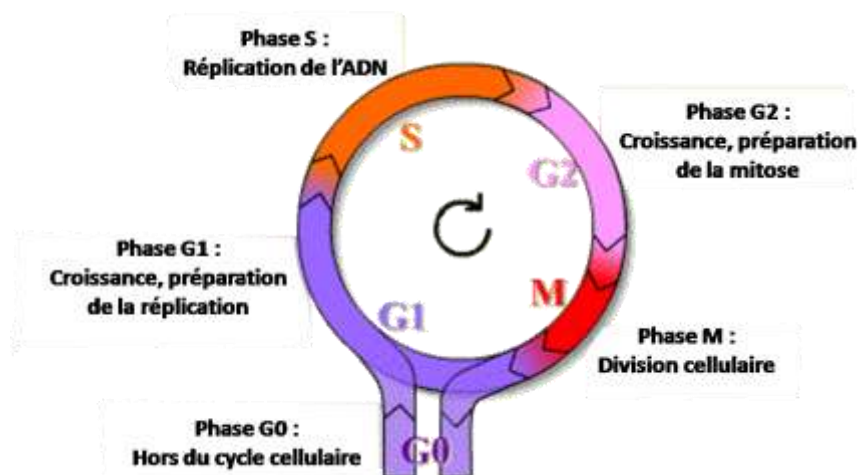


Figure 1 : Les différentes phases du cycle cellulaire (<http://www.snv.jussieu.fr/vie/dossiersBM/cyclecellVIE/index.htm>).

- Le stade G0 : ce stade est celui de la non division cellulaire ; ainsi, à la fin d'un cycle les cellules peuvent soit recommencer un cycle, soit passer en G0.
- Le stade G1 : les cellules préparent la phase S par la synthèse de protéines nécessaires à la synthèse d'ADN.
- Le stade S : durant cette phase de synthèse, l'ADN est répliqué et la cellule va passer de N chromosomes à 2N chromosomes.
- Le stade G2 : les cellules préparent la phase M par la synthèse de protéines nécessaires à la division cellulaire.
- Le stade M : phase durant laquelle une cellule-mère va se diviser en deux cellules-filles identiques ; c'est la mitose.

I.2.2.2. La régulation du cycle cellulaire

Différents mécanismes permettent la régulation du cycle cellulaire en agissant sur la transition entre les phases ou à des points de contrôles très précis (GOLIAS *et al.* 2004).

Les transitions entre les phases du cycle sont régulées par des complexes protéiques appelés kinases dépendantes des cyclines (cdk), qui doivent s'associer à une cycline pour être actifs. Le complexe protéique engagé sera différent en fonction de la phase du cycle à réguler. Par exemple, lors de la transition G1/S, c'est le complexe cdk2/cycline-E qui va intervenir ; en phase S, c'est le complexe cdk2/cycline-A (LEES 1995).

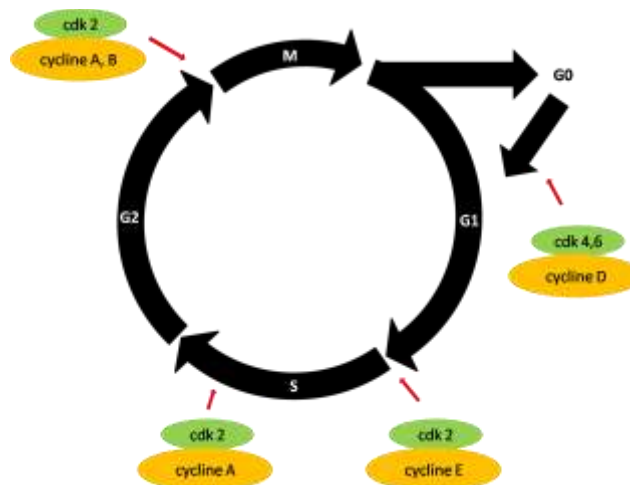


Figure 2 : Régulation du cycle cellulaire par les complexes kinases cyclines dépendantes (cdk)/cycline (Inspiré de Golias, Charalabopoulos *et al.* (2004)).

En plus de ce mécanisme qui régule les transitions entre les phases du cycle par les complexes cdk/cycline, il existe des mécanismes qui contrôlent le cycle cellulaire dans ses phases critiques :

- le mécanisme DDCP (*DNA Damage Checkpoint*) qui intervient avant et pendant la réplication de l'ADN,
- le mécanisme RCP (*Replication Checkpoint*) qui intervient lors de l'achèvement de la réplication avant de rentrer en mitose,
- le mécanisme MPC (*Mitotic Checkpoint*) qui intervient lors du positionnement des chromosomes sur la plaque mitotique.

I.2.3. La croissance des cellules cancéreuses

Un défaut dans un des systèmes de régulation va rendre la cellule cancéreuse virtuellement immortelle, soit par une division excessive, soit par une perte d'apoptose. Ce phénomène d'immortalité va conduire à la formation d'une tumeur, par prolifération anarchique de la cellule touchée. Avant de devenir cancéreuse la cellule doit subir, suivant les estimations, entre deux et sept altérations génétiques. Ces altérations interviennent sur une même cellule dans des gènes impliqués dans six fonctions cellulaires majeures présentés ci-dessous.

I.2.3.1. Les gènes de dépendance aux facteurs de croissance

Une cellule indépendante des facteurs de croissance va être constamment dans un état prolifératif, tout en étant insensible aux facteurs antiprolifératifs. Ainsi la cellule va gagner un pouvoir de multiplication autonome grâce à différents mécanismes :

- soit en synthétisant elle-même les facteurs de croissance auxquels elle est sensible (stimulation autocrine),
- soit en multipliant ses récepteurs aux facteurs de croissance ; ainsi elle sera activée par une plus faible quantité de facteurs,
- soit en inhibant ou en modifiant les récepteurs, ce qui rendra les facteurs antiprolifératifs inefficaces.

I.2.3.2. Les gènes de contrôle du cycle cellulaire

La cellule peut également modifier la régulation du cycle cellulaire. Si une mutation survient dans les kinases dépendantes des cyclines et/ou les cyclines, leur rôle dans les mécanismes de régulation inter- et/ou intra-phases pourra être modifié.

I.2.3.3. Les gènes de l'immortalisation cellulaire

Les gènes en cause ici permettent à la cellule d'acquérir un pouvoir répliatif illimité. Ce sont les gènes codant pour les télomérases. En effet, un dysfonctionnement des télomérases (surexpression) ne permettra plus le raccourcissement naturel des télomères. Ce raccourcissement

dans une cellule normale aboutit à la sénescence. Dans une cellule cancéreuse ce phénomène ne sera plus possible du fait de la surexpression des télomérases (Chapitre I : II.2.4. Les télomérases).

I.2.3.4. Les gènes de l'apoptose

L'apoptose est un phénomène clé de la vie cellulaire. La perte de ce mécanisme par mutation des gènes qui le régissent va entraîner un déséquilibre entre division et mort cellulaire. Il y aura donc une accumulation de cellules excédentaires.

I.2.3.5. Les gènes de néo-vascularisation

Cette caractéristique est très importante pour la tumeur, car la formation de vaisseaux sanguins va lui assurer un apport suffisant en oxygène et en éléments nutritifs permettant une croissance optimale.

I.2.3.6. Les gènes de la mobilité cellulaire

Ces gènes permettent à certaines cellules de se détacher de la tumeur primitive et ainsi de créer des métastases, qui représentent l'étape ultime de la cancérisation. Ces métastases migrent par le réseau lymphatique ou sanguin et vont créer des tumeurs secondaires.

L'acquisition par la cellule de plusieurs de ces propriétés tumorales peut être due à la mutation d'un seul gène, comme par exemple *p53* qui va jouer sur l'angiogenèse et la résistance à l'apoptose.

I.3. Les étapes de la cancérisation

La cancérisation est le phénomène qui va transformer une cellule saine en cellule tumorale. C'est un processus multi-étapes qui a été proposé pour la première fois par Armitage et Doll en 1954. À l'heure actuelle, les trois étapes de la cancérisation sont bien connues et caractérisées.

I.3.1. Les trois étapes de la cancérisation

I.3.1.1. L'initiation

L'initiation correspond au tout premier évènement que subit la cellule. Elle intervient de nombreuses années avant les premiers signes et symptômes de cancer et survient après une mutation irréversible au niveau de l'ADN, qui est le plus souvent due à un agent initiateur tel que les rayons (UV, X) ou les produits chimiques, pour ne citer que ces deux exemples.

La plupart du temps ces mutations interviennent dans de l'ADN non codant, ce qui n'a aucune incidence biologique. Plus rarement, une mutation peut toucher une partie codante de l'ADN

et donc possiblement un gène impliqué dans les six fonctions cellulaires citées précédemment (Chapitre I : 1.2.3. La croissance des cellules cancéreuses). La fréquence de survenue des cancers reste néanmoins faible en regard du nombre de divisions cellulaires se produisant dans la vie d'un individu, grâce à des systèmes enzymatiques de réparation de l'ADN qui réparent la majorité de ces altérations. Il peut arriver parfois que la mutation touche justement un gène codant pour une enzyme de réparation de l'ADN au niveau somatique ou constitutionnel.

Pour arriver à un processus cancéreux, c'est-à-dire à l'apparition d'un clone malin, il faut que la cellule subisse plusieurs événements rares non liés entre eux, ce qui explique l'augmentation de la fréquence d'apparition des cancers avec l'âge.

A ce stade, on ne peut pas encore parler de cellules tumorales, car elles n'ont pas encore acquis la capacité d'autonomie de croissance (WEINSTEIN *et al.* 1995).

I.3.1.2. La promotion

La promotion est due à un agent promoteur qui va exercer son action pendant des années, en facilitant la multiplication des cellules initiées. Il y aura ainsi un avantage sélectif de ces cellules. L'agent promoteur peut être considéré comme cancérigène, car dans la plupart des cas il ne va provoquer un cancer que si son action est combinée à la présence d'une cellule initiée. L'agent de promotion peut être :

- la nutrition,
- une substance toxique,
- une infection chronique,
- un traumatisme répété,
- une hormone (cancers du sein, de la prostate par exemple),
- l'âge,...

Une cellule qui arrive à ce stade de promotion n'est pas encore assurée d'acquérir la propriété d'autoreproduction nécessaire aux cellules cancéreuses. En effet, les cellules ayant subi ces mutations peuvent s'engager dans un processus d'apoptose, lié au trop grand nombre de mutations qui vont engendrer un ADN non viable. La capacité d'autoreproduction d'une cellule est acquise après deux à sept événements ; la prolifération d'une seule cellule va alors donner naissance à des clones cellulaires tous identiques génétiquement (WEINSTEIN *et al.* 1995).

I.3.1.3. L'invasion tumorale

Lors de cette ultime étape du processus de cancérisation, les cellules vont acquérir de nouvelles propriétés qui vont permettre la progression de la tumeur dans le tissu dont elle est issue,

puis dans l'organe correspondant et, en dernier lieu, dans des organes distants. Cette invasion va se faire de proche en proche, dans les trois dimensions de l'espace. Puis, au fur et à mesure de l'acquisition de nouvelles propriétés, la progression va prendre de l'ampleur.

Dans un premier temps, certaines cellules vont connaître de nouvelles mutations qui vont permettre à la tumeur de devenir un tissu bien organisé. D'autres cellules vont acquérir une certaine mobilité qui va leur permettre de se rapprocher des vaisseaux sanguins, sources d'éléments nutritifs nécessaires à la division cellulaire. Certaines cellules vont même acquérir la capacité de former de nouveaux vaisseaux sanguins. Ce phénomène, appelé angiogenèse, est très important pour la croissance de la tumeur, car il va permettre de créer un réseau de vaisseaux sanguins autour de la tumeur, lui apportant les éléments nutritifs et l'oxygène nécessaires à son métabolisme. Cette caractéristique, acquise par des mutations génétiques ou par l'hypoxie que subissent les cellules tumorales, apparaît dès que la tumeur atteint une taille de 1 à 2 mm de diamètre, grâce à la production de facteurs pro-angiogéniques.

D'autres cellules vont gagner la capacité de traverser les membranes basales. Ceci va conduire à l'invasion des organes voisins par dissémination au travers des vaisseaux sanguins et lymphatiques, mais également au travers des espaces naturels entre les organes. Cette caractéristique est à l'origine des métastases.

Lors de cette étape de cancérisation, les cellules deviennent de plus en plus instables génétiquement, du fait du grand nombre de divisions qu'elles subissent (WEINSTEIN *et al.* 1995).

La carcinogenèse colorectale suit les étapes décrites par Armitage et Doll (1954) qui se caractérisent par la morphologie des cellules (Figure 3).

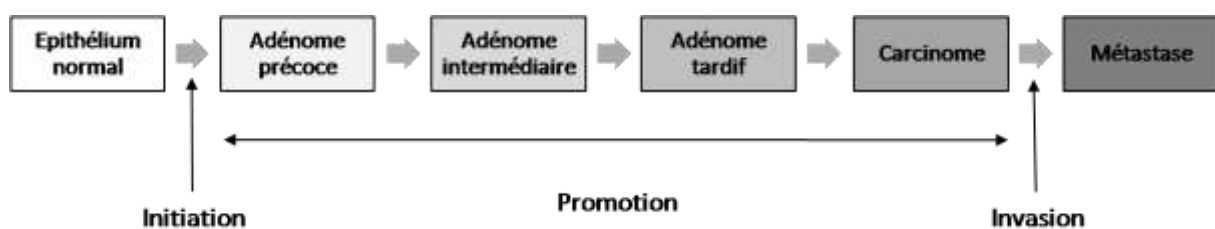


Figure 3 : Processus multi-étapes de la carcinogenèse colorectale en relation avec les changements morphologiques rencontrés.

Les cellules « normales » (épithélium normal) vont subir le premier évènement (initiation) suite à l'action de l'agent initiateur. Ce premier évènement, même s'il est bénin va entraîner un changement morphologique des cellules qui vont évoluer d'adénome précoce en adénome tardif, puis en carcinome. Chaque changement morphologique est le signe d'une progression dans l'évolution de l'étape de promotion. Après un certain temps, une dernière mutation va permettre à la cellule de gagner le phénotype invasif (métastase).

I.4. Illustration par le cancer colorectal

Il est nécessaire, avant d'aborder la carcinogenèse colorectale, de connaître l'anatomie et le rôle du colon, l'épidémiologie des CCR et nous pourrons enfin détailler les différents événements de cette carcinogenèse.

I.4.1. Caractéristiques anatomiques et fonctionnelles du gros intestin

I.4.1.1. Anatomie

Le côlon est situé dans le prolongement de l'intestin grêle dans la cavité abdominale et en avant des anses grêles. Son diamètre moyen est de 4 cm pour une longueur de 1,5 mètre. Il est divisé en plusieurs parties (Figure 4) :

- le cæcum fait la jonction avec la fin de l'intestin grêle. C'est dans cette partie du côlon droit que se situe l'appendice.
- le côlon droit, ou côlon ascendant, remonte jusqu'au niveau du foie et se termine par l'angle colique droit.
- le côlon transverse s'étend de l'angle colique droit à l'angle colique gauche.
- le côlon descendant, ou côlon gauche, descend de l'angle gauche jusqu'au niveau du bassin et se termine par le côlon sigmoïde.
- le rectum fait environ 13 cm de long et constitue la partie terminale du gros intestin.
- l'anus est le point de sortie du gros intestin. Il est formé d'un sphincter anal interne à motricité involontaire et un anneau externe à motricité volontaire.

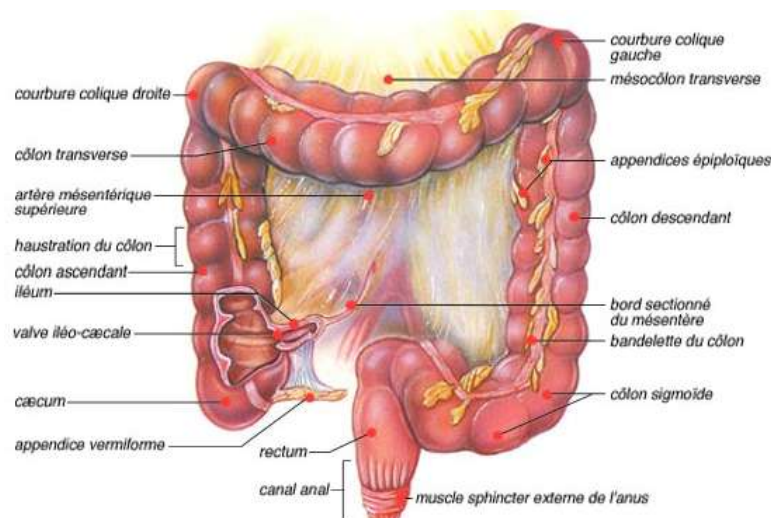


Figure 4 : Anatomie macroscopique du gros intestin (www.infocancer.org).

I.4.1.2. Fonctions

Le rôle du côlon est de concentrer les matières fécales par réabsorption de l'eau et des sels. Il n'intervient pas ou très peu dans la fonction de digestion, qui est effectuée par l'estomac, le duodénum, l'iléon et le jéjunum. Le côlon contient une flore bactérienne importante qui synthétise des vitamines et produit des gaz par fermentation.

Le côlon est animé de contractions qui font progresser les matières et les acheminent dans l'ampoule rectale. Les parois du rectum se distendent par l'arrivée des matières. Le réflexe de défécation entraîne le relâchement des deux sphincters, ainsi qu'une augmentation des contractions du côlon et du rectum.

I.4.2. Epidémiologie des cancers du côlon

L'incidence du cancer colorectal est très variable en fonction des régions du monde. En effet, elle varie d'un facteur 1 à 30 entre les régions à faible risque comme les régions rurales d'Afrique, les pays d'Amérique du Sud, la Chine, l'Inde et les régions à haut risque représentées par l'Australie, les pays d'Amérique du Nord (plus particulièrement la côte Ouest des Etats-Unis) et la plupart des pays d'Europe occidentale. Les pays d'Europe du Nord et de l'Est sont des régions à risque intermédiaire.

En France en 2005, 37413 nouveaux cas de cancers colorectaux ont été recensés, soit une incidence standardisée sur l'âge et pour 100000 habitants estimée à 37,7 chez les hommes et à 24,5 chez les femmes (BELOT *et al.* 2008). Ceci représente environ 15% de l'ensemble des cancers diagnostiqués chaque année, soit le troisième cancer le plus fréquent pour les deux sexes confondus, après le cancer de la prostate et celui du sein. Malgré l'amélioration du dépistage et du pronostic, la survie globale des sujets atteints de CCR est voisine de 50%. Ainsi pour l'année 2005, la France dénombre 16865 décès, ce qui correspond à un peu moins de la moitié du nombre de cas incidents, et place le cancer colorectal au deuxième rang de la mortalité par cancer pour les deux sexes confondus, après le cancer du poumon (BELOT *et al.* 2008).

I.4.3. Modèle de la carcinogenèse colorectale

A chacune des étapes de la transformation de l'épithélium intestinal normal en adénocarcinome, des altérations somatiques peuvent intervenir (Figure 5) (KINZLER and VOGELSTEIN 1996; LYNCH and DE LA CHAPELLE 2003). L'ordre chronologique de ces mutations est bien connu. Tous les évènements somatiques ne sont toutefois pas nécessaires à l'acquisition d'un phénotype malin, mais ils interviennent tous dans au moins un type de cancer colorectal. L'étude somatique de la carcinogenèse colorectale fait intervenir différents gènes (Figure 5).

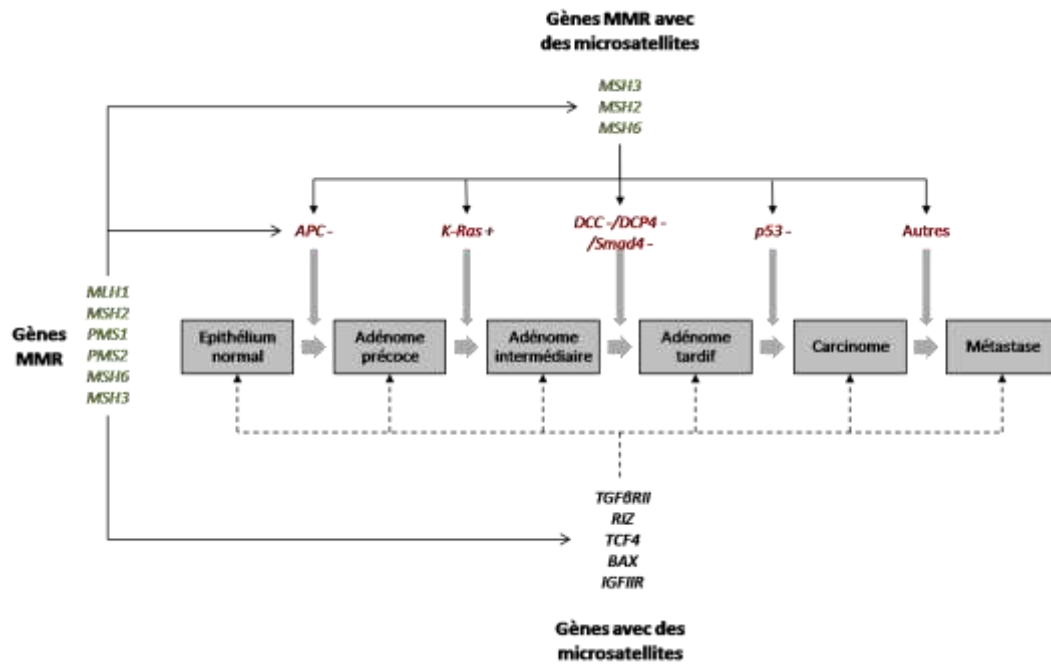


Figure 5 : Modèle de carcinogenèse colorectale adapté de Kinzler et Vogelstein (1996) ainsi que de Lynch et de la Chapelle (2003). Les gènes suppresseurs de tumeur et les oncogènes impliqués sont notés en rouge ; + indique une altération activatrice ; - une altération inactivatrice. En vert, figurent les gènes « *MisMatch Repair* » (MMR) qui peuvent être impliqués à chaque étape. Les altérations des gènes MMR sont toujours inactivatrices. En noir, sont représentés des gènes sur lesquels les gènes MMR ont une action et qui peuvent intervenir dans le processus de carcinogenèse colorectale.

Les gènes impliqués comprennent des oncogènes (*Ras*) et des gènes suppresseurs de tumeur (*APC*, *DCC*). Le gène *APC* se trouve altéré très tôt dans la progression tumorale et est donc considéré comme un gène qui peut initier le développement tumoral. Par comparaison, le gène *p53* n'est touché que très tard dans la progression du cancer colorectal. L'altération des gènes *MMR* (*MisMatch Repair*) accélère la progression des tumeurs, en augmentant le taux de mutations des gènes cibles.

L'évolution des cellules épithéliales normales en adénome bénin puis en carcinome malin suit un processus multi-étapes impliquant des altérations génétiques et épigénétiques (FEARON and VOGELSTEIN 1990), qui sont la conséquence d'évènements germinaux ou somatiques conférant à la cellule un avantage de croissance. D'un point de vue histologique, deux sortes de cryptes aberrantes existent : les cryptes hétéroplasiques associées aux muqueuses intestinales normales, qui seraient dues à une activation mutationnelle de l'oncogène *KRAS2* (SMITH *et al.* 1994), et les cryptes dysplasiques liées à une augmentation de la progression des cancers par une inactivation biallélique du gène suppresseur de tumeur *APC* (LAMLUM *et al.* 2000; SMITH *et al.* 1994).

II. Génétique des cancers

Les cancers sont tous dus à la survenue de mutations dans une cellule. A ce titre, on peut donc parler de maladie génétique de la cellule somatique. Toutefois, la première mutation qui va initier la transformation cellulaire peut être, selon le type de cancer, germinale ou somatique. Dans le cas d'une mutation germinale, celle-ci est transmise à la descendance ; le cancer sera alors héréditaire. Si la mutation est acquise après la naissance au niveau somatique, on parlera de cancer sporadique (non familial). La proportion des formes héréditaires est faible pour les cancers colorectaux.

II.1. Les différentes formes de cancers

II.1.1. Les cancers héréditaires

Le concept de cancer héréditaire est connu depuis longtemps, mais il est limité à des cas rares (*Xeroderma Pigmentosa*, polypose adénomateuse familiale, rétinoblastome,...) dont le mécanisme génétique a été compris récemment.

Ce type de cancer nécessite une première anomalie génétique ou un premier évènement (« *first hit* ») qui, par définition, doit être situé dans les cellules germinales. Il peut être hérité, ou apparaître *de novo* dans les gamètes.

Dans la majorité des cancers héréditaires, le premier évènement est une mutation dans un gène suppresseur de tumeur. Les gènes suppresseurs de tumeur codent pour des protéines qui interviennent dans différentes fonctions clés de l'organisme : maintenance de l'intégrité génomique, régulation du cycle cellulaire, apoptose, différenciation, transcription, traduction. Les mutations héritées dans ces gènes conduisent en général à une prédisposition moyenne à certains cancers, avec un risque cumulé important (BIGNON 2004).

L'immortalité des cellules malignes est une étape clé de la cancérisation, qui est acquise par accumulation de mutations dans les gènes. Cela peut expliquer pourquoi une seule mutation peut, quelques années ou quelques décennies plus tard, provoquer un cancer.

Le premier évènement (« *first hit* »), de nature variée, correspond à une mutation ponctuelle ou à une altération génétique de type méthylation, délétion ou insertion. Les évènements les plus souvent rencontrés sont les mutations non-sens et les insertions/délétions, qui donnent des protéines non fonctionnelles. L'interprétation de ce premier évènement est parfois difficile lorsqu'il s'agit d'une mutation faux-sens ou lorsqu'il s'agit d'une mutation qui touche un site d'épissage, car

ces différentes mutations peuvent donner des protéines dont la fonction est modifiée de manière variable (BIGNON 2004).

II.1.2. Les cancers non mendéliens

II.1.2.1. Les formes familiales non mendéliennes

Dans ces familles, la ségrégation n'est pas mendélienne, car il existe des sauts de génération, ou des défauts de pénétrance qui rendent difficile l'identification des gènes impliqués par les études de liaison classiques. On peut observer, dans ces familles, un nombre important de cancers similaires, sans qu'une altération génétique dans les gènes connus puisse être mise en évidence. Ces « formes familiales » représentent une petite proportion non négligeable de certains cancers (sein, côlon,...). Il est admis aujourd'hui qu'un grand nombre de ces formes familiales sont dues à des mécanismes oligogéniques, conduisant à des formes héréditaires non mendéliennes (BIGNON 2004).

II.1.2.2. Les formes sporadiques

Dans ces formes de cancers, le processus de cancérisation est encore différent, en ce sens où aucune mutation germinale n'intervient. Ces cancers résultent d'un nombre important de mutations exclusivement somatiques dans différents gènes impliqués dans le processus de cancérisation, mais pour lesquels une contribution génétique combinée à des facteurs environnementaux n'est pas exclue.

II.2. Les gènes impliqués dans les cancers

En fonction de leur rôle dans la cellule « normale », les gènes intervenant dans le cancer sont classés en oncogènes, gènes suppresseurs de tumeur, ou gènes de réparation de l'ADN. Alternativement, les gènes peuvent être classés en *gatekeeper*, *caretaker*, *landscaper* (KINZLER and VOGELSTEIN 1997; KINZLER and VOGELSTEIN 1998). Dans la suite de l'exposé nous utiliserons la première classification.

II.2.1. Les oncogènes

II.2.1.1. Définition

Les gènes qui stimulent de façon normale la division cellulaire sont appelés « proto-oncogènes ». Ils apparaissent hautement conservés au fil de l'évolution et ils régulent une cascade d'évènements qui maintiennent l'ordre de progression du cycle cellulaire, la division et la

différenciation cellulaire. Dans les cancers, cet ordre de progression est perdu quand un ou plusieurs des composants de cette voie est activé (PARK 1998).

Le mot « oncogène » désigne, quant à lui, une forme anormale d'un « proto-oncogène » qu'on retrouve dans les cancers. La mutation d'un seul allèle d'un oncogène ou d'un proto-oncogène (mutation dominante) suffit à donner une protéine possédant une propriété différente de la protéine normale. Cette nouvelle protéine, appelée « oncoprotéine », déclenchera une division cellulaire anormale. La mutation va jouer sur la structure, le niveau et le site d'expression de l'oncoprotéine.

Les premiers oncogènes qui ont été isolés à partir de virus correspondent à des formes mutées du gène cellulaire humain : le proto-oncogène. Dans les cancers humains, les proto-oncogènes sont souvent localisés à proximité de points de cassure chromosomique.

II.2.1.2. Fonctions et mécanismes d'action des proto-oncogènes

II.2.1.2.1. Mécanismes d'action d'un proto-oncogène

Il existe essentiellement trois mécanismes par lesquels les proto-oncogènes agissent.

Le premier mécanisme fait intervenir la phosphorylation des résidus sérine, thréonine, tyrosine des protéines, à partir d'un groupement phosphate de l'ATP. La phosphorylation va intervenir de deux manières sur le signal de transduction : soit elle va activer l'activité enzymatique « kinase » des protéines en modifiant leur conformation, soit générer un site de recrutement des protéines cibles qui seront à leur tour phosphorylées (HUNTER and COOPER 1985).

Le second mécanisme fait intervenir les GTPases. De la même manière que les kinases, ces protéines vont réguler les phosphorylations et donc la transmission du signal de transduction via le cycle GDP/GTP.

Le troisième mécanisme fait intervenir des facteurs de transcription nucléaires, qui contrôlent l'expression de proto-oncogènes impliqués dans la progression du cycle cellulaire (ARIGA *et al.* 1989; WASYLYK *et al.* 1990).

Par conséquent, une mutation survenant dans un proto-oncogène intervenant dans l'un de ces trois mécanismes d'action va modifier toute la cascade de réactions et avoir un effet sur la transduction du signal (Figure 6).

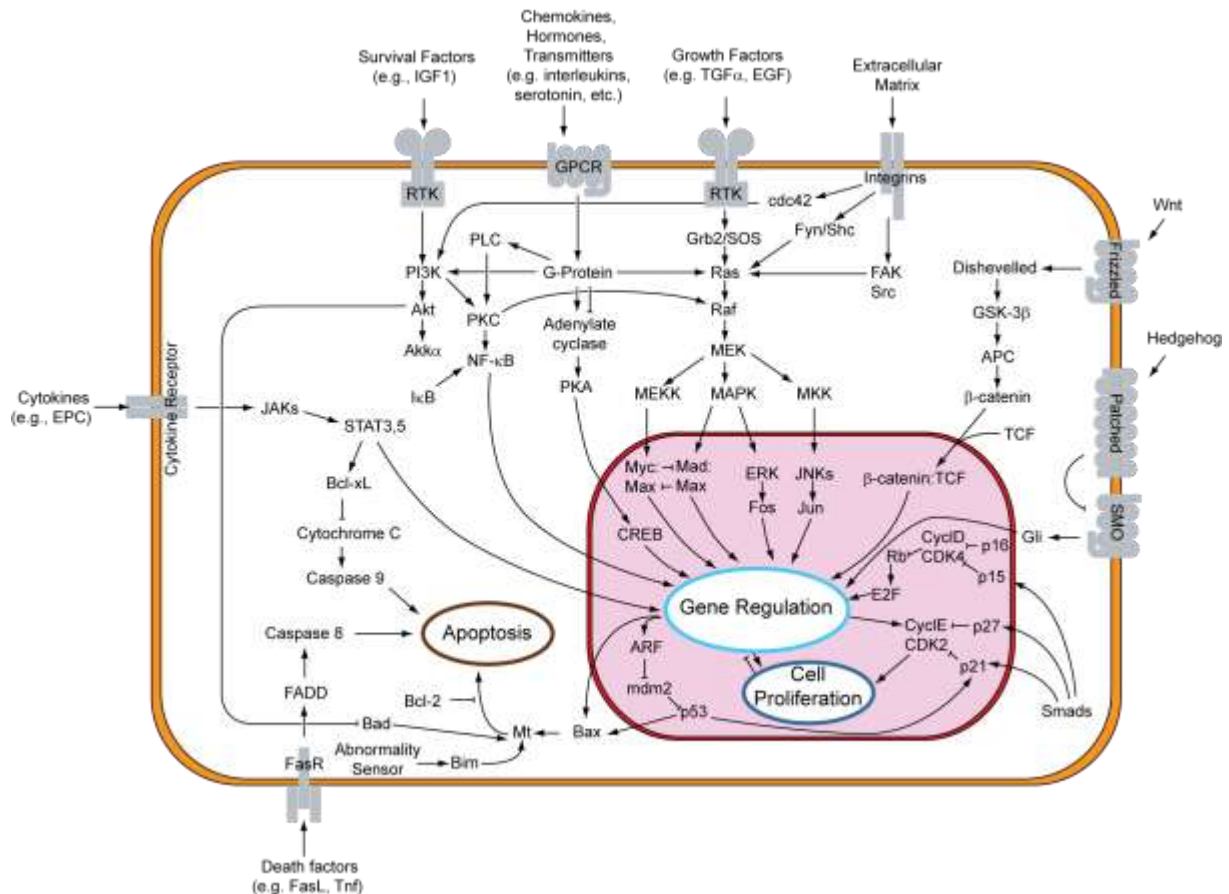


Figure 6 : Les différentes voies de transduction (LODISH 2003).

II.2.1.2.2. Les familles d'oncoprotéines impliquées dans la transduction des signaux de prolifération

a) Les facteurs de croissance

Les facteurs de croissance stimulent les cellules en G0 pour entrer dans le cycle cellulaire. Ces facteurs de croissance régulent finement les phénomènes biologiques de la cellule, car une quantité trop grande ou trop faible peut entraîner la mort cellulaire (PLEDGER *et al.* 1977). Le facteur « PDGF » (*Platelet Derived Growth Factor*), par exemple, active la prolifération des cellules en se liant à un récepteur spécifique exprimé ou non à la surface des cellules (EVA *et al.* 1982; ROSS *et al.* 1974).

b) Les récepteurs de facteurs de croissance

Un grand nombre de récepteurs dérivent d'oncoprotéines qui ont une activité tyrosine kinase. Les tyrosines kinases font partie d'une grande famille dont le rôle est de transférer un groupement phosphate de l'ATP à la tyrosine. La phosphorylation par les tyrosines kinases est un évènement rare (environ 0,05% de toutes les phosphorylations), mais qui régule des évènements clés dans la voie des signaux de transduction, qui contrôlent la croissance et la forme de la cellule. Les

tyrosines kinases ne sont pas forcément des récepteurs ; elles peuvent être cytoplasmiques (COOPER *et al.* 1982).

Ces protéines peuvent être mutées et se comporter anormalement. Par exemple, la protéine codée par l'oncogène *erb-B*, est un analogue muté du récepteur cellulaire du facteur de croissance épidermique « EGFR » (*Epidermal Growth Factor Receptor*). La stimulation de ce récepteur entraîne la transmission d'un signal qui permet de stimuler la division cellulaire.

c) Les protéines à activité GTPase

La première protéine de cette famille à avoir été découverte comme ayant un rôle dans la tumorigenèse est Ras (BOURNE 1987; BOURNE *et al.* 1990). Le gène *ras* très conservé code pour une protéine cytoplasmique, dont le rôle, en aval des récepteurs de facteurs de croissance, est crucial dans la régulation de la forme, de la mortalité et de la croissance cellulaire (BARBACID 1987). Il existe trois gènes humains codant pour les protéines Ras (*HRAS*, *KRAS* et *NRAS*).

d) Les oncoprotéines nucléaires

On peut citer comme exemple le gène *C-MYC*, connu pour son action de stimulation de la prolifération cellulaire. Son expression est dérégulée notamment dans la translocation t(8,14) observée dans le lymphome de Burkitt, par juxtaposition d'un promoteur fort.

e) Les autres oncoprotéines

Comme indiqué sur la Figure 6, les oncogènes ont bien d'autres activités, telles que :

- le contrôle du cycle cellulaire (cyclines, kinases dépendantes des cyclines, phosphatases),
- la régulation négative de l'apoptose (famille Bcl2, survivine),
- l'immortalisation des cellules, par une activité télomérase,
- l'implication dans la dissémination des cellules et la formation des métastases (rôle assuré par les métalloprotéinases et les protéines assurant l'adhésion cellulaire),
- la stimulation de l'angiogenèse (induite par les facteurs de croissance angiogéniques, comme le VEGF).

II.2.1.3. Les oncogènes dans les cancers héréditaires

Trois syndromes seulement sont attribués à des mutations constitutionnelles dans des oncogènes (*RET*, *MET*, *CDK4*) (FEARON 1997). Le gène *RET* est impliqué dans les formes familiales de cancers médullaires de la thyroïde et les néoplasies endocriniennes multiples de type 2A et 2B

(NEM2A, NEM2B). Le gène *MET* est activé dans certaines formes familiales de cancers papillaires rénaux et le gène *CDK4* dans les mélanomes héréditaires.

II.2.2. Les gènes suppresseurs de tumeur

II.2.2.1. Définition

Ces gènes ont pour fonction de freiner le déclenchement de la division cellulaire. Par extension, on considère maintenant comme gènes suppresseurs de tumeur tous les gènes dont la perte de fonction contribue à l'oncogenèse.

Les gènes suppresseurs de tumeur codent pour des protéines qui jouent un rôle dans la régulation de la croissance ou dans la différenciation cellulaire. La perte de leurs fonctions contribue directement au phénotype altéré des cellules cancéreuses. Contrairement aux oncogènes, ces gènes ont une action dite récessive, c'est-à-dire que les deux copies du gène doivent être mutées pour entraîner une perte de la fonction de la protéine (FEARON 1998).

II.2.2.2. Le modèle de Knudson et Comings

Les gènes suppresseurs de tumeur, par la nécessité de leur inactivation biallélique suivent le modèle élaboré par Knudson en 1971 et complété par Comings en 1973. Dans ce modèle, deux évènements sont nécessaires pour aboutir au cancer (théorie du *double hit*). Le premier évènement est hérité ou acquis, et une seconde mutation somatique est nécessaire sur le deuxième allèle pour initier la cancérisation. La première mutation, selon qu'elle soit germinale ou somatique, va permettre la distinction entre cancer familial et sporadique. En effet, à la suite du second évènement, toujours somatique, les mutations héréditaires mèneront vers des cancers familiaux, alors que les mutations acquises donneront des cancers sporadiques.

La différence dans la nature de la première mutation explique également pourquoi les cancers familiaux surviennent plus précocement que les cancers sporadiques. En effet, il faut plus de temps pour acquérir deux mutations somatiques sur les deux allèles d'un même gène, que pour acquérir une seule mutation somatique supplémentaire. En règle générale, le second évènement va inactiver le deuxième allèle du gène précédemment touché par délétion ou méthylation.

Le modèle de Knudson est applicable à un bon nombre de cancers à transmission autosomique dominante (KNUDSON 1996). Dans certains cas cependant, la mutation somatique n'est pas toujours mise en évidence en plus de la mutation germinale. Le second évènement ne serait donc pas toujours nécessaire, même si, techniquement, toutes les altérations possibles ne sont pas mises en évidence (BIGNON 2004).

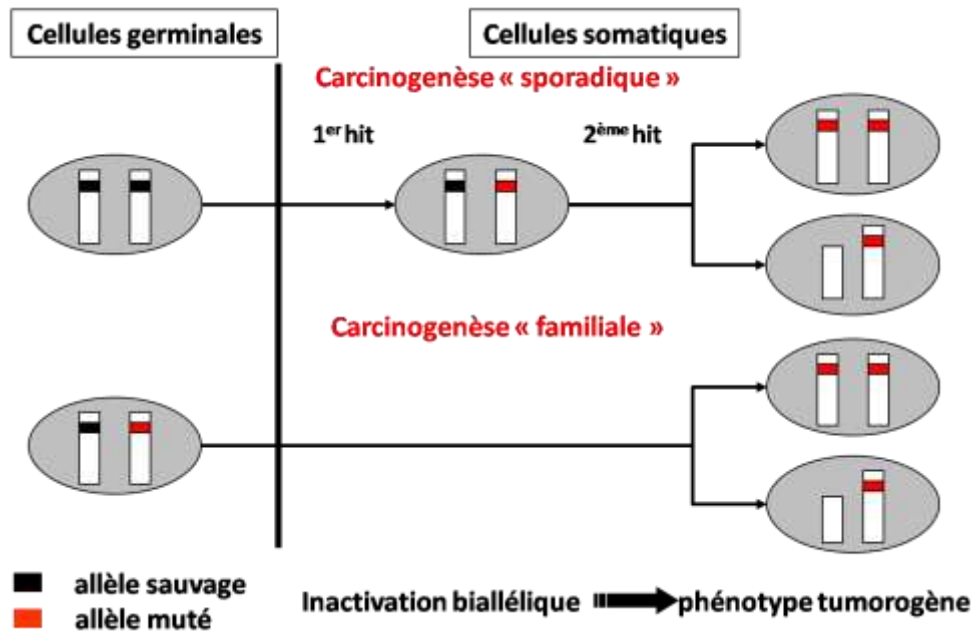


Figure 7 : La théorie de Knudson. Dans la carcinogenèse sporadique deux évènements somatiques sont nécessaires pour acquérir le phénotype tumorigène alors que dans la carcinogenèse familiale un seul évènement somatique est nécessaire, le premier étant hérité.

II.2.2.3. Exemples de gènes suppresseurs de tumeur impliqués dans des formes familiales de cancer

Quelques gènes suppresseurs de tumeur sont des antagonistes directs ou indirects des oncoprotéines dans la régulation de la croissance (FEARON 1998).

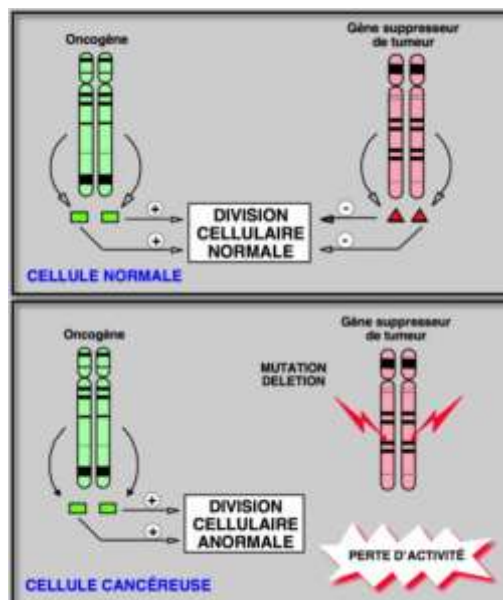


Figure 8 : Mécanisme d'action antagoniste d'un gène suppresseur de tumeur. Dans une cellule normale, les gènes suppresseurs de tumeur vont inhiber l'action des oncogènes. Lorsqu'une mutation survient dans un de ces gènes, l'inhibition est levée et la cellule présente une division cellulaire anormale.

II.2.2.3.1. Le gène *Rb*

Le gène suppresseur *Rb* est appelé ainsi par référence au rétinoblastome, qui est une tumeur rare de l'enfant affectant la rétine. La protéine pour laquelle il code appartient au système de contrôle de la prolifération cellulaire. C'est en se basant sur les formes familiales et sporadiques de rétinoblastome que Knudson a développé sa théorie.

La protéine RB agit lors de la transition G1/S. Cette dernière dépend des facteurs de transcription de la famille E2F, qui contrôlent l'expression de gènes indispensables à la synthèse de l'ADN, en phase S. Les protéines de la famille E2F existent soit sous forme libres, soit sous forme inactives complexées à la protéine RB. L'aptitude de la protéine RB à fixer les facteurs de transcription E2F dépend de son état de phosphorylation. En effet, lorsque la protéine RB est non phosphorylée, elle est active et peut fixer les facteurs E2F. Il en résulte un blocage de la transition G1/S. Lorsque la protéine RB est phosphorylée, elle devient inactive et incapable de fixer la protéine E2F qui ainsi libérée, permet la transition G1/S. La phosphorylation de RB est elle-même sous la dépendance des cdk associées à une cycline, jouant le rôle de verrous moléculaires au niveau de la transition entre les différentes phases du cycle cellulaire (ZHU 2005).

II.2.2.3.2. Les gènes *BRCA1* et *BRCA2*

Il s'agit des gènes responsables de la prédisposition aux cancers du sein, d'où leur nom *Breast Carcinoma Cancer Antigen*. Leur fonction montre qu'ils sont impliqués dans la réparation des anomalies nucléaires. Bien que l'expression de *BRCA1* et *BRCA2* soit ubiquitaire, les mutations de ces deux gènes n'augmentent fortement le risque de cancer qu'au niveau des tissus mammaire et ovarien. Cette spécificité tissulaire reste cependant une énigme. Monteiro (2003) a proposé une hypothèse selon laquelle la perte d'hétérozygotie de *BRCA1* serait plus fréquente dans les tissus mammaire et ovarien. L'influence des œstrogènes sur ces organes cibles pourrait également jouer un rôle majeur dans cette spécificité tissulaire.

II.2.2.3.3. Autres gènes suppresseurs

Il existe de nombreux autres gènes suppresseurs de tumeur responsables de formes familiales de cancers ou impliqués dans des cancers sporadiques. On peut citer *APC*, impliqué dans la polypose adénomateuse familiale, *VHL*, responsable du syndrome de von Hippel Lindau, ou encore *NF1*, responsable de la neurofibromatose de type 1. Il en existe plus d'une quarantaine aujourd'hui (Tableau 1).

Gène et synonyme (s)	Syndrome	Mode de transmission	Fréquence des porteurs dans la population générale	Voie	Localisation des tumeurs Principaux types de tumeurs
<i>APC</i>	Polypose familiale	Dominant	1/8000	Wnt	Côlon, estomac, duodénum
<i>AXIN2</i>	Polypose atténuée et anomalies dentaires	Dominant	?	Wnt	Côlon
<i>CDH1 (E-cadhérine)</i>	Carcinome gastrique familial	Dominant	1/20000 à 1/50000	Wnt	Estomac (forme diffuse)
<i>GPC3</i>	Syndrome de Simpson-Golabi-Behmel	Récessif lié à l'X	?	Wnt	Tumeur Wilms, hépatoblastome
<i>Cyld</i>	Cylindromatose familiale	Dominant	1/50000	Apoptose	Pilotrícholemmomas
<i>EXT1, 2</i>	Exostoses héréditaires multiples	Dominant	?	GLI	Os, cartilage
<i>PTCH</i>	Syndrome de Gorlin	Dominant	1/50000 à 1/100000	GLI	Peau, médulloblastome
<i>SUFU</i>	-	Dominant	?	GLI	Médulloblastome, peau (?)
<i>FH</i>	Leiomyomatose héréditaire	Dominant	?	HIF-1	Leiomyomas cutanés et utérins, rein
<i>SDHB, C, D</i>	Paragangliomes familiaux	Dominant	1/100000 à 1/300000	HIF-1	Paragangliome, phéochromocytome
<i>VHL</i>	Syndrome de von Hippel-Lindeau	Dominant	1/40000	HIF-1	Rein, phéochromocytome, cervelet, hémangiomes
<i>p53</i>	Syndrome de Li-Fraumeni	Dominant	1/40000	p53	Sein, sarcome, surrénale, SNC
<i>WT1</i>	WARG, Denys-Drash, Frasier	Dominant	1/100000	p53	Tumeur de Wilms, gonadoblastome
<i>STK11 (LKB1)</i>	Syndrome de Peutz-Jeghers	Dominant	1/50000 à 1/100000	PI3K	Côlon, pancréatique, ovaire
<i>PTEN</i>	Maladie de Cowden	Dominant	1/50000 à 1/100000	PI3K	Sein, thyroïde, utérus
<i>TSC1, TSC2</i>	Sclérose tubéreuse de Bourneville	Dominant	1/15000 à 1/10000	PI3K	Rein
<i>CDKN2A</i>	Mélanome familial	Dominant	1/5000	RB	Mélanome, pancréas
<i>RB1</i>	Rétinoblastome héréditaire	Dominant	1/40000	RB	Rétine, os
<i>NF1</i>	Neurofibromatose type 1	Dominant	1/3000	RTK	Neurofibromes, SNC
<i>BMPR1A, SMAD4 (DPC4)</i>	Polypose juvénile	Dominant	1/100000	SMAD	Côlon, estomac
<i>MEN1</i>	Néoplasie endocrine multiple de type 1	Dominant	1/40000 à 1/30000	SMAD	Parathyroïde, hypophyse, pancréas, tumeur carcinoïde
<i>BHD</i>	Syndrome de Birt-Hogg-Dube	Dominant	?	?	Rein, follicule de cheveux
<i>HRPT2</i>	Syndrome hyperparathyroïdique et tumeur de joue	Dominant	?	?	Parathyroïde, fibrome ossifiant de la joue
<i>NF2</i>	Neurofibromatose de type 2	Dominant	1/30000	?	Méningiome, neurinome VIII paire nerfs crâniens
<i>TCF1 (HNF1α), TCF2 (HNF1β)</i>	MODY3 et MODY5 (diabète de maturité du sujet jeune)	Dominant	?	HNF1, transcription	Foie, rein, ovaire
<i>SNF5/INI1</i>	Syndrome des tumeurs rhabdoïdes	Dominant	<1/100000	Chromatine	SNC, rein
<i>PHOX2B</i>	Syndrome d'Ondine	Dominant	?	Transcription	Neuroblastome

Tableau 1 : Exemples de gènes suppresseurs de tumeur. La fréquence des porteurs dans la population générale est donnée à titre indicatif et est susceptible de varier d'un facteur 2, voire plus pour les fréquences très faibles. Lorsque la fréquence est inconnue (notée ?), elle est *a priori* très faible. WAGR : *Wilms tumor-aniridia-genitourinary anomalies-mental retardation* ; HIF-1 : *hypoxemia induced factor-1* ; PI3K : phosphatidylinositol 3 kinase ; RB : protéine du rétinoblastome ; RTK : récepteurs des tyrosines kinases ; HNF1 : *hepatocyte nuclear factor 1* (STOPPA-LYONNET and LENOIR 2005).

II.2.3. Les gènes impliqués dans la réparation de l'ADN

Lors de la réplication de l'ADN ou lors de certains événements « agressifs » pour la cellule (stress oxydatif, rayons X,...) des mutations peuvent survenir. Ces mutations seront le plus souvent réparées, mais il peut arriver qu'elles se produisent directement dans les gènes de systèmes de réparation de l'ADN, ce qui va favoriser la progression des cancers. En effet, ces mutations vont induire une absence ou un dysfonctionnement d'une protéine de réparation de l'ADN. Celle-ci ne pouvant plus assurer sa fonction correctement, les aberrations vont s'accumuler au niveau de l'ADN, ce qui va entraîner rapidement la cellule vers un phénotype néoplasique (WOOD *et al.* 2001).

Il existe cinq grands systèmes de réparation de l'ADN qui seront détaillés plus en avant :

- le système de réparation par excision de nucléotides (NER, *Nucleotide Excision Repair*),
- le système de réparation des mésappariements de l'ADN (MMR, *MisMatch Repair*),
- le système de réparation des cassures doubles-brins de l'ADN (DSBR, *Double Strand Break Repair*),
- le système de réparation par excision de base (BER, *Base Excision Repair*),
- le système de réparation directe.

Système de réparations	Gènes principaux
Réparation par excision de nucléotides (NER, <i>Nucleotide Excision Repair</i>)	<i>ERCC1, CBS, RAD23B, CCNH, XPA, XPB, XPC, XPD, XPF, XPG, RPA, LIG1</i>
Réparation des mésappariements de l'ADN (MMR, <i>MisMatch Repair</i>)	<i>MLH1, MLH3, MSH2, MSH3, MSH6, PMS1, PMS2, EXO1</i>
Réparation des cassures doubles-brins de l'ADN (DSBR, <i>Double Strand Break Repair</i>)	<i>XRCC3, XRCC2, XRCC9, NBS1</i>
Réparation par excision de base (BER, <i>Base Excision Repair</i>)	<i>XRCC1, OGG1, MUTYH, MTH1, LIG3, APEX, POLB, PCNA</i>
Réparation directe	<i>MGMT</i>

Tableau 2 : Principaux gènes intervenant dans les différents systèmes de réparation de l'ADN (NACCARATI *et al.* 2007).

II.2.4. Les télomérases

Les télomérases sont responsables du pouvoir prolifératif indéfini des cellules tumorales du fait de leur surexpression, retrouvée dans 85% des cancers humains.

Les télomères sont des structures de chromatine spéciales situées en partie terminale des chromosomes linéaires, qui protègent ces régions de la dégradation et de la recombinaison. Ils sont composés de répétitions en tandem (TTAGGG) et sont liés à des protéines spécifiques (BLASCO 2005). La régulation des télomères est fortement associée à la régulation de la sénescence.

Dans les cellules « normales », les télomères se raccourcissent progressivement au cours des divisions cellulaires. Ce phénomène serait lié à l'incapacité des ADN polymérases à répliquer les extrémités des chromosomes. Dans les cellules cancéreuses, les télomères gardent une longueur

stable. Ces cellules ont donc acquis la capacité de restaurer et de maintenir la stabilité des séquences télomériques grâce aux télomérases (BRYAN and CECH 1999).

II.2.5. Les petits fragments d'ARN

Ils vont intervenir dans la régulation de l'expression génique au niveau de la transcription et après la transcription, c'est-à-dire lors de la synthèse des protéines. La spécificité de cette répression provient de petits segments d'ARN d'une vingtaine de nucléotides, les « siARN » et les « microARN » (miARN), considérés jusqu'à tout récemment comme des produits de dégradation métabolique. Ces entités, en induisant ou non la transcription de protéines, vont permettre à la cellule de s'adapter finement, notamment :

- au contexte génétique, en inactivant ou en surexprimant des gènes ;
- à l'environnement extérieur (nutriments, oxygénation) ou tissulaire, au stress ;
- au stade du développement du cycle cellulaire.

Les miARN sont issus de gènes qui ne codent pas pour des protéines, mais seulement pour des ARN simples-brins longs d'environ 21 à 24 nucléotides. Ils sont complémentaires des ARN messagers (ARNm). Lorsqu'ils se fixent à eux, ils vont soit les dégrader, soit empêcher leur traduction en protéines. Plusieurs centaines de microARN ont été découverts au cours de ces dernières années et certaines équipes ont montré notamment leur implication dans la cancérogénèse colorectale (SLABY *et al.* 2007).

De la même manière que les miARN, les petits ARN interférants (siARN, *small interfering RNA*) sont de petits ARN doubles-brins qui remplissent de nombreuses fonctions, dont celle d'inhibition post-transcriptionnelle des gènes.

II.3. Génétique des cancers colorectaux

II.3.1. Les différentes voies altérées dans les cancers colorectaux

II.3.1.1. La voie WNT/ β -caténine

La perte de la fonction d'APC dérégule la voie de signalisation WNT/ β -caténine qui est une des principales étapes limitantes de l'initiation des CCR (FODDE *et al.* 2001). La protéine principale de cette voie est la β -caténine, codée par le gène *CTNNB1*, qui a une localisation intracellulaire variable dans la membrane, le cytoplasme et le noyau des cellules épithéliales. Le taux de cette protéine est régulé par un complexe multi-protéique appelé « complexe de destruction » et composé d'APC, GSK3 β , de la kinase CKI et des protéines AXIN1 et AXIN2 (conductine). En l'absence du ligand WNT, la destruction du complexe déclenche la phosphorylation de la β -caténine par la GSK3 β et la CKI, ce qui entraîne la dégradation protéolytique. En présence de la glycoprotéine extracellulaire WNT, la formation du « complexe de destruction » est inhibée et la β -caténine échappe à la dégradation. La β -caténine interagit dans le noyau avec différents membres de la famille TCF/LEF qui vont moduler la transcription de différents gènes de prolifération (POLAKIS 1999).

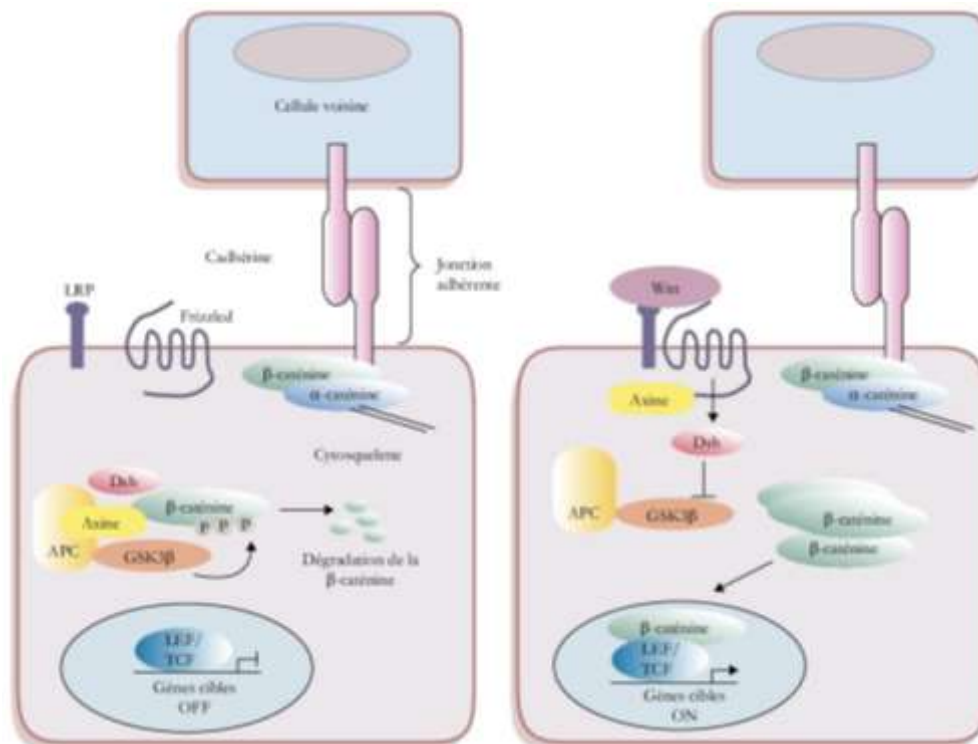


Figure 9 : Régulation de la β -caténine dans la voie Wnt. En l'absence de Wnt (schéma de gauche), la β -caténine est complexée à l'axine. APC et GSK3 phosphorylent la β -caténine avant qu'elle soit dégradée par le protéasome. En présence de ligand Wnt (schéma à droite), la β -caténine est découplée de ce complexe et est transloquée dans le noyau

où elle va s'associer aux facteurs TCF-LEF pour activer la transcription génique. La β -caténine existe aussi sous une forme liée à la cadhérine qui intervient dans l'adhésion cellule-cellule (REYA and CLEVERS 2005).

La grande majorité des cancers colorectaux sporadiques sont caractérisés par une activation constitutive de la voie WNT due soit aux mutations d'*APC*, soit à celles de *CTNNB1*. Les mutations tronquantes d'*APC* suppriment les motifs fonctionnels qui régulent le turnover de *CTNNB1*. L'activation constitutive de WNT/ β -caténine va activer le renouvellement des cellules et diminuer le potentiel de différenciation des cellules épithéliales intestinales, menant à la formation et à la progression des cryptes dysplasiques aberrantes pour donner des adénomes (REYA and CLEVERS 2005). Alternativement, les tumeurs colorectales avec un gène *APC* intact existent : elles présentent des mutations activatrices de la β -caténine qui affectent la fonctionnalité des sites de phosphorylation et qui vont ainsi induire une résistance à la dégradation protéolytique (MORIN *et al.* 1997; RUBINFELD *et al.* 1996; SPARKS *et al.* 1998). Outre les mutations de *CTNNB1* et d'*APC*, on retrouve des mutations dans d'autres membres de la famille WNT, comme le gène *AXIN2* (LAMMI *et al.* 2004).

Les mutations de la β -caténine sont rarement trouvées dans les cancers colorectaux possédant des instabilités microsatellitaires (SALAHSHOR *et al.* 1999).

II.3.1.2. La voie de RAS/RAF/MAPK

Les mutations activatrices de *KRAS2* sont retrouvées dans plus de 50% des cancers colorectaux (ANDREYEV *et al.* 2001; BOS *et al.* 1987; FORRESTER *et al.* 1987; VOGELSTEIN *et al.* 1988). En l'absence de mutations de *KRAS2*, 20% des cancers colorectaux contiennent des mutations activatrices de *BRAF* (DAVIES *et al.* 2002; RAJAGOPALAN *et al.* 2002). Ces deux gènes sont des composants essentiels de la voie de signalisation RAS/RAF/MAPK, une cascade de signalisation contrôlée par le GDP/GTP qui est connue pour moduler et contrôler la croissance cellulaire.

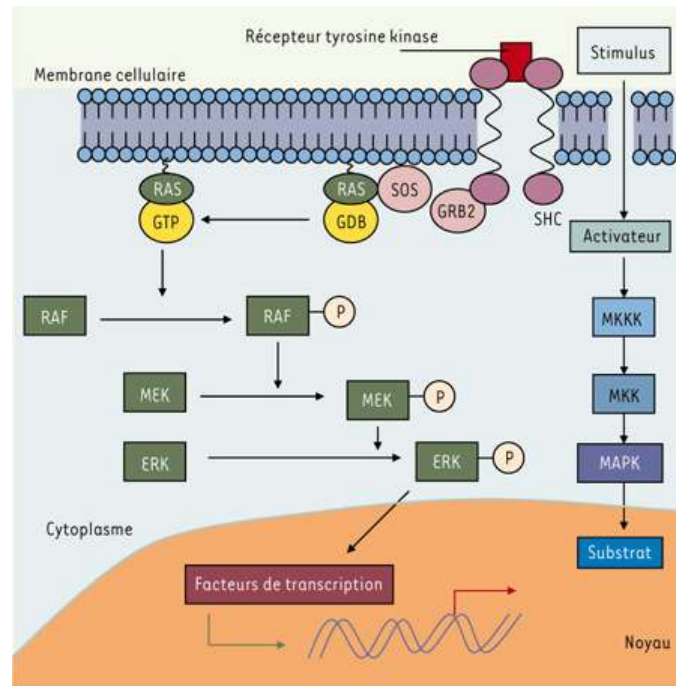


Figure 10 : Voie de signalisation des MAP-kinases, avec les voies RAS-RAF-MEK-ERK. (MAPK, *mitogen-activated protein kinases* ; MKK(K), *mitogen-activated protein kinase (kinase)* ; RTK, *receptor tyrosine kinase* ; SH2, *sequence homology 2*) (Extrait de Dalle *et al.* (2006)).

II.3.1.3. La voie TGF β

La progression tumorale des adénomes tardifs et des carcinomes précoces est accompagnée par des altérations de la voie de signalisation du TGF β (*Transforming Growth Factor β*). La résistance aux effets antiprolifératifs exercée par le TGF β est caractéristique de nombreuses cellules tumorales. La grande majorité des cancers colorectaux portent au moins une mutation inactivatrice dans les gènes de la voie TGF β tels que *TGF β 2*, *SMAD2*, *SMAD4* (EPPERT *et al.* 1996; TAKAGI *et al.* 1996; THIAGALINGAM *et al.* 1996), qui affectent l'angiogenèse, la prolifération et la différenciation cellulaire (BLOBE *et al.* 2000).

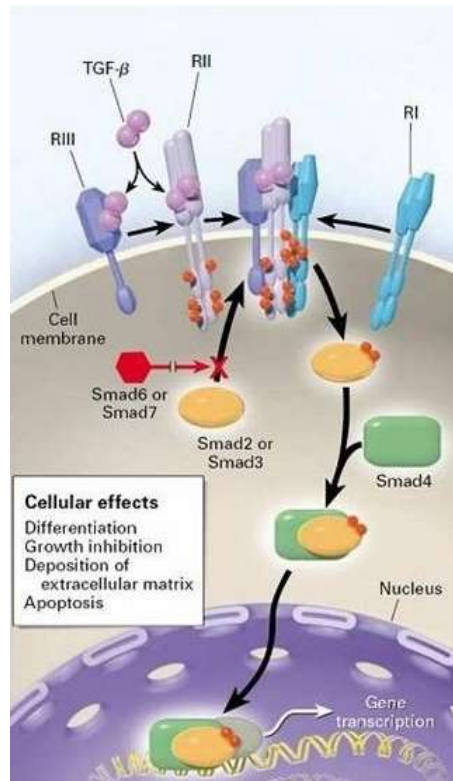


Figure 11 : Représentation simplifiée de la voie TGF- β faisant intervenir les protéines SMAD qui vont permettre de réguler la transcription (Extrait de Bolbe *et al.* (2000)).

II.3.1.4. La voie p53

L'inactivation biallélique du gène suppresseur de tumeur *p53* est retrouvée dans 45% des cancers colorectaux (BAKER *et al.* 1989; IACOPETTA 2003; PURDIE *et al.* 1991). La protéine p53 inhibe la croissance et stimule la mort cellulaire induite par le stress. Cette inactivation a des conséquences dramatiques pour le maintien de l'intégrité du génome dans les cellules intestinales tumorales.

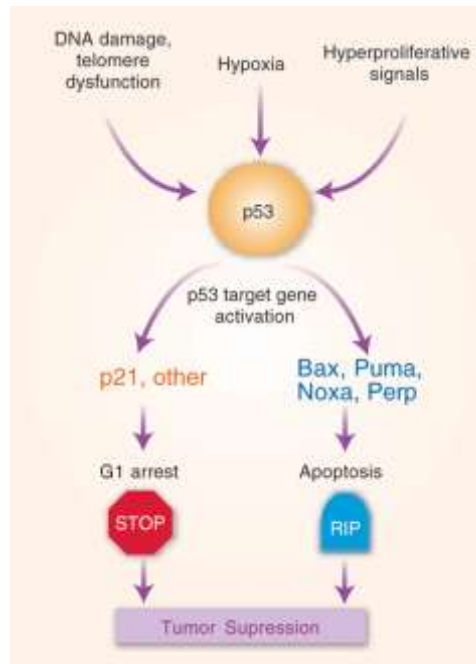


Figure 12 : Voie de p53 montrant les différents mécanismes pouvant affecter son expression, les gènes que p53 contrôle, et les processus sur lesquels ils vont jouer (Extrait de Attardi et DePinho (2004)).

Toutes ces voies, WNT/ β -caténine, RAS/RAF/MAPK, TGF β et p53, interviennent aussi bien dans les cancers colorectaux héréditaires que sporadiques ainsi que dans les différentes formes d'instabilités rencontrées qui caractérisent ces cancers.

II.3.2. Les différentes formes d'instabilités rencontrées

Les cancers colorectaux peuvent être classés par l'observation dans 15% des cas d'une instabilité génomique, ou pour les 85% restants par une instabilité chromosomique.

II.3.2.1. Les instabilités génomiques

Dans ce groupe de cancers présentant des instabilités génomiques, les cancers sont diploïdes et ne présentent pas de pertes alléliques (REMKOS *et al.* 1995; THIBODEAU *et al.* 1993). Ce groupe de cancers possède des instabilités microsatellitaires conférant aux tumeurs le statut MSI (microsatellite instable). Ces instabilités sont dues à des mutations dans les gènes de réparation des mésappariements de l'ADN (gènes MMR), ce qui inclut les gènes *MLH1*, *MSH2*, *PMS2*, *MLH3*, *MSH6* (NICOLAIDES *et al.* 1994; PELTOMAKI *et al.* 1993). Les protéines codées par les gènes MMR altérés ne peuvent plus réparer les mésappariements qui surviennent lors de la réplication. Cette absence de fonction, ou cette diminution de fonction de la protéine tronquée, sera transmise aux cellules-filles.

Dix à 20% des cancers sporadiques présentent une instabilité microsatellitaire (GRADY and CARETHERS 2008; ILYAS *et al.* 1999) due à une inactivation somatique biallélique d'un des gènes MMR (CUNNINGHAM *et al.* 1998; HERMAN *et al.* 1998; KANE *et al.* 1997; THIBODEAU *et al.* 1996; VEIGL *et al.*

1998), et 50 à 70% des tumeurs du syndrome HNPCC sont la conséquence de mutations germinales dominantes de *MLH1* ou *MSH2* (AALTONEN *et al.* 1993).

Les instabilités microsatellites des cancers sporadiques sont dues à une hyper-méthylation de la région promotrice de *MLH1* ou à une mutation ponctuelle d'un des gènes du système de réparation (FISHEL *et al.* 1994; PELTOMAKI *et al.* 1993). Une inactivation somatique de *MSH2* est un évènement rare dans les cancers sporadiques possédant des instabilités microsatellites (THIBODEAU *et al.* 1996), alors que les familles HNPCC présentent plus de 90% de leurs mutations détectables dans les gènes *MLH1* ou *MSH2* (PELTOMAKI 1997). La perte de fonction des gènes MMR va entraîner des mutations « secondaires » dans certains gènes comportant dans leur séquence codante des microsatellites (ESHLEMAN *et al.* 1995; JOHANNSDOTTIR *et al.* 2000). Ces mutations de type insertion/délétion vont entraîner un décalage du cadre de lecture et la synthèse d'une protéine tronquée non fonctionnelle.

La Figure 13 résume la chronologie des différentes altérations génétiques intervenant dans les cancers avec instabilités microsatellites.

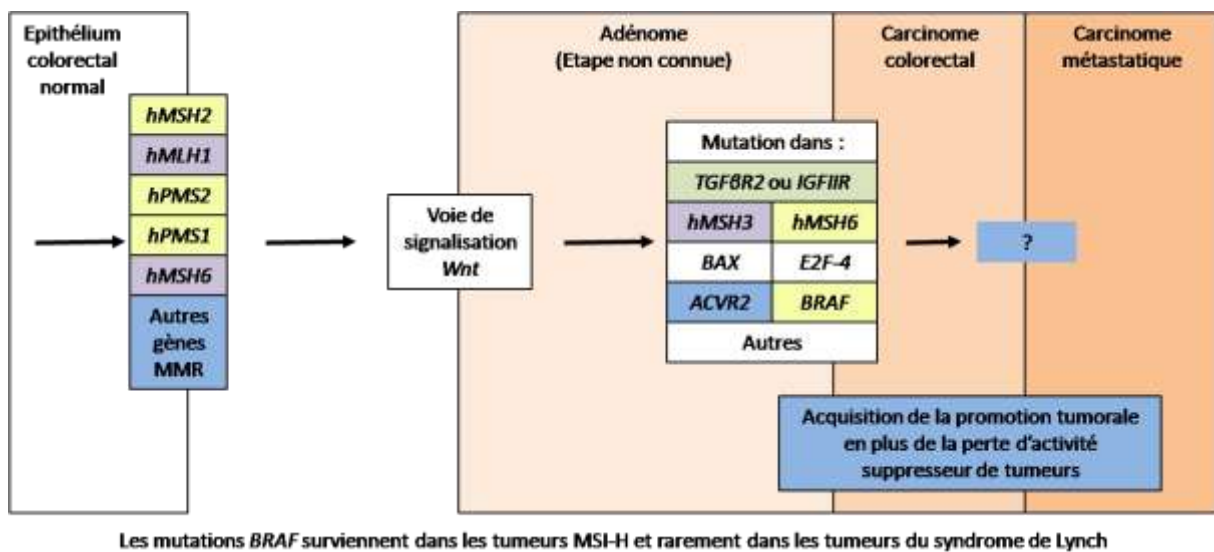


Figure 13 : Progression des tumeurs colorectales avec des instabilités microsatellites (MSI). Les tumeurs MSI, aussi bien sporadiques que celles des patients atteints du syndrome de Lynch, perdent leur fonction MMR au début de la progression polype→cancer. Les tumeurs sporadiques perdent pour la plupart leur fonction MMR par méthylation du promoteur de *MLH1*, alors que les patients atteints du syndrome de Lynch ont une mutation germinale dans l'un des gènes MMR. Il est généralement admis que la signalisation Wnt, le gardien de la néoplasie colorectale, est affectée ; mais les mécanismes ne sont pas clairs et incluent des mutations de *CTNNB1*, le gène codant pour la β-caténine, dans certains cas de syndrome de Lynch. En absence de réparations MMR, un phénotype hypermutable se développe dans lequel de multiples mutations surviennent dans l'ADN. Bien que la plupart des mutations se produisent dans les séquences non-codantes (microsatellites introniques), certains gènes tels que *TGFBR2*, *ACVR2*, *BAX*, *MSH3*, *MSH6*, et d'autres, qui ont des microsatellites dans leur séquence codante, seront atteints en absence de réparation MMR. Ces mutations favorisent la progression de la tumeur. Les mutations de *BRAF* sont principalement trouvées dans les tumeurs sporadiques avec une instabilité microsatellitaire, et non dans les tumeurs des patients atteints de syndrome de Lynch (Figure extraite de Grady et Carethers (2008)).

Les cancers MSI familiaux et sporadiques sont associés à une localisation proximale de cancer colorectal (THIBODEAU *et al.* 1993). Ces cancers sont généralement peu différenciés, avec une tendance à être multiples et à présenter des mucosécrétions abondantes. Malgré la multiplicité et la faible différenciation de ces cancers, leur pronostic est meilleur que pour les tumeurs sans instabilité génomique (GAFA *et al.* 2000; THIBODEAU *et al.* 1993). L'instabilité génétique des cellules possédant une instabilité microsatellitaire les rendrait plus sensibles à l'apoptose du fait de l'accumulation de mutations dans des gènes requis pour la croissance cellulaire (SHIBATA *et al.* 1994).

II.3.2.1.1. Les gènes cibles des instabilités microsatellitaires

a) Le gène *TGF β 2*

Ce gène, dont la séquence codante comporte une répétition poly-A10, est affecté dans 70 à 90% des cas de cancers colorectaux de type MSI. Cette instabilité va entraîner un dysfonctionnement du récepteur au TGF dans le contrôle de la prolifération cellulaire et de la progression tumorale (MARKOWITZ *et al.* 1995; WANG *et al.* 1995b).

b) Le gène *Bax*

Ce gène est un membre de la famille de gènes *BCL2*. Il est impliqué dans l'apoptose par la formation d'un hétérodimère de BAX et de BCL2. *Bax* comporte une séquence intragénique poly-G8, qui est retrouvée instable dans 50% des cancers colorectaux MSI (ABDEL-RAHMAN *et al.* 1999; RAMPINO *et al.* 1997; YAMAMOTO *et al.* 1997).

Les mutations de *Bax* et de *TGF β 2* apparaissent tardivement dans les dysplasies de haut grade, qui définissent la malignité (GRADY *et al.* 1998; YAGI *et al.* 1998).

c) Les gènes *MSH3* et *MSH6*

Ils comportent dans leur séquence codante des répétitions poly-A8 pour *MSH3* et poly-C6 pour *MSH6* qui sont retrouvées instables dans environ un tiers des cancers colorectaux MSI (MALKHOSYAN *et al.* 1996).

d) Les autres gènes présentant des instabilités microsatellitaires

Environ 15% des tumeurs MSI portent une mutation d'*IGF2R* dans leur séquence polyguanine qui ne peut plus interagir avec le régulateur de la prolifération et de la différenciation *TGF β 1*. Ce gène suppresseur de tumeur régule négativement la croissance cellulaire (SOUZA *et al.* 1996). *ACVR2* a également été montré muté dans la séquence polyadénine de son exon 10 dans les cancers du côlon MSI avec perte de la protéine ACVR2 (JUNG *et al.* 2004). On peut citer également les gènes *CDX-2* (WICKING *et al.* 1998), *CASP5* et *E2F4* (SOUZA *et al.* 1997), *SEC63*, *AIM2*, *PA2G4/EBP1*.

e) Autres gènes mis en cause

Le mécanisme mettant en cause le gène *BRAF* n'est pas connu, car celui-ci ne contient pas de microsatellites dans sa séquence codante. Ce gène code pour un composant en aval de *KRAS* dans la voie RAS/RAF/MAPK, qui est aussi souvent montré muté dans les tumeurs sporadiques MSI. Les mutations de *BRAF* le plus souvent rencontrées sont des mutations faux-sens (dont p.Val600Glu est la plus fréquente) qui vont activer constitutionnellement la protéine (DENG *et al.* 2004).

II.3.2.2. Les instabilités chromosomiques

Elles sont retrouvées dans 80 à 85% des cancers colorectaux. Les 2/3 sont localisées dans le côlon distal. Plusieurs formes d'instabilités chromosomiques sont observées (AGUILERA and GOMEZ-GONZALEZ 2008; LENGAUER *et al.* 1998) :

- des altérations dans le nombre de chromosomes (aneuploïdies*, appelées aussi CIN pour *Chromosome Instability*). L'aneuploïdie est un processus pathogène fondamental de la tumorigenèse qui est bien toléré.
- des réarrangements chromosomiques avec perte de matériel.
- des amplifications de gènes qui se produisent rarement dans les cancers du côlon, mais qui apparaissent jouer un rôle dans la transition des étapes de certains cancers colorectaux (BARDELLI *et al.* 2003; SAHA *et al.* 2001).

L'instabilité chromosomique intervient précocement dans la formation de la tumeur et en accélère la progression (Figure 14).

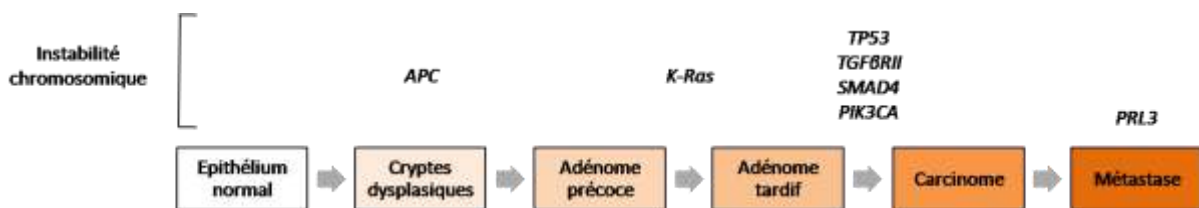


Figure 14 : Progression des tumeurs colorectales avec des instabilités chromosomiques (CIN). La caractéristique des instabilités chromosomiques est l'aneuploïdie. L'initiation de la néoplasie dans cette voie se produit avec l'interruption des composants de la voie de signalisation Wnt, y compris une mutation somatique dans un allèle et la perte d'hétérozygotie du deuxième allèle normal du gène *APC*. La progression est ensuite conduite par les vagues successives d'expansion clonale de cellules qui acquièrent les caractéristiques suivantes : une croissance accrue, des mutations activatrices du proto-oncogène *KRAS*, et des mutations de *p53* avec une perte d'hétérozygotie de l'allèle normal restant de *p53* induisant la formation du cancer. Dans certaines tumeurs du côlon, l'activation par des mutations de *PIK3CA* survient tard dans la séquence adénome→carcinome. Plusieurs molécules de signalisation TGF- β sont aussi touchées dans la progression des CIN, notamment les mutations du domaine kinase de *TGFBR2* et la perte d'hétérozygotie du bras long du chromosome 18 (18q) incluant les gènes *SMAD4* et *SMAD2*. Plusieurs gènes, dont *PRL3*, sont soupçonnés d'être impliqués dans les métastases et dans l'amplification génique (Adapté de Grady et Carethers (2008)).

* Caractéristique d'une cellule ne possédant pas un nombre normal de chromosomes

En moyenne, 20% des chromosomes impliqués ont subi des délétions (VOGELSTEIN *et al.* 1989). Les plus fréquentes (75%) sont localisées sur le bras court du chromosome 17 (17p) et le bras long du chromosome 18 (18q), viennent ensuite des délétions du bras long du chromosome 5, puis celles du bras court du chromosome 8 et du bras long du chromosome 22 (35%) (DELATTRE *et al.* 1989; VOGELSTEIN *et al.* 1989). Ces régions renferment des gènes majeurs, tels que le gène *p53* sur le chromosome 17p, les gènes *DCC* et *SMAD4/DPC4* et *SMAD2* sur le chromosome 18q et les gènes *APC* et *MCC* sur le chromosome 5q.

Les mécanismes responsables des instabilités chromosomiques sont identifiés dans certains cancers colorectaux. Les mutations de *p53* ont été initialement suspectées de causer les instabilités génomiques, mais il apparaît que seules des mutations spécifiques de *p53* interviennent.

II.3.2.2.1. Les gènes atteints dans les instabilités chromosomiques

a) Le gène *p53*

Le gène *p53* code pour la protéine du même nom qui a un rôle charnière dans le développement de tumeur, l'arrêt du cycle cellulaire, la différenciation et l'apoptose. L'altération de ce gène constitue l'évènement génétique le plus fréquent dans les cancers. Ce gène est en effet perdu ou muté dans environ la moitié des cancers humains.

L'un des principaux rôles de la protéine *p53* est de contrôler la réponse cellulaire aux dommages de l'ADN ainsi que la réponse aux signaux de stress comme l'hypoxie, la privation en facteurs de croissance, l'hyperthermie, l'activation d'oncogènes.

Les signaux de stress induisent l'accumulation de *p53* conduisant soit à un arrêt transitoire dans la progression du cycle cellulaire, permettant au processus de réparation d'intervenir, soit à l'induction de l'apoptose permettant d'éliminer les cellules altérées non réparées (Figure 15).

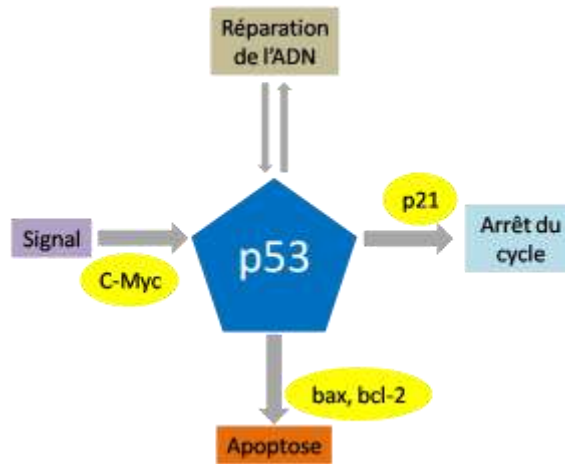


Figure 15 : p53, une molécule carrefour dans la réponse aux dommages à l'ADN*.

La protéine p53 activée régule la transcription de nombreux gènes impliqués dans l'apoptose et/ou le contrôle du cycle cellulaire. Elle peut ainsi induire l'arrêt du cycle cellulaire en augmentant la transcription de p21, un inhibiteur de kinases dépendantes des cyclines (cdk). Les cdk permettent la progression du cycle cellulaire par phosphorylation de substrats spécifiques comme la protéine RB. L'augmentation de p21 empêche cette phosphorylation et bloque la progression du cycle en phase G1/S (KIRSCH and KASTAN 1998).

L'orientation des cellules vers l'apoptose nécessite l'activation d'autres gènes tels que *bax*, *bcl-2*. Des mutations somatiques de *bax* ont été identifiées dans des tumeurs du côlon.

Les altérations de *p53* jouent un rôle important dans le développement des cancers. Les cellules tumorales ayant une mutation de *p53* ne sont plus capables d'assurer le maintien de l'intégrité génétique et accumulent les mutations permettant l'émergence de clones cellulaires de malignité accrue (KIRSCH and KASTAN 1998; SOUSSI 2000). Les anomalies de p53 sont également responsables de l'apparition de résistances cellulaires à la radiothérapie et à la chimiothérapie et constituent donc un marqueur de mauvais pronostic.

Des mutations constitutives (héritées) de ce gène constituent la base moléculaire du syndrome de Li-Fraumeni, syndrome prédisposant à un très large spectre de tumeurs (incluant en particulier des sarcomes, des ostéosarcomes, des tumeurs du système nerveux central, des cancers du sein,...).

b) Le gène *DCC*

Le gène *DCC* (*Deleted in Colon Cancer*) code pour une protéine transmembranaire de la super famille des immunoglobulines. Celle-ci est impliquée dans l'adhésion entre les cellules et la matrice extracellulaire. L'absence ou l'inactivation de la protéine DCC par mutation ponctuelle ou par

* <http://www.inrp.fr/Acces/biotic/gpe/dossiers/p53/html/mutasoma.htm>

délétion induirait une perte des propriétés adhésives des cellules tumorales, favorisant ainsi leur croissance et leur dissémination (métastase) (SHIBATA *et al.* 1996).

c) Les gènes *SMAD*

Ils codent pour des protéines impliquées dans la voie intracellulaire de transduction du signal TGF β . Des mutations inactivatrices des gènes *SMAD2*, *SMAD3* et *SMAD4* sont retrouvées dans 15% des cancers colorectaux (DERYNCK and FENG 1997; HELDIN *et al.* 1997; MASSAGUE 1998; SCHWARTZ-WALDHOFF *et al.* 1999; THIAGALINGAM *et al.* 1996). Les mutations de *SMAD4* joueraient un rôle dans la progression maligne des tumeurs colorectales (TAKAKU *et al.* 1998).

D'autres gènes sont également concernés par ce type d'instabilité, comme par exemple les gènes *MCC* (*Mutated in Colon Cancer*) et *APC*, qui sont perdus dans environ 20 à 40% des cas de CCR (MARTINEZ-LOPEZ *et al.* 1998).

II.3.2.2.2. Les mécanismes impliqués dans les instabilités chromosomiques

Les cancers aneuploïdes tendent à montrer des anomalies cytologiques durant la mitose incluant : un nombre anormal de centrosomes, des fuseaux multipolaires et des chromosomes isolés impliquant la dérégulation des processus intervenant dans les points de contrôle du fuseau mitotique (JALLEPALLI and LENGAUER 2001; SATO *et al.* 2001; SAUNDERS *et al.* 2000; YUEN *et al.* 2007).

La Figure 16 présente les différents points de contrôle du cycle cellulaire et les gènes intervenant dans les instabilités chromosomiques.

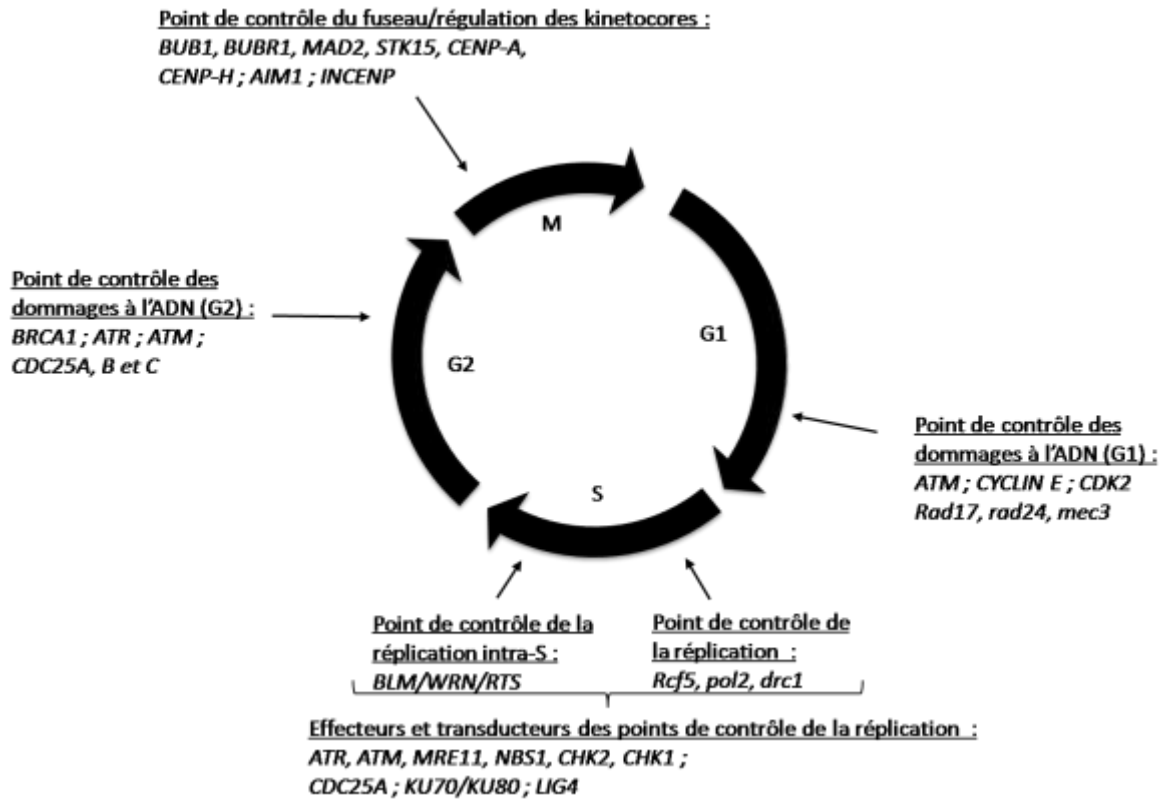


Figure 16 : Contrôle du cycle cellulaire. Les dommages de l'ADN, la réplication de l'ADN et les points de contrôle du fuseau mitotique sont indiqués avec la liste des gènes humains et des gènes de *S. cerevisiae* qui interviennent à chacun des points de contrôle (Extrait de Grady et Carethers (2008)).

Comme nous avons pu le voir, les mutations des gènes *APC*, *KRAS*, *PIK3CA*, *SMAD4*, *p53* et des autres gènes CIN sont des facteurs contribuant aux instabilités chromosomiques dans les cancers colorectaux (CHITTENDEN *et al.* 2008; WOOD *et al.* 2007).

a) Les points de contrôle du fuseau mitotique

Les gènes intervenant dans ce mécanisme codent pour des protéines qui contrôlent la séparation des chromatides lors de la transition de la métaphase à l'anaphase. Ces protéines peuvent être responsables d'aneuploïdies. Les gènes *MAD* et *BUB* sont hautement conservés ; leur activation cause l'arrêt du cycle cellulaire à travers l'inhibition du complexe de promotion de l'anaphase (*APC/C*). D'autres gènes faisant partie de cette catégorie ont été montrés comme intervenant dans les instabilités chromosomiques ; on peut citer : *PTTG*, *CDC4*, *hROD*, *hZW10*, *hZWILCH*, *DING* qui sont mutés dans les CCR (WANG *et al.* 2008; YUEN *et al.* 2007), et *CENP-A*, *CENP-H*, *HEC1*, *Aurora-B* (*AIM1*) et *INCENP* qui ont été retrouvés surexprimés dans les CCR (TATSUKA *et al.* 1998; TOMONAGA *et al.* 2005; TOMONAGA *et al.* 2003; YUEN *et al.* 2007).

b) Régulation des centrosomes

Le mécanisme de l'aneuploïdie pourrait faire intervenir un nombre anormal ou une fonction anormale des centrosomes (DOXSEY 1998). Ces derniers jouent un rôle central dans la ségrégation des

chromosomes durant la mitose. L'amplification des centrosomes observés dans les cancers augmenterait la fréquence des mitoses anormales et une mauvaise ségrégation des chromosomes (FUKASAWA 2007; LINGLE *et al.* 2002; PIHAN *et al.* 2001; ZHOU *et al.* 1999). Les gènes en 20q13 sont fréquemment amplifiés dans les cancers du côlon, notamment *STK15* (SEN *et al.* 1997). De la même manière les gènes *PLK2* et *PLK4* sont surexprimés et impliqués dans l'aneuploïdie et la fonction anormale des centrosomes (MACMILLAN *et al.* 2001). Les gènes *CDK4/CDK6* et *cycline-D* contribuent à la tumorigenèse en favorisant la génération des centrosomes amplifiés (ORTEGA *et al.* 2002).

c) Les points de contrôle de l'ADN

Une catégorie de gènes sont connus pour intervenir sur les points de contrôle qui régulent le maintien de la fidélité de l'ADN en réponse au stress génotoxique. Ces points de contrôle des dommages à l'ADN sont traditionnellement classés en deux catégories : ceux qui reconnaissent et répondent aux dommages de l'ADN dus aux agents génotoxiques et ceux qui contrôlent la fidélité de l'ADN répliqué (NYBERG *et al.* 2002). Différents gènes interviennent dans ce système : *p53*, *ATM*, *ATR*, *BRCA1*, *BRCA2* qui interagissent avec *Rad51* et *Rad17*.

Ces gènes interviennent durant la phase S du cycle cellulaire. Le gène *p53* joue notamment un rôle dans le développement des instabilités chromosomiques. *MER11* est retrouvé fréquemment muté dans les cancers du côlon MSI, même si sa fonction n'est pas clairement définie.

d) Les protéines du cycle cellulaire

Des mutations somatiques de *CDC4* sont retrouvées dans 11,5% des cancers du côlon (RAJAGOPALAN *et al.* 2004). *CDC4* est un effecteur en aval de *p53* qui permet la régulation de la stabilité du génome. Cette stabilisation dépend de la régulation par la cycline-E, mais également de l'effet de Notch et/ou c-Jun (PEREZ-LOSADA *et al.* 2005).

e) Les autres facteurs qui participent au développement des instabilités chromosomiques dans les cancers du côlon et du rectum

Les protéines contrôlant la réplication ont également été proposées pour jouer un rôle dans l'induction de l'instabilité chromosomique dans le cancer du côlon, par le biais d'effets sur la régulation du cycle cellulaire et la réplication de l'ADN. Ces mécanismes comprennent la résistance au TGF β , l'infection par le virus JC et la surexpression de *C-MYC* (GLICK *et al.* 1996; SHADAN *et al.* 2002).

Par exemple, la surexpression de *C-MYC* est impliquée dans l'atténuation des dommages à l'ADN lors de la phase G1/S et contribue à la tumorigenèse au travers de l'induction de l'aneuploïdie (FELSHER and BISHOP 1999).

f) Les pertes alléliques et les gènes associés

Le déséquilibre allélique s'observe dès les stades précoces de la carcinogenèse colorectale : des pertes chromosomiques sont ainsi détectables dès le stade adénome précoce (<3mm) dans plus de 90% des cas (SHIH *et al.* 2001). Les cancers LOH suivent une chronologie bien établie : la perte en 5q survient abruptement à la transition entre l'épithélium colique normal et l'adénome ; elle est suivie par la perte du chromosome 17p à la transition entre l'adénome et l'adénocarcinome, indiquant un rôle essentiel de ces pertes alléliques dans la progression tumorale (BOLAND *et al.* 1995). Le déséquilibre allélique croît ensuite sensiblement avec le stade tumoral, aussi bien qualitativement (perte précoce de 5q et plus tardive de 17p et 18q) que quantitativement (nombre de bras chromosomiques perdus) (CHOI *et al.* 2002). Des délétions des bras 8q et 22q sont également observées.

II.3.2.3. Les altérations épigénétiques et les instabilités épigénétiques

Les instabilités épigénétiques sont communes aux formes familiales et sporadiques de cancers colorectaux. La méthylation de l'ADN est présente dans la majorité du génome et est maintenue à un niveau relativement stable qui a été établi durant le développement (KONDO and ISSA 2004). Les zones concernées, appelées îlots CpG, sont présentes en 5' de 50 à 60% des gènes et sont normalement maintenues dans un état non-méthylé. Approximativement 70% de ces îlots portent des modifications épigénétiques. Dans les cancers, la plupart de ces îlots deviennent méthylés de façon aberrante et cette aberration peut être accompagnée par une répression de la transcription (BAYLIN and HERMAN 2000). La méthylation aberrante des gènes *MLH1*, *MGMT*, *HIC1* a été mise en évidence dans les CCR (CHEN *et al.* 2003; ESTELLER *et al.* 2001).

La Figure 17 présente les gènes touchés par la méthylation dans les îlots CpG.

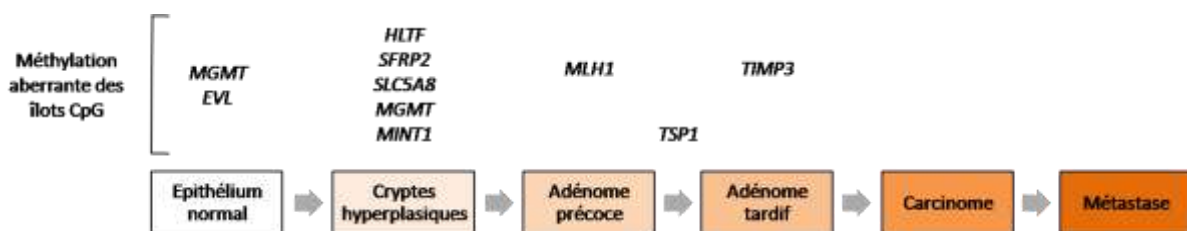


Figure 17 : Progression des tumeurs colorectales possédant un phénotype méthylé des îlots CpG (CIMP). La caractéristique de la voie CIMP est la méthylation anormale de plusieurs séquences de promoteurs de gènes suppresseurs de tumeur. Cette voie recouvre en partie les tumeurs sporadiques MSI, étant donnée l'hyperméthylation du promoteur de *MLH1* qui provoque la perte de sa fonction MMR. Bien que la voie CIMP ne soit pas encore bien comprise, Grady et Carethers (2008) proposent l'explication suivante pour la progression adénome→cancer. Les cryptes aberrantes hyperplasiques peuvent constituer la première lésion dans cette voie, à la place des cryptes aberrantes dysplasiques avec interruption de la signalisation (Chapitre I : II.3.2.2. Les instabilités chromosomiques). Les promoteurs méthylés de *MGMT*, *EVL*, *HLTF*, *SFRP2*, *SLCSA8* et *MINT1* se développent au cours de la phase d'initiation du développement de la tumeur colorectale. L'hyperméthylation du promoteur *MLH1* pourrait correspondre au

développement d'un adénome en dents de scie, avec la méthylation de *TSP1* et *TIMP3* qui contribue à la progression de la tumeur CIMP (GRADY and CARETHERS 2008).

La méthylation aberrante de *MLH1* est montrée dans plus de 80% des cancers colorectaux sporadiques MSI. L'expression et la fonction de *MLH1* sont restaurées lorsque la déméthylation des îlots CpG est réalisée dans les lignées cellulaires, ce qui indique que ce mécanisme intervient dans les cancers montrant cette instabilité (HERMAN *et al.* 1998; KANE *et al.* 1997; VEIGL *et al.* 1998). L'hyperméthylation a été mise en évidence pour d'autres gènes suppresseurs de tumeur, comme *CDKN2/p16*, *MGMT*, *p14*, *HLTF* (BAYLIN and HERMAN 2000; HERMAN *et al.* 1995; HERMAN *et al.* 1998; KANE *et al.* 1997; TOYOTA *et al.* 1999b; VEIGL *et al.* 1998). La méthylation de *CDKN2/p16* est retrouvée dans 40% des cancers colorectaux, mais également dans les adénomes colorectaux. La méthylation des gènes *HLTF*, *SLC5A8*, *MGMT*, *MINT1*, *MINT31* est retrouvée dans les stades précoces d'adénome (CHAN *et al.* 2002; MOINOVA *et al.* 2002). Les critères pour définir les instabilités dues aux méthylations (CIMP, *CpG Island Methylator Phenotype*) sont en constante évolution. Ces tumeurs pourraient être définies par une forte proportion de promoteurs méthylés de façon aberrante qui survient par un mécanisme unique comparé aux tumeurs non-CIMP (SAMOWITZ *et al.* 2005; TOYOTA *et al.* 1999a; TOYOTA *et al.* 1999b).

La méthylation de l'ADN peut agir directement sur la transcription, par inhibition directe entre les promoteurs méthylés et les facteurs de transcription comme AP-2, CREB, E2F, CBF et le facteur nucléaire kB (DENG *et al.* 2001; KONDO and ISSA 2004). La méthylation des îlots CpG peut également induire le silence transcriptionnel, en recrutant des protéines de liaison méthylées, telles que MeCP2, MBD2, MBD3, qui reconnaissent les séquences méthylées et recrutent les histones déacétylases. Ces dernières vont induire un changement dans la structure de la chromatine, ce qui va empêcher la fixation des facteurs de transcription sur les régions promotrices (DENG *et al.* 2001).

II.3.3. Prédisposition génétique aux cancers colorectaux

La prédisposition génétique aux cancers colorectaux est bien connue pour les formes héréditaires grâce à la découverte, dans plusieurs gènes, de mutations à forte pénétrance responsables de ces pathologies. La Figure 18 indique les gènes touchés par des mutations à forte pénétrance mises en cause dans la polypose adénomateuse familiale (PAF), le syndrome de Lynch, la polypose hamartomateuse familiale, la polypose associée à MYH (MAP, *MUTYH-Associated Polyposis*). Cependant les gènes responsables de certaines formes familiales restent à identifier. Les cancers sporadiques sont, quant à eux, dus à l'interaction entre des gènes de faible prédisposition, dont certains sont connus et d'autres restent à découvrir, et des facteurs environnementaux. Dans cette forme de cancer, plusieurs mutations de faible pénétrance peuvent contribuer de manière additive à la carcinogenèse colorectale, impliquant des interactions entre les gènes et des facteurs

environnementaux. La barrière entre cancers sporadiques et familiaux, ainsi qu'entre facteurs génétiques et environnementaux, reste cependant floue et doit être prise comme un continuum de risques de développer un cancer colorectal (DE LA CHAPELLE 2004).

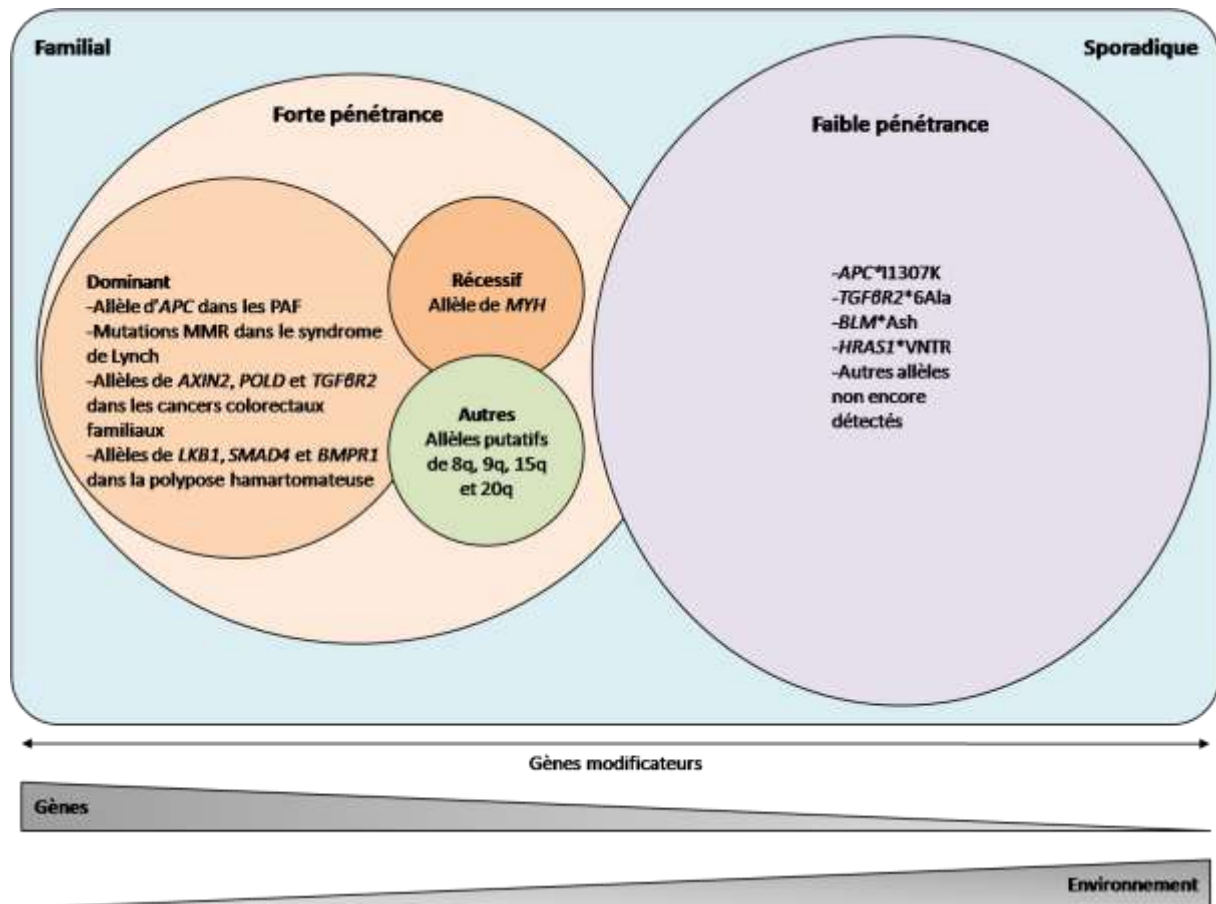


Figure 18 : Vue d'ensemble de la contribution génétique et environnementale dans les cancers colorectaux. L'influence de l'environnement est inversement proportionnelle à la force de la position génétique. Du côté des formes familiales, sont recensés les différents gènes de forte pénétrance connus comme jouant un rôle dans ces pathologies. Dans la catégorie « autres » sont indiqués les loci supposés retrouvés dans les cancers colorectaux qui ont une forte composante héréditaire, avec peu d'influence environnementale mais dont le gène causal reste à identifier. Du côté des formes sporadiques sont recensés les gènes et les variants connus comme prédisposant à la pathologie (DE LA CHAPELLE 2004).

A partir des critères génétiques, on distingue différentes formes de cancers colorectaux héréditaires ou sporadiques. Les formes héréditaires regroupent les cancers familiaux de base moléculaire encore inconnue et tous les syndromes dont les gènes de prédisposition sont déjà connus (HNPCC, PAF,...) (Figure 19).

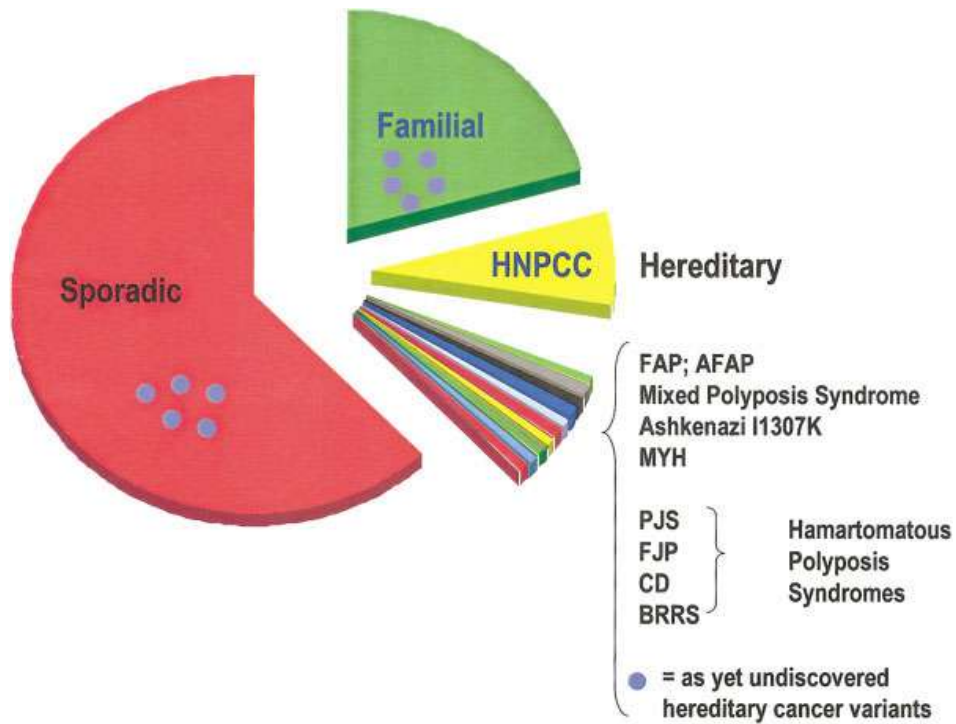


Figure 19 : Répartition des différentes formes de cancers colorectaux (LYNCH *et al.* 2004a). (HNPCC, *hereditary non polyposis colorectal cancer* ; FAP, *Familial adenomatous polyposis* ; AFAP, *attenuated familial adenomatous polyposis* ; PJS, *Peutz-Jeghers Syndrome* ; FJP, *familial juvenile polyposis* ; CD, *Cowden disease* ; BRRS, *Bannayan-Ruvalcaba-Riley syndrome* ; •, variants encore inconnus de cancer héréditaire).

II.4. Les formes familiales

Les formes familiales représentent environ 35% de la totalité des cancers colorectaux. On distingue cinq grands groupes syndromiques familiaux, incluant chacun une ou plusieurs entités cliniques dont la cause génétique est le plus souvent connue. L'identification de mutations dans les gènes incriminés chez des patients atteints de ces syndromes permet de proposer des tests de prédisposition génétique aux membres des familles concernées (Tableau 3).

Syndromes de polyposes familiales	Gènes mis en cause
Polypose Adénomateuse Familiale (PAF)	
PAF sévère	<i>APC</i>
PAF atténuée	<i>APC</i>
Syndrome de Turcot	2/3 avec des mutations germinales d' <i>APC</i> 1/3 avec des mutations germinales des gènes MMR
Syndrome de Gardner	<i>APC</i>
Polypose associée à <i>MUTYH</i>	<i>MUTYH (MYH)</i>
Syndrome de Lynch (HNPCC)	
Syndrome de Lynch de types 1 et 2	<i>MSH2, MLH1, MSH6, PMS2</i>
Syndrome de Muir-Torre	
Cancer colorectal familial de type X	Base génétique inconnue
Syndromes de Polypose Hamartomateuse	
Syndrome de Peutz-Jeghers (PJ)	<i>LKB1, STK11</i>
Polypose Juvénile familiale (PJF)	<i>SMAD4, BMPR1A, ENG</i>
Syndrome de Cowden	<i>PTEN</i>
Syndrome de Bannayan-Ruvalcaba-Riley (BRR)	<i>PTEN</i>
Syndrome de Gorlin	<i>PTCH</i>
Syndrome héréditaire de polypose mixte	Base génétique inconnue
Syndrome de polypose hyperplasique	<i>K-RAS, B-RAF</i>

Tableau 3 : Causes génétiques des polyposes familiales (Gène MMR : gènes de réparation des mésappariements de l'ADN) (Tableau inspiré de Rustgi (2007) et de Davidson (2007)).

Dans la partie suivante, nous développerons seulement les formes familiales le plus souvent rencontrées, c'est-à-dire la polypose adénomateuse, la polypose associée à *MUTYH* et le syndrome de Lynch.

II.4.1. La polypose adénomateuse familiale (PAF)

II.4.1.1. Contexte historique

C'est en 1721 qu'a été décrit, pour la première fois, le cas d'un patient présentant un grand nombre de polypes dans le gros intestin (MEZELIO 1721). Dans les années 1880, Cripps (1882) suggère le caractère héréditaire de la polypose adénomateuse familiale (PAF). Les caractéristiques cliniques et l'histoire complète de la PAF sont rapportés dans le livre de Bussey (1975). La PAF est une des formes familiales les moins fréquentes, avec moins de 1% des cas de carcinomes colorectaux. La manifestation la plus importante de ce syndrome est l'apparition, dès l'enfance, dans le côlon et le rectum, d'une multitude de polypes adénomateux (d'une centaine à plus d'un millier) (Figure 20). Dans presque 100% des cas la polypose évolue en adénocarcinome vers l'âge précoce de 35-40 ans

(BUSSEY 1975; PENNA *et al.* 1992). La PAF concerne une naissance sur 7000 à 10000 aux Etats-Unis, avec un mode de transmission autosomique dominant à forte pénétrance (DE LA CHAPELLE 2004).



Figure 20 : Polypose adénomateuse classique d'un patient de 20 ans présentant de petits polypes sessiles en tapis (plusieurs centaines) (GATALICA and TORLAKOVIC 2008).

La PAF se manifeste classiquement par des polypes sessiles présents sur la totalité de la muqueuse colique ; des polypes pédonculés sont également retrouvés, mais en moins grand nombre. En microscopie, l'adénome montre une architecture tubulaire identique à celle retrouvée chez les patients atteints de cancers sporadiques (GATALICA and TORLAKOVIC 2008).

D'autres manifestations digestives et extradiigestives peuvent s'ajouter aux polypes colorectaux : des tumeurs desmoïdes (10% à 15% des patients), des polypes gastriques et de l'intestin grêle (au niveau du duodénum), une hypertrophie congénitale de l'épithélium pigmentaire de la rétine (*CHRPE, Congenital Hypertrophy of Retinal Pigment Epithelium*, 70% à 85% des patients), des dents surnuméraires (70% à 90% des patients), des tumeurs de la thyroïde, ... (NIEUWENHUIS and VASEN 2007; RUSTGI 2007; WALON *et al.* 1997).

Des syndromes plus rares présentent des manifestations extra-colorectales spécifiques. Ainsi, le syndrome de Turcot montre une association assez rare de polypose recto-colique familiale et de tumeur du système nerveux central. De la même manière, le syndrome de Gardner est caractérisé par des ostéomes, des fibromes de la peau et des kystes épidermoïdes qui viennent se greffer à la polypose (DE LA CHAPELLE 2004; GARDNER 1951).

II.4.1.2. Polypose adénomateuse familiale atténuée

La forme atténuée de PAF est le phénotype variant le plus courant de polypose. Elle se distingue de la forme classique par un nombre plus faible de polypes (inférieur à 100), mais garde intact son potentiel de transformation en cancer colorectal. Cette forme de PAF intervient à un âge plus élevé d'environ 15 ans par rapport à la forme sévère, soit un âge d'apparition de 50 ans (GATALICA and TORLAKOVIC 2008). Les manifestations extra-coliques et l'histoire familiale des patients

atteints ne sont pas très bien renseignées. L'incidence réelle de cette polypose atténuée n'est pas connue, mais on peut l'estimer à environ 10% des PAF (KNUDSEN *et al.* 2003; VASEN 2000).

II.4.1.3. Les variants génétiques d'APC associés à la PAF

En 1986 une délétion interstitielle du bras long du chromosome 5 a été observée par cytogénétique chez un patient atteint de multiples polypes coliques. Un an plus tard le gène « *FAP gene* », désormais dénommé *adenomatous polyposis coli (APC)*, était localisé en 5q21-22 précisément dans la région chromosomique délétée. Depuis sa découverte, de nombreuses recherches ont été effectuées pour comprendre le lien entre l'hétérogénéité des phénotypes de PAF et les génotypes d'APC (BULOW *et al.* 2006; KINZLER and VOGELSTEIN 1996; LYNCH *et al.* 2004b).

La protéine APC « normale » va mener à la dégradation de la β -caténine, en coopération avec la protéine kinase 3β -glycogène synthétase et la phosphorylase β -caténine cytoplasmique. Lorsqu'une mutation germinale ou somatique intervient dans le gène *APC*, la protéine correspondante ne dégrade plus la β -caténine cytoplasmique qui devient stable. Il va en résulter une translocation de la β -caténine dans le noyau cellulaire qui va activer la transcription des gènes cibles, tels que *c-myc*, *cycline D1*, *MMP-7*, *Axine2/conductine*, par la voie Wnt. Cette voie de signalisation possède de nombreuses fonctions (adhésion cellulaire, interaction intracellulaire, transmission intracellulaire de signaux à l'épithélium, régulation du cycle cellulaire et apoptose) qui pourront être activées (GOSS and GRODEN 2000; SIEBER *et al.* 2000) (Chapitre I : II.3.1.1. La voie WNT/ β -caténine).

Le gène *APC* est composé de 15 exons codant pour 2843 acides aminés avec l'exon 15 qui représente 75% de la séquence codante (6.5kb) et qui est divisé en plusieurs domaines de fonctions différentes, incluant notamment la liaison à la β -caténine, la liaison à l'ADN et EB1 (DE LA CHAPELLE 2004; KORINEK *et al.* 1997; SU *et al.* 1993). De par sa longueur, l'exon 15 représente une cible privilégiée pour les mutations germinales et somatiques (Figure 21).

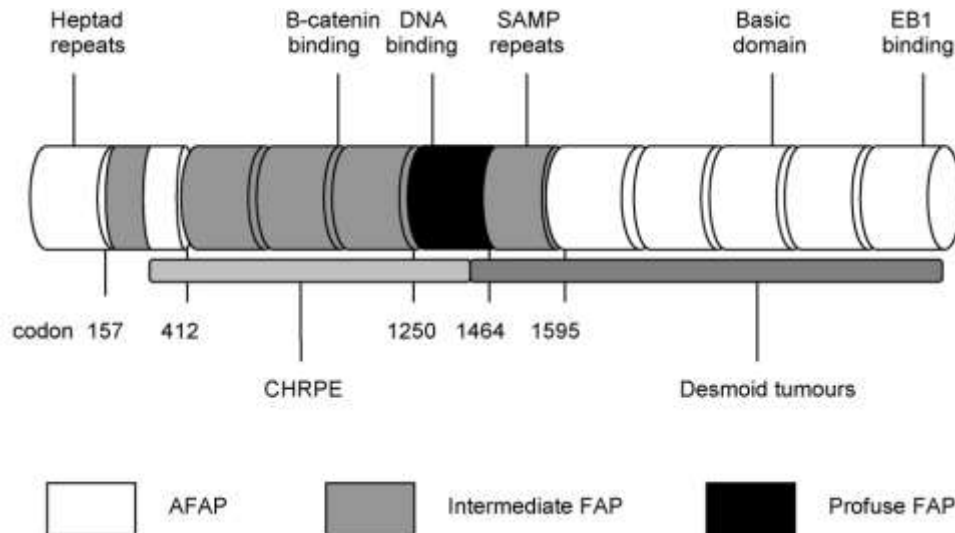


Figure 21 : Représentation schématique de la protéine APC (NIEUWENHUIS and VASEN 2007) indiquant les domaines structuraux clés qui interviennent dans ses fonctions, ainsi que l'effet des mutations en fonction de leur localisation sur le gène. Le domaine *Heptad repeats* joue un rôle dans le contrôle de l'adhésion cellulaire. Les domaines *β-catenin binding* et *DNA binding*, zones de liaison à la β-caténine et à l'ADN jouent un rôle de régulateurs négatifs du niveau de β-caténine et ainsi contrôlent les gènes intervenant dans l'entrée et dans la progression du cycle cellulaire. Les domaines *SAMP repeats* et *BASIC* sont associés à la migration et la division cellulaire. *EB1 binding* maintient la stabilité du chromosome pendant la mitose. *CHRPE*, zone dans laquelle les mutations vont entraîner le phénotype d'hypertrophie congénitale de l'épithélium pigmentaire de la rétine. *Desmoid tumors*, zone dans laquelle les mutations vont entraîner des tumeurs desmoïdes. *AFAP*, polypose adénomateuse familiale atténuée (<100). *Intermediate FAP*, polypose adénomateuse familiale intermédiaire (100<x<5000). *Profuse FAP*, polypose adénomateuse familiale abondante (>5000).

Depuis la découverte et l'identification des mutations du gène *APC* (gène suppresseur de tumeur) dans les PAF, en 1991 par Groden (GRODEN *et al.* 1991), plus de 1000 mutations germinales ont été reportées (LAURENT-PUIG *et al.* 1998). Environ 85% des patients atteints de PAF classique (DAVIDSON 2007) et 20% de patients atteints de PAF atténuée (KNUDSEN *et al.* 2003) montrent une mutation dans *APC*. La majorité (95%) de ces mutations (non-sens, insertion, délétion) conduit à une protéine tronquée (DE LA CHAPELLE 2004).

Les polyposes adénomateuses familiales répondent à la théorie de Knudson, selon laquelle une mutation germinale existe et nécessite un second évènement somatique, soit par mutation ponctuelle, soit par perte d'hétérozygotie (LOH) (DE LA CHAPELLE 2004; KINZLER and VOGELSTEIN 1996; LAMLUM *et al.* 1999; NIEUWENHUIS and VASEN 2007). Classiquement, les mutations somatiques d'*APC* surviennent précocement dans la carcinogenèse et sont réparties sur tout le gène, mais plus particulièrement dans un point chaud de mutation (*MCR*, *mutation cluster region*) localisé entre les codons 1309 et 1061 (BEROUD and SOUSSI 1996). La mutation héritée et la mutation somatique présentent toute deux une forte corrélation entre la localisation du premier évènement (mutation germinale) et du second (mutation somatique), ce qui diffère du modèle de Knudson pur, dans lequel les évènements ne sont pas liés. En effet, quand la mutation germinale intervient proche du codon 1300, la seconde mutation est souvent une perte d'hétérozygotie de la région *APC*, et lorsque la mutation germinale est distante de cette région, la mutation somatique se situe dans le *MCR* et

donne une protéine tronquée non fonctionnelle (BERTARIO *et al.* 2003; FRIEDL *et al.* 2001; LAMLUM *et al.* 1999).

Deux mutations faux-sens (I1307K et E1317Q) sont retrouvées comme étant associées à la PAF. La mutation I1307K n'altère pas la fonction de la protéine APC ; elle induit un état d'hypermutabilité somatique chez les juifs Ashkénazes (LAKEN *et al.* 1997). La seconde mutation (E1317Q) est également trouvée associée aux cancers colorectaux de type PAF ou PAF atténuée, mais sa pathogénicité n'est pas encore élucidée (KNUDSEN *et al.* 2003).

II.4.1.4. Corrélation génotype-phénotype dans les PAF

Les polyposes adénomateuses familiales sont de bons exemples pour illustrer la relation entre génotype et phénotype. En effet, les différentes mutations d'APC ont montré des phénotypes différents en fonction de leur localisation. Spirio *et al.* (1993) ont établi que le phénotype atténué des PAF était dû à une mutation localisée en 5' ou en 3' du gène, voire dans un site d'épissage de la région de l'exon 9 d'APC. De la même manière, une polypose sévère (>5000 polypes) peut être associée à une mutation entre les codons 1250-1464 (WALON *et al.* 1997), notamment le codon 1309 qui a été associé, dans de nombreuses publications, à un phénotype sévère de polypose (CASPARI *et al.* 1994; FICARI *et al.* 2000; FRIEDL *et al.* 2001; GEBERT *et al.* 1999).

Les mutations d'APC peuvent également être associées à des manifestations extra-coliques. Les génotypes liés aux tumeurs desmoïdes sont généralement situés en aval du codon 1400, alors que les mutations situées entre 311-1444, ou 1465 donneraient le phénotype CHRPE (hypertrophie congénitale de l'épithélium pigmentaire de la rétine) (ENOMOTO *et al.* 2000; FICARI *et al.* 2000; GEBERT *et al.* 1999; GIARDIELLO *et al.* 1997; WALON *et al.* 1997).

II.4.2. Polypose associée à *MUTYH* (MAP)

II.4.2.1. Histoire et caractérisation

Cette polypose a été découverte récemment par Al-Tassan *et al.* (2002) dans une famille présentant le même phénotype qu'une polypose adénomateuse familiale, mais avec un mécanisme récessif supposé. Les trois individus atteints de la fratrie ne présentaient aucune mutation dans le gène APC. Toutefois, des mutations somatiques (transversions G:C→T:A) du gène APC ont été observées dans les tumeurs, suggérant une déficience héréditaire du gène (*MUTYH*) de réparation de la 7,8-dihydro-2'-deoxyguanosine (8-oxoG) (AL-TASSAN *et al.* 2002). Cette transversion est particulièrement délétère si elle n'est pas réparée, car elle va entraîner le changement d'une séquence GAA en codon stop TAA (SAMPSON *et al.* 2005).

Cette nouvelle polypose, appelée polypose associée à *MUTYH* (MAP pour *MUTYH associated polyposis*), représente entre 0,5% et 1% de la totalité des cancers colorectaux (DE LA CHAPELLE 2004; FLEISCHMANN *et al.* 2004). Elle doit son nom à la découverte de mutations germinales dans le gène *MUTYH* (*MYH*). Son mode de transmission est autosomique récessif avec une pénétrance forte et probablement complète quand les mutations causales sont bialléliques (AL-TASSAN *et al.* 2002; CHEADLE and SAMPSON 2007; DE LA CHAPELLE 2004; LIPTON and TOMLINSON 2004; SAMPSON *et al.* 2005).

En règle générale, les publications tendent à comparer la MAP à la polypose adénomateuse familiale atténuée (SAMPSON *et al.* 2005). En effet, les patients atteints de MAP et ceux atteints de PAF sont assez difficiles à distinguer, car les données cliniques entre ces deux maladies sont très semblables (SIEBER *et al.* 2003). Suivant les publications, le nombre de polypes décrits varie de 10 à 1000, avec un âge d'apparition un peu plus tardif que pour une polypose adénomateuse familiale atténuée (DE LA CHAPELLE 2004; GATALICA and TORLAKOVIC 2008; LIPTON and TOMLINSON 2004; SIEBER *et al.* 2004). Les polypes observés sont moyennement dysplasiques, tubulaires et localisés préférentiellement dans le côlon gauche (LIPTON *et al.* 2003; LIPTON and TOMLINSON 2004). Contrairement à la PAF, les manifestations extra-coliques sont peu nombreuses (LIPTON and TOMLINSON 2004), même si des polypes peuvent être retrouvés couramment dans l'estomac et le duodénum (18% des patients) (GATALICA and TORLAKOVIC 2008), ainsi que des polypes hyperplasiques dans le rectum (CHOW *et al.* 2006).

II.4.2.2. Le gène *MUTYH* et les mutations associées à la MAP

Le gène *MUTYH* est localisé sur le chromosome 1 entre les bandes cytogénétiques p34.3 et p32.1. Il s'agit d'un homologue du gène *MutY* d'*Escherichia coli*, cloné et séquencé en 1996 (SLUPSKA *et al.* 1996). Il intervient dans le système de réparation par excision de bases (*BER*, *Base Excision Repair*) en codant pour une protéine responsable de l'excision des adénosines appariées avec une 8-oxoG (DE LA CHAPELLE 2004; SIEBER *et al.* 2003). La 8-oxoG est un produit des dommages à l'ADN occasionnés par un stress oxydatif (ARETZ *et al.* 2006; JASS 2008) qui est éliminé par le système BER.

II.4.2.2.1. Mécanisme d'action du système BER

Les ADN glycosylases reconnaissent les bases anormales de l'ADN et catalysent le clivage hydrolytique de la liaison N-glycosyl qui relie la base au sucre. La génération d'un site apurinique/aprimidinique (AP) va entraîner l'intervention du système BER, qui, dans la plupart des cas peut fonctionner en utilisant un ensemble de protéines communes. La fonction de l'AP endonucléase et de la phosphodiesterase est de générer une cassure d'un seul nucléotide contenant les groupements 3'hydroxyle et 5'phosphate terminal, qui va pouvoir être comblée par l'ADN polymérase. Enfin, une ADN ligase peut sceller le reste de la cassure. Ce mécanisme est le principe

global de réparation, mais, suivant le nombre de nucléotides manquants, il sera quelque peu différent. Ainsi, le premier mécanisme, appelé « *long patch* » est impliqué pour les cassures de plus d'un nucléotide (Figure 22). Ce mécanisme semble être prédominant lorsque la réparation est initiée sur un site AP, oxydé ou réduit, généré par des rayons X ou des agents chimiques. La réparation d'une cassure d'un seul nucléotide produite par la voie « *short patch* » est réalisée lorsqu'un site AP « ordinaire » est généré (Figure 22) (NORBURY and HICKSON 2001).

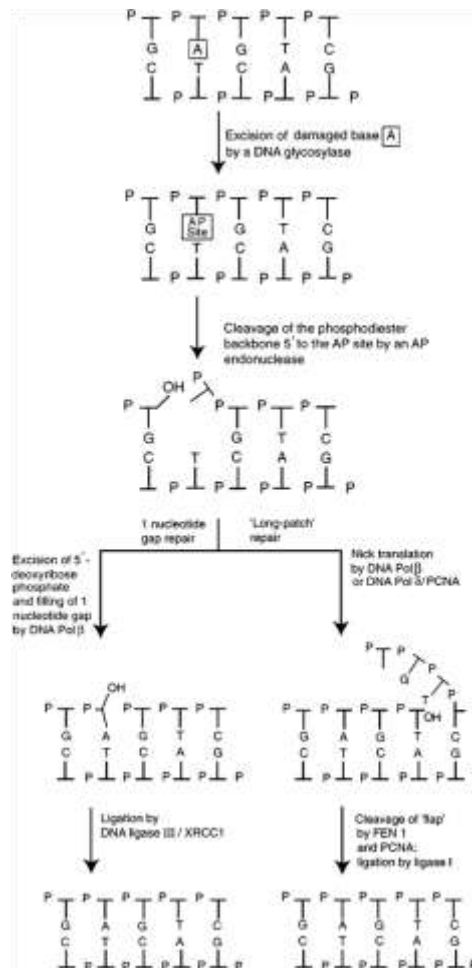


Figure 22 : Voie de réparation de l'ADN par le système BER (*Base Excision Repair*). La génération d'un site apurique/aprimidique sur l'ADN va entraîner l'intervention du système BER qui, selon le type de dommage, va suivre la voie « *short patch* » pour les cassures d'un seul nucléotide ou « *long patch* » pour des cassures plus importantes (NORBURY and HICKSON 2001).

Une des altérations fréquemment rencontrée dans les cancers colorectaux est la présence sur l'ADN d'une guanine oxydée (8-oxoG), due à un stress oxydatif, qui va favoriser les transversions G→T. Ces transversions sont particulièrement délétères, car elles peuvent potentiellement engendrer un codon stop (SAMPSON *et al.* 2005). Le système BER va assurer le remplacement des 8-oxoG par des guanines non oxydées. La Figure 23 montre les différentes enzymes et le mécanisme simplifié permettant l'élimination de la 8-oxoG (DAVID *et al.* 2007).

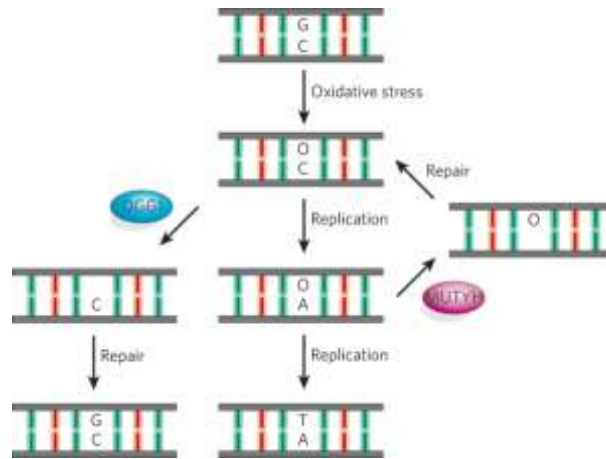


Figure 23 : Voie « *short-patch* » délimination des 8-oxoG par le système BER. Dans ce système, deux enzymes montrées comme prédisposant au CCR sont impliquées. MUTYH va cliver l'adénosine qui est appariée à la 8-oxoG et la remplacer par une cytosine. OGG1 va ensuite remplacer la 8-oxoG par une guanine non oxydée. Ainsi la séquence nucléotidique sera réparée (DAVID *et al.* 2007).

II.4.2.2.2. Mutations du gène *MUTYH*

Ce gène comprend de nombreuses mutations (Figure 24), mais deux mutations, Y165C dans l'exon 7 et G382D dans l'exon 13, représentent environ 80% des mutations de *MUTYH* impliquées dans les MAP (KNUDSEN *et al.* 2003; RUSTGI 2007). Huit à 30% des cas de PAF ou PAF atténuée sans mutation du gène *APC* seraient dus à des mutations dans *MUTYH* (AL-TASSAN *et al.* 2002; ARETZ *et al.* 2006; CHEADLE and SAMPSON 2007; DALLOSSO *et al.* 2008; DE LA CHAPELLE 2004; JONES *et al.* 2002; KAIRUPAN *et al.* 2005; SAMPSON *et al.* 2005; SIEBER *et al.* 2003) et environ 6% des personnes possédant des mutations bialléliques de *MUTYH* développent un ou plusieurs cancers colorectaux avec des phénotypes plutôt atténués (LIPTON and TOMLINSON 2004; SIEBER *et al.* 2003). Si les mutations bialléliques de *MUTYH* sont responsables d'une pénétrance élevée, la question de la pénétrance chez les porteurs hétérozygotes d'une mutation *MUTYH* a été posée. Les résultats sont assez contradictoires. Si Farrington *et al.* (2005) ont montré une augmentation du risque de cancers colorectaux chez les patients porteurs de la mutation G382D à l'état hétérozygote âgés de plus de 55 ans, d'autres études n'ont pas mis en évidence cette association (BALAGUER *et al.* 2007; SAMPSON *et al.* 2005).

Les mutations Y165C et G382D surviennent dans des régions très conservées de *MUTYH* (LIPTON and TOMLINSON 2004) et représentent plus de 50% des mutations de *MUTYH* trouvées chez des patients atteints de cancers colorectaux (DE LA CHAPELLE 2004). Des deux mutations, la mutation Y165C est la plus délétère pour le système BER (LIPTON and TOMLINSON 2004). D'autres mutations spécifiques de populations ont été identifiées ; ainsi, la mutation Y90X a été retrouvée au Pakistan, la mutation E466X en Inde, la délétion nt1395-7delGGA en Italie (JASS 2008), et plus de 20 autres mutations ont été identifiées comme causales (Figure 24) (CHEADLE and SAMPSON 2007).

Chez plusieurs patients, une seule mutation monoallélique de *MUTYH* a été détectée, ce qui laisse présager de l'action d'un mécanisme oligogénique (DE LA CHAPELLE 2004). C'est pourquoi, d'autres gènes candidats ont été investigués dans les MAP. Il a été montré que les instabilités micro-satellitaires (MSI) n'interviennent pas dans les tumeurs MAP, et que pour initier la mutation de transversion C:G→A:T d'*APC*, certains adénomes MAP montrent une mutation faux-sens de *KRAS* qui va donner une substitution spécifique et exclusive de la glycine en position 12 par une cystéine (G12C) (LIPTON *et al.* 2003). Par contre ni *BRAF*, ni *SMAD4*, ni *TGFβ* n'ont été trouvés mutés dans les adénomes et cancers MAP (LIPTON *et al.* 2003; LIPTON and TOMLINSON 2004). De la même manière, les autres gènes du système BER ont été investigués (*OGG1*, *NEIL1*, *NEIL2*, *NEIL3*, *NUDT1*, *NTH1*) et aucune mutation n'a été mise en évidence. Ils n'interviennent donc pas dans cette polypose autosomale récessive, par contre il n'est pas exclu que ces gènes contribuent à la détermination du phénotype polyposique récessif ou dominant (DALLOSSO *et al.* 2008).

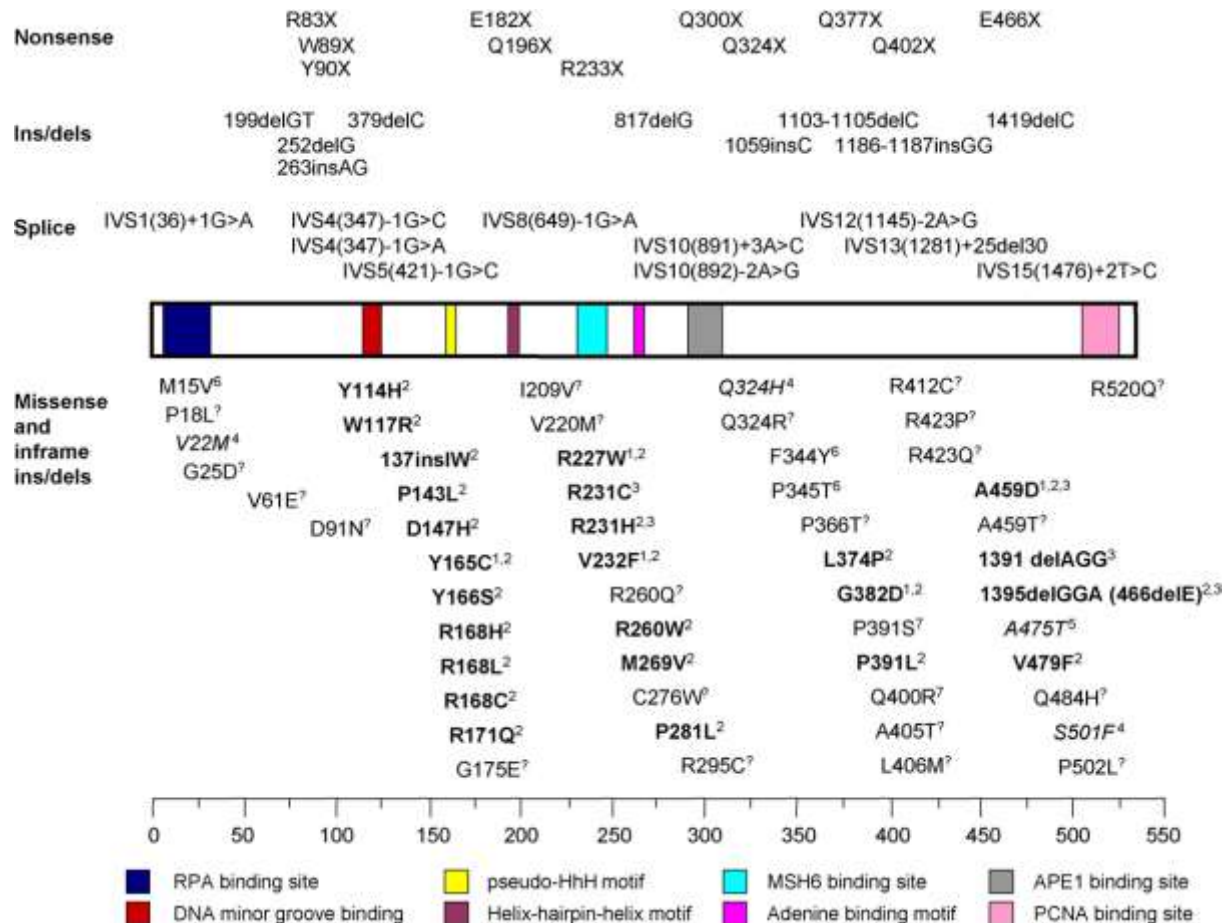


Figure 24 : Distribution et localisation des mutations identifiées à ce jour dans le gène *MUTYH*. Les positions présumées des domaines fonctionnels est indiquée par rapport à la région codante *MUTYH*. Les variants faux-sens éprouvés ou susceptibles d'être pathogènes sont indiqués en gras (1, données montrant une fonctionnalité ; 2, variant rare trouvé en association avec une mutation *MUTYH* constatée chez un patient ; 3, variant rare dans un état homozygote chez un patient), ceux qui sont peu susceptibles d'être pathogènes sont en italique (4, polymorphisme commun ; 5, dans les cas de mutations bi-alléliques *MUTYH*) et ceux qui sont indéfinis sont en caractères ordinaires (6, variant rare trouvé en association avec une autre mutation *MUTYH* chez un patient avec le CCR de moins de 60 ans et avec plus de deux

membres de sa famille touchés ; 7, une mutation somatique dans le cancer de l'estomac ; ?, pas de preuve évidente de pathogénicité) (CHEADLE and SAMPSON 2007).

II.4.3. Le syndrome de Lynch

II.4.3.1. Historique et caractérisation

Le syndrome de Lynch, qui à l'origine était appelé « *cancer family syndrome* », a été décrit pour la première fois dans une publication de Warthin (1913). C'est en 1971 que Lynch et Krush ont réétudié cette même famille et ont décidé d'appeler ce syndrome « HNPCC », acronyme pour *Hereditary Non Polyposis Colorectal Cancer* (cancer colorectal héréditaire sans polypose). Cette pathologie avait été déterminée dans les années 1960 comme étant transmise de manière autosomique dominante (LYNCH *et al.* 1966). Ce caractère autosomique dominant est confirmé par l'augmentation du risque de 2 à 4 fois de développer un cancer pour les individus possédant une histoire familiale positive (présence de cancers chez les parents au premier degré) (MECKLIN 2008).

Le nom de ce syndrome a été soumis à de nombreux débats, car après s'être longtemps appelé HNPCC, le nom est redevenu « syndrome de Lynch ». Le nom HNPCC évoque trop l'absence de polypes, ce qui ne correspond pas à la réalité clinique de cette pathologie. La mise en place des critères d'Amsterdam I et II en 1991 puis 1999 a permis de définir un mode de sélection des patients dans le cadre d'un dépistage sur des critères cliniques (Tableau 4).

Critères d'Amsterdam I
<p>Pour répondre à ces critères, il faut au moins trois parents avec un cancer colorectal vérifié par histologie qui répondent à ces différents points :</p> <ul style="list-style-type: none"> • 1 sujet doit être parent au premier degré des 2 autres ; • Au moins 2 générations successives doivent être affectées ; • Au moins un des parents avec un cas de cancer colorectal détecté avant 50 ans ; • La polypose adénomateuse familiale doit être exclue.
Critères d'Amsterdam II
<p>Pour répondre à ces critères, il faut au moins trois sujets parents avec un cancer colorectal héréditaire non polyposique ou un cancer associé (cancer colorectal, de l'endomètre, de l'estomac, des ovaires, de l'uretère/rein, cerveau, intestin grêle, des voies hépatobiliaires et de la peau (tumeur sébacée)) :</p> <ul style="list-style-type: none"> • 1 sujet doit être parent au premier degré des 2 autres ; • Au moins 2 générations successives doivent être affectées ; • Au moins un des cancers colorectaux non polyposique, ou cancers associés doivent être détectés avant 50 ans ; • La polypose adénomateuse familiale doit être exclue pour les cancers colorectaux ; <p>Les tumeurs doivent être vérifiées quand c'est possible.</p>

Tableau 4 : Critères d'Amsterdam I (VASEN *et al.* 1991) et II (VASEN *et al.* 1999).

Les critères d'Amsterdam de type I sont très sélectifs et permettent d'identifier un syndrome de Lynch chez 77% des patients retenus. En revanche, la sensibilité n'était pas bonne et a nécessité d'élargir les critères (Amsterdam II) aux cancers extra-coliques du spectre HNPCC. L'application de ces nouveaux critères ont diminué le taux de mutations identifiées chez les patients retenus mais ont considérablement augmenté la sensibilité de détection des cas.

Depuis la découverte du rôle des gènes de réparation des mésappariements de l'ADN (gènes MMR) dans cette pathologie, le terme syndrome de Lynch (LS) est utilisé lorsque des mutations germinales des gènes MMR sont mises en cause. Jass (2006) a proposé que le terme HNPCC pourrait être utilisé pour indiquer certaines formes cliniques de cancers colorectaux héréditaires (hors PAF) possédant des instabilités chromosomiques, mais dont aucune transmission verticale de mutations des gènes MMR n'est montrée (JASS 2006; MECKLIN 2008). D'autres chercheurs ont appelé ces cancers colorectaux familiaux non polyposiques (hors PAF) dont la base génétique n'est pas connue, « cancer colorectal familial de type X » (DAVIDSON 2007; LINDOR *et al.* 2005; LYNCH *et al.* 2006), ce qui correspond à 40% des familles avec des cancers colorectaux qui ne présentent pas de mutations germinales des MMR, ni d'instabilité chromosomique (BOLAND 2006; LINDOR *et al.* 2005).

Le syndrome de Lynch est le plus fréquent des cancers colorectaux avec une incidence de 1 à 2 pour 1000 (BOLAND 2006; VASEN 2000). Il représenterait de 0,5 à 5% de la totalité des cancers colorectaux, les chiffres variant en fonction des publications et de la définition du syndrome : 2 à 4% (KINZLER and VOGELSTEIN 1996) ; 0,5 à 5% (DE LA CHAPELLE and PELTOMAKI 1995; MECKLIN *et al.* 1995; MOORE and COWLED 1999) ; 1 à 5% (VASEN 2000) ; 1 à 6% (LYNCH and DE LA CHAPELLE 1999; SALOVAARA *et al.* 2000) ; 5% (DE LA CHAPELLE 2004) ; 3 à 4% (BOLAND 2006) ; 3 à 5% au USA (SAMOWITZ *et al.* 2001) ; 3 à 4% (RUSTGI 2007) ; 1 à 3% (DIONIGI *et al.* 2007) ; 2 à 5% (MECKLIN 2008). Cette variation peut également venir de la performance des études moléculaires réalisées. Par exemple, Houlston *et al.* (1992) ont déterminé par analyse de ségrégation que le syndrome de Lynch représentait 13% des cancers colorectaux. En outre, plusieurs études ont montré que la prévalence de cette pathologie était due à la fois aux facteurs génétiques, mais également à des facteurs environnementaux et notamment l'alimentation (DE LA CHAPELLE 2004; LYNCH and DE LA CHAPELLE 2003; VASEN 2000).

Le syndrome de Lynch se caractérise par une carcinogenèse accélérée avec apparition d'un carcinome en 2 à 3 ans, au lieu de 8 à 10 ans chez les autres patients (JASS and STEWART 1992; LYNCH and DE LA CHAPELLE 1999; LYNCH *et al.* 2008). En effet, dans ce syndrome, une cellule adénomateuse avec une déficience en protéine MMR acquiert des mutations 2 à 3 fois plus rapidement qu'une cellule sans ce déficit. Il en résulte une accumulation de mutations dans les oncogènes et les gènes suppresseurs de tumeur, ce qui va mener à une progression rapide de la malignité (KINZLER and VOGELSTEIN 1996).

La Figure 25 compare l'évolution du syndrome de Lynch et des PAF.

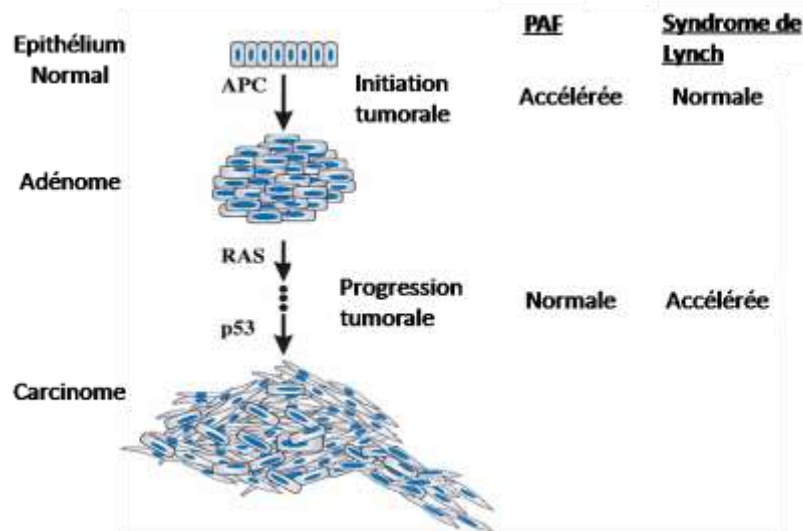


Figure 25 : Comparaison entre le développement de la polypose adénomateuse familiale (PAF) et celui du syndrome de Lynch. La PAF est le résultat d'une augmentation du taux d'initiation tumorale due à l'atteinte du gène APC alors que le syndrome de Lynch, quant à lui, montre une augmentation de la progression tumorale due à l'atteinte des gènes MMR, alors que l'initiation tumorale est normale (Inspiré de Kinzler et Vogelstein (1996)).

En règle générale, le syndrome de Lynch est caractérisé par un âge précoce de survenue du cancer (inférieur à 45 ans pour 80% des patients) (DAVIDSON 2007; DIONIGI *et al.* 2007; LYNCH and DE LA CHAPELLE 2003; LYNCH *et al.* 2008; RUSTGI 2007; VASEN 2000), même si le cancer colorectal peut parfois se développer à un âge plus tardif (DAVIDSON 2007). La localisation est proximale (côlon droit ou transverse) dans 70% des cas (LYNCH *et al.* 2008; VASEN 2000), avec une prédominance de survenue dans le côlon droit (LYNCH and DE LA CHAPELLE 2003; RUSTGI 2007). Il débute par de petits adénomes plats difficilement visibles lors de la coloscopie – à laquelle les gastroentérologues préfèrent la chromoendoscopie – et possède une architecture villositaire et une dysplasie de haut grade. Parfois, il y a présence de polypes adénomateux hyperplasiques confirmant la désuétude du terme HNPCC pour cette pathologie (BOLAND 2006; HURLSTONE *et al.* 2005; JASS 2006; LAMMI *et al.* 2004; VASEN 2000), même si par leur nombre restreint, on ne peut pas parler de polypose (MECKLIN 2008).

Ces cancers colorectaux sont pauvrement différenciés et présentent un excès de mucus et des lymphocytes infiltrant la tumeur dans la majorité des cas (DIONIGI *et al.* 2007; JASS 2004; LYNCH and DE LA CHAPELLE 2003; LYNCH *et al.* 2008). De plus, ces cancers ont la particularité d'être synchrones* dans 45% des cas (DIONIGI *et al.* 2007) mais aussi métachrones† après 10 ans dans 25 à 30% des cas (LYNCH and DE LA CHAPELLE 1999; LYNCH *et al.* 2008; VASEN 2000).

En plus de cette accumulation de cancers colorectaux, les familles présentent un risque accru de développer d'autres cancers tels qu'un second cancer colorectal (risque de 70 à 85%), un cancer de l'endomètre (risque de 50%) et d'autres cancers (risque inférieur à 15%) (VASEN 2000) : cancer des

* Synchrones : survenue d'un cancer colorectal dans les 6 mois qui suivent la résection de la première tumeur.

† Métachrone : survenue d'un cancer colorectal plus de 6 mois après la résection de la première tumeur.

ovaires, de l'estomac, de l'intestin grêle, des reins/uretère, du cerveau, des voies hépatobiliaires, de la peau (tumeur sébacée) (FUSARO *et al.* 1996; GATALICA and TORLAKOVIC 2008; LYNCH and DE LA CHAPELLE 1999; LYNCH *et al.* 2008; MECKLIN 2008; RUSTGI 2007; VASEN 2000). Lorsque certaines tumeurs ou cancers sont associés de manière fréquente, une dénomination nouvelle est donnée, c'est le cas du syndrome de Muir-Torre qui est l'association d'un cancer colorectal et d'un cancer de la peau se manifestant par des tumeurs des glandes sébacées (LYNCH *et al.* 2008; MECKLIN 2008; RUSTGI 2007).

II.4.3.2. Mutations et polymorphismes mis en cause dans le syndrome de Lynch

Les premiers gènes qui ont été associés à ce syndrome sont les gènes *MSH2* et *MLH1*, localisés sur les chromosomes 2p16 et 3p21 et qui appartiennent à la famille des gènes de réparation des mésappariements de l'ADN (gènes MMR). Cette association a été découverte grâce à une analyse de liaison réalisée dans une famille, confirmant la transmission mendélienne de ce syndrome (LINDBLOM *et al.* 1993; PELTOMAKI *et al.* 1993). Une autre étude a montré cette association grâce à la perte allélique d'un marqueur microsatellite localisé en 2p16 qui est lié à la susceptibilité aux cancers colorectaux (KINZLER and VOGELSTEIN 1996).

Le système MMR répare et corrige les mésappariements de bases et de petits nucléotides (de 1 à 4 bp), ainsi que les insertions/délétions arrivant pendant la réplication (DIONIGI *et al.* 2007; RUSTGI 2007) (Chapitre I : II.5.3.2.5.d)(1)Mécanisme). Les gènes appartenant à ce système sont des gènes suppresseurs de tumeur (DE LA CHAPELLE 2004; JASS 2006) qui ont de multiples fonctions. Certaines de leurs mutations, délétions, inversions, duplications sont assez courantes d'après Lynch et de la Chapelle (1999) et peuvent être délétères, car elles vont perturber les points de contrôle du cycle cellulaire et/ou l'apoptose (CHUNG and RUSTGI 2003; EDELMANN and EDELMANN 2004; JASS 2006). De plus, cette perte de fonction des gènes MMR va entraîner des mutations dans les régions codantes de gènes tels qu'*APC*, *KRAS*, *p53*, *TGFBR1I*, *Bax*,..., intervenant dans l'initiation et la progression tumorale (RUSTGI 2007; VASEN 2000), tout en sachant que cette perte de fonction n'est pas observée pour des mutations hétérozygotes de ces gènes (PARSONS *et al.* 1993).

De récentes études ont montré que le syndrome de Lynch peut être expliqué dans 30 à 80% des cas (DE LA CHAPELLE 2004) par la présence d'une mutation germinale héréditaire dans un des 4 gènes principaux du système MMR mis en cause dans cette pathologie (*MLH1*, *MSH2*, *MSH6*, *PMS2*) (BOLAND 2006; DE LA CHAPELLE 2004; DIONIGI *et al.* 2007; GATALICA and TORLAKOVIC 2008; JASS 2006; LYNCH and DE LA CHAPELLE 1999; LYNCH and DE LA CHAPELLE 2003; MECKLIN 2008; RUSTGI 2007; VASEN 2000). Les autres gènes de ce système, *MLH3*, *MSH3* et *PMS1* n'ont pas encore une implication très claire dans ce syndrome, même si dans de rares cas des mutations causales ont pu être identifiées.

La difficulté réside dans l'interprétation de l'effet de certaines mutations, ou plus exactement de variants inconnus identifiés dans les familles. La base de donnée INSIGHT (www.insight-group.org) comporte plus de 500 variants décrits (OU *et al.* 2008). Certaines mutations sont clairement délétères (stops, épissages, insertions, ou délétions), alors que l'interprétation de variants faux-sens est beaucoup plus délicate (DE LA CHAPELLE 2004). Les mutations germinales des gènes *MLH1* et *MSH2* représentent à elles seules de 60 à 95% des mutations causales du syndrome de Lynch, alors que celles du gène *MSH6* vont intervenir pour 7 à 10% des cas, et celles de *PMS2* pour 1% (DAVIDSON 2007; LYNCH and DE LA CHAPELLE 2003; MECKLIN 2008; RUSTGI 2007; TRUNINGER *et al.* 2005).

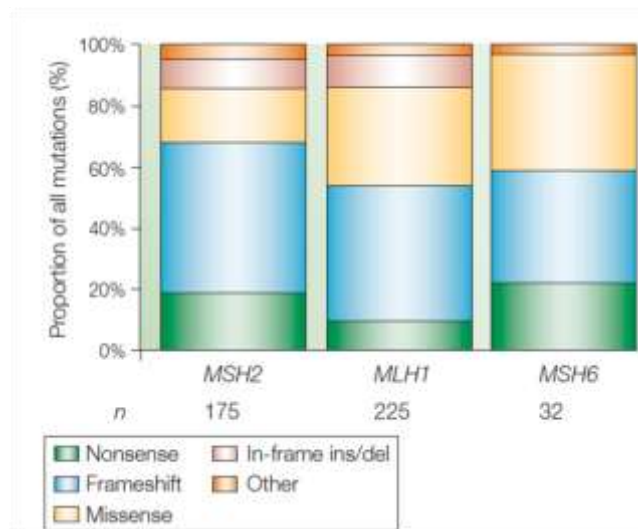


Figure 26 : Distribution des différentes mutations de *MSH2*, *MLH1* et *MSH6* selon de la Chapelle *et al.* (2004). (Nonsense, mutation non-sens ; Frameshift, mutation qui décale le cadre de lecture ; Missense, mutation faux-sens ; In-frame ins/del, mutation dans le cadre de lecture insertion/délétion ; Other, autres mutations).

Les mutations de *MLH1* et *MSH2* sont sur la totalité de la séquence des gènes sans réel point chaud de mutations (*hot-spot*) (PELTOMAKI and DE LA CHAPELLE 1997). Ces mutations *MLH1* et *MSH2* sont retrouvées dans 45 à 86% des familles correspondant aux critères d'Amsterdam. Ce pourcentage est de moins de 8 à 30% pour les familles à critères élargis. Des mutations de *MLH1* (29%) et de *MSH2* (16%) sont non-sens (LYNCH and DE LA CHAPELLE 2003) et 10 à 30% des mutations de *MSH2* sont des délétions, ce qui est moins courant dans *MLH1* et exceptionnel dans *MSH6* (DE LA CHAPELLE 2004; WIJNEN *et al.* 1998). Une mutation assez fréquente (de 5 à 10% des mutations du syndrome de Lynch) dans le gène *MSH2* est une transversion A→T dans le site donneur d'épissage de l'intron 5 de *MSH2*, qui va conduire à une perte de l'exon 5 (DE LA CHAPELLE 2004). Des mutations germinales de *PMS2*, *MSH6* et même plusieurs mutations non-sens et une mutation faux-sens de *MLH3* ont été retrouvées associées à des patients développant des cancers colorectaux (MIYAKI *et al.* 1997; TRUNINGER *et al.* 2005; WU *et al.* 2001b).

II.4.3.3. Corrélation phénotype/génotype

Différentes études ont permis de mettre en relation un génotype des gènes MMR avec des phénotypes particuliers de la maladie, ou des prédispositions accrues. En effet, pour les trois gènes les plus souvent associés au syndrome de Lynch (*MLH1*, *MSH2*, *PMS2*), une pénétrance de 80% pour le cancer colorectal, de 60% pour ceux de l'endomètre et de 20% pour les autres cancers est retrouvée (DE LA CHAPELLE 2004).

La présence de mutations dans les cinq premiers exons de *MSH2* va diminuer le risque de développer un cancer de l'endomètre (VASEN 2000), alors que, de manière générale, les mutations de *MSH2* vont augmenter la survenue de cancers extra-coliques et du syndrome de Muir-Torre (LYNCH *et al.* 2008). Lin *et al.* (1998b) ont montré que 33% des mutations *MSH2* et 12% de celles de *MLH1* interviennent dans les cancers extra-coliques.

Les mutations de *MSH6*, par leur faible pénétrance, vont quant à elles provoquer une forme atténuée de syndrome de Lynch, caractérisée par un âge de survenue plus élevé (LYNCH and DE LA CHAPELLE 2003; MIYAKI *et al.* 1997). De plus, *MSH6* est moins fréquemment impliqué dans les cancers colorectaux, mais montre une forte association avec le cancer de l'endomètre (JASS 2006; LYNCH *et al.* 2008; VASEN 2000). Les mutations de *PMS2* vont donner des syndromes de Lynch plus précoces, ainsi que des tumeurs localisées au niveau cérébral (NICOLAIDES *et al.* 1998).

II.4.3.4. Instabilité microsatellitaire (MSI)

L'inactivation de l'allèle sauvage des gènes de réparation des mésappariements de l'ADN va entraîner une perte d'expression rapide de ceux-ci, ce qui va augmenter les erreurs spontanées (mutations somatiques) lors de la réplication de l'ADN (RUSTGI 2007). Ces erreurs se produisent notamment dans les motifs répétés de l'ADN (microsatellites), mais peuvent aussi perturber les gènes suppresseurs de tumeur possédant de petites répétitions dans leur partie codante (*ACVR2*, *PTHLH*, *TGFBR1*, *MARCKS*, *MSH3*, *TCF4*, *RAD50*, *CASP5*, *BAX*, *RIZ*, *MBD4*, *MSH6*, *BLM*, *IGF2R*, *PTEN*, *AXIN2*, *WISP3*, *CDX2*) (DIONIGI *et al.* 2007; YAMAGUCHI *et al.* 2006). Si on amplifie par PCR un certain nombre de marqueurs microsatellites sur le génome, on observera une apparente instabilité de ces marqueurs dans la tumeur. On ne retrouvera pas cette instabilité si on amplifie parallèlement ces marqueurs à partir d'ADN extrait de tissu sain ou de lymphocytes. Cette instabilité permet de mettre en évidence les erreurs de réplication (RER) liées à la déficience en gènes MMR et confère à la tumeur un statut MSI (Microsatellite Instable). Il est nécessaire de savoir cependant que, si la majorité des syndromes de Lynch présentent un statut MSI (85 à 95%) (ALEXANDER *et al.* 2001; BOLAND 2006; KINZLER and VOGELSTEIN 1996; LYNCH and DE LA CHAPELLE 1999), ces instabilités ne sont cependant pas spécifiques de ce syndrome. On en retrouve à hauteur de 15% dans les cancers colorectaux

sporadiques (LYNCH *et al.* 2008; THIBODEAU *et al.* 1998; VASEN 2000). Afin de distinguer les syndromes de Lynch des cancers colorectaux sporadiques MSI-H (MSI-High), la recherche de la mutation V600E dans le gène *BRAF* permet, par sa présence, d'exclure une mutation dans les gènes MMR (GATALICA and TORLAKOVIC 2008). La fréquence des mutations V600E est de l'ordre de 5,7% dans les tumeurs MSI (KADIYSKA *et al.* 2007).

Du fait de la relative facilité du test RER par rapport au séquençage complet des gènes MMR, il a été proposé d'élargir le test par rapport aux critères d'Amsterdam, selon les critères de Bethesda.

Critères Originaux de Bethesda
<ul style="list-style-type: none"> • Individus avec des cancers dans leur famille qui répondent aux critères d'Amsterdam. • Individus avec deux cancers en relation avec HNPCC, incluant les cancers colorectaux synchrones et métachrones, ou les cancers extra-coliques associés (endomètre, ovaire, gastrique, hépatobiliaire, intestin grêle, rein, uretère). • Individus avec un cancer colorectal et un parent au premier degré avec un cancer colorectal et/ou cancer extra-colique associé à HNPCC et/ou un adénome colorectal ; un des cancers diagnostiqué avant 50 ans et les adénomes avant 40 ans. • Individus avec un cancer colorectal ou de l'endomètre diagnostiqué avant 50 ans. • Individus avec un cancer colorectal du coté droit avec une structure histopathologique indifférenciée diagnostiqué avant 50 ans. • Individus avec un cancer colorectal avec un type cellulaire en bague à chaton diagnostiqué avant 50 ans. • Individus avec un adénome diagnostiqué avant 40 ans.
Critères Révisés de Bethesda
<ul style="list-style-type: none"> • Cancer colorectal diagnostiqué chez des patients de moins de 50 ans. • Présence de tumeurs colorectales synchrones ou métachrones ou d'autres tumeurs associées HNPCC (colorectal, endomètre, estomac, ovaire, pancréas, uretère, rein, voie biliaire, intestin grêle, cerveau, adénome des glandes sébacées) indifféremment de l'âge. • Cancer colorectal avec une histologie MSI-High (présence de lymphocytes infiltrant la tumeur, réaction lymphocytaire identique à la maladie de Crohn, différenciation mucineux/bague à chaton, croissance médullaire) diagnostiqué chez des patients de moins de 60 ans. • Cancer colorectal diagnostiqué chez un ou plusieurs parents au premier degré avec des tumeurs en relation avec HNPCC, avec un des cancers diagnostiqué avant 50 ans. • Cancer colorectal diagnostiqué chez deux parents, ou plus, du premier ou second degré avec des tumeurs en relation avec HNPCC, indifféremment de l'âge.

Tableau 5 : Critères de Bethesda permettant une meilleure identification des cancers colorectaux (RUSTGI 2007).

De 50 à 60% des patients de moins de 30 ans avec des tumeurs colorectales sont MSI-H, contre seulement 12% des patients de plus de 35 ans. Sur la totalité des patients présentant une forte instabilité (MSI-H), seulement 50% portent une mutation germinale d'un gène MMR, ce qui laisse présager de l'intervention d'autres mécanismes (DAVIDSON 2007; VASEN 2000).

La détection de ce syndrome peut être effectuée par l'analyse de l'instabilité microsatellitaire, qui est moins coûteuse que le séquençage systématique des gènes MMR (VASEN 2000) et plus précise que les évaluations morphologiques seules par histopathologie ; ces dernières ne permettent en effet de détecter que 40% des tumeurs MSI-H (ALEXANDER *et al.* 2001). A l'heure actuelle, l'immunohistochimie (IHC) est utilisée en complément des tests d'instabilité microsatellitaire, car elle permet d'identifier 95% des syndromes de Lynch, en montrant l'absence sélective d'une protéine MMR dans la tumeur (BOLAND 2006). L'immunohistochimie permet, en complément du test RER, d'orienter le séquençage des gènes au niveau constitutionnel.

II.5. Les formes sporadiques de cancers colorectaux : des maladies complexes

II.5.1. Définition d'une maladie complexe ou multifactorielle

Cette forme de maladie s'oppose aux maladies mono-factorielles dues à l'action d'un seul et unique facteur. Les maladies complexes, dont le cancer colorectal sporadique est un exemple, sont dues à des facteurs environnementaux (alimentation, activité physique, traitements hormonaux, prise médicamenteuse, habitudes de vie comme la consommation de tabac,...) et des facteurs génétiques qui, par leur action combinée, vont mener au phénotype malin (HEAVEY *et al.* 2004).

II.5.2. Les facteurs environnementaux dans les cancers colorectaux

Des études épidémiologiques ont été réalisées sur les populations migrant des zones à faible risque vers les zones à fort risque de cancer colorectal. Elles montrent que le risque de CCR chez la population migrante est similaire à celui de la population implantée depuis des générations dans la zone à haut risque, ce qui indique que le risque de CCR est largement déterminé par l'exposition à l'environnement (TOMATIS and BARTSCH 1990).

II.5.2.1. Les macro-constituants alimentaires qui protègent du CCR

II.5.2.1.1. Les fruits, les légumes et les autres micronutriments

Les études épidémiologiques suggèrent une association inverse entre la consommation de légumes et le risque de CCR (FRANCESCHI 1999) ; pour les fruits, cette association est moins évidente (SLATTERY *et al.* 1997).

L'effet protecteur des fruits et des légumes est attribué à différents composants de ceux-ci : les fibres (que nous développerons dans la partie suivante : Chapitre I : II.5.2.1.2. Les fibres alimentaires) (WCRF 1997) mais également les antioxydants, incluant les micronutriments, tels que les vitamines C et E, le sélénium (Chapitre I : II.5.2.3.4. Le sélénium), les caroténoïdes, le β -carotène, le lycopène et la lutéine, les flavonoïdes et les iso-flavonoïdes (COMA 1998; LEVI *et al.* 2000). D'autres composants ont été montrés comme ayant un effet protecteur, tels que les glucosinolates, les isothiocyanates, les composés d'allium et les phytoœstrogènes (VERHOEVEN *et al.* 1996; VOORRIPS *et al.* 2000).

Plusieurs études *in vivo* montrent que les glucosinolates, composants des crucifères (les légumes à feuilles vertes, les brocolis, les choux de Bruxelles, les choux), induisent l'activité des GST* (LIN *et al.* 1998a; NASTRUZZI *et al.* 2000). Un effet protecteur de la consommation de brocolis est observé pour le géotype nul de *GSTM1* (COTTON *et al.* 2000). Il a été montré que les glucosinolates avaient une activité chémopréventive, en régulant certaines enzymes du foie et de la muqueuse intestinale intervenant dans la détoxification des xénobiotiques.

II.5.2.1.2. Les fibres alimentaires

La première explication avancée pour expliquer la différence d'incidence des CCR entre les populations était une hypothèse basée sur la consommation des fibres alimentaires (BURKITT 1971). Cette explication repose sur l'alimentation des populations rurales d'Afrique et d'Asie qui consomment une grande quantité d'aliments contenant des polysaccharides non-digestibles provenant des parois cellulaires de plantes non raffinées (TROWELL and BURKITT 1986). Cette hypothèse est maintenant abandonnée.

Les fibres alimentaires auraient un grand rôle dans la protection contre les cancers colorectaux. Différentes études ont montré que les fibres alimentaires sont associées à un risque diminué de CCR. En effet, une consommation doublée en fibres réduit les risques de CCR de 40% par rapport à une population dont la consommation de fibres est réduite (BINGHAM *et al.* 2003; PETERS *et al.* 2003).

Peters *et al.* (2003) ont mis en avant la capacité des fibres alimentaires à diminuer non seulement le risque de cancer, mais également à protéger de la formation d'adénomes colorectaux. Cette association est d'autant plus forte pour les fibres issues des graines, des céréales et des fruits.

Toutefois, d'autres études, dont celle de Fuchs *et al.* (1999), ne révèlent pas d'effet protecteur des fibres alimentaires. Ces variations sont dues aux recrutements hétérogènes des populations utilisées pour les études. Dans certaines études, par exemple, le nombre d'individus recruté n'est pas assez conséquent pour obtenir des tests statistiques fiables. La différence de localisation géographique des populations étudiées peut également expliquer les différences de résultats d'une étude à l'autre. En effet, la géographie va influencer fortement les résultats de ces études, car les habitudes alimentaires sont différentes d'un pays à l'autre, voire même d'une région à l'autre. C'est pourquoi, dans ces études épidémiologiques, il faut prendre un grand soin dans la constitution des cohortes.

* GST : Glutathion S-transférase, enzyme jouant un rôle dans la voie de détoxification de l'organisme.

Le mécanisme d'action des fibres n'est pas connu. Il semblerait toutefois qu'il mette en jeu les propriétés physiques des fibres, responsables d'une rétention d'eau et d'une augmentation de la croissance des bactéries commensales, via une fermentation anaérobie (EASTWOOD and MORRIS 1992). Les effets bénéfiques des fibres sont connus : la dilution des matières fécales ou l'augmentation du temps du transit intestinal,... (CUMMINGS 1981). D'autres effets peuvent également être attribués aux produits de fermentation de ces fibres, notamment aux acides gras à courte chaîne (acide butyrique,...). Dans l'intestin, ces acides gras vont devenir une source d'énergie pour la flore locale. Ils vont également permettre de faire diminuer le pH colique et fécal, ce qui est bénéfique pour l'hôte (SILVI *et al.* 1999). De plus, plusieurs publications ont montré que ces acides gras induisent l'apoptose *in vivo* et *in vitro* (CUMMINGS 1981; ROEDIGER 1990; ROWLAND *et al.* 2000), par un mécanisme dépendant des caspases (CHAI *et al.* 2000).

II.5.2.1.3. Les acides gras polyinsaturés n-3

Des études prospectives ont montré un effet protecteur de la consommation de poisson (NORAT *et al.* 2005; WILLETT 2001). Bien que cet effet protecteur soit encore loin d'être définitivement prouvé, ces résultats ont suscité un intérêt, en montrant qu'une absorption élevée d'acide gras polyinsaturés n-3 provenant d'huile de poisson pouvait exercer un effet anti-carcinogène sur la muqueuse colique. Ces acides gras polyinsaturés n-3 pouvaient réduire la production d'eicosanoïdes pro-inflammatoires et inhiber l'expression et l'activité de l'enzyme COX2*, à la manière de l'aspirine et d'autres anti-inflammatoires non-stéroïdiens (HALL *et al.* 2007).

Les effets protecteurs du poisson ont été associés à des polymorphismes génétiques affectant l'expression de protéines clés, comme dans la signalisation de PPAR et l'expression de COX2. Ces données épidémiologiques sont néanmoins encore confuses. Par exemple, les personnes portant un polymorphisme mineur dans *PPARδ* ont un risque accru d'adénome, mais pas de tumeur. Cet effet protecteur du poisson, vis-à-vis du développement tumoral est plus important pour les individus portant un polymorphisme de *COX2* (SIEZEN *et al.* 2006; SIEZEN *et al.* 2005). Les gènes *COX2* et *PPARδ* sont associés au contrôle et au métabolisme des acides gras, ainsi qu'à la formation des eicosanoïdes inflammatoires. Ces observations soutiennent l'hypothèse que la composition en acides gras du poisson est la clé de l'effet protecteur. Cette hypothèse est également appuyée par des publications montrant l'effet antiprolifératif et pro-apoptotique de l'acide eicosapentaénoïque (EPA) dans les cellules épithéliales du côlon (LATHAM *et al.* 2001). Un autre argument est la suppression de la mitose (ANTI *et al.* 1992) et l'augmentation de l'apoptose par l'EPA ainsi que l'acide

* COX2 : Cyclooxygénase de type 2 jouant un rôle dans la voie de l'inflammation.

docosahexaénoïque (DHA), dans les cryptes de la muqueuse rectale humaine (CHENG *et al.* 2003; COURTNEY *et al.* 2007).

Ce qui semble important dans l'association entre le CCR et ces acides gras n-3, ce n'est pas la quantité d'acides gras n-3 contenus dans l'alimentation, mais le rapport de la quantité d'acide gras n-3 (notamment dérivés du poisson tels l'EPA (C20:5) et la DHA (C22:6)) sur la quantité d'acides gras n-6 (tel l'acide arachidonique (C20:4)) (BUSSTRA *et al.* 2003; NKONDJOCK *et al.* 2003). L'estimation actuelle de ce rapport dans les régimes alimentaires occidentalisés est de 30 fois plus d'acides gras n-6 que d'acides gras n-3, alors qu'un ratio proche de 1/2 serait plus optimal pour la santé (SIMOPOULOS 2006).

II.5.2.2. Les macro-constituants alimentaires qui prédisposent au CCR

II.5.2.2.1. Les graisses animales et la viande rouge

Une forte consommation de graisses animales et/ou de viande rouge sont des facteurs de risque de CCR largement reconnus (BOYLE and LANGMAN 2000; LIPKIN *et al.* 1999) mais les résultats restent divergeants.

Concernant les graisses, les résultats vont dans le sens d'une prédisposition au CCR, même si les résultats ne sont pas forcément significatifs. En effet, Beresford *et al.* (2006) ont montré dans leur étude qu'une réduction de 10% de graisses ingérées avec une augmentation de la consommation de fruits, de légumes et de grains permettait de mettre en évidence une diminution du risque de CCR chez près de 20000 femmes post-ménopausées. Une autre étude a mis en évidence un risque augmenté de CCR pour un groupe d'individus consommant beaucoup de graisses provenant de produits riches en énergie et ayant une faible activité physique (WILLETT 2001).

Les données épidémiologiques sur la consommation de viande montrent une corrélation positive avec le risque de CCR (BREUER-KATSCHINSKI *et al.* 2001; NORAT *et al.* 2002). Les résultats de la majorité des études indiquent une plus forte corrélation positive entre le risque de CCR et une forte consommation de viande rouge ou de viande ayant subi un traitement (salaison, cuisson à la poêle, au four, avec matière grasse,...) que pour une consommation de viande blanche. Toutefois le mode de cuisson ne montre pas d'association significative avec le CCR comme la consommation de viande crue.

Sandhu *et al.* (2001) ont réalisé une méta-analyse de 13 études d'association portant sur la consommation de viande et le risque de CCR. Ils ont mis en avant une association positive entre ces deux éléments. Deux autres études (LARSSON and WOLK 2006; NORAT *et al.* 2005) mettent en avant un risque relatif (RR) de CCR de 1,28 pour une consommation de viande rouge lorsqu'ils opposent les catégories de gros consommateurs et de faibles consommateurs, et un risque relatif de 1,20 pour les

consommateurs de viandes ayant subi un traitement. Il faut noter que ces études ont été réalisées sur des populations de pays industrialisés dont le risque de CCR est déjà élevé.

Les mécanismes qui lient la consommation de viande rouge et/ou cuite au risque de CCR ne sont pas encore élucidés. La consommation de viande est associée à l'augmentation de l'ingestion de composés carcinogènes, tels que les amines hétérocycliques (SUGIMURA *et al.* 1990), les hydrocarbures polycycliques aromatiques dérivés des réactions thermiques survenant durant la cuisson (CROSS and SINHA 2004), ainsi que la génération de radicaux libres par réaction de Fenton catalysée par le fer de l'hème (LUND *et al.* 1999; WURZELMANN *et al.* 1996). De plus, la plupart des viandes traitées contiennent un taux élevé de nitrites et de nitrates, qui, combinés à l'hème contribuent à la production de composés nitrosamines mutagènes (BINGHAM *et al.* 1996; CROSS *et al.* 2003). Les bactéries intestinales jouent également un rôle dans cette prédisposition, en générant des composés N-nitroso* potentiellement carcinogènes (HUGHES *et al.* 2001). Tous ces composés, nitrites, nitrates (WARD *et al.* 2007), hétérocycles aminés et tous les composés mutagènes dérivés de la viande (WU *et al.* 2006) contribuent modestement, mais significativement, au risque de polypes adénomateux associé à une forte consommation de viande.

La viande rouge est une source riche en fer qui indépendamment est associée à une augmentation du risque de CCR (NELSON 2001) et agit comme pro-oxydant dans la lumière intestinale (LUND *et al.* 1999). L'hypothèse par laquelle l'hème dérivé directement de la viande rouge dégrade la muqueuse colorectale (SESINK *et al.* 1999) est confirmée par le fait que la viande de volaille pauvre en hème (SANJOAQUIN *et al.* 2004) n'est pas associée à une augmentation du risque de CCR (NORAT *et al.* 2005).

Plusieurs études ont montré une association entre le risque de CCR et la consommation de viande rouge chez les acétylateurs rapides[†] (NAT1 et NAT2) et pour le phénotype de CYP1A2[‡] (CHEN *et al.* 1998b; KAMPMAN *et al.* 1999; LADERO *et al.* 1991; LANG *et al.* 1994; MARCHAND 1999; ROBERTS-THOMSON *et al.* 1996; SHIBUTA *et al.* 1994; SLATTERY *et al.* 2000b). En effet, les résultats préliminaires de le Marchand *et al.* (1999) montrent que les génotypes des acétylateurs rapides NAT et le phénotype rapide de CYP1A2 ont une activité augmentée chez les consommateurs de viandes rouges, ce qui prédisposerait ces individus à un risque élevé de CCR.

En conclusion il existe une relation très forte entre une forte consommation de viandes rouges cuites et le risque de CCR. Cette association est confirmée lorsqu'on compare l'incidence des

* N-nitroso : composés alkylants dont le plus connu est le N-Nitroso-N-Méthylurée (MNU).

† Les acétylateurs rapides jouent un rôle dans la détoxification des xénobiotiques.

‡ CYP1A2, membre de la famille des cytochromes P450 qui intervient dans la détoxification de l'organisme.

CCR dans les pays industrialisés (grands consommateurs de viande) et dans les pays d'Afrique et d'Asie en voie de développement (faibles consommateurs) (JOHNSON and LUND 2007).

II.5.2.2.2. Les glucides hautement assimilables et la charge glycémique importante

Dans de nombreuses publications, les patients diabétiques ont montré un risque modérément élevé pour le CCR (LE MARCHAND *et al.* 1997). Les patients non-diabétiques avec un taux de glucose postprandial élevé ont un risque plus élevé de cancer colorectal que les patients avec un taux de glucose normal (SCHOEN *et al.* 1999). Une explication de cette association entre diabète et CCR est que les deux maladies sont liées à un syndrome métabolique (BOWERS *et al.* 2006) et à des phénomènes inflammatoires (JOHN *et al.* 2006). On peut penser qu'une consommation très régulière (chronique) d'aliments riches en glucides rapidement assimilés pourrait promouvoir directement la carcinogenèse colorectale, probablement en prolongeant les phases d'hyperglycémie et d'hyperinsulinémie (GIOVANNUCCI 2001a).

II.5.2.3. Les micronutriments

En parallèle des études sur la relation entre le cancer colorectal et les principaux constituants de l'alimentation, de très nombreuses recherches ont été menées sur les micro-constituants, tels que les micronutriments essentiels et biologiquement actifs, dont la plupart sont des métabolites secondaires de plantes ou des métabolites « phytochimiques ».

II.5.2.3.1. Le calcium

Il a été montré à maintes reprises qu'une consommation relativement élevée en calcium et/ou en produits laitiers protégerait modestement contre le cancer colorectal (CHO *et al.* 2004; McCULLOUGH *et al.* 2003; PARK *et al.* 2007). La supplémentation en calcium par voie orale modifierait la prolifération des cellules des cryptes d'une manière consistant en la suppression de néoplasies (BOSTICK 1997). Au contraire, une alimentation pauvre en calcium pourrait mener à une inhibition de l'apoptose et à une augmentation de la prolifération cellulaire (MOBARHAN 1999; PIETINEN *et al.* 1999). L'effet du calcium serait soit direct, soit indirect, par précipitation des acides gras et des acides biliaires fécaux toxiques et cytotoxiques (WARGOVICH *et al.* 1983). A l'heure actuelle, le mécanisme d'action du calcium au niveau moléculaire n'est pas élucidé.

II.5.2.3.2. La vitamine D

Les chercheurs ont longtemps pensé que la vitamine D protégeait du CCR. Cette affirmation est sans doute un peu simpliste, car le statut en vitamine D d'un individu dépend à la fois de l'apport alimentaire et de la synthèse endogène en réponse à la lumière du soleil. Garland et Garland (1980) ont suggéré que les différences d'incidence du CCR liées à la géographie pourraient être attribuées

aux variations d'exposition au soleil. Cette explication ne tient toutefois pas lorsqu'on considère la forte incidence du cancer colorectal, en Australie et en Nouvelle-Zélande. Il a été démontré que la vitamine D réprimait la prolifération et favorisait la différenciation cellulaire (DIAZ *et al.* 2000; XUE *et al.* 1999), ce qui permet de croire en son action protectrice sur les CCR.

II.5.2.3.3. L'acide folique

L'acide folique, qui est obtenu uniquement à partir de l'alimentation ou de compléments alimentaires, est essentiel pour la synthèse et la réparation de l'ADN, grâce à l'enzyme MTHFR. Un déficit en acide folique peut mener à une mauvaise incorporation de l'uracile dans l'ADN et à un excès de cassures doubles-brins de l'ADN (WICKRAMASINGHE and FIDA 1994). Il peut également provoquer des anomalies de méthylation de la cytosine, tant au sein des îlots CpG, que tout au long de l'ADN non-codant. La méthylation anormale de l'ADN modifie l'expression des gènes. Ce mécanisme est une des principales causes de dommages génétiques conduisant au cancer (AMES 2006).

La diminution du risque de CCR associée aux légumes peut être attribuée en partie à la présence d'acide folique et de vitamine B₁₂, qui interviennent dans les voies de méthylation de l'ADN. Différentes études ont montré qu'une faible consommation d'acide folique augmentait le risque de cancers et d'adénomes colorectaux (BENITO *et al.* 1993; GIOVANNUCCI *et al.* 1995b; GIOVANNUCCI *et al.* 1998; TSENG *et al.* 1996). Cet effet de l'acide folique est mis en avant par Chen *et al.* (1996) et par Su et Arab (2001) qui ont montré une augmentation du risque de CCR chez les individus ayant une faible consommation d'acide folique, de méthionine et une forte consommation d'alcool.

II.5.2.3.4. Le sélénium

Le sélénium est présent dans les séléno-protéines, qui, par leur action enzymatique, interviennent dans une composante essentielle du système de défense antioxydant des mammifères et des glutathions peroxydases exprimées dans les tissus de l'organisme. La carence en sélénium chez les animaux est associée à des signes de dommages oxydatifs. Les composés du sélénium organique peuvent également inhiber la formation d'adduits à l'ADN covalent (EL-BAYOUMY *et al.* 1992). Il a également été montré que le sélénium pouvait induire l'apoptose dans les cellules tumorales *in vitro* (LU *et al.* 1995) et supprimer les aberrations de méthylation de l'ADN, en inhibant l'activité de l'ADN méthyltransférase (FIALA *et al.* 1998). Ces effets peuvent être considérés comme des mécanismes de suppression tumorale. Ils sembleraient ainsi qu'un faible taux en sélénium augmenterait le risque de nombreux cancers (WHANGER 2004).

II.5.2.3.5. Les antioxydants

Des radicaux libres très réactifs tels que le radical hydroxyle ($\cdot\text{OH}$), l'oxyde nitrique ($\text{NO}\cdot$) et le peroxydinitrite ($\text{ONOO}\cdot$) endogène sont générés par une variété de mécanismes, comprenant les processus inflammatoires (HUSSAIN *et al.* 2003), la radiolyse de l'eau par les rayonnements ionisants ou encore par des coproduits du métabolisme oxydatif ordinaire. Les cellules ont développé un arsenal de mécanismes antioxydants permettant de protéger les macromolécules des dommages des radicaux libres (HELBOCK *et al.* 1998). Les effets cumulatifs des dommages oxydatifs surviennent dans les protéines impliquées dans la régulation de la prolifération cellulaire et de l'apoptose (HOFSETH *et al.* 2003b), ainsi que sous la forme de cassures des brins d'ADN et de mutations (BJELLAND and SEEBERG 2003). Les défenses corporelles dépendent des nutriments antioxydants, comme l'acide ascorbique, l' α -tocophérol et le β -carotène, provenant de tous les aliments végétaux. Les fruits, les légumes et les boissons à base de plantes sont également riches en antioxydants phytochimiques comme les caroténoïdes et les polyphénols, dont beaucoup vont donner la saveur et la couleur aux aliments.

II.5.2.3.6. Les métabolites secondaires des plantes

Les plantes sont riches en substances biologiquement actives qui présentent une activité anti-carcinogénique *in vitro* et dans des modèles animaux de cancer (JOHNSON *et al.* 1994). Beaucoup de ces substances ont probablement évolué comme pesticides naturels permettant de repousser les invertébrés, mais elles peuvent interagir avec les voies biochimiques de nombreuses espèces de vertébrés, dont l'homme fait partie. Les produits de dégradation du glucosinolate, comme beaucoup d'autres composés phytochimiques, ont d'abord attiré l'attention des chercheurs par leur potentiel toxique. Il a ensuite été constaté que beaucoup de ces composés pourraient exercer des effets potentiellement bénéfiques, en agissant comme antioxydants naturels (SCALBERT *et al.* 2005). Ils pourraient être également impliqués dans les voies biochimiques de l'inflammation, de la prolifération cellulaire, de la différenciation et de l'apoptose (JOHNSON 2007). De nombreux métabolites secondaires des plantes, parmi lesquels le resvératrol, le limonène, le glycyrrhizin, le gingérole, l'indole-3-carbinol, la génistéine et l'apigénine inhiberaient différentes étapes de la voie de régulation NF- κ B (AGGARWAL and SHISHODIA 2006), un facteur de transcription impliqué fortement dans la carcinogenèse (KARIN *et al.* 2002). Les effets en aval des inhibiteurs naturels et synthétiques de NF- κ B réduisent l'expression d'enzymes pro-inflammatoires clés telles que l'iNOS et la COX2.

L'exemple le plus célèbre d'une activité biologique d'un métabolite secondaire de plantes est l'acide salicylique, qui a été initialement isolé à partir du saule (*Salix alba*). Le salicylate et les dérivés commerciaux de l'aspirine inhibent de façon irréversible les enzymes COX1 et COX2 et suppriment l'activité de NF- κ B (YIN *et al.* 1998). D'autres composés phytochimiques, tels que les flavonoïdes

représentés par exemple par la quercétine agissent en tant qu'inhibiteurs de l'enzyme COX2 (BAUMANN *et al.* 1980). Compte tenu de l'effet protecteur de l'aspirine contre le cancer colorectal, on pourrait très bien imaginer qu'un inhibiteur naturel de la COX2 pourrait exercer un effet similaire ; il n'a cependant pas été démontré que ces composés pouvaient atteindre des niveaux physiologiques élevés, que ce soit dans le plasma ou dans les matières fécales.

La cascade de signalisation Wnt a également été présentée comme étant sensible aux différentes substances phytochimiques d'origine alimentaire, notamment à la quercétine. L'interruption de la voie de signalisation Wnt est un élément clé lors du développement du CCR et d'autres cancers du tube digestif (PARK *et al.* 2005a).

Les quelques exemples cités ici ne donnent qu'un bref aperçu de la variété des constituants biologiquement actifs de plantes. Au-delà des effets constatés *in vitro*, on ignore toujours si ces composés ou leurs métabolites peuvent atteindre leur cible supposée à des concentrations physiologiquement actives. En effet, beaucoup de composés phytochimiques sont mal absorbés, rapidement métabolisés et excrétés dans la lumière de l'intestin (PETRI *et al.* 2003).

II.5.2.4. Les facteurs non-alimentaires

II.5.2.4.1. Les anti-inflammatoires non stéroïdiens

Plusieurs publications montrent que les anti-inflammatoires non stéroïdiens (AINS), dont l'aspirine, ont un effet protecteur vis-à-vis du cancer colorectal (GIOVANNUCCI *et al.* 1994; GREENBERG *et al.* 1993; LOGAN *et al.* 1993). Les AINS ont été montrés comme ayant une action anti-tumorigène grâce à la surexpression de l'activité de COX2 (DIMBERG *et al.* 2001; TUYNMAN *et al.* 2004), mais les doses auxquelles ils agissent ne sont pas connues. En effet, Gann *et al.* (1993) ont montré que les doses adéquates pour prévenir des infarctus du myocarde n'étaient pas suffisantes pour réduire l'incidence du CCR. Pour ces auteurs, l'effet de l'aspirine était statistiquement significatif pour une dose d'aspirine de plus de 300 mg/jour, prise pendant plus d'un an. Leur étude a permis de constater qu'une telle prise médicamenteuse faisait diminuer le risque d'adénomes colorectaux de 40% (GARCIA RODRIGUEZ and HUERTA-ALVAREZ 2000). Cette constatation a également été observée chez les individus avec une histoire familiale positive pour la pathologie (COOGAN *et al.* 2000), notamment chez les patients atteints de polypose adénomateuse familiale (PAF) prenant du celecoxib® (AINS) chez qui une réduction significative du nombre d'adénomes a été montrée (PHILLIPS *et al.* 2002).

L'effet anticancéreux des AINS peut être exercé via plusieurs voies, mais il a été particulièrement montré par l'inhibition des enzymes COX (cyclooxygénase) et LOX (lipoxygénase). Ces deux groupes d'enzymes jouent des rôles clés dans le métabolisme de l'acide arachidonique et

des autres acides gras polyinsaturés, produisant un arsenal de molécules à activité biologique qui interviennent dans divers processus tels que la réponse à l'inflammation. Il a été montré que cette dernière était étroitement associée à l'incidence des cancers dans de nombreux tissus (SHACTER and WEITZMAN 2002). L'inhibition des enzymes des voies COX et LOX entraverait ainsi le processus d'inflammation et empêcherait par là même la croissance tumorale.

Les cyclooxygénases existent sous plusieurs formes dont COX1 et COX2 qui convertissent l'acide arachidonique en prostaglandines et en tromboxanes. COX1 est exprimée normalement dans les tissus et joue un rôle normal dans la régulation, alors que COX2 est quasi-indétectable dans les cellules et tissus normaux (DANNHARDT and KIEFER 2001). COX2 a une expression inductible par une batterie de cytokines, de facteurs de croissance et de promoteurs tumoraux. Ainsi la production excessive de prostaglandines qui découle de l'expression de COX2 va mener à l'inflammation chronique, un facteur de risque de CCR unanimement reconnu (PRESCOTT and FITZPATRICK 2000).

Le mécanisme de COX2 n'est pas totalement connu dans la promotion de la tumorigenèse, mais il a été montré qu'elle pourrait inhiber l'apoptose (SUN *et al.* 2002) et promouvoir l'angiogenèse (INIGUEZ *et al.* 2003). De plus, une association entre une expression de COX2 élevée et des mutations d'APC a été suggérée ce qui met en avant le fait que COX2 est une cible en aval de la voie APC/ β -caténine/ TCF (DIMBERG *et al.* 2001). Cette voie métabolique est montrée comme étant atténuée lors de la prise d'AINS (DIHLMANN *et al.* 2001).

Le rôle des lipoxygénases et de leurs métabolites dérivés a également été proposé dans le développement des CCR et d'autres cancers. Les enzymes LOX travaillent en parallèle de la voie des COX, en convertissant l'acide arachidonique en leucotriènes et en acide hydroxyeicosatétraoïque. Plusieurs formes de LOX ont été montrées comme ayant un effet pro-carcinogène, suggéré par leur expression élevée dans les cancers (SHUREIQI and LIPPMAN 2001).

II.5.2.4.2. L'activité physique et l'indice de masse corporelle

Un certain nombre d'études indiquent que le surpoids et l'obésité sont positivement associés au risque de cancer du côlon. Une des premières études menées sur le sujet par Lee et Paffenbarger (1992) a montré une association positive entre le surpoids – calculé grâce à l'indice de masse corporelle (IMC) – et le risque de développer un cancer. En règle générale, les effets néfastes du surpoids et du manque d'activité physique ont été confirmés dans de nombreuses études (GIOVANNUCCI *et al.* 1995a; SLATTERY *et al.* 2003; THUNE and LUND 1996; WANNAMETHEE *et al.* 2001), qui montrent une plus grande significativité de cette association pour les hommes que pour les femmes (GERHARDSSON DE VERDIER *et al.* 1990; LE MARCHAND *et al.* 1997).

L'indice de masse corporelle est une mesure assez grossière de la surcharge pondérale. La répartition de la graisse corporelle serait en réalité un facteur plus important dans la pathogénèse de maladies impliquant l'obésité (JANSSEN *et al.* 2004). Les graisses viscérales abdominales représentent environ 18% de l'ensemble du tissu adipeux (ROSS *et al.* 1992) et exercent un effet sur divers aspects du métabolisme lipidique ainsi que l'étiologie de la maladie (KISSEBAH 1996). Le tour de taille est souvent utilisé pour fournir une mesure simple de l'adiposité abdominale. Une étude récente (PISCHON *et al.* 2006), menée pendant 6 ans sur environ 400000 hommes et femmes de neuf pays européens (population EPIC), a permis de déterminer que le poids et l'IMC étaient significativement associés à un risque de cancer du côlon chez les hommes (risque relatif (RR)=1,55), mais pas chez les femmes. Cette association a été confirmée en utilisant le tour de taille comme indicateur chez les hommes et les femmes pré-ménopausées mais pas chez les femmes ménopausées utilisant l'hormonothérapie substitutive.

Cette population EPIC a permis également de montrer l'effet de l'activité physique sur le risque de cancer colorectal (FRIEDENREICH *et al.* 2006). Cette étude a montré qu'environ deux heures d'activité physique modérée par jour ou une heure d'activité plus vigoureuse seraient associées à une réduction du risque de cancer du côlon de 20 à 25% et à une réduction d'environ 35% du risque de cancer du côlon droit. Ces effets ont été observés de manière plus marquée chez les individus ayant un IMC inférieur à 25 et un apport énergétique relativement faible. Pour la population en surpoids, cette association est tout de même réelle mais statistiquement moins significative.

Les preuves d'une relation causale entre l'obésité et le cancer du côlon s'accumulent et impliqueraient l'activité inflammatoire chronique asymptomatique de la muqueuse colique. Les maladies inflammatoires de l'intestin sont une cause reconnue de la carcinogenèse colorectale (MUNKHOLM 2003). Le tissu adipeux est une source importante de facteurs endocriniens appelés « adipokines », qui comprennent la leptine, l'adiponectine, et les cytokines (exemple : le facteur de nécrose tumorale alpha (TNF- α) et l'interleukine 6 (IL-6) (RONDINONE 2006; TRAYHURN and BEATTIE 2001). Chez les personnes obèses, il a également été montré que le tissu adipeux recruterait un grand nombre de macrophages qui sécrètent une grande variété de molécules-signal pro-inflammatoires et de cytokines (FANTUZZI 2005). L'obésité est de plus en plus considérée comme un état inflammatoire de bas grade dans lequel les effets néfastes sont exercés sur une variété d'organes cibles parmi lesquels le côlon (GUNTER and LEITZMANN 2006; JOHN *et al.* 2006). Cette hypothèse d'inflammation de bas grade est retenue dans l'étiologie des cancers colorectaux sporadiques (GUNTER *et al.* 2006; ITO *et al.* 2005; OTANI *et al.* 2006). Une hypothèse est que le milieu pro-inflammatoire associé à l'obésité induirait une cascade de signalisation dans la muqueuse, activant le facteur de transcription NF- κ B et augmentant l'expression de iNOS et de COX2 (JOHN *et al.*

2006). Ce mécanisme est censé supposer jouer un rôle important dans la suppression de l'apoptose (KARIN *et al.* 2002), qui est une condition préalable à la tumorigenèse (JOHNSON 2001).

Une preuve directe d'un lien entre l'état métabolique et les processus inflammatoires de la muqueuse colique est fournie par Poullis *et al.* (2004). Ces auteurs ont mesuré le taux de calprotectine fécale qui se lie au calcium et se retrouve dans les neutrophiles, les monocytes et les macrophages. La calprotectine fécale jouerait un rôle dans les processus inflammatoires comprenant ceux liés à la néoplasie épithéliale (GEBHARDT *et al.* 2006). Dans cette étude, le taux de calprotectine fécale a été trouvé positivement corrélé à l'âge, à l'obésité et à l'inactivité physique et inversement proportionnel à la consommation de fibres, de fruits et de légumes.

Les niveaux anormalement élevés de protéine réactive-C (CRP), de TNF- α et d'IL-6 qui sont observés dans le plasma des personnes obèses sont associés au développement de la résistance à l'insuline qui semble être elle-même associée à la vulnérabilité accrue au cancer du côlon (WEI *et al.* 2005). En revanche, cette observation n'a pas été confirmée pour le cancer du rectum (PALMQVIST *et al.* 2002). Cet état de résistance à l'insuline est caractérisé par une intolérance au glucose, une élévation du taux plasmatique d'insuline et d'IGF-1, mais également par un faible niveau des protéines se liant à IGF. L'insuline pourrait causer une hyper-prolifération des cellules épithéliales coliques *in vivo* (TRAN *et al.* 2006) et *in vitro*, en augmentant le taux de prolifération et d'apoptose (GIOVANNUCCI 2001a). Ces deux observations sont en faveur de la tumorigenèse. La leptine, qui est un produit de sécrétion des adipocytes, est également présente à des niveaux anormalement élevés dans le plasma des sujets obèses. Elle favoriserait la mitose et supprimerait l'apoptose dans l'épithélium colique (AMEMORI *et al.* 2007; HODA *et al.* 2007; OGUNWOBI and BEALES 2007).

II.5.2.4.3. Le tabac

Les fumeurs de tabac, en particulier ceux avec une longue histoire de fumeur, semblent avoir un risque augmenté d'adénomes et de cancers colorectaux (GIOVANNUCCI 2001b). La fumée de tabac est une source majeure de carcinogènes, tels que les nitrosamines, les hétérocycles aminés et les hydrocarbures polycycliques, qui sont métabolisés par les enzymes de détoxification telles que GSTM1 (LIN *et al.* 1995). Il a été estimé qu'environ 21% des instabilités microsatellitaires des tumeurs colorectales sont attribuables à la fumée de cigarettes (SLATTERY *et al.* 2000a). Wu *et al.* (2001a) ont montré que le risque d'instabilité dans les cancers du côlon était augmenté pour les patients qui consomment de la viande rouge et pour ceux qui fument. Les voies mises en jeu par l'organisme pour éliminer ces composés carcinogènes sont les mêmes que celles intervenant dans la prédisposition au CCR associé à la viande rouge.

II.5.2.4.4. L'alcool

La relation entre la consommation d'alcool et le risque de CCR reste controversée. En effet, des études rapportent une association positive (BARDOU *et al.* 2002; SHARPE *et al.* 2002) entre cette consommation et le CCR, alors que d'autres rapportent le contraire (BREUER-KATSCHINSKI *et al.* 2000; YE *et al.* 2003). Slattery *et al.* (2001) ont montré que la consommation d'alcool sur une longue durée augmentait la probabilité d'avoir des instabilités microsatellites, contrairement aux autres facteurs associés à l'augmentation de CCR.

Le mécanisme d'action de l'alcool dans les CCR n'est pas élucidé, mais l'alcool pourrait inhiber la réparation de l'ADN et les formes acétaldéhydes d'adduits à l'ADN. Il a aussi été émis que l'alcool pourrait influencer la prolifération des cellules muqueuses et être en relation avec les changements histologiques survenant dans les tissus tumoraux (KUNE and VITETTA 1992).

La consommation d'alcool, comme celle des autres facteurs alimentaires, peut également interagir avec des variants génétiques. En effet, l'association positive entre la consommation d'alcool et le variant hétérozygote du gène *ALDH2* a été mise en avant dans la population japonaise (MURATA *et al.* 1999). Cette étude montre également que la consommation d'alcool est souvent liée à une déficience en nutriments, tels que les vitamines, qui protègent du CCR. Une autre association entre la consommation d'alcool et le risque de CCR a été mise en avant pour un génotype variant de *MTHFR* (ULRICH *et al.* 1999).

II.5.2.4.5. L'hormonothérapie substitutive

Il apparaîtrait que l'hormonothérapie substitutive diminuerait le risque de cancer colorectal, d'après des études menées sur des utilisatrices d'hormonothérapie substitutive post-ménopausale (CRANDALL 1999; FRANCESCHI and LA VECCHIA 1998; NANDA *et al.* 1999). L'hormonothérapie substitutive augmenterait le taux d'œstrogènes endogènes, qui décline chez la femme post-ménopausée, et inhiberait par conséquent les récepteurs d'œstrogènes par hyperméthylation.

Ces différentes observations fournissent une base mécanistique plausible de la vulnérabilité accrue du cancer du côlon dans le monde industrialisé, avec des origines de stress métabolique associées à une forte consommation d'énergie et une faible activité physique. Le cancer du côlon pourrait donc être considéré comme une maladie occidentale au même titre que le diabète non-insulino dépendant, les maladies cardiovasculaires et le syndrome métabolique (HOTAMISLIGIL 2006).

II.5.3. Les facteurs génétiques de prédisposition

Comme nous avons pu le voir dans la définition des maladies complexes (Chapitre I : II.5.1. Définition d'une maladie complexe ou multifactorielle), le CCR sporadique fait intervenir des facteurs de prédisposition environnementaux que nous venons de détailler. La partie suivante va exposer le deuxième facteur de prédisposition, le facteur génétique et notamment les polymorphismes.

II.5.3.1. Définition d'un variant ou polymorphisme de prédisposition

Les formes sporadiques de cancers font intervenir des gènes de susceptibilité qui participent à la pathologie de manière cumulative mais avec un poids qui varie selon les formes polygéniques. Ces cancers sont à mettre en opposition avec les formes familiales à transmission mendélienne classique dont le terrain génétique est assez bien cerné aujourd'hui. Sans être responsables de la maladie elle-même, ces gènes de prédisposition peuvent précipiter ou venir contrecarrer le processus de tumorigenèse grâce à des allèles aggravants ou protecteurs. Ces allèles sont présents chez tous les sujets (indemnes ou atteints de la maladie) avec une fréquence dans la population générale supérieure à 1%. Ceux-ci ne vont pas entraîner obligatoirement une pathologie par leur présence contrairement aux variants à pénétrance forte et transmission mendélienne rencontrés dans les cancers familiaux. Ces variants de prédisposition vont, suivant le(s) gène(s) touché(s) et/ou leur localisation, influencer sur la voie réactionnelle dans laquelle ils se trouvent. Cet effet ne sera pas suffisant pour engendrer à lui seul la pathologie, il ne fera qu'augmenter le risque de survenue de la maladie. La pathologie sera donc due à l'effet cumulé de plusieurs polymorphismes de prédisposition et de facteurs environnementaux. Généralement, un polymorphisme de prédisposition va engendrer un risque relatif d'apparition de la maladie d'environ 1,2 contrairement aux variants de forte prédisposition qui donnent un risque relatif supérieur à 1,5. Chez les sujets malades, les polymorphismes de prédisposition de ces gènes sont capables de moduler la précocité d'apparition, la gravité ou l'expression de sous-types particuliers de la maladie. Les conséquences physiologiques peuvent n'apparaître que lorsque la fonction du gène devient indispensable pour compenser la maladie ou permettre la réponse adéquate à un substrat endogène ou à un médicament (JUNIEN 2001).

II.5.3.2. Les polymorphismes de prédisposition impliqués dans les cancers colorectaux

Les polymorphismes de prédisposition rencontrés dans les cancers colorectaux appartiennent à différentes voies biochimiques jouant un rôle essentiel dans ce type de cancers.

Les voies exposées ci-dessous représentent les voies majoritairement étudiées dans ces pathologies, il est bien entendu que d'autres voies pourraient être investiguées.

II.5.3.2.1. La voie de l'inflammation

Le développement des cancers dû à l'inflammation est un processus conduit par les cellules inflammatoires et qui fait intervenir une variété de médiateurs, tels que les cytokines, les chimiokines, et les enzymes, ce qui va créer un microenvironnement inflammatoire (COUSSENS and WERB 2002). Bien que cette réponse de l'hôte doit naturellement supprimer les tumeurs, elle peut également faciliter le développement du cancer par de multiples voies de signalisation (YANG *et al.* 2005).

a) L'inflammation

L'inflammation est un processus physiologique, en réponse aux dommages tissulaires résultant d'infection par des pathogènes microbiens, d'irritation chimique et/ou de blessures (PHILIP *et al.* 2004). Au stade très précoce de l'inflammation, les neutrophiles sont les premières cellules à migrer vers les sites inflammatoires sous la régulation des molécules produites par les macrophages et les mastocytes déjà présents dans les tissus (COUSSENS and WERB 2002; NATHAN 2002). La progression de l'inflammation va faire intervenir divers types de leucocytes (les lymphocytes) et de cellules inflammatoires qui seront activés et attirés sur le site de l'inflammation par un réseau de signalisation impliquant un grand nombre de facteurs de croissance, cytokines et chimiokines (COUSSENS and WERB 2002; NATHAN 2002).

b) L'inflammation et le cancer

L'association entre l'inflammation et le cancer a été mise en avant par les études épidémiologiques et cliniques : par exemple, le risque de CCR est 10 fois plus important pour les individus atteints de maladies inflammatoires de l'intestin comme les colites ulcéraives ou la maladie de Crohn (ITZKOWITZ and YIO 2004; SERIL *et al.* 2003) que pour les individus non atteints par ces pathologies.

Les causes de l'inflammation sont diverses, elles peuvent être microbiennes, ou physiques non-infectieuses et/ou irritantes par un produit chimique (PHILIP *et al.* 2004).

c) Les mécanismes mis en cause

L'inflammation chronique est caractérisée par des lésions tissulaires durables, dommages induits par la prolifération cellulaire et la réparation des tissus. La prolifération des cellules dans ce contexte est généralement corrélée avec la « métaplasie », un changement réversible de type

cellulaire. La « dysplasie » est un trouble de la prolifération cellulaire conduisant à la production de cellules atypiques (CORDON-CARDO and PRIVES 1999).

(1) Le potentiel mutagène de l'inflammation

Le microenvironnement de l'inflammation chronique est dominé par les macrophages (COUSSENS and WERB 2002; NATHAN 2002). Ces macrophages, en collaboration avec d'autres leucocytes, génèrent des niveaux élevés d'oxygène réactif (ROS) et des espèces azotées pour combattre l'infection (MAEDA and AKAIKE 1998). Toutefois, la persistance de ces agents de lutte contre l'infection ainsi que la continuité des lésions tissulaires et de prolifération cellulaire est délétère (OKADA 2002) par la production d'agents mutagènes, tels que le peroxy-nitrite, qui réagissent avec l'ADN et provoquent des mutations lors de la prolifération des cellules épithéliales et du stroma (MAEDA and AKAIKE 1998). Les macrophages et les lymphocytes T peuvent libérer du TNF- α et des facteurs inhibiteurs de la migration des macrophages qui aggravent les dommages à l'ADN (POLLARD 2004). Ces facteurs inhibant la migration sont dépendants de p53 ce qui entraîne l'accumulation de mutations oncogéniques (HUDSON *et al.* 1999). De plus, ils contribuent également à la tumorigenèse en interférant avec la voie Rb-E2F (PETRENKO and MOLL 2005).

(2) Rôle de l'inflammation dans le développement tumoral

Une grande population de leucocytes ainsi que d'autres cellules du système immunitaire infiltrent la tumeur et mettent en place son microenvironnement inflammatoire (YANG *et al.* 2005). Celui-ci peut faciliter la rupture de la membrane basale, un processus nécessaire pour l'invasion et la migration des cellules tumorales (COUSSENS and WERB 2002). Les macrophages, les neutrophiles, les éosinophiles, les cellules dendritiques, les mastocytes et les lymphocytes sont également reconnus pour être des éléments clés des tumeurs d'origine épithéliale (COUSSENS and WERB 2001; MACARTHUR *et al.* 2004; YANG *et al.* 2005).

L'infiltration de cellules immunitaires dans les tumeurs peut réprimer la croissance tumorale (CLARK *et al.* 1989; CLEMENTE *et al.* 1996) ou au contraire, avoir un rôle dans la promotion des tumeurs dans le cas des cancers associés à l'inflammation (COUSSENS and WERB 2002; KHONG and RESTIFO 2002; SMYTH *et al.* 2004). Les mutations accumulées dans les cellules épithéliales conduisent au dérèglement de leur croissance et de leur migration et vont alors provoquer le recrutement des leucocytes (LIN and POLLARD 2004). De plus, les cellules tumorales peuvent aussi produire des cytokines et chimiokines pour attirer les cellules du système immunitaire, telles que les macrophages (LEEK and HARRIS 2002; LIN *et al.* 2002; SAJI *et al.* 2001), afin de faciliter le développement du cancer (COUSSENS and WERB 2002; LIN and POLLARD 2004; YANG *et al.* 2005).

d) Les molécules clés**(1) Les cytokines**

Les cytokines, comprenant les interleukines (IL), le TNF- α , les facteurs de croissance et de différenciation, sont sécrétées ou liées à une molécule membranaire. Elles vont jouer un rôle de régulation de la croissance, de la différenciation et de l'activation des cellules immunitaires (DRANOFF 2004). La signalisation des cytokines pourrait contribuer à la progression tumorale suivant deux aspects : la stimulation de la croissance et la différenciation cellulaire ainsi que l'inhibition de l'apoptose des cellules modifiées sur le site inflammatoire (HUDSON *et al.* 1999; POLLARD 2004).

(2) Le facteur nucléaire κ B (NF- κ B)

NF- κ B est un terme général faisant référence à des facteurs de transcription dimériques de la famille Rel (KARIN and BEN-NERIAH 2000). Dans le cytoplasme, NF- κ B existe sous forme d'un complexe inactif NF- κ B-I κ B dans lequel I κ B inhibe NF- κ B (VIATOUR *et al.* 2005). En réponse à des stimuli extracellulaires, tels que les cytokines, I κ B est soumis à la phosphorylation, l'ubiquitination et la dégradation protéolytique et va libérer ainsi NF- κ B qui sera transloqué vers le noyau (BALDWIN 1996), où il se liera aux régions promotrices de ses gènes cibles (SWEENEY *et al.* 2004).

Ces cibles sont les gènes de la médiation immunitaire, de la réaction inflammatoire, des mécanismes anti-apoptotiques, de régulation de la prolifération cellulaire ainsi que les gènes codant pour les régulateurs antagonistes de NF- κ B (KARIN *et al.* 2002). NF- κ B agit également comme un médiateur de la progression de l'inflammation en régulant l'expression d'un large éventail de molécules inflammatoires, telles que des cytokines et les facteurs d'adhérence (PERKINS 2000; TAK and FIRESTEIN 2001).

(3) L'oxyde nitrique synthétase inductible (iNOS)

iNOS est une enzyme catalysant la production de monoxyde d'azote (NO) qui a été retrouvée surexprimée dans les maladies inflammatoires chroniques et divers types de cancers (KIM *et al.* 2005). Le monoxyde d'azote est une molécule importante de la régulation de la réponse inflammatoire (HUSSAIN *et al.* 2004) et du développement du cancer (HOFSETH *et al.* 2003a). iNOS est soumise à une induction par les cytokines pro-inflammatoires, telles que le TNF- α et l'IL-1 β (HUSSAIN *et al.* 2004), et à la transactivation par NF- κ B (LI and VERMA 2002). iNOS peut donc être un effecteur en aval des cytokines et de NF- κ B en reliant l'inflammation au cancer.

(4) La cyclooxygénase-2 (COX2)

L'expression de COX2 peut être induite par une large gamme de stimuli tels que le lipopolysaccharide, les cytokines pro-inflammatoires (IL-1, TNF), et les facteurs de croissance (KARIN *et al.* 2002; WILLIAMS *et al.* 1999). Les produits de COX2 sont les prostaglandines, qui sont des médiateurs de l'inflammation (NATHAN 2002; STEELE *et al.* 2003). Divers anti-inflammatoires non stéroïdiens (AINS) affectent l'activité de COX2 par modification covalente de celle-ci ou par concurrence pour un site de fixation du substrat. L'utilisation à long terme d'AINS a été montrée comme pouvant réduire le risque de plusieurs cancers dont le cancer colorectal (BUSKENS *et al.* 2002; FARROW *et al.* 1998; GIARDIELLO *et al.* 1995; WILLIAMS *et al.* 1999). COX2 est surexprimée dans de nombreux types de cancers et est impliquée dans divers processus tels que la prolifération cellulaire, l'angiogenèse, l'augmentation des métastases. Elle a également été montrée comme ayant une activité anti-apoptotique (EBERHART *et al.* 1994; PRESCOTT and FITZPATRICK 2000; TSUJII *et al.* 1997).

L'acide arachidonique est le substrat de la COX2 pour produire les prostaglandines. Cet acide arachidonique peut être transformé par une autre enzyme, la lipoxigénase, en leucotriènes. Ceux-ci ont été proposés comme étant un des chaînons manquants entre l'inflammation et le cancer (DuBOIS 2003).

(5) Le facteur-1 α inductible par l'hypoxie (HIF-1 α)

HIF-1 α est un facteur de transcription hétérodimérique largement accepté comme médiateur de l'homéostasie de l'oxygène (WANG *et al.* 1995a). L'hypoxie est un caractère résultant de modifications métaboliques au cours de l'inflammation (KONG *et al.* 2004) que l'on retrouve sur les sites de liaisons inflammatoires. En réponse à l'hypoxie, HIF-1 α active un large éventail de molécules réactives à l'hypoxie, telles que l'érythropoïétine, iNOS, VEGF, le transporteur 1 du glucose, et d'autres enzymes glycolytiques (SEMENZA 1999).

L'induction de NF- κ B par l'hypoxie est tributaire de l'existence de HIF-1 α (SEMENZA 1999). HIF-1 α est activé par les cytokines pro-inflammatoires (TNF- α et de l'IL-1 β) d'une manière dépendante de NF- κ B (JUNG *et al.* 2003a; JUNG *et al.* 2003b; ZHOU *et al.* 2003). Il a également été constaté que COX2 intervient dans un mécanisme faisant intervenir HIF-1 α par son produit, la prostaglandine E2 (JUNG *et al.* 2003b). HIF-1 α joue également un rôle essentiel dans le développement tumoral, parce qu'il facilite le phénotype des cellules cancéreuses grâce à un renforcement de l'activité glycolytique et l'augmentation de la transcription de VEGF, un puissant facteur angiogénique, important dans les métastases et la croissance tumorale (JAIN 2002).

La Figure 27 montre les différents mécanismes et molécules mis en jeu dans la voie de l'inflammation menant au cancer.

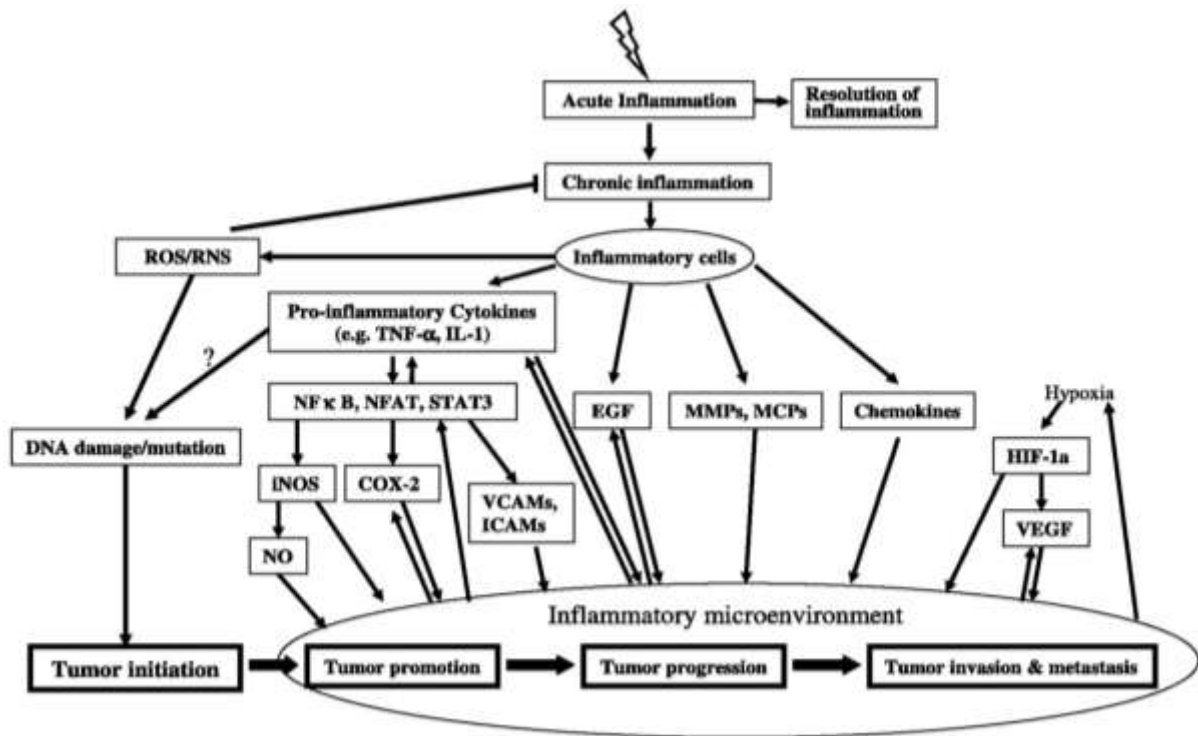


Figure 27 : Mécanismes participant à l'inflammation dans le développement du cancer (infections microbiennes, irritations chimiques, blessures des tissus,...). « *Tumor promotion* » indique le processus au cours duquel se développent les lésions bénignes dans les cellules. « *Tumor progression* » définit le processus au cours duquel les tumeurs bénignes progressent en carcinomes malins (Lu *et al.* 2006).

e) Les polymorphismes des gènes la voie de l'inflammation

Différents polymorphismes des gènes de cette voie ont été montrés comme pouvant avoir un rôle dans le cancer colorectal. C'est ainsi que les gènes *COX* (*PTG*) ont été investigués dans différentes études. Goodman *et al.* (2004) ont montré que l'allèle variant du polymorphisme Val444Ile du gène *PTGS1* avait un OR de 0,52 pour le génotype hétérozygote mais les résultats ne sont pas significatifs. De la même manière, Landi *et al.* (2006; 2003) ont étudié deux SNP (p.Leu237Met et p.Val481Ile) dans les gènes *COX1* avec une prédisposition significative pour le polymorphisme Leu237Met, et pour deux autres polymorphismes (c.-6457T et p.Val511Ala) du gène *COX2* qui ne sont pas retrouvés associés au CCR. Dans cette même publication, l'auteur a également étudié d'autres polymorphismes dans des gènes appartenant à la voie de l'inflammation tels que le rs1801282 (p.Pro12Ala) du gène *PPARγ* qui est retrouvé comme étant protecteur de manière significative dans les CCR (OR=0,55), mais également un polymorphisme du gène *IL6* (rs1800795) et du gène *TNFα* (rs3091256) qui sont retrouvés respectivement prédisposant avec un OR de 1,4 et protecteur avec un OR de 0,85, mais ces associations ne sont toujours pas significatives. Goodman *et al.* (2004) dans la même étude que le gène *PTGS1* ont étudié d'autres polymorphismes dans les lipoxgénases 5 et 12 (*ALOX5* et *ALOX12*) montrés comme protecteurs pour les deux polymorphismes d'*ALOX5* (c.-1752G>A, OR=0,6 ; c.-1699G>A, OR=0,56), et comme n'ayant pas d'effet

sur le CCR pour le SNP Gln261Arg du gène *ALOX12*. Tomlinson *et al.* (1996) ont également étudié un gène de prédisposition appartenant à cette voie, le gène *PLA2G2A*, codant pour une phospholipase A2. Les deux SNP étudiés (p.Cys964Gly et p.Gly1073Cys) n'ont pas montré d'association avec le cancer colorectal.

Le Tableau 6 récapitule les différents effets des SNP recensés dans les publications.

Gène	SNP	Effet	Significatif Oui/Non	Référence bibliographique
<i>PTGS1</i> (<i>COX1</i>)	p.Val444Ile	Protecteur pour les hétérozygotes	Non	Goodman <i>et al.</i> (2004)
	p.Leu237Met	Prédisposant	Oui	Landi <i>et al.</i> (2006; 2003)
	p.Val481Ile		Non	Landi <i>et al.</i> (2006; 2003)
<i>PTGS2</i> (<i>COX2</i>)	c.-6457T	N.D	Non	Landi <i>et al.</i> (2006; 2003)
	p.Val511Ala	N.D	Non	Landi <i>et al.</i> (2006; 2003)
<i>PPARγ</i>	p.Pro12Ala p.Pro12Ala	Protecteur	Oui	Landi <i>et al.</i> (2006; 2003)
<i>IL6</i>	rs1800795	Prédisposant	non	Landi <i>et al.</i> (2006; 2003)
<i>TNFα</i>	rs3091256	Protecteur	Non	Landi <i>et al.</i> (2006; 2003)
<i>ALOX5</i>	c.-1752G>A	Protecteur	Oui	Goodman <i>et al.</i> (2004)
	c.-1699G>A	Protecteur	Oui	Goodman <i>et al.</i> (2004)
<i>ALOX12</i>	p.Gln261Arg	N.D	Non	Goodman <i>et al.</i> (2004)
<i>PLA2G2A</i>	p.Cys964Gly	N.D	Non	Tomlinson <i>et al.</i> (1996)
	p.Gly1073Cys	N.D	Non	Tomlinson <i>et al.</i> (1996)

Tableau 6 : Polymorphismes des gènes de la voie l'inflammation (N.D : Non Défini).

II.5.3.2.2. La voie de détoxication

a) Les xénobiotiques

Les xénobiotiques sont des molécules étrangères à l'organisme de faible masse moléculaire, tels que des médicaments, des polluants, des additifs alimentaires mais également certains composés naturellement présents dans les aliments. Tous les êtres vivants ont donc dû développer des mécanismes pour faire face à certaines de leurs propriétés potentiellement néfastes pour le corps humain comme leur réactivité chimique avec certains constituants cellulaires, ou l'hydrophobie qui empêchera leur élimination (BEAUNE and LORiot 2000).

b) Mécanisme de détoxication mettant en jeu les enzymes du métabolisme des xénobiotiques (EMX)

Les xénobiotiques étant d'une grande diversité, les êtres vivants ont développé un arsenal d'enzymes pouvant réaliser un grand nombre de réactions chimiques sur des substrats ayant des structures très diverses. Ces enzymes ont été classées en différentes catégories selon leur fonction dans la voie de détoxication comme le montre la Figure 28.

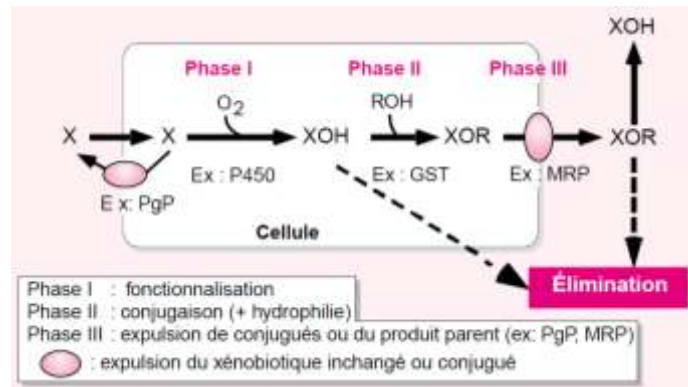


Figure 28 : Illustration des différentes phases du métabolisme des xénobiotiques. La pénétration des xénobiotiques dans la cellule se fait grâce à leur caractère hydrophobe. Ils peuvent soit être expulsés de la cellule par une pompe P-glycoprotéine (PgP), soit être fonctionnalisés par des enzymes de phase I (CYP, monooxygénases à cytochromes P450). Le produit obtenu sera soit conjugué par des enzymes de phase II, soit éliminé. Le nouveau produit pourra à son tour être éliminé de la cellule soit directement, soit par des protéines dites de phase III comme le transporteur *mrp* (*multidrug related protein*). Les produits finaux seront éliminés dans la bile ou les urines (BEAUNE and LORiot 2000).

Les xénobiotiques pénètrent facilement dans la cellule grâce à leur hydrophobicité, mais des pompes telles que la PgP (P-glycoprotéine) peuvent également les expulser. Une fois dans la cellule, ils sont pris en charge par les enzymes de phase I, dites de fonctionnalisation, dont les plus importantes sont les cytochromes P450 (CYP) catalysant des réactions de monooxygénation (BEAUNE 1999). Les enzymes de phase II vont neutraliser un groupement réactif (thiol, amine, aldéhyde) ou rendre hydrophile le xénobiotique, fonctionnalisé ou non, afin de faciliter son élimination en le conjuguant à un groupement glutathion, acide glucuronique, méthyle, acétyle,... La dernière étape de ce processus fait intervenir les enzymes de phase III pour les métabolites très hydrophiles. Ces enzymes sont des protéines telles que le transporteur *mrp* (*multidrug related protein*) (KEPPLER 1999; STIEGER and MEIER 1998). Le métabolite final de ce processus pourra être éliminé dans la bile ou les urines, ou, après transport, être métabolisé une nouvelle fois dans de nouveaux tissus possédant un autre spectre d'enzymes du métabolisme des xénobiotiques (EMX) (phase IV) (BADAWI *et al.* 1996). Toutes les étapes de cette voie ne sont pas indispensables, car le métabolisme de chaque composé est particulier.

Les nombreux métabolites produits peuvent parfois être plus ou moins toxiques que le produit de départ. Par exemple, le métabolite d'un médicament peut être plus ou moins actif que la molécule initiale. La Figure 29 montre l'équilibre qui existe entre les métabolites toxiques et non toxiques, actifs et inactifs. Cet équilibre dépend de la nature, de la quantité et de la qualité des enzymes du métabolisme des xénobiotiques dont l'expression varie fortement en fonction de divers facteurs, génétiques, environnementaux et physiopathologiques (BEAUNE and LORiot 2000).

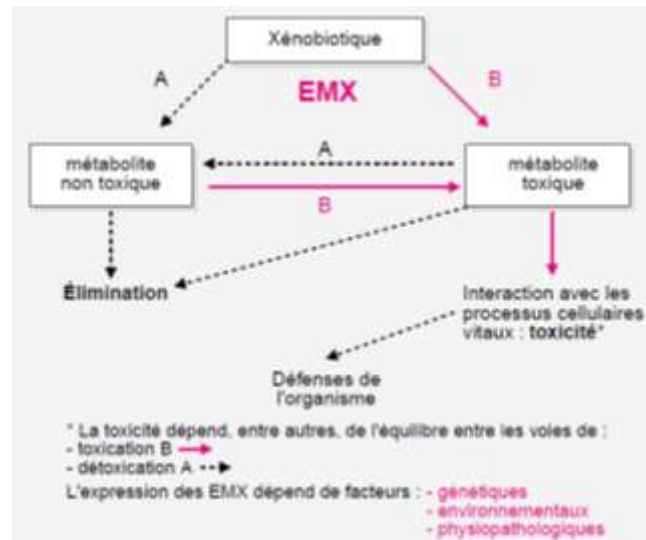


Figure 29 : Équilibre entre l'activité toxique des métabolites et la détoxification par les enzymes. Les xénobiotiques sont transformés en métabolites plus ou moins toxiques. L'équilibre entre ces voies toxiques et de détoxification dépend des caractéristiques des enzymes du métabolisme des xénobiotiques (nature, quantité, qualité). Cette expression est très variable, en fonction des facteurs génétiques, environnementaux et physiopathologiques qui vont intervenir dans la susceptibilité individuelle aux xénobiotiques (BEAUNE and LORiot 2000).

La Figure 30 montre durant quelle phase de la détoxification et sur quel composé les gènes de cette voie exercent leur activité. Il est ainsi plus facile de se rendre compte de l'effet qu'aura un polymorphisme qui modulerait l'expression d'un des gènes de cette voie.

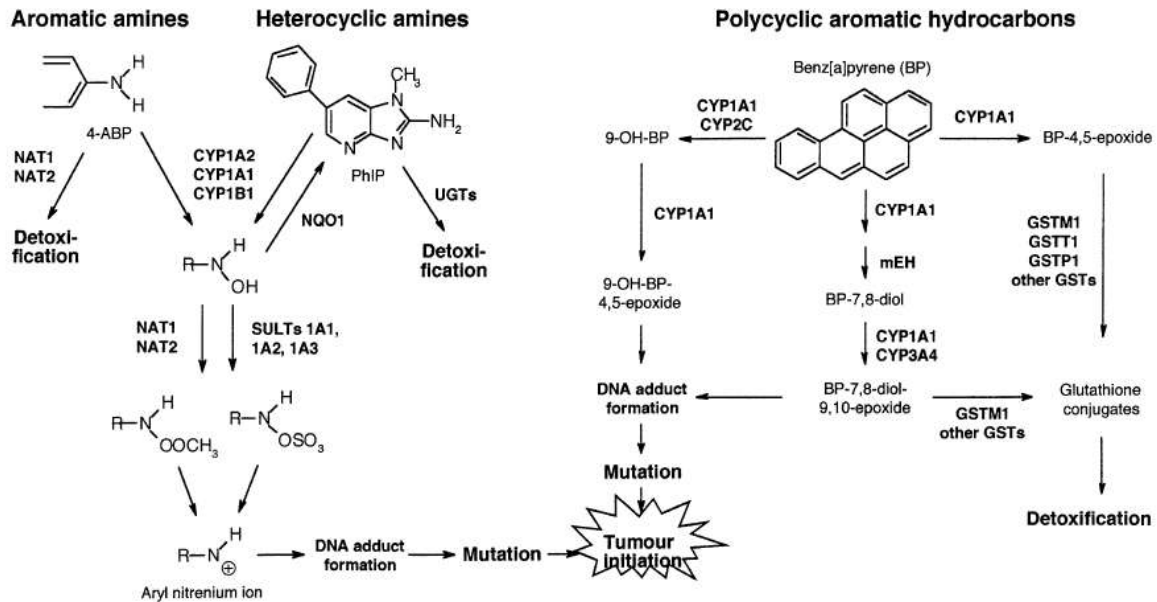


Figure 30 : Les différentes voies de détoxification qu'utilise l'organisme pour éliminer les métabolites carcinogènes. Par souci de simplification, les voies utilisant les enzymes UGT et NAT ne sont pas indiquées ici (SACHSE *et al.* 2002).

c) Les polymorphismes des gènes de la voie de détoxification des xénobiotiques

Les polymorphismes appartenant aux gènes de cette voie ont été très largement étudiés. En effet, la seule publication de Sachse *et al.* (2002) montre l'étude de 14 SNP appartenant au cytochrome P450 (CYP), quatre SNP dans les glutathion S-transférases (GST) et neuf autres

polymorphismes appartenant à divers gènes de cette voie tels que : *NQO1*, *NAT2*, *SULT1A1*, *SULT1A2*, *EPHX1*. Ainsi, deux SNP des cytochromes P450 ont montré un effet significatif, prédisposant pour *CYP1A1*2B* (m2) (OR=1,84) et protecteur pour *CYP2C19*2* (OR=0,68). Des polymorphismes de *GSTM1* et *EPHX1* (p.Tyr113Leu) ont été montrés prédisposants de manière significative avec des OR=1,33 et OR=1,15. Le polymorphisme G1259C du gène *CYP2E1* est également retrouvé comme prédisposant pour le CCR par Kiss *et al.* (2000) (OR=1,91). Ce même effet prédisposant significatif est retrouvé pour plusieurs polymorphismes des *GST* et des *NAT* dans différentes publications : allèle nul de *GSTM1* par Zhong *et al.* (1993) OR=1,78 et Gawronska-Szklarz *et al.* (1999) OR=2,53 ; codon 105 de *GSTP1* par Harris *et al.* (1998) OR=1,78 ; allèle nul de *GSTT1* par Deakin *et al.* (1996) OR=1,88, Zhang *et al.* (1999) OR=4,49 et Butler *et al.* (2001) OR=2,18. Pour l'acétylateur rapide *NAT1* un OR de 1,92 est retrouvé par Bell *et al.* (1995) et pour *NAT2* des OR de 2,48, 2,41 et 1,8 sont retrouvés respectivement par Lang *et al.* (1986), Ilett *et al.* (1987) et Roberts-Thomson *et al.* (1996).

Les polymorphismes exposés ici ont été trouvés significatifs dans les publications citées, mais d'autres études ont montré ces mêmes polymorphismes avec une significativité nulle.

Le Tableau 7 récapitule les différents effets des SNP recensés dans les publications.

Gène	SNP	Effet	Significatif Oui/Non	Référence bibliographique
<i>CYP1A1*2B</i>	c.2455A>G	Protecteur	Oui	Sachse <i>et al.</i> (2002)
<i>CYP2C19*2</i>	c.681G>A	Protecteur	Oui	Sachse <i>et al.</i> (2002)
<i>CYP2E1</i>	c.1259G>C	Prédisposant	Oui	Kiss <i>et al.</i>
<i>GSTM1</i>	Allèle nul	Prédisposant	Oui	Sachse <i>et al.</i> (2002) Zhong <i>et al.</i> (1993) Gawronska-Szklarz <i>et al.</i> (1999)
<i>GSTP</i>	Codon 105	Prédisposant	Oui	Harris <i>et al.</i> (1998)
<i>GSTT1</i>	Allèle nul	Prédisposant	Oui	Deakin <i>et al.</i> (1996) Zhang <i>et al.</i> (1999) Butler <i>et al.</i> (2001)
<i>NAT1</i>	Phénotype rapide	Prédisposant	Oui	Bell <i>et al.</i> (1995)
<i>NAT2</i>	Phénotype rapide	Prédisposant	Oui	Lang <i>et al.</i> (1986) Ilett <i>et al.</i> (1987) Roberts-Thomson <i>et al.</i> (1996)
<i>EPHX1</i>	p.Tyr113Leu	Prédisposant	Oui	Sachse <i>et al.</i> (2002)

Tableau 7 : Polymorphismes des gènes de la voie de détoxification des xénobiotiques.

II.5.3.2.3. La voie des radicaux mono-carbonés

a) Les folates

Le terme « folate » est le terme générique utilisé pour la famille des vitamines B composées d'un cycle aromatique ptéridine lié à l'acide p-aminobenzoïque et d'un résidu de glutamate (SHANE 1995). Après absorption dans l'intestin grêle, le folate contenu dans les aliments est soumis à une hydrolyse en méthyltétrahydrofolate (méthylTHF) qui est la forme prédominante de folate dans le

plasma (SHANE 1995). L'acide folique subit également une conversion en méthylTHF. Ce processus est saturé à des doses proches de 270 µg. A des niveaux plus élevés, l'acide folique est transporté directement dans le plasma (SWEENEY *et al.* 2007). Par conséquent, l'ingestion quotidienne de 400 µg, une dose couramment utilisée dans les suppléments, produit un niveau plasmatique important de l'acide folique.

b) Le mécanisme des radicaux mono-carbonés

Les folates cellulaires agissent comme des donneurs et des accepteurs de groupements méthyles dans la biosynthèse des précurseurs de nucléotides utilisés lors de la synthèse de l'ADN, et comme réserve de groupements méthyles pour la méthylation de l'ADN, de l'ARN et des protéines (Figure 31). Ces processus cellulaires importants sont le résultat du métabolisme des folates reliés par la méthylènetétrahydrofolate réductase (MTHFR), qui catalyse la conversion irréversible du 5,10-méthylèneTHF en 5-méthylTHF. Le substrat de MTHFR (5,10-méthylèneTHF) est également un substrat pour la thymidylate synthétase (TS), enzyme de la méthylation de la désoxyuridine monophosphate (dUMP) en désoxythymidine monophosphate (dTMP), qui est la seule source *de novo* de thymidine. Cette étape est limitante dans la synthèse de l'ADN des cellules de mammifères (CHOI and MASON 2002). Le 5,10-méthylèneTHF est aussi utilisé dans la production de formylTHF, qui est à son tour utilisé dans la synthèse *de novo* de purines. Le produit de MTHFR, le 5-méthylTHF, va donner un groupe méthyle à l'homocystéine grâce à la méthionine synthétase (MTR) qui deviendra la méthionine. La méthionine sera alors adénylée pour former la S-adénosylméthionine (SAM), qui est le donneur de groupe méthyle dans les réactions de méthylation de l'ADN et de l'ARN, alors que SAM inhibe la MTHFR par une boucle de rétroaction négative (HUBNER and HOULSTON 2009).

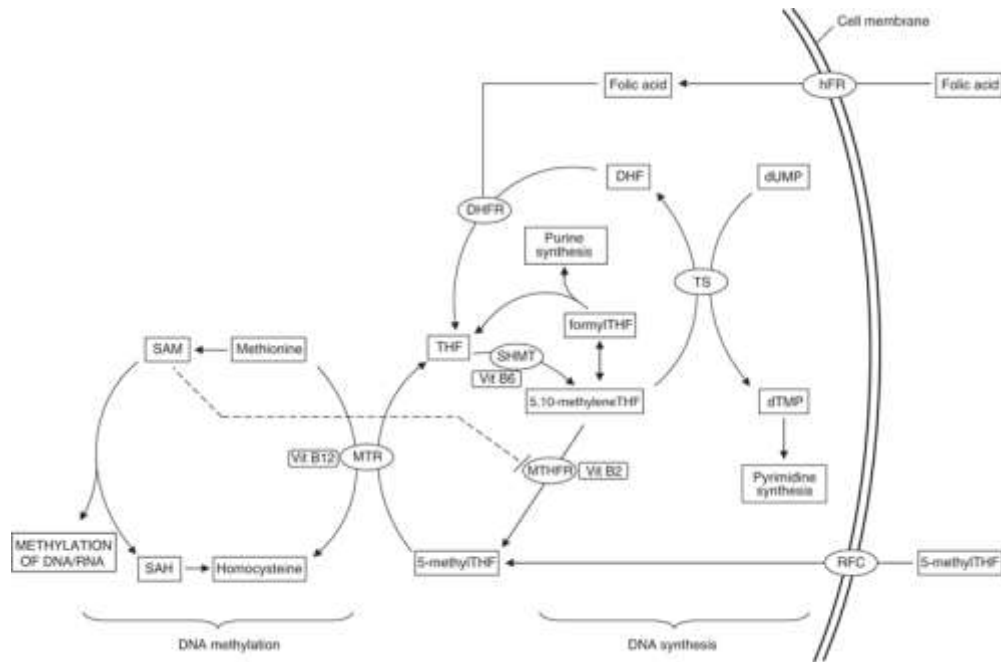


Figure 31 : Représentation schématique du métabolisme des folates illustrant l'entrée des folates et de l'acide folique dans la voie, et le flux des groupements méthyles, soit vers la synthèse de l'ADN soit vers sa méthylation. (RFC, transporteur réduit de folates ; hFR, récepteur des folates humain ; MTR, méthionine synthétase ; MTHFR, méthylène-tétrahydrofolate réductase ; SHMT, sérine hydroxyméthyltransférase ; TS, thymidylate synthétase ; THF, tétrahydrofolate ; DHF, dihydrofolate ; SAM, S-adenosylméthionine ; SAH, S-adenosylhomocystéine ; dUMP, désoxyuridine monophosphate ; dTMP, désoxythymidine monophosphate) (HUBNER and HOULSTON 2009).

La variation dans la distribution des groupements méthyles lorsque le métabolisme des folates est altéré a un impact sur la méthylation et la synthèse de l'ADN, ces deux processus cellulaires étant essentiels dans la transformation néoplasique (CHOI and MASON 2002). La méthylation des résidus cytosine des îlots CpG est un processus épigénétique qui est important dans l'expression des gènes et qui joue également un rôle dans le maintien de la stabilité de l'ADN. L'hyperméthylation des régions promotrices des gènes se traduit par une perte de fonction des gènes suppresseurs de tumeur, alors qu'une réduction globale de la méthylation donne des instabilités chromosomiques et une augmentation des événements mutationnels (EDEN *et al.* 2003).

La réduction de la synthèse de désoxythymidylate conduisant à une mauvaise incorporation d'uracile au cours de la synthèse de l'ADN est un second mécanisme par lequel une carence en folates peut perturber l'intégrité de l'ADN et promouvoir la cancérogenèse. Une augmentation en uracile de l'ADN et des cassures chromosomiques surviennent chez les individus déficients en folates ce qui peut être restauré par l'augmentation de la prise de folates (BLOUNT *et al.* 1997).

c) Les polymorphismes des gènes de la voie des radicaux monocarbonés

Dans cette voie le gène le plus étudié est *MTHFR*, et notamment le polymorphisme Cys677Thr qui a été démontré comme diminuant l'activité de l'enzyme (FROSST *et al.* 1995). Cette diminution d'activité n'est pas forcément liée à un effet protecteur ou prédisposant de ce

polymorphisme pour le CCR. En effet, différentes publications montrent un effet protecteur par exemple, Park *et al.* (1999) (OR=0,81) et Slattery *et al.* (1999) (OR=0,83), mais d'autres études montrent un effet prédisposant comme celles de Chen *et al.* (1998a) (OR=1,44) et de Marugame *et al.* (2000) (OR=1,24). Cette grande différence dans l'effet de ce polymorphisme doit être due au biais inhérent aux études d'association (cohortes trop petites, mixité des populations). D'autres gènes ont également montré des polymorphismes associés au risque de CCR, c'est le cas d'un SNP (p.Ala2756Gly) du gène *MTR* qui est montré protecteur dans différentes études (CHEN *et al.* 1998a; MA *et al.* 1999), mais également le SNP Ala66Gly du gène *MTRR* qui montre un risque relatif de 1,4 dans une étude cas-témoins réalisée aux USA (LE MARCHAND *et al.* 2002a). Des polymorphismes complexes des gènes *TYMS* et *CBS* ont également été étudiés. La répétition en tandem de 28pb du gène *TYMS* est montrée comme protectrice (OR=0,9) (ULRICH *et al.* 2002) et la délétion de 6pb du gène *TYMS* comme prédisposante (OR=1,13) (ULRICH *et al.* 2002), alors que l'insertion de 68pb dans le gène *CBS* est montrée comme étant en association avec le CCR par le Marchand *et al.* (2002a) et Shannon *et al.* (2002).

Le Tableau 8 récapitule les différents effets des SNP recensés dans les publications.

Gène	SNP	Effet	Significatif Oui/Non	Référence bibliographique
<i>MTHFR</i>	p.Cys677Thr	Protecteur	Non	Park <i>et al.</i> (1999) Slattery <i>et al.</i> (1999)
		Prédisposant	Non	Chen <i>et al.</i> (1998a) Marugame <i>et al.</i> (2000)
<i>MTR</i>	p.Ala2756Gly	Protecteur	Non	Chen <i>et al.</i> (1998a) Ma <i>et al.</i> (1999)
<i>MTRR</i>	p.Ala66Gly	Prédisposant	Non	Le Marchand <i>et al.</i> (2002a)
<i>TYMS</i>	Répétition de 28pb	Protecteur	Non	Ulrich <i>et al.</i> (2002)
	Délétion de 6pb	Prédisposant	Non	Ulrich <i>et al.</i> (2002)
<i>CBS</i>	Insertion de 68pb	Prédisposant	Non	Marchand <i>et al.</i> (2002a) Shannon <i>et al.</i> (2002)

Tableau 8 : Polymorphismes des gènes de la voie des radicaux monocarbonés.

II.5.3.2.4. Voie de l'insuline

a) Mécanisme

L'insuline et les IGF (*insulin-like growth factors*) jouent chacun un rôle important et complémentaire dans le métabolisme et la croissance. L'insuline influence le métabolisme à court terme (par exemple, après un repas), alors que les IGF exercent un effet à plus long terme sur la croissance. IGF-1 inhibe l'apoptose et est nécessaire pour la progression du cycle cellulaire (AARONSON 1991). Plus de 90% de l'IGF-1 circulant est complexé avec l'IGFBP-3 (*insulin-like growth factor binding protein-3*). La plupart des IGF-1 et IGFBP-3 présents dans la circulation sont produits par le foie, et sont régulés par des hormones de croissance (JONES and CLEMMONS 1995). Les actions des IGFBP peuvent s'opposer à celles de l'IGF-1 en partie en se liant à l'IGF-1 et en réduisant le taux

d'IGF-1 libre (RECHLER 1997). En outre, IGFBP-3 peut inhiber directement les cellules cibles (RAJAH *et al.* 1997). Dans les tissus, les protéases IGFBP améliorent la disponibilité d'IGF-1 par clivage de l'IGFBP, augmentant ainsi la concentration d'IGF-1 libre (JONES and CLEMMONS 1995). L'activité biologique de l'IGF est déterminée par les actions d'IGF-1 et IGFBP circulant, et par la production locale de l'IGF, IGFBP et de protéases IGFBP.

Les cellules de l'épithélium colorectal normal et cancéreux expriment *in vitro* les récepteurs de l'IGF-1 ; lorsqu'il est activé par l'IGF-1, le complexe récepteur-ligand inhibe l'apoptose et permet la progression du cycle cellulaire (BJORK *et al.* 1993; GUO *et al.* 1992; KOENUMA *et al.* 1989; POLLAK *et al.* 1987; WATKINS *et al.* 1990). Ainsi, l'IGF-1 peut influencer sur les étapes précancéreuses et cancéreuses. La carcinogenèse colorectale résulte de l'accumulation de modifications moléculaires spécifiques (VOGELSTEIN *et al.* 1988) et de l'augmentation du renouvellement cellulaire pouvant augmenter la vitesse à laquelle ces modifications s'accumulent. Les IGF pourraient influencer la carcinogenèse en étant l'un des déterminants les plus importants du turnover cellulaire. Dans des lignées cellulaires de cancer du côlon, l'IGF-1 augmente la production du facteur de croissance vasculaire endothélial, un facteur angiogénique qui soutient la croissance tumorale (WARREN *et al.* 1996).

De la même manière qu'IGF-1, l'insuline stimule la croissance des cellules coliques normales et cancéreuses *in vitro* (BJORK *et al.* 1993; KOENUMA *et al.* 1989; WATKINS *et al.* 1990). Bien que le tissu colique cancéreux possède les récepteurs de l'insuline et d'IGF-1 (GUO *et al.* 1992; MACDONALD *et al.* 1993), les propriétés mitogènes de l'insuline, contrairement à ses propriétés de contrôle de la glycémie, peuvent être médiées par les récepteurs d'IGF-1 (MACAULAY 1992). Toutefois, les effets mitogènes surviennent à un taux supraphysiologique d'insuline (FLIER *et al.* 1986; MOLLER and FLIER 1991); ainsi, d'autres médiateurs des effets mitogènes de l'insuline sont susceptibles d'être opérationnels. L'insuline augmente l'IGF-1 bioactif par le biais de divers mécanismes. L'hormone de croissance est le principal régulateur de la production hépatique d'IGF-1, et le nombre de récepteurs hépatiques de l'hormone de croissance est en partie régi par l'insuline (JONES and CLEMMONS 1995; SMITH *et al.* 1995; UNDERWOOD *et al.* 1994).

La voie de transduction du signal de l'insuline et d'IGF-1, impliqués dans la régulation de l'expression des gènes et l'effet mitogène, est médiée par l'activation de la protéine Ras (BURGERING *et al.* 1991; JHUN *et al.* 1994). Les mutations *ras*, qui augmentent l'activité de la protéine Ras, surviennent dans environ la moitié des cancers du côlon (BOS 1988) et peuvent favoriser la croissance des adénomes (VOGELSTEIN *et al.* 1988). En plus d'un effet mitogène direct, l'insuline peut préparer les cellules aux effets de certains facteurs de croissance en influençant la farnésylation* de Ras

* Farnésylation : Addition d'un groupement C15-prenyl (farnésyl) sur le résidu carboxy-terminal d'une protéine.

(GUTIERREZ *et al.* 1989). La farnésylation de Ras, catalysée par une farnésyl-transférase, détermine sa localisation dans la membrane plasmique et est nécessaire pour la fonctionnalisation et la transformation des protéines Ras. L'insuline augmente le pool de protéines Ras farnésylées *in vitro* (LEITNER *et al.* 1997) et *in vivo* (GOALSTONE *et al.* 1999), et ainsi amorce la réponse cellulaire des facteurs de croissance qui utilisent cette voie Ras, dont IGF-1. Parmi les différents facteurs de croissance examinés, l'insuline semble capable de stimuler la farnésyl-transférase.

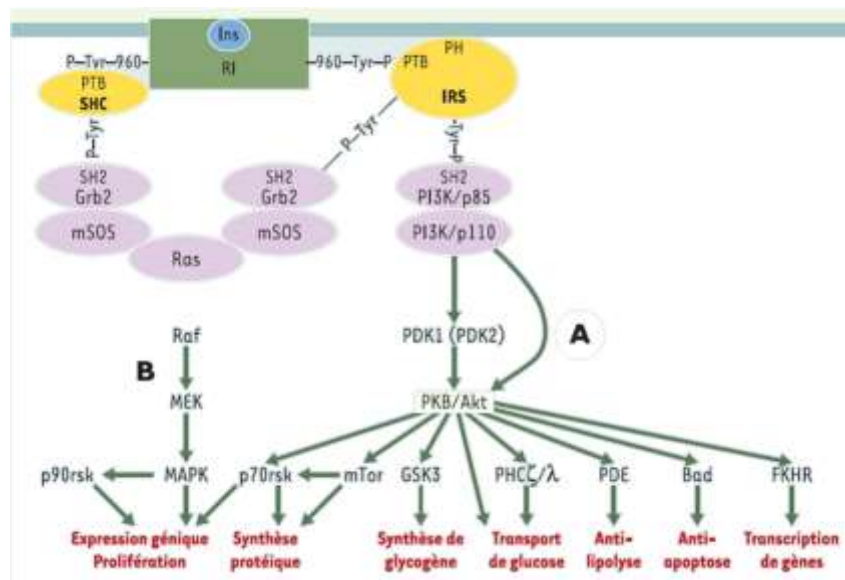


Figure 32 : Principales voies de signalisation de l'insuline : la voie PI3 kinase et la voie MAP kinase. A. L'insuline en se fixant sur son récepteur (RI) va entraîner une cascade de phosphorylation/déphosphorylation de diverses protéines qui par leur action sur l'organisme vont engendrer des effets anti-apoptotique et anti-lipolytique mais également la transcription de gènes, le transport du glucose, la synthèse protéique et de glycogène. En outre, l'expression génique et la prolifération peuvent être activées par la voie PIK3 (A) mais également par la voie des MAP kinases (B). B. Cette voie est également activée par la fixation de l'insuline à son récepteur qui va activer une cascade de phosphorylation en passant par les protéines Ras et Raf puis MAPK (CAPEAU 2003).

b) Les polymorphismes des gènes de la voie de l'insuline

Un des polymorphismes étudiés du gène *IGF1* est une répétition de CA en position 192 qui a été montrée comme prédisposante dans les CCR par Slattery *et al.* (2004a) (OR=1,2). Un autre polymorphisme c.-202A>C du gène *IGFBP3* ne montre pas d'effet particulier sur la prédisposition au CCR, car son OR est de 1,1 avec une absence de significativité de ces résultats (SLATTERY *et al.* 2004a). Le SNP Gly972Arg du gène *IRS1* a été montré par Slattery *et al.* (2004a) comme étant significativement prédisposant (OR=1,4) pour le CCR, alors que le SNP Gly1057Asp du gène *IRS2* ne montre aucun effet sur le CCR (SLATTERY *et al.* 2004a) (OR=1,0).

Le même auteur a étudié les mêmes polymorphismes dans une étude où l'association entre l'effet du polymorphisme sur la pathologie et la prise d'AINS a été observée (SLATTERY *et al.* 2004b). Les auteurs ont mis en évidence que la prise médicamenteuse modifiait l'effet du polymorphisme lorsque ce facteur n'était pas étudié. Ainsi le polymorphisme du gène *IGF1* montre une association

protectrice plus importante qu'à l'ordinaire lorsqu'elle est accompagnée d'une prise régulière d'anti-inflammatoires non-stéroïdiens (AINS). De la même façon, le polymorphisme d'*IGFBP3* ne montre un effet (OR=0,64) que lorsqu'il est accompagné d'une prise d'AINS, et celui d'*IRS1* montre un effet protecteur significatif qui est exacerbé par la prise d'AINS (SLATTERY *et al.* 2004b).

Le Tableau 9 récapitule les différents effets des SNP recensés dans les publications.

Gène	SNP	Effet	Significatif Oui/Non	Référence bibliographique
<i>IGF1</i>	Répétition de CA	Prédisposant	Non	Slattery <i>et al.</i> (2004a)
<i>IGFBP3</i>	c.-202A>C	N.D	Non	Slattery <i>et al.</i> (2004a)
<i>IRS1</i>	p.Gly972Arg	Prédisposant	Oui	Slattery <i>et al.</i> (2004a)
<i>IRS2</i>	p.Gly1057Asp	N.D	Non	Slattery <i>et al.</i> (2004a)

Tableau 9 : Polymorphismes des gènes de la voie de l'insuline (N.D : Non Déterminé).

II.5.3.2.5. Voie de réparation de l'ADN

a) Le système de réparation par excision de bases (BER)

Ce système de réparation qui intervient dans la forme de cancer colorectal familiale MAP a déjà été détaillé, dans la section « Chapitre I : II.4.2.2.1.Mécanisme d'action du système BER », nous ne nous intéresserons ici qu'aux polymorphismes de cette voie.

Pour la voie BER, la majorité des études portent sur l'analyse de variants (p.Arg194Trp, p.Arg280His, p.Arg399Gln) du gène *XRCC1*. Seul le SNP Arg399Gln montre une légère augmentation du risque de CCR par la présence du variant mais elle est statistiquement non significative (ABDELRAHMAN *et al.* 2000; MORENO *et al.* 2006; YEH *et al.* 2005). Moreno *et al.* (2006) trouvent une interaction significative des allèles variants des polymorphismes Arg194Trp et Arg280His avec l'âge. Ces allèles seraient associés à une diminution du risque de CCR chez les patients de moins de 54 ans comparé aux patients de plus de 70 ans. Une étude coréenne (HONG *et al.* 2005) a montré quant à elle une association significative entre les allèles 194Arg-280Arg-399Gln et une augmentation du risque de CCR (OR=1,59).

Le gène *OGG1* est le second gène à être fréquemment investigué dans les associations avec le CCR et plus particulièrement le SNP Ser326Cys. Ce SNP donne des résultats non concluants lorsque l'on regarde la totalité des études portant sur ce SNP. En effet, Moreno *et al.* (2006) ainsi que Goodman *et al.* (2006) montrent une augmentation du risque de CCR lorsque le variant est présent alors que Hensen *et al.* (2005) montrent une diminution.

Le Tableau 10 récapitule les différents effets des SNP recensés dans les publications.

Gène	SNP	Effet	Significatif Oui/Non	Référence bibliographique
XRCC1	p.Arg194Trp	Protecteur	Oui en fonction de l'âge	Moreno <i>et al.</i> (2006)
	p.Arg280His	Protecteur	Oui en fonction de l'âge	Moreno <i>et al.</i> (2006)
	p.Arg399Gln	Prédisposant	Non	Abdel-Rahman <i>et al.</i> (2000)
OGG1	p.Ser326Cys	Prédisposant	Non	Moreno <i>et al.</i> (2006)
		Protecteur	Non	Goodman <i>et al.</i> (2006) Hensen <i>et al.</i> (2005)

Tableau 10 : Polymorphismes du système de réparation BER.

b) Le système de réparation par excision de nucléotide (NER)

(1) Mécanisme

Le système NER est le mécanisme le plus polyvalent de réparation de l'ADN, en supprimant un grand nombre de lésions structurellement indépendantes de l'ADN : lésions volumineuses telles que les dimères de pyrimidine, les photoproduits, les grands adduits chimiques. Il existe deux voies de réparation dans le système NER : le système NER du génome global, détectant et éliminant les lésions dans l'ensemble du génome, et le système NER couplé à la transcription, assurant une réparation plus rapide des lésions situées sur le brin transcrit. Le système NER comporte au moins quatre étapes (FRIEDBERG 2003; GILLET and SCHARER 2006) :

- la reconnaissance des dommages par un complexe de protéines dont XPC,
- le déroulage de l'ADN par le complexe TFIIH comprenant XPD,
- la suppression des fragments simples-brins endommagés (entre 27 et 30 paires de bases) par des molécules telles qu'ERCC1 et le complexe XPF,
- la synthèse de l'ADN polymérase.

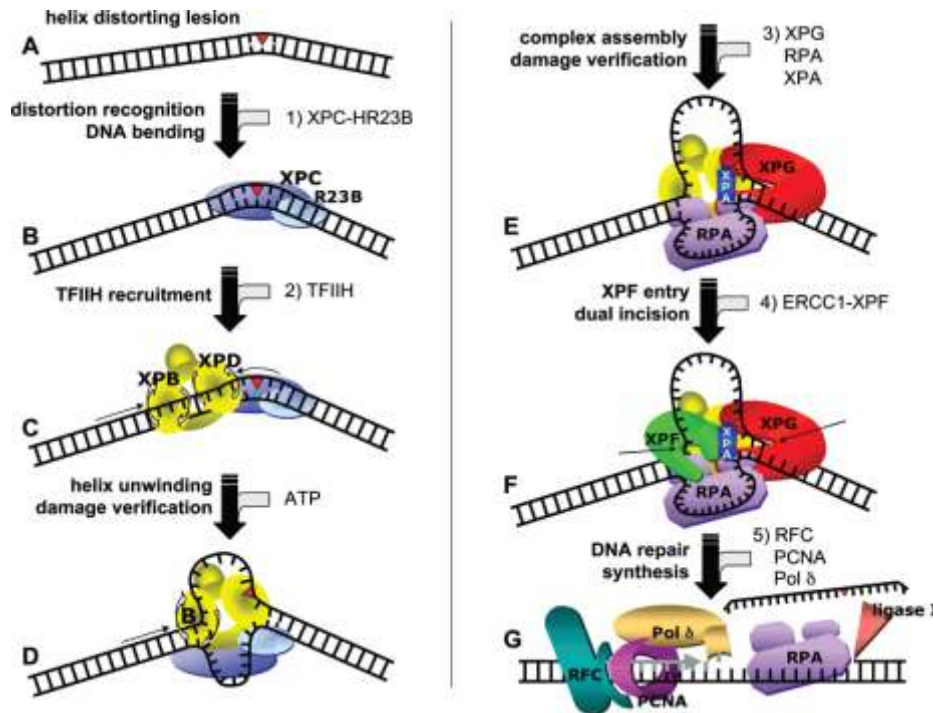


Figure 33 : Réparation d'une lésion de l'ADN par le système NER. (A) Une lésion de l'ADN induit des distorsions de l'hélice. (B) XPC-HR23B détecte des déformations de l'hélice et stabilise l'ADN. (C) XPC-HR23B recrute TFIIH sur le site de la lésion. (D) TFIIH déroule l'hélice d'ADN jusqu'à ce qu'une des sous-unités de l'hélicase (ici XPD) rencontre une base chimiquement modifiée, la deuxième sous-unité de l'hélicase (ici XPB) va sur le déroulement de l'ADN pour créer une ouverture en « bulle » de 20bp ; (E) RPA, XPA et XPG sont ensuite recrutés pour rassembler le complexe de « pré-incision ». (F) XPF-ERCC1 rejoint le complexe et la double incision se produit (en 5' par ERCC1-XPF et en 3' par XPG). (G) RPA reste lié à l'ADN simple-brin et facilite la transition vers la réparation par Pol δ (ou ϵ), soutenu par la RFC et PCNA ; et la ligase I termine la réparation (GILLET and SCHARER 2006).

(2) Les polymorphismes des gènes de réparation du système NER

Les résultats des études d'association des polymorphismes du système NER ne montrent aucune association forte avec le cancer colorectal. Le polymorphisme le plus étudié est Lys751Gln du gène *XPD* qui montre une association seulement avec les adénomes (BIGLER *et al.* 2005; SKJELBRED *et al.* 2006).

Par exemple, les allèles 1213Gly et 1097Val du gène *CSB* ont été montrés par Berndt *et al.* (2006) comme augmentant le risque de CCR par rapport à l'allèle sauvage. Goodman *et al.* (2006) ont également montré que cinq SNP dans les gènes *XPD*, *XPF* et *XPG* du système NER ne sont pas associés au risque de cancer, mais cette étude peut être prise en défaut sur la taille des cohortes utilisées et le recrutement qui fait intervenir différentes ethnies. Pour conclure, Mort *et al.* (2003) ont analysé plusieurs SNP dans les gènes *XPD*, *XPG*, *XPF*, et *ERCC1*, et n'ont pas trouvé d'association entre ces polymorphismes et le CCR mais encore une fois les populations utilisées avaient des effectifs très limités.

Le Tableau 11 récapitule les différents effets des SNP recensés dans les publications.

Gène	SNP	Effet	Significatif Oui/Non	Référence bibliographique
<i>XPD</i>	p.Lys751Gln	Prédisposant sur les adénomes	Oui	Bigler <i>et al.</i> (2005) Skjelbred <i>et al.</i> (2006)
<i>CSB</i>	1213Gly	Prédisposant	Non	Berndt <i>et al.</i> (2006)
	1097Val	Prédisposant	Non	Berndt <i>et al.</i> (2006)

Tableau 11 : Polymorphismes du système de réparation NER.

c) Le système de réparation des cassures doubles-brins (DSBR)

(1) Mécanisme

Les cassures des doubles-brins (DSB) sont produites par des erreurs de réplication et par des agents exogènes, tels que les rayonnements ionisants. La réparation des cassures des doubles-brins est intrinsèquement plus particulière puisqu'aucun modèle intact n'est disponible (KHANNA and JACKSON 2001). Au moins deux voies de réparation des cassures des doubles-brins sont reconnues : la recombinaison homologue et la voie de réparation des extrémités par jointure non homologue (NHEJ). Dans les recombinaisons homologues, les terminaisons de l'ADN sont amputées en 5' et 3' sous la direction de nucléases. La queue simple-brin en 3' résultante va ensuite envahir la double hélice d'ADN homologue, et sera prolongée par l'action de l'ADN polymérase. Le résultat donnera deux molécules d'ADN intactes. Ces voies impliquent plus de 16 protéines telles que BRCA1 et BRCA2 et XRCC3. Le système NHEJ implique directement la ligation des deux extrémités par l'intervention de diverses protéines (Figure 34) (O'DRISCOLL and JEGGO 2006).

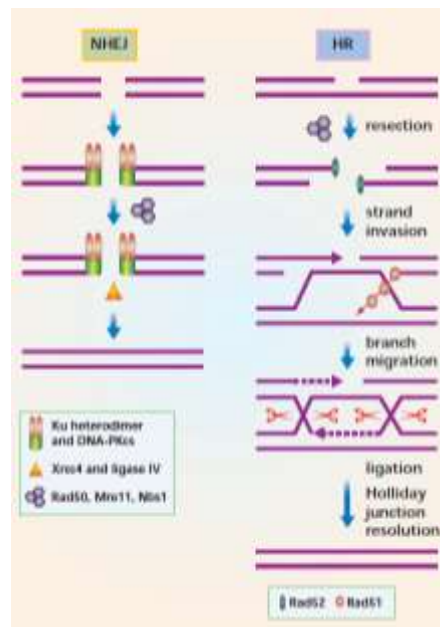


Figure 34 : Réparation des cassures doubles-brins d'ADN. A droite, la recombinaison homologue (HR) et à gauche, la réparation des cassures par jointure non homologue (NHEJ). Le système NHEJ fait intervenir diverses protéines qui vont stabiliser l'ADN puis l'intervention de XRCC4 et de la ligase IV va permettre la réparation de cette cassure. La réparation homologue va permettre de synthétiser deux brins d'ADN identiques grâce à l'intervention de molécules telles que Rad51 et Rad52 (KHANNA and JACKSON 2001).

(2) Les polymorphismes des gènes de réparation du système DSBR

Les polymorphismes de cette voie ne montrent pas de forte association avec le cancer ou les adénomes colorectaux. Le polymorphisme Thr241Met du gène *XRCC3* est associé au CCR mais de manière opposée suivant les publications. En effet, Mort *et al.* (2003) montrent un effet prédisposant pour l'allèle T (OR=1,52) alors que Krupa et Balasiak (2004) et Jin *et al.* (2005) ont montré une prédisposition pour l'allèle 241Met (respectivement OR=9,45 et 3,13). Ces données conflictuelles entre les études d'association et le cancer sont communes et résultent d'un effectif insuffisant (moins de 100 patients pour Mort *et al.* (2003) et Krupa et Balasiak (2004)).

D'autres polymorphismes appartenant aux gènes de cette voie ont été investigués, mais aucune association n'a été retrouvée. On peut citer comme exemple, c.4541A>G et c.17893A>G du gène *XRCC3* (TRANAH *et al.* 2004), p.Arg188His du gène *XRCC2* (MORENO *et al.* 2006; TRANAH *et al.* 2004), p.Thr297Ile du gène *XRCC9* (GOODMAN *et al.* 2006) et p.Glu185Gly du gène *NBS1* (GOODMAN *et al.* 2006).

Le Tableau 12 récapitule les différents effets des SNP recensés dans les publications.

Gène	SNP	Effet	Significatif Oui/Non	Référence bibliographique
<i>XRCC3</i>	p.Thr241Met	Prédisposant	Non	Mort <i>et al.</i> (2003)
		Protecteur	Non	Krupa et Balasiak (2004) Jin <i>et al.</i> (2005)
	c.4541A>G	N.D	Non	Tranah <i>et al.</i> (2004)
	c.17893A>G	N.D	Non	Tranah <i>et al.</i> (2004)
<i>XRCC2</i>	p.Arg188His	N.D	Non	Tranah <i>et al.</i> (2004) Moreno <i>et al.</i> (2006)
<i>XRCC9</i>	p.Thr297Ile	N.D	Non	Goodman <i>et al.</i> (2006)
<i>NBS1</i>	p.Glu185Gly	N.D	Non	Goodman <i>et al.</i> (2006)

Tableau 12 : Polymorphismes du système de réparation DSBR (N.D : Non Déterminé).

d) Le système de réparation des mésappariements de l'ADN (MMR)

(1) Mécanisme

Des motifs de séquence spécifique composée des répétitions de dinucléotides sont connus pour être instables dans certains cancers. Ce phénotype d'« instabilité des microsatellites » est causé par des défauts dans la réparation des mésappariements de l'ADN dans le syndrome de Lynch et dans certains cancers sporadiques. La réparation des mésappariements de l'ADN (MMR) supprime les nucléotides mal appariés par une ADN polymérase et les boucles d'insertion/suppression qui résultent de glissement au cours de la réplication des séquences répétées ou durant la recombinaison. Des défauts de ce système augmentent de façon spectaculaire le taux de mutations, en accélérant le processus de l'oncogenèse (HOEIJMAKERS 2001). Plusieurs gènes sont impliqués dans cette voie de réparation MMR tels que *hMLH1*, *hMSH2*, *hPMS2* et *hMSH6* (DE LA CHAPELLE 2004).

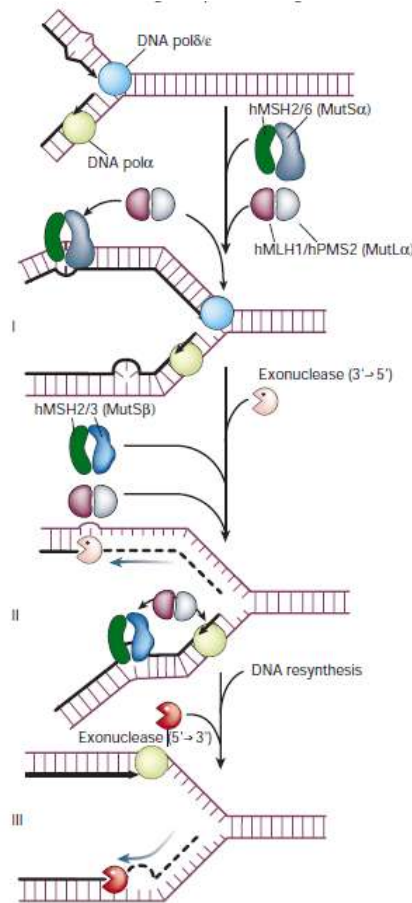


Figure 35 : Réparation des mésappariements de l'ADN. Les hétérodimères de hMSH2/6 (appelée hMutS α) se fixent sur le mésappariement et les boucles d'ADN simples-brins (première étape), alors que les dimères hMSH2/3 (hMutS β) reconnaissent les boucles d'insertion/suppression (étape 2). Les complexes hétérodimériques hMutL comme les protéines hMLH1/hPMS2 (hMutL α) et hMLH1/hPMS1 (hMutL β) interagissent avec les complexes MSH et les facteurs de réplication. La discrimination des brins peut être fondée sur le contact avec la machinerie de réplication qui est à proximité. Un certain nombre de protéines sont impliquées dans l'excision du nouveau brin passées les étapes de mésappariements et de resynthèse telles que Pol δ/ϵ , RPA, PCNA, RFC, l'exonuclease 1 et l'endonuclease FEN1 (étape 2 et 3) (HOEIJMAKERS 2001).

(2) Les polymorphismes des gènes de réparation du système MMR

Berndt *et al.* (2007) ont analysé des polymorphismes dans les gènes MMR et ils ont montré que l'allèle variant 1036Ala de *hMSH3* est associé au CCR en augmentant le risque de celui-ci. Après stratification en fonction de la localisation tumorale, l'allèle 940Gln de *hMSH3* est trouvé associé à une augmentation du risque de CCR dans les cancers du côlon proximal alors que l'allèle 39Glu de *hMSH6* montre une augmentation du risque pour le cancer du rectum. Kim *et al.* (2004) n'ont pas mis en avant d'association entre le risque de CCR sporadique et quatre SNP répartis dans deux gènes de cette voie (p.Ile219Val et p.Val384Asp de *hMLH1* et p.Leu390Phe et gIVS12 de *hMSH2*). De la même manière Yamamoto *et al.* (2005) ont investigué deux SNP (p.Thr439Met et p.Pro757Leu) du gène *EXO1* qui interagit directement avec les protéines MSH2, MSH3 et MLH1. L'allèle 439Met est montré comme augmentant le risque de CCR (OR=2,37) et le génotype homozygote variant (757LL) du SNP Pro757Leu comme réduisant le risque de CCR (OR=0,40).

Le Tableau 13 récapitule les différents effets des SNP recensés dans les publications.

Gène	SNP	Effet	Significatif Oui/Non	Référence bibliographique
<i>hMSH3</i>	1036Ala	Prédisposant	Non	Berndt <i>et al.</i> (2007)
	940Gln	Prédisposant dans les CCR proximaux	Non	Berndt <i>et al.</i> (2007)
<i>hMSH6</i>	39Glu	Prédisposant pour le cancer du rectum	Non	Berndt <i>et al.</i> (2007)
<i>hMLH1</i>	p.Ile219Val	N.D	Non	Kim <i>et al.</i> (2004)
	p.Val384Asp	N.D	Non	Kim <i>et al.</i> (2004)
<i>hMSH2</i>	p.Leu390Phe	N.D	Non	Kim <i>et al.</i> (2004)
	g.IVS12	N.D	Non	Kim <i>et al.</i> (2004)
<i>EXO1</i>	p.Thr439Met	Prédisposant	Non	Yamamoto <i>et al.</i> (2005)
	p.Pro757Leu	Protecteur	Non	Yamamoto <i>et al.</i> (2005)

Tableau 13 : Polymorphismes du système de réparation MMR (N.D : Non Déterminé).

e) Le système de réparation directe

(1) Mécanisme

La méthyltransférase O⁶-méthylguanine (MGMT) est une protéine ubiquitaire de réparation, vitale dans la réduction des effets mutagènes des agents alkylants, des liaisons covalentes à la position O⁶ de la guanine. MGMT agit comme une seule protéine qui renverse les dommages d'alkylation, en catalysant le transfert de groupes alkyls de la guanine à un site actif de la cystéine (MARGISON and SANTIBANEZ-KOREF 2002). MGMT participe dans une seule réaction seulement, et est inactivée de manière irréversible. La MGMT encode une protéine de réparation appelée l'alkyltransférase O⁶-alkylguanine (AGT) (GERSON 2004).

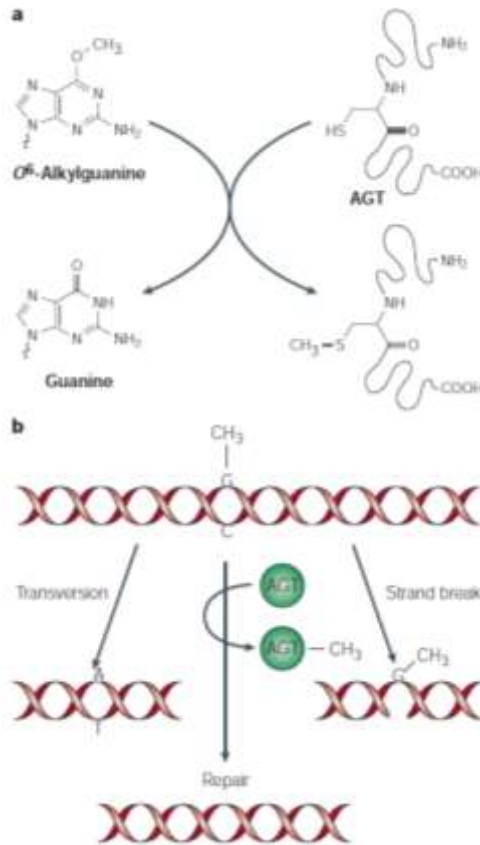


Figure 36 : Mécanisme d'action de l'alkyltransférase O⁶-alkylguanine (AGT). (a) L'AGT recherche sur l'ADN double-brin les alkylations en position O⁶ de la guanine. Un transfert covalent du groupe alkyl (dans cette figure un groupe méthyl, CH₃) au site actif conservé de la cystéine inactive l'AGT et restaure une guanine normale. (b) Si la réparation de la lésion G-CH₃ ne fonctionne pas, il peut en résulter une mutation par transition G→A ou cassure de brin (GERSON 2004).

(2) Les polymorphismes des gènes du système de réparation directe

L'étude de Tranah *et al.* (2006) est la seule à avoir montré des associations entre les polymorphismes de MGMT et le risque de CCR. Les autres études ne montrent aucune association (BIGLER *et al.* 2005; GOODMAN *et al.* 2006; MORENO *et al.* 2006). Tranah *et al.* (2006) ont donc montré un effet protecteur de l'allèle variant du polymorphisme Ile143Val du gène *MGMT* seulement chez les femmes atteintes de CCR et aucune association n'a été montrée pour le SNP Leu84Phe de *MGMT*. Après stratification de la population, l'allèle variant 84Phe de ce polymorphisme est retrouvé comme augmentant le risque de CCR chez les femmes consommant des boissons alcoolisées tous les jours. Chez les femmes post-ménopausées l'allèle 143Val du gène *MGMT* réduit significativement le risque de CCR et augmente l'effet protecteur des hormones post-ménopause prises par ces femmes.

Le Tableau 14 récapitule les différents effets des SNP recensés dans les publications.

Gène	SNP	Effet	Significatif Oui/Non	Référence bibliographique
<i>MGMT</i>	p.Ile143Val	Protecteur chez les femmes	Non	Tranah <i>et al.</i> (2006)
	p.Leu84Phe	N.D	Non	Tranah <i>et al.</i> (2006)

Tableau 14 : Polymorphismes du système de réparation directe (N.D : Non Déterminé).

Les polymorphismes présentés dans cette partie ne sont que des exemples, bien d'autres publications ont été réalisées sur des associations entre des SNP et le risque de cancer colorectal. Les exemples présentés ici sont issus pour la plupart des publications de Naccarati *et al.* (2007), de de Jong *et al.* (2002) et de Sachse *et al.* (2002).

III. Méthodes d'identification des gènes et des polymorphismes de prédisposition dans les maladies complexes

III.1. Une méthode de choix : L'étude d'association cas-témoins

Les études de liaison utilisées pour identifier des gènes de forte pénétrance dans les formes familiales ne sont pas efficaces pour la recherche de gènes de prédisposition aux formes sporadiques. Les études d'association sont plus pertinentes pour l'étude des variants génétiques qui sont communs dans la population générale (fréquence supérieure à 1%), car elles ont une puissance statistique bien supérieure aux études de liaison pour la détection de gènes de faible prédisposition (RISCH and MERIKANGAS 1996). A l'origine, ces études étaient limitées à l'analyse des gènes survenant dans des voies connues de la pathologie. A l'heure actuelle, et avec l'avènement des puces à ADN pangénomiques, ces études ne sont plus confinées à l'analyse des seuls gènes candidats (BISHOP 2004).

III.1.1. Une étape clé : la constitution des cohortes

Ce type d'étude nécessite d'être en possession de plusieurs centaines de patients atteints de la pathologie à étudier, et un nombre similaire ou supérieur de témoins. Ces deux cohortes « patients » et « témoins » doivent être le plus possible similaires aux niveaux de l'âge, de la région géographique, de la proportion hommes/femmes et d'un point de vue ethnique. À la suite du recrutement, les études des variants et/ou des gènes de prédisposition seront réalisées en comparant la distribution des génotypes dans la population « témoins » et « patients ». Une différence de distribution importante donne une indication sur le rôle possible qu'ils peuvent jouer dans la pathologie étudiée (BISHOP 2004).

III.1.2. Précautions d'analyse

Les résultats générés par ce type d'étude sont à prendre avec beaucoup de précautions, car des difficultés peuvent être rencontrées lors de l'interprétation des résultats.

III.1.2.1. Défaillance du recrutement

III.1.2.1.1. L'homogénéité des cohortes

La population générale est constituée de groupes ou stratifications qui peuvent avoir une distribution différente dans la fréquence des génotypes et donc interférer sur l'évaluation du risque de la maladie. Un exemple simple pour illustrer ces propos, est l'incorporation dans une cohorte d'individus d'ethnies différentes. Les résultats seront faussés, car la distribution de certains SNP entre ethnies différentes peut varier énormément, allant de la présence d'un allèle, à son absence totale. C'est pour cette raison que le recrutement des cohortes est une étape très importante, qui va conditionner la qualité des résultats obtenus (BISHOP 2004; RISCH 2000).

III.1.2.1.2. La taille des cohortes

Un des points critiques de ces études d'association qui a longtemps été ignoré est la taille des cohortes utilisées. En effet, de nombreuses études épidémiologiques ont mis en avant des facteurs (génétiques, environnementaux,...) influençant une pathologie en se basant sur des cohortes de taille insuffisante pour obtenir une puissance statistique significative. Le Tableau 15 donne les effectifs théoriques, dans une étude d'association cas-témoins, pour la population « patients », en fonction des fréquences alléliques et de l'OR attendu.

OR	Fréquence de l'allèle de susceptibilité chez les témoins					
	1%	5%	10%	20%	30%	40%
1,1	221927	46434	24626	13987	10759	9505
1,2	58177	12217	6509	3730	2896	2581
1,3	27055	5702	3051	1763	1380	1240
1,5	10604	2249	1213	712	566	516
2,0	3193	687	377	229	188	177
4,0	598	134	78	52	46	47

Tableau 15 : Taille approximative de l'effectif des patients dans une étude d'association cas-témoins pour différentes fréquences alléliques et en fonction des OR attendus. L'effectif de la population « témoins » doit être au moins égal à celui de la population « patients » et être similaire du point de vue du recrutement (âge, genre, ethnie,...) (HATTERSLEY and McCARTHY 2005).

III.1.2.2. Notion de SNP causal ou de marqueur de la pathologie

Les études peuvent montrer l'association d'un polymorphisme à une pathologie alors qu'il n'intervient nullement dans le processus de la pathologie. Ce polymorphisme peut être un simple marqueur de la maladie et être en déséquilibre de liaison avec un autre gène ou polymorphisme qui sera causal (Figure 37). Toute la difficulté réside dans l'identification du gène responsable, par l'identification de variants fonctionnels reliés au marqueur génétique (BISHOP 2004).

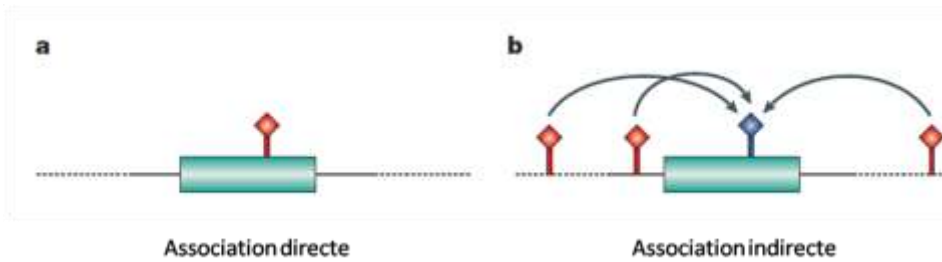


Figure 37 : Association directe ou indirecte d'un polymorphisme à un gène. (a) Dans ce cas, le SNP testé (en rouge) est directement situé dans le gène causal, donc l'association à la pathologie est favorisée (association directe). (b) Dans le cas d'une association indirecte, le ou les SNP trouvé(s) associé(s) au gène causal ne sont pas dans celui-ci, et peuvent être situé(s) dans un autre gène à proximité, ou dans de l'ADN non codant. Dans ce cas, le SNP est un marqueur de la pathologie (HIRSCHHORN and DALY 2005).

III.1.3. Les paramètres à analyser

Les études d'association sont basées sur l'analyse directe d'un polymorphisme en comparant la fréquence des génotypes des patients atteints de la pathologie à celle des témoins appariés. Pour cela deux calculs sont réalisés (EASTON 2004).

III.1.3.1. Le seuil critique de significativité

Il est calculé par la loi du χ^2 . Cette loi est calculée à partir d'un test de χ^2 qui ne peut être réalisé que pour les populations en équilibre d'Hardy-Weinberg, c'est-à-dire que les génotypes doivent être distribués de façon aléatoire dans la population.

Cette loi du χ^2 permet de comparer les valeurs obtenues aux valeurs attendues. Le résultat de cette loi, appelé p-value, est compris entre 0 et 1 et correspond à la chance d'obtenir par hasard le même résultat sous l'hypothèse nulle. Un résultat est significatif pour des valeurs inférieures à 5% avec un intervalle de confiance de 95% et un degré de liberté égal à 1.

III.1.3.2. L'estimation du risque relatif

Cette estimation du risque relatif est effectuée par le calcul de l'odds ratio (OR) dont un exemple est donné dans le Tableau 16.

	Génotypes		
	AA	AB	BB
Témoins	a	c	e
Patients	b	d	f

Tableau 16 : Distribution des effectifs des génotypes AA, AB et BB pour les patients et les témoins.

L'OR va mesurer le risque, par exemple pour les individus ayant le génotype AB, de développer la pathologie par rapport à ceux possédant le génotype AA. Le calcul de l'OR sera le suivant :

$$OR = \frac{a \times d}{b \times c}$$

Ce calcul du risque relatif est accompagné d'un calcul des bornes de l'intervalle de confiance qui est classiquement fixé à 95% (EASTON 2004).

III.2. Alternative d'étude d'association

Pour s'affranchir des inconvénients inhérents aux études d'association cas-témoins classiques (stratification et mauvais appariement des populations « témoins » et « patients »), une étude d'association impliquant les deux parents et l'individu atteint a été proposée. Pour l'analyse entre parents et enfant atteint, le test TDT (*Transmission Disequilibrium Test*) est utilisé pour détecter le gène de prédisposition. Ce test va permettre de comparer la fréquence de transmission d'un allèle étudié, à celle attendue d'après les lois de Mendel. En d'autres termes, ce test consiste à comparer la distribution des allèles d'un marqueur biallélique polymorphe, comme un SNP, transmis par les parents à l'individu malade, à la distribution des allèles parentaux non transmis. L'avantage de ce test est de constituer un groupe « témoins » parfaitement apparié pour la population d'origine au groupe de malades étudiés : la comparaison se fait entre les allèles parentaux transmis et non transmis (SPIELMAN *et al.* 1993).

III.3. Les différentes approches permettant l'identification des gènes de prédisposition

Il existe deux grandes approches différentes pour la détection de gènes de prédisposition : l'analyse du génome complet (GWA, *Genome-Wide Association*), ou l'approche « gènes candidats ». Chacune de ces deux approches possède ses propres avantages et défauts que nous allons évoquer dans la partie suivante.

III.3.1. L'approche « gènes candidats »

Les gènes candidats sont généralement des gènes avec des fonctions biologiques connues régulant directement ou indirectement les processus de développement des pathologies investiguées, ce qui pourrait être confirmé par l'évaluation de l'effet des variants des gènes des voies menant à la pathologie dans une étude d'association (Tabor *et al.* 2002).

Cette approche a cependant été critiquée en raison de la faiblesse de la réplication des résultats et de sa capacité limitée à inclure tous les gènes causaux. Cette approche est très subjective dans le processus de sélection des candidats à partir du nombre de possibilités potentielles. Le

principal inconvénient est qu'elle est basée sur des connaissances acquises (physiologiques, biochimiques ou fonctionnelles) (ZHU and ZHAO 2007).

L'approche « gènes candidats » possède également ses limites dans le choix des SNP à tester dans le gène d'intérêt. En général, l'étude se limite à l'investigation d'un nombre réduit de SNP dans le gène.

Malgré ses désavantages, cette approche permet la détection de gènes de prédisposition avec un coût d'analyse restreint.

III.3.2. L'approche « génome complet »

Cette approche est possible depuis l'apparition des puces à ADN de génotypage. Elle ne se base sur aucune présupposition quant à l'importance des caractéristiques fonctionnelles du polymorphisme étudié. En général, cette approche sert à associer des régions chromosomiques avec la pathologie étudiée grâce des marqueurs génétiques. Le plus souvent, ces régions intègrent un grand nombre de gènes candidats qu'il faudra sélectionner. Les principaux inconvénients de cette technique sont le coût et la capacité d'analyse des données générées (ZHU and ZHAO 2007).

III.3.2.1. Méthodes pour augmenter l'efficacité des études d'association pangénomiques

Comme ces études sont coûteuses, différentes approches ont été développées pour garder la puissance statistique de celles-ci tout en faisant des économies budgétaires.

III.3.2.1.1. L'approche multi-étapes

Dans cette approche, le but va être de génotyper dans un premier temps un nombre d'individus moins important qu'une étude d'association classique, mais la multiplication des étapes (deux ou trois étapes) va permettre de renforcer la puissance statistique.

Dans la première étape, un seuil de détection des SNP candidats est fixé pour permettre de détecter de manière assez large les loci recherchés tout en gardant à l'esprit qu'un grand nombre de résultats faux-positifs ont pu passer ce seuil. Le seuil est bien entendu fixé de manière à ce que tous les SNP susceptibles d'avoir une importance dans la pathologie le dépassent, et que la majorité des SNP non liés soient éliminés. Tous les marqueurs qui ont passé cette première étape de sélection sont ensuite testés dans une seconde population indépendante de la première, qui doit être aussi grande, ou plus grande que la population initiale. Lors de cette deuxième étape, le génotypage est effectué par une autre technologie plus adaptée à l'analyse d'un nombre de SNP moins conséquent

(exemple : *5'-nuclease assay*). Les marqueurs restants peuvent être répliqués lors d'une troisième étape afin d'éliminer les faux-positifs encore présents (HIRSCHHORN and DALY 2005).

Les étapes successives permettent de minimiser la présence de fausses associations, et d'obtenir un coût d'analyse convenable tout en gardant une puissance très satisfaisante. La Figure 38 illustre les propos ci-dessus.

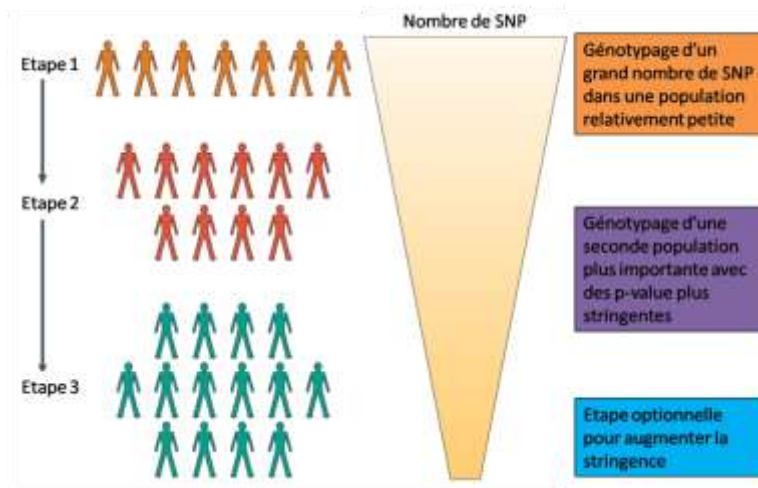


Figure 38 : Illustration de l'approche multi-étapes permettant de génotyper par GWA un nombre moins important de SNP. Pour cela, cette approche est basée sur la multiplication des étapes. La première étape consiste à génotyper une petite population et à fixer un seuil de détection permettant la détection des SNP d'intérêt ; toutefois un nombre assez important de faux positifs passeront cette première étape. La seconde étape va consister à faire une réplique des SNP sélectionnés dans la première étape par une méthode moins coûteuse et plus appropriée à l'analyse d'un nombre plus restreint de SNP (*5' nuclease assay*). Le génotypage est réalisé sur une population indépendante de la première et de taille identique voire plus importante. Cette seconde étape va permettre d'éliminer un grand nombre de faux positifs. La troisième étape est facultative et permet d'éliminer les derniers faux positifs restants en génotypant les SNP qui ont passé les deux étapes précédentes (HIRSCHHORN and DALY 2005).

D'autres techniques ont été proposées pour augmenter la significativité des résultats. On peut citer, par exemple, les tests de permutation qui consistent à mélanger les données des patients et des témoins, et par des tests statistiques à vérifier la validité des résultats précédemment trouvés.

III.3.2.1.2. L'approche requérant des populations avec un effet fondateur

L'utilisation de population avec effet fondateur* permet de diminuer le nombre de marqueurs analysés. Ces populations offrent de puissants avantages pour localiser de manière efficace les gènes qui sont liés à une pathologie mendélienne, mais elles ont moins d'avantages dans l'étude des maladies communes.

En effet, dans ces populations, les variations de fréquences alléliques seront dues essentiellement à la pathologie, car tous les individus ont un fond génétique commun que l'on ne

* Les populations avec effet fondateur sont issues d'un nombre limité d'individus qui se sont reproduits entre eux depuis 100 générations, ou moins.

retrouve pas dans une population issue de mixité géographique, ethnique. Les populations qui ont été utilisées avec succès dans la plupart de ces études sont des groupes tels que les Finlandais, les Canadiens du Québec et Juifs Ashkénazes.

Ces études ont montré une bonne corrélation pour les variants possédant une faible fréquence allélique dans la population, mais cette corrélation est beaucoup moins bonne pour les allèles plus fréquents dans la population, cela étant dû à la « jeunesse » de la population (HIRSCHHORN and DALY 2005).

III.3.2.1.3. L'approche par mélange d'ADN

Dans cette approche ce n'est pas le nombre de marqueurs analysés qui va diminuer, mais le nombre de puces utilisées. En effet, cette approche consiste à mélanger des quantités égales d'ADN de plusieurs individus avant le génotypage. Dans des pools, la détermination de l'effet d'un allèle est réalisée en comparant les fréquences de cet allèle entre la population « témoins » et « patients » ; par analogie au génotypage on va appeler cette technique l'allélotypage. Toutefois, la détermination d'une fréquence allélique d'un pool ne permet pas, en l'absence de génotypes individuels, d'identifier des mécanismes génétiques récessifs.

Les inconvénients de cette approche résident dans le fait que l'on ne puisse pas étudier les phénotypes indépendamment, ni les interactions gène-gène et gène-environnement.

Une fois que le polymorphisme de prédisposition est découvert par les différentes techniques citées ci-dessus, il est nécessaire de connaître l'impact de celui-ci. La partie suivante va présenter quelques techniques permettant l'analyse de l'effet d'un polymorphisme.

III.4. Impact de la fonction d'un SNP lié à une prédisposition

Un SNP de prédisposition aura un impact direct sur la pathologie étudiée, ou sera un marqueur de cette pathologie par sa localisation dans le génome (BROOKS 2003).

III.4.1. Corrélation entre la localisation d'un SNP et son effet prédisposant

Un SNP peut être localisé n'importe où dans le génome : dans une région intergénique, dans un gène, exonique, intronique ou encore dans un site d'épissage. Les SNP de prédisposition intergéniques ou introniques posent un problème d'interprétation, en ce sens où leur impact sur la fonction des gènes est encore très mal connu. Dans cette catégorie, on peut citer le SNP intergénique rs10505477 dans le locus 8q24.21, qui a été montré associé à de nombreux cancers, dont le cancer

colorectal (ZANKE *et al.* 2007), mais dont la fonction à ce jour n'a pas été élucidée. A l'inverse, les SNP exoniques ou situés dans les promoteurs, voire encore dans les sites canoniques d'épissage, peuvent, en fonction de la variation générée (faux-sens ou silencieuse) avoir un effet direct sur la pathologie, en modifiant l'organisation structurale et la conformation de la protéine ; ceci est surtout vrai pour les variants faux-sens (BROOKS 2003).

III.4.2. Analyse du caractère fonctionnel d'un SNP

Sur le plan méthodologique et technique, il est assez facile de mettre en avant l'effet d'un SNP au niveau transcriptionnel, alors que cet effet au niveau post-transcriptionnel est très difficilement observable.

III.4.2.1. Analyse bioinformatique

La première méthode d'analyse est une méthode bioinformatique basée sur des logiciels qui permettent de prédire la pathogénicité d'un SNP en fonction de la variation qu'il va générer. Parmi ces logiciels on peut citer par exemple les logiciel Sift (NG and HENIKOFF 2006) et Tcoffee (NOTREDAME *et al.* 2000; POIROT *et al.* 2003). Le premier logiciel prédit l'effet au niveau de la protéine, alors que le second logiciel va réaliser des alignements de séquences et prédit le caractère délétère ou activateur en fonction de la conservation inter-espèces du nucléotide perdu. Un autre logiciel a été développé spécifiquement pour l'analyse des polymorphismes survenant dans les sites d'épissage afin de prédire un épissage alternatif. Ce logiciel nommé Alamut (Interactive Biosoftware, Rouen, France) utilise différents logiciels de prédictions tels SpliceSiteFinder-like (ZHANG 1998), MaxEntScan (YEO and BURGE 2004) et GeneSplicer (PERTEA *et al.* 2001). Il est à noter que ce logiciel intègre les données de conservation phylogénétique pour ses prédictions.

III.4.2.2. Analyse de l'expression de l'ARN messenger

Les autres méthodes d'identification sont basées sur des analyses expérimentales et non virtuelles comme la précédente.

La première des méthodes, et la plus couramment utilisée, est une méthode qui nécessite l'ARN messenger (ARNm) des patients. Dans cette technique, l'ARNm servira à déterminer un taux d'expression par PCR quantitative dans le tissu atteint et à le comparer à celui du tissu sain. Le rapport calculé pourra ensuite être corrélé avec le génotype du SNP et ainsi déterminer en fonction de cette corrélation la fonctionnalité du SNP.

III.4.2.3. Analyse par minigène

L'approche minigène est utilisée essentiellement pour l'étude de l'effet des SNP contenus dans les sites d'épissage. L'exon variant est cloné dans un vecteur plasmidique et encadré par deux exons avec leurs séquences introniques permettant l'épissage. Ce vecteur est introduit dans des cellules qui vont permettre l'épissage. La présence d'un transcrit alternatif éventuel est mise en évidence par la migration sur gel, des produits de RT-PCR réalisés à partir des ARNm nouvellement transcrits (Figure 39) (Stoss *et al.* 1999).

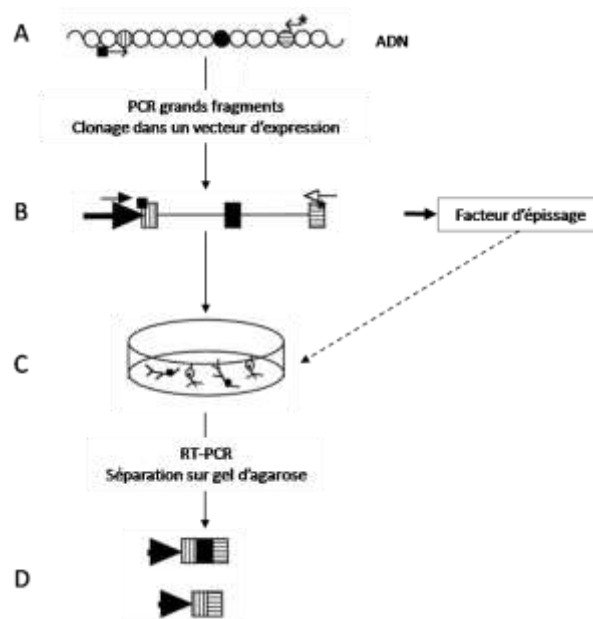


Figure 39 : Principe du système minigène. (A) À l'aide d'une PCR grands fragments, l'exon épissé alternativement (cercle noir) et les exons consécutifs (cercles rayés), ainsi que les régions intergéniques (cercles) sont amplifiés à partir d'un clone d'ADN génomique. Les sites de restriction introduits par les amorces PCR sont indiqués avec une étoile et un carré. (B) Après sous-clonage dans un vecteur, le minigène est cloné une nouvelle fois dans un vecteur d'expression eucaryote en utilisant les sites de restriction unique introduits par PCR (étoile et carré). Le promoteur eucaryote est indiqué par une flèche épaisse. Après transfection, les ARN sont analysés par RT-PCR en utilisant une amorce antisens (flèche blanche) et une amorce sens (flèche noire). (C) Le minigène peut être cotransfecté avec des facteurs d'épissage putatifs afin de tester leur capacité d'épissage. (D) Les produits de PCR peuvent être discriminés en fonction de la taille ou de l'hybridation, en raison de la présence ou de l'absence de l'exon épissé alternativement (Stoss *et al.* 1999).

III.4.2.4. Analyses par cassette d'expression

Les vecteurs d'expression permettent d'étudier l'effet d'un SNP localisé dans le promoteur d'un gène. Dans cette méthode, on clone en amont d'un gène rapporteur (β -galactosidase) le promoteur contenant le SNP. Ce vecteur est ensuite transfecté dans des cellules qui permettront de comparer l'expression du gène rapporteur entre le promoteur sauvage et le promoteur variant.

IV. Conclusion de la partie bibliographique

Les cancers colorectaux familiaux ont fait l'objet de très nombreuses études scientifiques qui ont permis de déterminer les gènes impliqués et les altérations génétiques responsables. Pour les CCR sporadiques, de nombreuses études épidémiologiques ont permis de suspecter un rôle protecteur ou délétère de différents aliments ou composés alimentaires. Ces différents facteurs interagissent vraisemblablement avec le « fond génétique » des individus et notamment avec certains gènes dits de « susceptibilité mineure », c'est-à-dire des gènes de faible pénétrance, qui agissent de manière cumulative et constituent la base de la prédisposition génétique aux cancers sporadiques. Malheureusement, les résultats de ces études ont été souvent remis en cause par l'absence de reproductibilité.

L'identification des facteurs de susceptibilité génétique aux CCR sporadiques représente un enjeu majeur dans l'espoir éventuel de proposer des mesures de surveillance dans la population, ou dans la perspective plus probable de mieux comprendre la carcinogénèse colorectale au niveau génétique en relation avec les facteurs environnementaux.

Dans ce but, une étude d'association portant sur plus de 1000 témoins et 1000 patients a été mise en place au laboratoire. Dans un premier temps par, une approche « gènes candidats » nous avons testé la répliation des polymorphismes décrits comme associés dans la littérature. Dans un second temps, par une approche pangénomique sur puces à ADN, nous avons cherché à identifier de nouveaux loci candidats.

CHAPITRE II :

ÉTUDE EXPERIMENTALE

I. Introduction

Cette étude axée sur les polymorphismes de prédisposition aux cancers colorectaux sporadiques est un projet initié par les docteurs Stéphane BÉZIEAU et Bruno BUECHER, qui a bénéficié d'un financement d'un Programme Hospitalier de Recherche Clinique (PHRC) régional. Le recrutement effectué dans six hôpitaux publics et cinq cliniques répartis dans la région des Pays de la Loire a permis d'inclure, après plus de quatre ans d'inclusion, 1023 patients atteints de cancers colorectaux sporadiques et 1121 témoins indemnes de toute pathologie colorectale. Chaque personne ayant accepté de participer à cette étude s'est vue prélever deux tubes de sang veineux à partir desquels l'ADN a été extrait. Les médecins, lors de l'inclusion, ont rempli un questionnaire épidémiologique standardisé précisant les habitudes de vie des patients et des témoins (Annexe 1). A chaque dossier d'un patient a été ajoutée une fiche de recueil clinique, remplie par le médecin traitant apportant des précisions sur la nature du cancer (Annexe 2). La banque d'ADN ainsi constituée a été le point de départ de ces travaux débutés au sein du Laboratoire d'Etude du Polymorphisme de l'ADN (LEPA) dirigé par le docteur Stéphane BÉZIEAU qui a rejoint début 2008 l'équipe d'accueil Biomarqueur, Métaplasie, Dysplasie (EA BioMétaDys) du professeur Jean-François MOSNIER.

En 2006, les technologies disponibles au laboratoire nous ont permis de réaliser une étude d'association par une approche « gènes candidats » sélectionnés par rapport aux publications et de réaliser des corrélations entre des associations de génotypes, ou d'allèles, mais également avec des facteurs environnementaux.

Les avancées technologiques survenues ces dernières années ont permis de rendre accessibles des approches sur le génome entier au plus grand nombre. L'approche GWA (*Genome-Wide Association*), ou approche pangénomique, a permis l'analyse concomitante de 10000 à 500000 SNP et elle permet maintenant d'analyser simultanément jusqu'à un million de SNP répartis sur la totalité du génome. Les études de ces milliers de SNP, contrairement à une étude « gènes candidats », ne se limitent plus aux seuls SNP fonctionnels et aux gènes connus comme impliqués dans la pathologie étudiée, mais elles ouvrent de nouvelles perspectives dans la découverte de facteurs de prédisposition. C'est début 2007 que nous avons lancé notre propre étude d'association GWA en utilisant cette toute nouvelle technologie en analysant plus de 250000 SNP par des puces de génotypage (Affymetrix®).

C'est en suivant cet ordre chronologique que la partie suivante va être exposée. Dans un premier temps les différents articles issus de l'approche « gènes candidats » vont être introduits et explicités brièvement :

- l'article 1 étudiant une combinaison de SNP des cytochromes P450 augmentant le risque de CCR sporadique lorsqu'il est associé à une consommation importante de viande rouge ;
- l'article 2 mettant en évidence 5 SNP prédisposants ou protecteurs vis-à-vis du CCR sporadique parmi une étude menée sur 52 polymorphismes ;
- l'article 3 répondant à la question suivante : est-ce que les individus de la population générale porteurs hétérozygotes des polymorphismes associés aux formes familiales de polyposes modérées MAP ont un risque accru de développer un cancer colorectal ?

Dans un second temps, une introduction sera réalisée sur l'approche GWA que nous avons menée, puis seront exposés le matériel et les méthodes utilisés ainsi que les résultats obtenus.

Dans un troisième temps, et avant de conclure et de présenter les perspectives, une discussion générale sera réalisée sur différents points rencontrés dans les sections précédentes, mais également, d'un point de vue plus général, sur les différentes approches utilisées.

II. Approche « gènes candidats »

II.1. Données de départ

Cette approche « gènes candidats » a été menée à partir des études publiées jusqu'à la fin 2005. En effet, ces études avaient mis en avant différents variants (SNP) jouant un rôle dans la prédisposition au cancer colorectal sporadique. Cette analyse effectuée, nous avons choisi, parmi tous les variants retrouvés dans la bibliographie, 57 polymorphismes répartis dans 35 gènes et déjà montrés comme associés au CCR, ou mis en évidence par dHPLC (denaturing high performance liquide chromatography) dans le laboratoire sur un panel de 50 patients et 50 témoins choisis dans notre cohorte. Tous les variants responsables ou ayant un rôle dans les formes familiales de cancer colorectal ont été éliminés.

II.1.1. Les polymorphismes choisis

Les polymorphismes, ainsi que les voies auxquelles ils appartiennent et les publications dans lesquelles ils ont été montrés associés au CCR sont présentés dans la partie suivante.

II.1.1.1. Les polymorphismes des gènes de la voie de l'inflammation

L'inflammation intervient dans le processus de carcinogenèse de différentes manières : soit par le recrutement des cellules immunitaires qui vont, par leurs médiateurs (cytokines et chimiokines) favoriser le développement tumoral, soit par la production d'espèces réactives à l'oxygène (ROS) très mutagènes, émises principalement par les macrophages (Chapitre I : II.5.3.2.1. La voie de l'inflammation).

Les différents polymorphismes retenus dans les gènes de cette voie sont présentés dans le Tableau 17.

Nom du gène	Changement nucléotidique	Changement protéique	dbSNP	Référence bibliographique
<i>ALOX5</i>	c.760G>A	p.Glu254Lys	rs2228065	Goodman <i>et al.</i> (2004)
<i>ALOX12</i>	c.782G>A	p.Gln261Arg	rs1126667	Goodman <i>et al.</i> (2004)
<i>IL6</i>	c.-239G>C	-	rs1800795	Landi <i>et al.</i> (2003)
<i>IL8</i>	c.-352T>A	-	rs4073	Landi <i>et al.</i> (2003)
<i>PLA2G2A</i>	c.-859C>G	-	rs11573156	dHPLC
<i>PLA2G2A</i>	c.132C>T	p.Tyr44Tyr	rs4744	dHPLC
<i>PLA2G2A</i>	c.185+88G>A	-	rs2236772	dHPLC
<i>PLA2G2A</i>	c.435+230C>T	-	rs11677	dHPLC
<i>PLD2</i>	c.1731C>T	p.Thr577Ile	rs1052748	Yamada <i>et al.</i> (2003)
<i>PTGS1</i>	c.22C>T	p.Trp8Arg	rs1236913	dHPLC, Halushka <i>et al.</i> (2003), Ulrich <i>et al.</i> (2004)

Nom du gène	Changement nucléotidique	Changement protéique	dbSNP	Référence bibliographique
<i>PTGS1</i>	c.50C>T	p.Pro17Leu	rs3842787	dHPLC, Halushka <i>et al.</i> (2003), Ulrich <i>et al.</i> (2004)
<i>PTGS1</i>	c.639C>A	p.Gly213Gly	rs5788	dHPLC, Halushka <i>et al.</i> (2003)
<i>PTGS1</i>	c.123G>A	p.Gln41Gln	rs3842788	dHPLC, Halushka <i>et al.</i> (2003)
<i>PTGS2</i>	c.-646C>T	-	rs20420	dHPLC, Goodman <i>et al.</i> (2004)
<i>PTGS2</i>	c.306G>C	p.Val102Val	rs5277	dHPLC, Cox <i>et al.</i> (2004)
<i>PTGS2</i>	c.1815+427T>C	-	rs5275	dHPLC, Cox <i>et al.</i> (2004)
<i>PTGS2</i>	c.1815+1912A>G	-	rs4648298	dHPLC, Cox <i>et al.</i> (2004)
<i>PPARG</i>	c.36C>G	p.Pro12Ala	rs1801282	Landi <i>et al.</i> (2003)
<i>PPARG</i>	c.1431C>T	p.His477His	rs3856806	Jiang <i>et al.</i> (2005)

Tableau 17 : Les différents polymorphismes de la voie de l'inflammation retenus dans notre approche « gènes candidats » (dHPLC indique que le polymorphisme a été trouvé au sein de notre laboratoire par criblage dHPLC d'une cohorte de 50 patients et 50 témoins).

II.1.1.2. Les polymorphismes des gènes de la voie de détoxication des xénobiotiques

Cette voie intervient dans la prédisposition au CCR, car les xénobiotiques (molécules exogènes de faible masse moléculaire) lors de leur métabolisation produisent des métabolites qui peuvent être toxiques, et qui, par leur accumulation, pourront entraîner un cancer (Chapitre I : II.5.3.2.2. La voie de détoxication).

Les différents polymorphismes retenus dans les gènes de cette voie sont présentés dans le Tableau 18.

Nom du gène	Changement nucléotidique	Changement protéique	dbSNP	Référence bibliographique
<i>CYP1A2</i>	c.-163A>C	-	rs762551	de Jong <i>et al.</i> (2002), Landi <i>et al.</i> (2005), Sachse <i>et al.</i> (2002)
<i>CYP1A2</i>	c.1548T>C	p.Asn516Asn	rs2470890	Landi <i>et al.</i> (2005), Sachse <i>et al.</i> (2002)
<i>CYP2E1</i>	c.-1293G>C	-	rs3813867	Landi <i>et al.</i> (2005)
<i>CYP2E1</i>	c.-1053C>T	-	rs2031920	Landi <i>et al.</i> (2005)
<i>CYP1B1</i>	c.1294C>G	p.Leu432Val	rs1056836	de Jong <i>et al.</i> (2002), Landi <i>et al.</i> (2005), Sachse <i>et al.</i> (2002)
<i>CYP2C9</i>	c.430C>T	p.Arg144Cys	rs1799853	de Jong <i>et al.</i> (2002), Landi <i>et al.</i> (2005), Sachse <i>et al.</i> (2002)
<i>EPHX1</i>	c.337T>C	p.Tyr113His	rs1051740	de Jong <i>et al.</i> (2002), Landi <i>et al.</i> (2005), Sachse <i>et al.</i> (2002)
<i>GSTA1</i>	c.-4605G>A	-	rs3957356	van der Logt <i>et al.</i> (2004)

Nom du gène	Changement nucléotidique	Changement protéique	dbSNP	Référence bibliographique
<i>GSTM1</i>	Délétion complète du gène	-	-	de Jong <i>et al.</i> (2002), Landi <i>et al.</i> (2005), Sachse <i>et al.</i> (2002)
<i>GSTM3</i>	c.468+21delAGG	-	-	de Jong <i>et al.</i> (2002), de la Chapelle (2004)
<i>GSTP1</i>	c.313A>G	p.Ile105Val	rs1695	de Jong <i>et al.</i> (2002), Landi <i>et al.</i> (2005), Sachse <i>et al.</i> (2002)
<i>GSTP1</i>	c.341C>T	p.Ala114Val	rs1138272	de Jong <i>et al.</i> (2002), Landi <i>et al.</i> (2005), Sachse <i>et al.</i> (2002)
<i>GSTT1</i>	Délétion complète du gène	-	-	de Jong <i>et al.</i> (2002), Landi <i>et al.</i> (2005), Sachse <i>et al.</i> (2002)
<i>NQO1</i>	c.415C>T	p.Arg139Trp	rs4986998	Landi <i>et al.</i> (2005), Sachse <i>et al.</i> (2002)
<i>NQO1</i>	c.559C>T	p.Pro187Ser	rs1800566	de Jong <i>et al.</i> (2002), Sachse <i>et al.</i> (2002)
<i>SULT1A2</i>	c.714A>C	p.Asn235Thr	rs1059491	Sachse <i>et al.</i> (2002)
<i>UGT1A1</i>	c.-3279T>G	-	rs4124874	Innocenti <i>et al.</i> (2002)
<i>UGT1A1</i>	c.-3156G>A	-	rs10929302	Innocenti <i>et al.</i> (2002)
<i>UGT1A6</i>	c.541A>G	p.Thr181Ala	rs2070959	van der Logt <i>et al.</i> (2004)
<i>UGT1A6</i>	c.552A>C	p.Arg184Ser	rs1105879	van der Logt <i>et al.</i> (2004)

Tableau 18 : Les différents polymorphismes de la voie de détoxication des xénobiotiques retenus dans notre approche « gènes candidats ».

II.1.1.3. Les polymorphismes des gènes de la voie des radicaux monocarbonés

Cette voie intervient dans le processus de carcinogenèse comme une étape limitante de la synthèse de l'ADN, car elle est initiée par un seul métabolite (l'acide folique ou les folates), dont la carence entraînera des mauvaises incorporations de bases et des cassures dans l'ADN. Le second mécanisme de cette voie pouvant intervenir dans le processus cancéreux est la méthylation de l'ADN, qui a déjà été montrée comme ayant un rôle dans les cancers colorectaux sporadiques (Chapitre I : II.5.3.2.3. La voie des radicaux mono-carbonés).

Le Tableau 19 présente les polymorphismes dans les gènes de cette voie que nous avons étudiés.

Nom du gène	Changement nucléotidique	Changement protéique	dbSNP	Référence bibliographique
<i>CBS</i>	c.844ins68	-	-	Le Marchand <i>et al.</i> (2002a), Sharp et Little (2004)
<i>MTHFD1</i>	c.1958G>A	p.Arg653Gln	rs2236225	Chen <i>et al.</i> (2004)
<i>MTHFR</i>	c.665C>T	p.Ala222Val	rs1801133	Le Marchand <i>et al.</i> (2002a), Sharp et Little (2004)
<i>MTHFR</i>	c.1286A>C	p.Ala429Glu	rs1801131	Le Marchand <i>et al.</i> (2002a), Sharp et Little (2004)
<i>MTR</i>	c.2756A>G	p.Asp919Gly	rs1805087	Le Marchand <i>et al.</i> (2002a), Sharp et Little (2004)
<i>MTRR</i>	c.66A>G	p.Ile22Met	rs1801394	Le Marchand <i>et al.</i> (2002a), Sharp et Little (2004)

Nom du gène	Changement nucléotidique	Changement protéique	dbSNP	Référence bibliographique
<i>TYMS</i>	c.943+447del TTAAAG	-	-	Sharp et Little (2004)

Tableau 19 : Les différents polymorphismes de la voie des radicaux monocarbonés retenus dans notre approche « gènes candidats ».

II.1.1.4. Les polymorphismes des gènes de la voie de l'insuline

La voie de l'insuline intervient dans le processus cancéreux en relation avec plusieurs voies distinctes impliquées directement dans l'activation de la prolifération cellulaire et de l'expression génique (Chapitre I : II.5.3.2.4.Voie de l'insuline).

Le Tableau 20 montre les polymorphismes en association avec le CCR que nous avons retenus pour notre étude.

Nom du gène	Changement nucléotidique	Changement protéique	dbSNP	Référence bibliographique
<i>GH1</i>	c.456+90T>A	-	rs2665802	Le Marchand <i>et al.</i> (2002b)
<i>IGF1</i>	c.-1006CA(19)	-	-	Slattery <i>et al.</i> (2004b)
<i>IGFBP3</i>	c.-336A>C	-	rs2854744	Slattery <i>et al.</i> (2004b)
<i>IRS1</i>	c.2911G>A	p.Gly971Arg	rs1801278	Slattery <i>et al.</i> (2004b)
<i>VDR</i>	c.1024+283G>A	-	rs1544410	Slattery <i>et al.</i> (2004b)

Tableau 20 : Les différents polymorphismes de la voie de l'insuline retenus dans notre approche « gènes candidats ».

II.1.1.5. Les polymorphismes des gènes de la voie de réparation de l'ADN

Les gènes de cette voie jouent un rôle direct dans les processus de carcinogenèse par leur action sur l'ADN. En effet, un défaut de fonctionnement d'une de ses enzymes peut entraîner des dommages à l'ADN irréversibles favorisant la carcinogenèse (Chapitre I : II.4.2.2.1.Mécanisme d'action du système BER).

Le Tableau 21 recense les différents polymorphismes que nous avons retenus pour notre étude.

Nom du gène	Changement nucléotidique	Changement protéique	dbSNP	Référence bibliographique
<i>OGG1</i>	c.977C>G	p.Cys326Ser	rs1052133	Hansen <i>et al.</i> (2005)
<i>MUTYH</i>	c.494A>G	p.Tyr165Cys	-	Al-Tassan et al (2002)
<i>MUTYH</i>	c.1145G>A	p.Gly382Asp	-	Al-Tassan et al (2002)
<i>MUTYH</i>	c.1435G>T	p.Val479Phe	-	Seq
<i>MUTYH</i>	c.694G>A	-	-	Seq
<i>MUTYH</i>	c.12639G>A	-	-	Seq

Tableau 21 : Les différents polymorphismes de la voie de réparation de l'ADN choisis pour notre approche « gènes candidats » (Seq, polymorphisme retrouvé par le séquençage de 100 patients de notre cohorte).

II.1.2. Méthode d'analyse

Au total, sur les 57 polymorphismes choisis dans les cinq voies, 46 d'entre eux sont des SNP. Parmi eux, cinq ont une fréquence inférieure à 1% et six sont des polymorphismes complexes (délétion, répétition, insertion).

II.1.2.1. Analyse des SNP

Tous les SNP ont été analysés par *5'-nuclease-assay*. Cette technique permet un génotypage haut débit des polymorphismes portant sur un seul nucléotide. Il est réalisé grâce à des sondes Taqman®, désignées et marquées par Applied Biosystems. Chaque sonde émettra une fluorescence différente en fonction du génotype du SNP. Ainsi le logiciel SDS v2.1 (Applied Biosystems) pourra faire une discrimination entre les individus porteurs homozygotes de l'allèle variant, porteurs hétérozygotes et porteurs homozygotes de l'allèle sauvage (ou allèle ancestral).

La Figure 40 montre le principe de la détection du génotype par la technologie de *5'-nuclease-assay*.

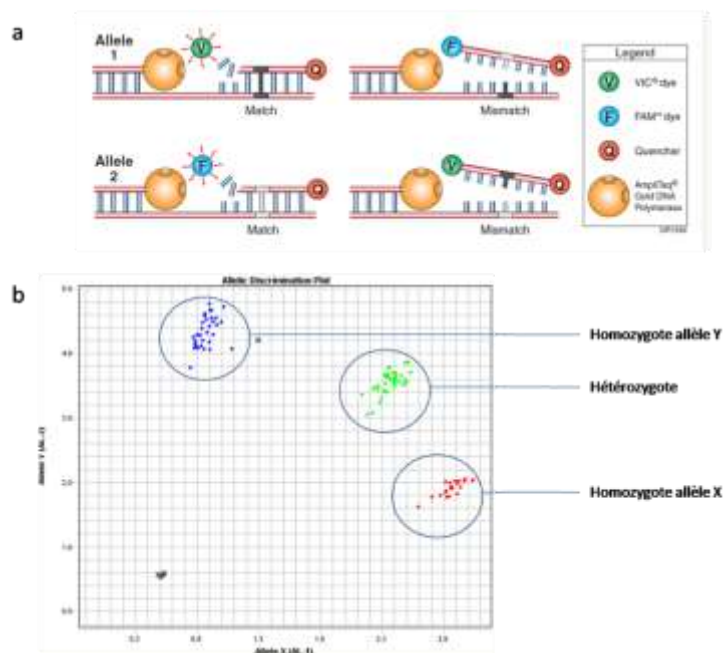


Figure 40 : Principe et résultats de la technique de génotypage par *5'-nuclease assay*. (a) pour chaque SNP à tester, deux sondes allèles spécifiques ont été désignées par le fournisseur. Les deux sondes sont marquées en 5' par une molécule fluorescente (le « reporter ») dont la nature diffère en fonction de l'allèle : la sonde discriminant l'allèle majoritaire est marquée par du VIC, tandis que celle discriminant l'allèle minoritaire est marquée par du FAM. En outre, les deux sondes sont liées en 3' à une autre molécule, le « quencher », qui lorsqu'elle est proche de la molécule fluorescente VIC ou FAM stabilise l'énergie de cette dernière et empêche l'émission de fluorescence. Lorsqu'une des deux sondes s'apparie parfaitement à l'ADN cible, la polymérase va la dégrader par sa fonction exonucléasique et ainsi va séparer la molécule reporter (VIC ou FAM) de la molécule quencher et provoquer, par conséquent, une émission de fluorescence. Lorsque la sonde possède un mésappariement, au niveau du variant, par rapport à l'ADN cible, elle ne s'apparie pas, et reste intacte ce qui n'engendre pas d'émission de fluorescence. (b) Les résultats apparaissent sous la forme de nuage de points colorés en bleu ou rouge correspondant à la fluorescence émise ; pour les individus hétérozygotes, la fluorescence émise sera un mélange de celle des deux allèles (www.appliedbiosystems.com).

II.1.2.2. Analyse des polymorphismes complexes

Ces polymorphismes ont été étudiés par deux types distincts de PCR multiplexes utilisant des amorces marquées par une molécule fluorescente (FAM ou HEX). Les fragments générés ont été séparés par une électrophorèse réalisée sur un séquenceur capillaire 3130 (Applied Biosystems). Les données ont été analysées par les logiciels Genescan 3.1 et Genotyper 2.5.2 (Applied Biosystems).

La première PCR est une PCR multiplexe classique permettant l'analyse de petites délétions, insertions et répétitions. La seconde, nommée QMPSF (*Quantitative Multiplexe PCR of Short Fluorescent Fragments*), est plus complexe et rajoute une notion de quantitatativité, en comparant l'intensité des fragments obtenus à ceux d'un gène témoin non lié au gène à tester, qui ne doit pas, *a priori*, être délété. Dans notre cas, nous avons choisi le gène *MLH1* comme gène témoin. Cette technique permet donc de connaître le nombre d'allèles dont dispose un gène. Lors de l'analyse, l'amplitude des pics du gène à tester va être évaluée comparativement à celle de l'exon amplifié du gène témoin, ici *MLH1*. La première étape de l'analyse va donc consister à ajuster artificiellement la hauteur du pic d'amplicon *MLH1* de tous les individus d'une série de QMPSF. La taille relative des pics d'amplicons du gène à tester va ainsi être comparée facilement et directement entre tous les individus de la série. Une duplication ou une délétion hétérozygote seront objectivées respectivement par une multiplication ou une division par deux de la hauteur du pic (Figure 41).

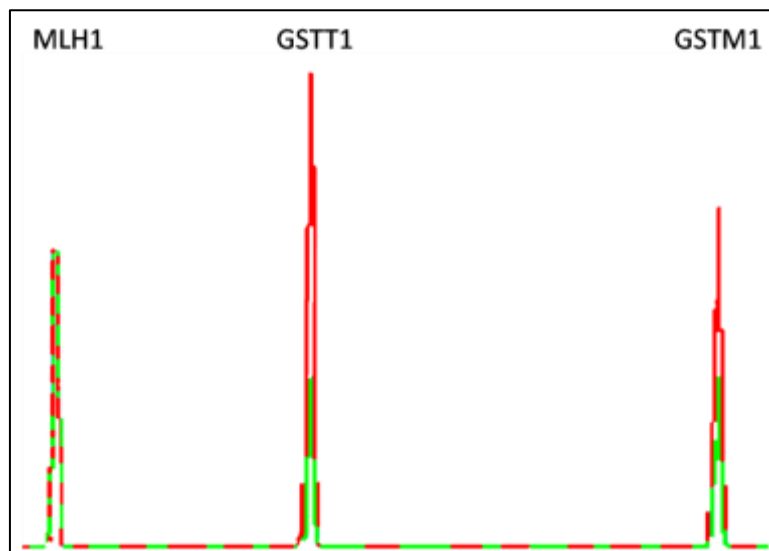


Figure 41 : Résultat de QMPSF pour deux individus après ajustement des pics *MLH1*. L'individu correspondant à la courbe rouge n'est pas délété pour les gènes *GSTT1* et *GSTM1* alors que l'individu représenté par la courbe verte est délété hétérozygote pour les deux gènes.

II.2. Etude de 6 polymorphismes des cytochromes P450

(CYP) (Article 1)

Il a été montré que la prédisposition au cancer colorectal sporadique serait due à une accumulation de facteurs aussi bien génétiques qu'environnementaux (DE LA CHAPELLE 2004; LICHTENSTEIN *et al.* 2000). C'est pourquoi nous nous sommes attachés, dans cette étude, à mettre en évidence cette association de facteurs de risque, non pas en étudiant un polymorphisme isolé avec un facteur environnemental comme il était courant de le voir dans la littérature, mais en étudiant simultanément plusieurs polymorphismes appartenant à la même voie et 16 facteurs environnementaux, allant du type de cuisson utilisé, à la consommation d'alcool,... (Annexe 1 ou Tableau 2 de l'article). En l'occurrence, nous avons choisi pour cela six polymorphismes des cytochromes P450 (*CYP1A2* c.-163A>C, *CYP1A2* c.1458T>C, *CYP2E1* g.-1293G>C, *CYP2E1* g.-1053C>T, *CYP1B1* c.1294C>G et *CYP2C9* c.430C>T).

Les cytochromes P450 font partie d'une superfamille d'enzymes comprenant de nombreuses isoformes comme le montre le Tableau 22. Elle compte en effet, 18 familles et 43 sous-familles, dans lesquelles sont répartis 57 gènes et 59 pseudogènes (NELSON *et al.* 2004). La plupart des CYP peuvent métaboliser plusieurs substrats et beaucoup peuvent catalyser de nombreuses réactions, ce qui explique leur importance dans le métabolisme d'un très grand nombre de molécules endogènes et exogènes (SACHSE *et al.* 2002). La plupart des réactions catalysées par les CYP sont des monooxygénations qui vont intervenir durant la première phase de la détoxification des xénobiotiques, afin de fonctionnaliser les métabolites potentiellement carcinogènes (BEAUNE 1999). Ces réactions vont intervenir dans le métabolisme de l'acide arachidonique, la biosynthèse des eicosanoïdes, du cholestérol, des stérols, de l'acide biliaire, la synthèse et le catabolisme des stéroïdes et de la vitamine D₃, l'hydroxylation de l'acide rétinoïque et le métabolisme de la neuroamine (NELSON *et al.* 2004).

Familles	Fonctions	Membres	Noms
CYP1	Métabolisme des médicaments et des stéroïdes (spécialement l'œstrogène)	3 sous-familles 3 gènes 1 pseudogène	CYP1A1, CYP1A2, CYP1B1
CYP2	Métabolisme des médicaments et des stéroïdes	13 sous-familles 16 gènes 16 pseudogènes	CYP2A6, CYP2A7, CYP2A13, CYP2B6, CYP2C8, CYP2C9, CYP2C18, CYP2C19, CYP2D6, CYP2E1, CYP2F1, CYP2J2, CYP2R1, CYP2S1, CYP2U1, CYP2W1
CYP3	Métabolisme des médicaments et des stéroïdes (incluant la testostérone)	1 sous-famille 4 gènes 2 pseudogènes	CYP3A4, CYP3A5, CYP3A7, CYP3A43
CYP4	Métabolisme de l'acide arachidonique ou des acides gras	6 sous-familles 11 gènes 10 pseudogènes	CYP4A11, CYP4A22, CYP4B1, CYP4F2, CYP4F3, CYP4F8, CYP4F11, CYP4F12, CYP4F22, CYP4V2, CYP4X1, CYP4Z1
CYP5	Synthèse du thromboxane A2	1 sous-famille 1 gène	CYP5A1

Familles	Fonctions	Membres	Noms
CYP7	Biosynthèse des acides biliaires Hydroxylase 7-alpha du noyau stéroïde	2 sous-familles 2 gènes	CYP7A1, CYP7B1
CYP8	Variées	2 sous-familles 2 gènes	CYP8A1 (synthétase de la prostacycline), CYP8B1 (biosynthèse de l'acide biliaire)
CYP11	Biosynthèse des stéroïdes	2 sous-familles 3 gènes	CYP11A1, CYP11B1, CYP11B2
CYP17	Biosynthèse des stéroïdes Hydroxylase 17-alpha	1 sous-famille 1 gène	CYP17A1
CYP19	Biosynthèse des stéroïdes	1 sous-famille 1 gène	CYP19A1
CYP20	Fonctions inconnues	1 sous-famille 1 gène	CYP20A1
CYP21	Biosynthèse des stéroïdes	2 sous-familles 2 gènes 1 pseudogène	CYP21A2
CYP24	Dégradation de la vitamine D	1 sous-famille 1 gène	CYP24A1
CYP26	Hydroxylase de l'acide rétinoïque	3 sous-familles 3 gènes	CYP26A1, CYP26B1, CYP26C1
CYP27	Variées	3 sous-familles 3 gènes	CYP27A1 (biosynthèse de l'acide biliaire), CYP27B1 (hydroxylase 1-alpha de la vitamine D3, activation de la vitamine D3), CYP27C1 (fonctions inconnues)
CYP39	Hydroxylation en 7-alpha du 24-hydroxycholestérol	1 sous-famille 1 gène	CYP39A1
CYP46	Hydroxylase de 24-cholestérol	1 sous-famille 1 gène	CYP46A1
CYP51	Biosynthèse du cholestérol	1 sous-famille 1 gène 3 pseudogènes	CYP51A1

Tableau 22 : Isoformes des cytochromes P450 (Nelson *et al.* 2004)*.

Les six polymorphismes que nous avons sélectionnés interviennent tous dans la détoxification des amines aromatiques hétérocycliques et des hydrocarbures polycycliques aromatiques. Ils ont été sélectionnés par méta-analyse bibliographique, sur la base de données PubMed, pour leur action dans le CCR sporadique. Les facteurs environnementaux ont, quant à eux, été sélectionnés également par méta-analyse bibliographique, mais seuls les plus pertinents ont intégré le questionnaire.

La chronologie de cette étude comprend tout d'abord l'analyse indépendante de l'ensemble des facteurs environnementaux et génétiques, puis l'analyse combinée des facteurs environnementaux et, pour finir, l'analyse combinée des six SNP avec les différents facteurs environnementaux.

Le premier résultat obtenu dans cette étude nous a confortés dans la qualité du recrutement de notre population. En effet, l'analyse indépendante des facteurs environnementaux nous a permis de confirmer, pour un grand nombre d'entre eux, les associations déjà mises en évidence par des études épidémiologiques antérieures et soumises à des recommandations d'ordre nutritionnel. On

* <http://drnelson.utmem.edu/human.P450.table.html>

peut citer par exemple, l'augmentation du risque de CCR induit par une faible activité physique ou la protection vis-à-vis du CCR d'une consommation régulière de pain complet. Au contraire, d'autres résultats se sont avérés plus étonnants, telle que l'absence de consommation d'alcool qui prédisposerait au CCR sporadique. L'analyse combinée des facteurs environnementaux nous a, quant à elle, indiqué que l'habitude alimentaire type induisant le risque minimal de CCR consistait en une consommation quotidienne de fruits (plus de deux fois par jour), de légumes (plus de deux fois par jour), de produits laitiers (plus de trois fois par jour) et de pain complet.

Par opposition, l'analyse des polymorphismes seuls n'a montré aucune association statistiquement significative avec le cancer colorectal sporadique que ce soit pour les polymorphismes analysés de manière indépendante ou combinée. En revanche, l'étude concomitante des SNP et des facteurs exogènes (environnementaux) a mis en évidence une combinaison d'allèles de SNP qui semble exacerber le pouvoir carcinogène intrinsèque de la viande rouge. Ainsi, alors que le risque de CCR associé à une forte consommation de viande rouge (>5 fois/semaine) correspond à un OR de 3,41, ce risque atteint un OR de 43,78 chez les individus forts consommateurs de viande rouge qui portent la combinaison allélique « *CYP1A2* c.163A/*CYP1A2* c.1548T/*CYP2E1* c.1293G/*CYP2E1* c.-1053C/*CYP1B1* c.1294G/ *CYP2C9* c.430T », par rapport aux individus consommant de la viande rouge moins de quatre fois par semaine qui présentent la combinaison de six allèles ancestraux (supposée être la combinaison allélique associée au risque minimal de CCR).

Cette étude souligne l'importance de l'étude simultanée de multiples variants d'une même voie métabolique dans le cadre de la prédisposition au CCR. En effet, la même combinaison d'allèles, trouvée précédemment associée au CCR sporadique lors d'une forte consommation de viande rouge présente également un risque très fort de CCR chez les individus obèses ayant un indice de masse corporelle supérieur à 30 (OR=12,55 ; 95%CI [6,28-25,09]) par rapport aux individus ayant un IMC faible ou normal. Ce résultat est un bon exemple, car il permet de mettre en avant qu'une combinaison allélique peut intervenir dans l'augmentation du risque chez une sous-catégorie particulière de patients voire pour certains endophénotypes. De plus, ce résultat permet de montrer l'importance de l'étude de plusieurs facteurs ensemble, car pris isolément, certains des allèles de ces SNP étaient trouvés comme protecteurs dans d'autres études, alors que la combinaison prise dans son ensemble est prédisposante.

ARTICLE 1

Combinations of cytochrome P450 gene polymorphisms enhancing the risk for sporadic colorectal cancer related to red meat consumption.

KÜRY S^{*}, BUECHER B^{*}, ROBIOU-DU-PONT S^{*}, SCOUL C, SEBILLE V, COLMAN H, LE HOUÉROU C, LE NEEL T, BOURDON J, FAROUX R, OLLIVERY J, LAFRAISE B, CHUPIN LD, BEZIEAU S

Cancer Epidemiol Biomarkers Prev **16**: 1460-1467 (2007)

^{*} Contribution équivalente au travail

Combinations of Cytochrome P450 Gene Polymorphisms Enhancing the Risk for Sporadic Colorectal Cancer Related to Red Meat Consumption

Sébastien Kury,^{1,2} Bruno Buecher,^{2,3} Sébastien Robiou-du-Pont,^{1,2} Catherine Scoul,^{1,2} Véronique Sébille,⁴ Hélène Colman,^{1,2} Claire Le Houérou,² Tanguy Le Neel,^{1,5} Jérémie Bourdon,⁶ Roger Faroux,⁷ Jean Ollivry,⁹ Bernard Lafraisse,¹⁰ Louis-Dominique Chupin,⁸ and Stéphane Bézieau^{1,2}

¹Laboratoire d'Etude de l'ADN, Faculté de Médecine de Nantes, France; ²Institut des Maladies de l'Appareil Digestif et CIC Institut National de la Santé et de la Recherche Médicale and ³Service d'Hépatogastroentérologie, CHU de Nantes; ⁴Laboratoire de Biomathématiques-Biostatistique, Faculté de Pharmacie, Université de Nantes; ⁵Biofortis Nantes; ⁶Laboratoire d'Informatique de Nantes Atlantique, Université de Nantes; ⁷Service d'Hépatogastroentérologie, CHD de La Roche-Sur-Yon, France; ⁸Centre d'Examen de Santé, La Roche-Sur-Yon; ⁹Association des Gastroentérologues de Vendée, Challans, France; and ¹⁰Centre d'Examen de Santé de Saint-Nazaire, France

Abstract

Susceptibility to sporadic colorectal cancers (CRC) is generally thought to be the sum of complex interactions between environmental and genetic factors, all of which contribute independently, producing only a modest effect on the whole phenomenon. However, to date, most research has concealed the notion of interaction and merely focused on dissociate analyses of risk factors to highlight associations with CRC. By contrast, we have chosen a combinative approach here to explore the joint effects of several factors at a time. Through an association study based on 1,023 cases and 1,121 controls, we examined the influence on CRC risk of environmental factors coanalyzed with combinations of six single nucleotide polymorphisms located in cytochrome P450 genes (c.-163A>C

and c.1548T>C in *CYP1A2*, g.-1293G>C and g.-1053C>T in *CYP2E1*, c.1294C>G in *CYP1B1*, and c.430C>T in *CYP2C9*). Whereas separate analyses of the SNPs showed no effect on CRC risk, three allelic variant combinations were found to be associated with a significant increase in CRC risk in interaction with an excessive red meat consumption, thereby exacerbating the intrinsic procarcinogenic effect of this dietary factor. One of these three predisposing combinations was also shown to interact positively with obesity. Provided that they are validated, our results suggest the need to develop robust combinative methods to improve genetic investigations into the susceptibility to CRC. (Cancer Epidemiol Biomarkers Prev 2007;16(7):1460-7)

Introduction

Given its high prevalence and poor prognosis, colorectal cancer (CRC) is very much a public health issue in industrialized Western countries. In France, it is the third most frequent cancer (>36,000 new cases were diagnosed in 2000) and the second cause of cancer-related death (16,000 deaths were caused by CRC in 2000 for both sexes; ref. 1). Moreover, the current aging of the French population tends to drastically increase the number of incident cases, which warrants research studies on strategy development for CRC prevention.

The vast majority of CRCs are sporadic disorders, the genesis of which includes both environmental and heritable factors (2). Epidemiologic studies on Western populations have emphasized the large contribution of food and lifestyle to sporadic CRC risk (3-9). High-fat and low-fiber diets, as well as alcohol, tobacco, and red or processed meat consumption, have been shown to produce high levels of polycyclic aromatic hydrocarbons and heterocyclic aromatic amines. These pro-

carcinogenic agents are potentially very harmful and may play a key role in the malignant transformation of cells by interacting with DNA (10, 11). However, their biological deleterious effects cannot be exerted without an oxidative activation by phase I cytochrome P450 (CYP) isoenzymes (12). CYPs are encoded by highly polymorphic genes, which results in a remarkable interindividual variability of enzymatic activity. Therefore, the level of procarcinogens activated by an individual depends on genetic background, which determines the degree of susceptibility to procarcinogen exposure and, consequently, to cancer.

Genetic association studies have reported several examples of single nucleotide polymorphisms (SNP) in CYP genes correlated to modulations of CRC risk, whether alone (13-22) or following interaction with environmental factors such as red meat or cigarette consumption (12, 23-26). Despite laying the foundations for further genetic investigations, these studies still provide too partial a view of the cancer susceptibility phenomenon. In fact, they do not take into account the functional redundancy of CYP enzymes and they usually only analyze the modifications of cancer risk associated with an environmental factor and a unique SNP at a time. In fact, the same alimentary compound can be the substrate of numerous CYP enzymes, the polymorphisms of which may all interfere with CRC risk. Rather than unique genetic polymorphisms, these are combinations of polymorphisms that determine the individual genetic background that will finally potentiate the intrinsic promoting or protective effect of xenobiotics on colorectal carcinogenesis (27). Thus, a clue to the understanding of susceptibility to CRC probably involves an examination of these combinations of polymorphisms conducted in the environmental setting.

Received 3/22/07; revised 4/30/07; accepted 5/10/07.

Grant support: Regional Hospital Clinical Research Program and the Groupement des Entreprises Françaises dans la Lutte contre le Cancer, the Association Anne de Bretagne Génétique, and the Ligue Régionale Contre le Cancer.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Note: Supplementary data for this article are available at Cancer Epidemiology Biomarkers and Prevention Online (<http://cebp.aacrjournals.org/>).

S. Kury, B. Buecher, and S. Robiou-du-Pont contributed equally to this work.

Requests for reprints: Sébastien Kury, Laboratoire d'Etude de l'ADN, Faculté de Médecine de Nantes, 1 rue Gaston Veil, 44035 Nantes Cedex, France. Phone: 33-240-41-29-69. E-mail: Sébastien.Kury@univ-nantes.fr

Copyright © 2007 American Association for Cancer Research.

doi:10.1158/1055-9965.EPI-07-0236

This is precisely the target we set ourselves through the population-based case-reference study that we report here. In the present work, we coanalyzed six SNPs of CYP genes—c.-163A>C and c.1548T>C in *CYP1A2*, g.-1293G>C and g.-1053C>T in *CYP2E1*, c.1294C>G in *CYP1B1*, and c.430C>T in *CYP2C9*—and various environmental factors selected from a bibliographical meta-analysis among the factors presumably the most relevant to colorectal carcinogenesis. We investigated combinations of CYP allelic variants that could serve as biomarkers for CRC prevention and/or prognosis, regarding their possible influence on the relative risk for developing CRC when interacting with environmental factors.

Materials and Methods

Experimental Design and Assembly of the Cohorts. To investigate potential associations between sporadic CRCs and both environmental and genetic factors, we conducted a case-controlled association study. From December 2002 to March 2006, we assembled two cohorts of 1,023 patients and 1,121 controls. The recruitment of these individuals was carried out in the Pays de la Loire region, which is located in the west of France and incorporates the administrative districts of Loire Atlantique, Maine et Loire, Mayenne, Sarthe, and Vendée. Six public hospitals and five clinics situated within these districts gave their agreement to collaborate in this project, thereby constituting a network of hepatogastroenterologists and oncologists specialized in CRCs. All patients with a personal history of CRC diagnosed at an age ≥ 40 years were eligible for the case cohort provided that there was no suspicion of major genetic predisposition (familial form of CRC).

Most of the control individuals were selected during a routine check-up by physicians working in one of two Health Examination Centers of region Pays de la Loire. Recruitment of individuals ≥ 70 years was completed in the departments of internal medicine and hepatogastroenterology of the University Hospital of Nantes. In this way, we made up a control cohort that matched the case cohort according to sex, 5-year age groups, and geographic origins. Control individuals were eligible to be included in the study if they were ≥ 40 years of age, lived in the study region, spoke French, and were physically and mentally able to participate in a personal interview of ~ 15 to 20 min. Unaffected individuals with a familial history of CRC or polyps were noneligible.

To meet the requirements of the local (Comité de Protection des Personnes dans la Recherche Biomédicale) and national (Commission nationale de l'Informatique et des Libertés) ethics committees that approved the study, a written informed consent was obtained from each individual included, after having been clearly explained the research protocol by a physician. During the examination, two venous blood samples of 4 mL each were collected in EDTA tubes from the participant and were used later on for extraction of genomic DNA. All information regarding participants was made anonymous after the collection of blood samples.

In the presence of the physician, each participant answered the same one-page standardized questionnaire that was especially drawn up to provide valuable information on demographic characteristics, regular diet, weight and height, personal history of diabetes mellitus and dyslipidemia, physical activity, and smoking habits (see Fig. 1). In this multiple-choice questionnaire, we did not take into account every potential risk factor reported in the literature but we focused more on some of the environmental factors among the most relevant to colorectal carcinogenesis. It is worth noting that we used the grandparents' place of residence—or by default, the parents' place of residence—to define the geographic origin of each participant. Questions about eating habits had been developed with the help of a nutritionist to

cover the main categories of food components found in the French diet. Participants were questioned only about the average consumption frequency of foods, whereas quantitative questions were deliberately avoided to keep the questionnaire clear and simple.

For cases, we additionally requested endoscopy and histology reports from the study oncologist or hepatogastroenterologist. To obtain homogeneous data, a standardized report model designed as a multiple-choice questionnaire was provided to each physician, which notably included questions about tumor aspect and location, presence of metastasis, and tumor infiltration.

All data provided by both the participants and the physicians through the different questionnaires were stored in a common electronic SQL database set up for the sole purpose of the present study. This database also enables the collection of genetic data generated as explained below.

Genotype Analysis. Genomic DNA was extracted from collected blood samples using the Nucleon BACC2 kit (GE Healthcare). DNA concentrations were calculated using PicoGreen technology (Molecular Probes) and diluted to 10 ng/ μ L in 96-well format after a customized automated procedure developed by Microlab Star (Hamilton Robotics).

All study participants were genotyped for six SNPs related to drug-metabolizing enzymes CYP involved in heterocyclic aromatic amines and polycyclic aromatic hydrocarbons carcinogen metabolism: SNPs rs762551 (c.-163A>C; *CYP1A2*^F) and rs2470890 (c.1548T>C; p.N516N) in *CYP1A2*, rs3813867 (g.-1293G>C; *CYP2E1*^{5A} and *CYP2E1*^{5B}) and rs2031920 (g.-1053C>T; *CYP2E1*^{5A} and *CYP2E1*^{5B}) in *CYP2E1*, rs1056836 (c.1294C>G; p.V432L; *CYP1B1*³) in *CYP1B1*, and rs1799853 (c.430C>T; p.R144C; *CYP2C9*^{2A}, *CYP2C9*^{2B}, and *CYP2C9*^{2C}) in *CYP2C9*. These six SNPs were selected for their relevance to sporadic CRC according to a bibliographical meta-analysis done through PubMed.

Genotypes were determined using high-throughput TaqMan allelic discrimination tests. Primers and dye-labeled MGB-NFQ probes were designed and synthesized by Applied Biosystems (Supplementary Table S1). Reactions were set up in a 384-well plate in a 6 μ L final volume, including 2.25 μ L of 2 \times Universal Master Mix, 0.11 μ L of 40 \times Assay Mix, 1.64 μ L of water, and 20 ng of genomic DNA (10 ng/ μ L). Reaction plates were thermocycled in an i-cycler (Bio-Rad Laboratories): An initial 10 min denaturation at 95°C was followed by 45 cycles including a denaturation step at 92°C for 15 s and an annealing/extension step at 60°C for 1 min. After amplification, end-point fluorescence readings were conducted on an Applied Biosystems ABI 7900HT sequence detection system. Genotypes were assigned using the allelic discrimination software SDS v2.1 (Applied Biosystems).

Statistical Analysis

Determination of CRC Risk Associated with Environmental Factors Analyzed Independently. The statistical analysis was done with SAS 9.1 statistical software (SAS Institute, Inc.). Continuous variables were reported as median and range and compared between the case and control groups by Mann-Whitney *U* test. Categorical variables were reported as number of patients (percentages) and compared by χ^2 or Fisher's exact tests when appropriate. *P* values < 0.05 were considered to be statistically significant.

Univariate analyses were done using conditional logistic regression to account for the matching of cases and controls. An appraisal was made on the association between environmental variables adjusting for age followed by an assessment of the odds ratio (OR) of cancer cases together with 95% confidence intervals (95% CI). Multivariate analyses were then done using an iterative stepwise selection procedure to select the variables that were significantly associated with CRC, as

1462 CYP Gene Polymorphisms and CRC Related to Red Meat

assessed by the likelihood ratio test (variable candidates for the model were those associated with CRC in univariate analyses with the $P < 0.15$ criterion; variables were selected in the model using the $P < 0.05$ criterion).

Determination of CRC Risk Associated with CYP Allelic Variants Analyzed Independently. To assess the risk of CRC associated with each of the six CYP SNPs listed above, we used SNPStats, a free web-based tool designed for genetic epidemiology

purposes (28). SNPStats has a user-friendly interface thanks to the PHP server programming language and uses the free software environment R for statistical computing implemented by the algorithm of the package *genetics*. As regards our study, the association with cancer was estimated for each SNP by OR of cancer cases along with 95% CI; calculations were adjusted for age and sex, and followed an unconditional logistic regression model. For each SNP, homozygous carriers of the

QUESTIONNAIRE – LIFE HABITS to be completed by the patient

Name:

Grands-parents' birthplace (city or region or district):
By default, parents' birthplace:

Caution : a unique answer is required for each question

Are you:

Non-smoker? Current smoker? Former smoker? ⇒ stop date:

What amount of time do you spend on physical activity (e.g. walking, biking, or other sports):

Less than 1 hour / week? 1 to 3 hours / week? More than 3 hours / week?

How often do you drink alcohol:

Never? Sometimes? 2 to 3 times / week? Everyday?

How often do you eat red meat (beef, lamb...):

Once / week or less? 2 to 4 times / week? More than 5 times / week?

How often do you eat poultry and/or white meat (chicken, guinea fowl, turkey, rabbit, calf...):

Once / week or less? 2 to 4 times / week? More than 5 times / week?

How often do you eat cold cuts:

Once / week or less? 2 to 4 times / week? More than 5 times / week?

How often do you eat fish:

Once / week or less? 2 to 4 times / week? More than 5 times / week?

What is your favorite cooking method (for meat, poultry and fish):

Roasting or frying with fats? Roasting or frying without fats?
 Grilling? Steaming?

How often do you eat fruits:

Less than once / day? Once / day? Twice / day or more?

How often do you eat vegetables:

Less than once / day? Once / day? Twice / day or more?

How often do you eat pastries, viennoiseries, ice:

Less than once / week? Once / week? Twice / week or more?

How often do you eat dairy products (milk, cheese, yoghurt, custard dessert...):

Once / day or less? Twice / day? 3 times / day or more?

Do you rather eat:

White bread? Wholemeal bread?

Please, be sure that you answered to all the questions and thank you for your participation.

Figure 1. Standardized questionnaire on life habits fulfilled by the study participants.

Table 1. General characteristics of study patients

	Case (n = 1,023)	Control (n = 1,121)
Age, y	67 (40-99)*	62 (41-101)
Body mass index, kg/m ²		
<18.5	21 (2)*	11 (1)
18.5-24.9	411 (42)	544 (49)
25-29.9	393 (40)	405 (36)
>30.0	152 (16)	157 (14)
Sex ratio		
Male	632 (62)*	609 (54)
Female	391 (38)	512 (46)

NOTE: Data are median (range) for continuous variables or numbers of patients (percentages) for categorical variables.

*P ≤ 0.01.

most frequent allele in the cohorts were used as the reference for the analyses. Calculations of ORs were double checked by statistical analysis done with SAS 9.1 statistical software (SAS Institute).

Analyses of gene-environment interactions were then done with SNPStats, which proposes statistical models for calculation of risk associated with one pair of interacting variables. For each of the six CYP SNPs, we investigated all the possible interactions with the environmental covariables of the questionnaire completed by the study participants, and we calculated 95% CI ORs of cancer cases to determine if any "SNP-environmental factor" pair could modify the risk for CRC in our cohorts. A likelihood ratio test was used to investigate interaction among variables of interest.

Analysis of CRC Risk Associated with Multiple-SNPs Combinations. A special option of the software SNPStats is the haplotype analysis done through R package *haplo.stats*.

Using genotypes obtained at different points of polymorphisms (SNPs) located on the same chromosomal region, the software makes it possible to predict and rebuild the haplotypes characterizing a group of individuals through a log-additive model. Although the six CYP SNPs chosen for our study are not all located on the same chromosome and, by definition, cannot belong to a common haplotype, we nevertheless used this SNPStats option to predict preferential combinations of these six SNPs. Based on the genotyping results determined for the six single SNPs in the patient and control cohorts, an evaluation was made of the relative frequencies of multiple-SNP combinations present in the cohorts and the corresponding samples sizes were calculated. Logistic regression was then applied to these data to analyze the associations of the SNP combinations to CRC risk translated into 95% CI ORs.

By use of SNPStats, we investigated the possible SNP combinations interacting with environmental factors and we estimated their associations to CRC risk with 95% CI ORs.

Results

The present work is based on 1,023 patients with sporadic CRC and 1,121 unaffected controls whose main general characteristics are indicated in Table 1. The mean age is slightly higher for cases than for controls (67 versus 62 years old; P ≤ 0.01), mainly because of the overrepresentation of individuals ≥70 years old in the case cohort [404 individuals (39%) versus 199 (18%) in the control cohort] despite our efforts to match the two cohorts by age. Mean body mass indexes are approximately equal between the two cohorts (26 kg/m² in patients versus 25 kg/m² in controls; P ≤ 0.01). As regards gender distribution, males are more numerous in both cohorts but they are slightly underrepresented in the control cohort (62%

Table 2. Univariate and multiple variable models of promoting and protective factors independently associated with CRC

Variable	Univariate analyses		Multivariate analyses*	
	OR (95% CI)	P	OR (95% CI)	P
Female gender [†]	0.71 (0.59-0.85)	<0.001	0.76 (0.61-0.93)	0.009
Physical activity, <1 h/wk [‡]	1.82 (1.44-2.29)	<0.001	1.77 (1.43-2.17)	<0.001
Tobacco, former and current smokers [§]	1.04 (0.86-1.24)	0.705	NA	
Alcohol consumption, never	1.66 (1.25-2.21)	0.001	2.08 (1.57-2.77)	<0.001
Cold cuts, ≥twice a week [¶]	1.43 (1.19-1.71)	<0.001	1.25 (1.02-1.53)	0.034
Red meat, ≥5 times a week ^{**}	3.41 (1.93-6.04)	<0.001	2.81 (1.52-5.21)	0.001
White meat, ≥2 times a week ^{††}	1.04 (0.86-1.26)	0.701	NA	
Fish, ≥2 times a week ^{†††}	0.91 (0.76-1.08)	0.279	NA	
Vegetables ≥once a day ^{††††}	0.92 (0.67-1.26)	0.597	NA	
Fruits, <twice a day	1.33 (1.11-1.60)	0.002	1.23 (1.01-1.51)	0.040
Dairy products, ≤once a day	1.50 (1.24-1.81)	<0.001	1.41 (1.14-1.74)	0.001
Cooking, grilling, roasting or frying without fats ^{**}	0.87 (0.58-1.31)	0.507	NA	
Cooking, roasting or frying with fats ^{**}	1.08 (0.71-1.64)	0.716	NA	
Bread, wholemeal bread ^{***}	0.56 (0.44-0.70)	<0.001	0.63 (0.49-0.82)	0.001
Diabetes, yes ^{†††}	2.31 (1.62-3.28)	<0.001	2.02 (1.37-2.99)	<0.001
Dyslipidemia, yes ^{††††}	0.74 (0.60-0.92)	0.007	0.65 (0.51-0.83)	0.001
Pastry, once a week ^{†††††}	1.22 (0.93-1.60)	0.146	1.23 (1.00-1.52)	0.049

Abbreviation: NA, not analyzed.

*Multivariate analyses were done only on variables that were selected for their significant association with CRC according to univariate analyses (using the P < 0.05 criterion).

[†]References used for the variables were male gender.

[‡]One to 3 h or >3 h a week.

[§]Nonsmokers.

^{||}Sometimes or twice to thrice a week or everyday.

[¶]Once a week or less.

^{**}Once a week at most or two to four times a week.

^{††}Once a week at most.

^{†††}Less than once a day.

^{|||}Twice a day or more.

^{|||}Twice a day or three times a day or more.

^{***}Steaming.

^{††††}White bread.

^{†††††}No.

^{†††††}Less than once a week.

1464 CYP Gene Polymorphisms and CRC Related to Red Meat

Table 3. Analysis of single-SNP association with CRC ($n = 2,131$; OR calculations were adjusted by sex and age)

Gene	SNP	Genotype	Controls	Patients	OR (95% CI)	P
CYP1A2	c.-163A>C (rs762551)	A/A	553 (49.5%)	514 (50.7%)	1.00	0.85
		A/C	480 (42.9%)	420 (41.5%)	0.98 (0.81-1.17)	
		C/C	85 (7.6%)	79 (7.8%)	1.08 (0.77-1.51)	
CYP1A2	c.1548T>C (rs2470890)	T/T	454 (40.6%)	428 (42.2%)	1.00	0.95
		T/C	523 (46.8%)	464 (45.8%)	0.98 (0.81-1.18)	
		C/C	141 (12.6%)	121 (11.9%)	0.96 (0.72-1.27)	
CYP2E1	g.-1293G>C (rs3813867)	G/G	1,029 (92%)	944 (93.2%)	1.00	0.39
		G/C	88 (7.9%)	67 (6.6%)	0.81 (0.58-1.14)	
		C/C	1 (0.1%)	2 (0.2%)	2.17 (0.18-25.47)	
CYP2E1	g.-1053C>T (rs2031920)	C/C	1,027 (91.9%)	940 (92.8%)	1.00	0.067
		C/T	90 (8.1%)	67 (6.6%)	0.80 (0.57-1.12)	
		T/T	1 (0.1%)	6 (0.6%)	5.77 (0.67-49.51)	
CYP1B1	c.1294C>G (rs1056836)	C/C	368 (32.9%)	317 (31.3%)	1.00	0.22
		C/G	576 (51.5%)	507 (50%)	0.99 (0.81-1.20)	
		G/G	174 (15.6%)	189 (18.7%)	1.22 (0.94-1.58)	
CYP2C9	c.430C>T (rs1799853)	C/C	821 (73.4%)	745 (73.5%)	1.00	0.61
		C/T	280 (25%)	247 (24.4%)	0.96 (0.78-1.17)	
		T/T	17 (1.5%)	21 (2.1%)	1.34 (0.69-2.60)	

in patients versus 54% in controls; $P \leq 0.01$). Overall, study participants were very compliant as the global response rate to the questionnaire on life habits reached 97.5%; the only relatively weak response rate was in relation to the "method of cooking" item, giving a response of 88.2%.

Table 2 describes the more significant results obtained with the univariate and multivariate analyses of environmental factors. In the study cohorts, the consequence of red meat consumption on CRC risk was very noticeable and quite distinguishable from those of other environmental factors: Compared with the reference group of moderate eaters, high consumers (≥ 5 times a week) were at increased risk with an OR of 3.414 (95% CI, 1.93-6.04). To a lesser extent, diabetes status, lack of physical activity, as well as frequent cold cuts and pastry consumption, or, on the contrary, poor-quality fruit intake and low dairy product and fruit consumption were also associated with an increased risk of CRC. Paradoxically, the

lack of alcohol consumption was also associated with a decreased CRC risk. By contrast, female gender, dyslipidemia status, and frequent consumption of wholemeal bread decreased the risk of CRC. In addition, we found no effect at all for fish, white meat and cigarette consumption, or cooking methods.

From these observations, we also examined combinations of environmental factors found to have a more obvious bearing on CRC risk. As expected, we observed that individuals with the lowest risk of CRC were those who combined the highest consumption of fruit (more than twice a day), vegetables (more than twice a day), dairy products (more than thrice a day), and wholemeal bread. They were associated with an OR of 0.78 (95% CI, 0.62-0.98, $P = 0.036$) compared with reference individuals who had the same nutritional habits except that they ate white bread instead of wholemeal bread. People with the highest risk of CRC were

Table 4. Analysis of association to CRC risk of multiple CYP SNPs combinations interacting with the environmental variable "red meat consumption"

Combination	No. individuals exhibiting the combination	Red meat consumers ≤ 4 times/wk		Red meat consumers ≥ 5 times/wk	
		OR (95% CI)		OR (95% CI)	P
A. Multiple-SNPs combination and red meat cross-classification interaction table ($n = 2,087$, adjusted by age and sex)					
ATGCC	659 (31.6%)	1.00		13.72 (8.53-22.06)	<0.001
ATGCCG	461 (22.1%)	1.09 (0.88-1.36)		4.81 (2.08-11.14)	<0.001
CCGCC	280 (13.4%)	0.95 (0.73-1.24)		3.88 (1.20-12.55)	0.024
CCGCCG	204 (9.8%)	0.87 (0.66-1.15)		14.20 (7.64-26.42)	<0.001
ATGCCGT	92 (4.4%)	0.84 (0.54-1.32)		36.87 (18.93-71.81)	<0.001
B. Multiple-SNPs combinations within red meat ($n = 2,087$, adjusted by age and sex)					
ATGCC	659 (31.6%)	1.00		1.00	
ATGCCG	461 (22.1%)	1.09 (0.88-1.36)		0.35 (0.17-0.72)	0.004
CCGCC	280 (13.4%)	0.95 (0.73-1.24)		0.28 (0.12-0.65)	0.003
CCGCCG	204 (9.8%)	0.87 (0.66-1.15)		1.04 (0.76-1.41)	0.804
ATGCCGT	92 (4.4%)	0.84 (0.54-1.32)		2.69 (1.72-4.19)	<0.001
C. Red meat within multiple-SNPs combinations ($n = 2,087$, adjusted by age and sex)					
ATGCC	659 (31.6%)	1.00		13.72 (8.53-22.06)	<0.001
ATGCCG	461 (22.1%)	1.00		4.42 (1.92-10.15)	<0.001
CCGCC	280 (13.4%)	1.00		4.08 (1.26-13.21)	0.019
CCGCCG	204 (9.8%)	1.00		16.34 (9.47-28.21)	<0.001
ATGCCGT	92 (4.4%)	1.00		43.78 (27.05-70.88)	<0.001

NOTE: Allelic variants composing the combinations of six SNPs are ordered as follows: first CYP1A2 c.-163A>C, second CYP1A2 c.1548T>C, third CYP2E1 g.-1293G>C, fourth CYP2E1 g.-1053C>T, fifth CYP1B1 c.1294C>G, and sixth CYP2C9 c.430C>T. Only the five more frequent combinations (>4%) are represented. (A) Cross-classification using a common reference category for both interacting variables. The group of moderate red meat consumers (≤ 4 times/wk) exhibiting the most frequent combination of SNPs is selected as the reference category. ORs are estimated, together with 95% CI, for all other combinations distributed between the groups of moderate and great (≥ 5 times/wk) red meat consumers, according to a log-additive model. Significant associations to modification of CRC risk are typed in bold characters. (B) The influence of SNP combination on modification of CRC risk is estimated by ORs calculation within each group of moderate and great red meat consumers. (C) Red meat consumption nested within SNP combinations: The influence of the frequency of red meat consumption is estimated by OR calculation within each combination of SNPs.

those eating fruit and vegetables once a day at the most, dairy products twice a day at the most, and white bread (OR, 1.36, 95% CI, 1.15-1.60, $P = 0.0004$).

Table 3 shows the results of the univariate analyses done with SNPStats for the six single SNPs selected: rs762551 (c.-163A>C) and rs2470890 (c.1548T>C) in *CYP1A2*, rs3813867 (g.-1293G>C) and rs2031920 (g.-1053C>T) in *CYP2E1*, rs1056836 (c.1294C>G) in *CYP1B1*, and rs1799853 (c.430C>T) in *CYP2C9*. We did not find any association with CRC for any of the allelic variants tested, as ORs never reached statistical significance either for proximal and distal colon, or rectal cancer. Moreover, analyses of interactions between single SNPs and environmental factors of the questionnaire did not show any interacting pair of variables that would modify the risk of CRC (Supplementary Table S2). The results obtained with SNPStats were confirmed with SAS 9.1, by performing univariate analyses adjusted for the covariate SNP of the environmental factors found to be potential risk factors (Table 3). As an example, ORs associated with a high consumption of red meat (more than five times a week) remained steady and statistically significant at about 3.4 ($P < 0.001$) when adjusted for any of the six SNPs.

We then analyzed the association between CRC and combinations of multiple SNPs. We tested combinations of two to six SNPs: None was found to be associated with a modification of CRC risk. Therefore, we investigated interactions between these combinations and environmental factors of the questionnaire. The integration of multiple-SNP combinations to the statistical analyses barely changed the ORs calculated above for most environmental factors, especially for the potential risk factors listed in Table 2. However, as summarized in Table 4, striking results were once again obtained with red meat consumption. Part A of Table 4 indicates that three multiple-SNP combinations—ATGCGT, CCGCGC, and ATGCCC (allelic variants in the combinations are ordered as follows: *CYP1A2* c.-163A>C, then *CYP1A2* c.1548T>C, *CYP2E1* g.-1293G>C, *CYP2E1* g.-1053C>T, *CYP1B1* c.1294C>G, and *CYP2C9* c.430C>T)—were associated with a strong elevation of CRC risk in significant red meat consumers when compared with a common reference group of moderate red meat consumers presenting the most frequent combination ATGCCC. Part B shows that the risk of CRC did not vary significantly between moderate red meat consumers (<4 times a week) whatever SNP combination they exhibited; on the other hand, in significant red meat consumers, the influence of SNP combination was more perceptible as individuals with combination ATGCGT were associated with a significantly increased risk (OR, 2.69; 95% CI, 1.72-4.19) compared with combination ATGCCC; by contrast, combinations of ATGCGC and CCGCCC were at lower risk with ORs dropping to 0.35 (0.17-0.72) and 0.28 (0.12-0.65). Part C of Table 4 takes into account the influence of red meat consumption in the different SNPs combinations; it shows that the intrinsic CRC risk of red meat consumption expressed in Table 2 (OR, 3.414 in great red meat consumers) was slightly increased for combinations (OR, 4.42 and 4.08) ATGCGC and CCGCCC, but considerably enhanced for combinations CCGCGC (OR, 16.34) and ATGCCC (OR, 13.72), and extremely high for combination ATGCGT (OR, 43.78).

Further analyses of SNPs combinations revealed an interaction between combination ATGCGT and body weight (Supplementary Table S3). Compared with individuals with low to normal weight (body mass index <18.5-24.9; see Table 1), obese individuals (body mass index ≥ 30) were at a statistically significantly increased risk of CRC (OR, 12.55; 95% CI, 6.28-25.09), and overweight individuals (body mass index 25-29.9) were associated with a barely significantly increased risk (OR, 2.24; 95% CI, 0.94-5.31). None of the other most frequent SNP combinations (>4%) showed an interaction with body weight.

Discussion

Colorectal carcinogenesis has been the subject matter of many studies, all of which confirm an environmental and genetic bicomponent for carcinogenesis of sporadic CRC (2). Numerous population-based studies have underlined associations between risk for CRC and SNPs—notably including the SNPs of the CYP genes we have chosen in our study—interacting with at least one environmental factor. Nevertheless, most of these studies do not produce a sufficient statistical power, and are therefore not reproducible, because they rely upon a too small number of nonhomogeneous or even badly designed cohorts (usually 40-300 individuals; ref. 29). To circumvent this problem, we based our association study on two large cohorts including more than 1,000 individuals each. To date, only a few population-based association studies as large as ours have been reported on genetic susceptibility to CRC and, to our knowledge, none of them have comprised uniquely sporadic cases. Moreover, no comparable studies have been reported on a French population, which highlights the interest and originality of our cohorts, as genetic predisposition is definitively a population-specific phenomenon. In addition, we were especially careful about the quality of the recruitment, whether this was the clinical criteria for patients (every doubtful familial case was excluded systematically) or the homogeneity of both cohorts (i.e., matching as well as possible for age, sex, and geographic origins). We are aware that a weak point of our recruitment resides in the imbalance between patients and controls of over 70 years old; however, simulations of OR calculations on population samples showed that this imbalance did not seem to alter our estimations of CRC risk.

As mentioned in Results, response rate to the questionnaire on life habits was very good, which was translated by a high statistical significance of the results we obtained for the univariate analyses of environmental factors ($P < 0.001$ for many analyses). These results were consistent with previous descriptions of environmental risk factors in other epidemiologic studies (30-35). Thus, physical activity, fruit, dairy products, and wholemeal bread appeared as protective factors regarding the risk of CRC. However, above all, the most striking result concerned the high consumption of red meat (more than five times a week), which we found associated with a sizable increase in CRC risk (OR, 3.414; 95% CI, 1.93-6.04) compared with a more reasonable consumption (four times a week at the most). Of course, this observation would be moderated by the fact that individuals with this habit constitute only a minor group within both cohorts (69 individuals; i.e., 3.2% of the whole study population). Nevertheless, the procarcinogenic role of red meat is no longer a novelty because it has already been reported several times in the literature and it has been biologically explained by the metabolism of red meat into deleterious polycyclic aromatic hydrocarbons and heterocyclic aromatic amines (36). In any case, the size of the two study cohorts leads us to believe that the number of frequent red meat eaters—although they are small in number—is representative of the local French population we examined. Therefore, the results we obtained may be a little overestimated but they reflect a true tendency.

In the same way, certain other results deserve to be discussed here. For example, the observation of a reduced risk of CRC in patients with dyslipidemia was unexpected. As dyslipidemia was taken into account only in patients requiring medical therapy, we may assume that most of them were treated with inhibitors of hydroxyl-methylglutaryl CoA reductase (also called statins), a widely prescribed class of lipid-lowering agents. This may be a confounding factor as these agents exert a documented protective effect against CRC (37). As regards the lack of effect relating to fish and vegetables, this can be ascribed to an excessively poor representation of the "high consumer categories" for these two covariates. Despite

Public Health communications on the virtues of these two products against cancer, they have failed to become essential elements in today's French diet. In contrast, the absence of influence of tobacco on CRC risk is not that surprising because its deleterious effect was essentially described in lung and upper digestive tract cancers (38, 39). Moreover, the participants' answers may have been distorted, notably in patients for whom tobacco consumption appears to be a deliberate risk habit. The same remark could be applied to alcohol consumption, the relative protective effect of which rather contradicted the data reported in the literature and is illogical in terms of detoxification. As regards covariate "cooking methods," the absence of significant results can be correlated to the relatively low response rate (88.2%), which certainly translates the difficulty encountered by the participants to choose a single answer definitively representative of their habits. Indeed, this problem probably comes from the formulation of the question.

The results of the analysis of multiple environmental factor combinations turned out to be extremely logical. Frequent eating of meals rich in fruit, vegetables, dairy products, and wholemeal bread appeared as the most effective combination against CRC. Yet, the relatively weak difference found between single or combined environmental factors underline a barely detectable additive effect or even an independent effect of these covariates on carcinogenesis in the study population.

For the genetic part of our work we focused on six SNPs of four CYP genes that all belong to phase I of the xenobiotic detoxification metabolic pathway, and that are not localized on the same chromosome. We therefore expected to find an additive or rather interactive effect not due to real haplotypes but to variations in the activity of putative interdependent encoded enzymes. According to our bibliographical meta-analysis of previous association studies, five of the six SNPs of CYP genes that we analyzed had already been found to be associated with a rise in risk for colorectal adenomas or cancer, whether alone or in interaction with an environmental factor. In general, the increased risk had been correlated to allelic variants increasing the activity of encoded enzymes, whether for SNPs c.-163A>C (rs762551) in *CYP1A2* (13, 19-21, 40), g.-1053C>T (rs2031920) and g.-1293G>C (rs1056836) in *CYP2E1* (14, 15), c.1294C>G (rs1056836) in *CYP1B1* (17), or c.430C>T (rs1799853) in *CYP2C9* (16, 18). However, we were unable to reproduce these results in our cohorts, as we found no association with modulation of CRC risk for any of the six single SNPs analyzed independently or together with environmental factors in our lifestyle questionnaire. An explanation for the divergence with prior studies probably lies in the smaller populations tested: For instance, for *CYP1A2* c.-163A>C, 94 individuals (OR, 3.7; 95% CI, 1.3-10.7) in one study (20) and 377 cases/326 controls in yet a second more comprehensive study (OR, 1.53; 95% CI, 1.13-2.08; ref. 21), and 163 patients and 163 controls (OR, 1.91; 95% CI, 1.05-3.52) for *CYP2E1* g.-1053C>T (15).

On the other hand, analyses of multiple-SNP combinations showed a strong increase in the intrinsic risk of excessive red meat consumption for three different combinations of allelic variants, ATGCCC, CCGCGC, and ATGCGT, found in 31.6%, 13.4%, and 4.4% of the study individuals respectively (see Table 4). As opposed to these three allelic variant combinations, two other frequent ones, ATGCGC (22.1%) and CCGCCC (9.8%), appeared "neutral" as they barely enhanced the basal CRC risk due to red meat consumption alone (see Table 2). Interaction analyses with other environmental factors revealed that, among the combinations above, ATGCGT was the one that predisposed the most to CRC, as it was the only one found to be additionally associated with a perceptible increase in the risk of CRC in obese individuals (~14.4% of the study population). This last observation is a good example of gene-environment interaction: In a way, the genetic background seems to exacerbate the procarcinogenic effect of the environmental factor, because, in our cohorts, obesity alone

was not found to be associated with an increased CRC risk as reported in the literature (41). On the other hand, genetic expression seems to be influenced by environmental factors, as independent analysis of the ATGCGT combination showed no association with CRC risk. Interestingly, obesity was shown to increase expression and thereby oxidative activities of CYP enzymes, especially CYP1E2, which exacerbates the procarcinogenic effect of xenobiotics (42, 43).

A hypothesis for the marked interaction of combination ATGCGT with obesity and red meat consumption would be that the allelic variants composing this combination contribute to the simultaneous enhancement of the metabolic activities of *CYP1A2*, *CYP2E1*, *CYP1B1*, and *CYP2C9* enzymes. Yet, if we postulate from previous works the theoretical combination leading to the greatest enzymatic activity of these four enzymes, we obtain ATGCGC (i.e., a combination we found not to be associated with an effect on CRC risk in our cohorts). In fact, the first four allelic variants of this combination perfectly match the observed ATGCGT combination and they correspond to allelic variants of *CYP1A2* (c.-163A and c.1548T) and of *CYP2E1* (g.-1293G and g.-1053C) that have been shown to induce an increased enzymatic activity *in vitro*, and which have been found to be associated with an increase in CRC risk by interacting with red or processed meat (14, 19, 20, 44). This first part of the combination should be associated with an elevated production of polycyclic aromatic hydrocarbons and heterocyclic aromatic amines, and thereby to an increased cancer risk. In the same way, the "neutral" effect on CRC risk observed for combination CCGCCC can be explained as the result of the concomitant decreased activity of *CYP1A2* and increased activity of *CYP2E1*. In fact, the divergence between observed and theoretical predisposing hexanucleotide combinations resides in the last two variants corresponding to c.1294C>G of *CYP1B1* and c.430C>T of *CYP2C9*. Aklillu et al. (45) showed that the allelic variant c.1294C>G of *CYP1B1* increases enzymatic activity, but, to date, it has only been found to be associated with lung and head-and-neck cancers (46-48). The allelic variant c.430C>T of *CYP2C9* was found to be associated with a decrease in CRC risk (16). However, the size of the population studied in this report was small and the significance of these results may be disputable. Therefore, this divergence between theory and observation does not discredit the predisposing bearing on CRC that we found for the ATGCGT combination. On the contrary, it illustrates the difficulty in extrapolating a general biological model from the combination of independent analyses. This contrasts with environmental factors for which theory perfectly matched observations. It is precisely the very notion of dependence or rather interactivity, between enzymes of the same pathway in this case, which creates all the complexity of the susceptibility phenomenon.

In conclusion, we observed three combinations of CYP allelic variants that strongly predispose to CRC by interacting and enhancing the intrinsic predisposing potential of excessive red meat consumption. One of these combinations, ATGCGT, has a particularly strong effect on CRC risk and also interacts positively with obesity, another known promoting factor of CRC. Our present work illustrates the need to go beyond the analysis of single SNPs proposed in most of the current studies on genetic predisposition for complex diseases like cancer. Studies on predisposition to colorectal and prostate cancer have already started to show the power of combinative approaches (49-52). In addition, our present data highlight the very interest of analyzing SNP combinations in connection with environmental risk factors to a better appreciation of the complexity of *in vivo* events accounting for cancer susceptibility.

We are aware that our approach can be considerably improved. The most significant results presented here relate indeed to a small subgroup of the cohort [69 great red meat consumers (3.2%) of 2,144 individuals], and they would need to be reproduced and confirmed on larger cohorts. In addition,

investigation of a more comprehensive panel of SNPs would certainly lead to even more conclusive observations. Therefore, we now plan to examine additional polymorphic genes involved in the metabolic pathway of xenobiotic detoxification, such as genes of glutathione S-transferases, UDP-glucuronosyl transferases, or sulfotransferases. Nevertheless, in the present work, we deliberately reduced the number of studied SNPs in a didactical perspective to better illustrate the phenomenon of interaction between genes.

Although our combinative approach contains room for improvement, we believe that it is more meaningful than association studies dealing with one risk factor at a time. The identification of SNP combinations modulating CRC risk cancer in a precise environmental context is a first step toward the determination of genetic profiles of susceptibility, which should enable the establishment of personal nutritional recommendations for patients.

Acknowledgments

We thank all the patients, control individuals, and physicians who participated in this work and made it possible.

References

- Remontet L, Esteve J, Bouvier AM, et al. Cancer incidence and mortality in France over the period 1978-2000. *Rev Epidemiol Sante Publique* 2003;51:3-30.
- Lichtenstein P, Holm NV, Verkasalo PK, et al. Environmental and heritable factors in the causation of cancer-analyses of cohorts of twins from Sweden, Denmark, and Finland. *N Engl J Med* 2000;343:78-85.
- Gerhardsson de Verdier M, Hagman U, Steineck G, Rieger A, Norell SE. Diet, body mass and colorectal cancer: a case-referent study in Stockholm. *Int J Cancer* 1990;46:832-8.
- Cope GF, Wyatt JJ, Pinder IF, Lee PN, Heatley RV, Kelleher J. Alcohol consumption in patients with colorectal adenomatous polyps. *Gut* 1991;32:70-2.
- Levin B. Nutrition and colorectal cancer. *Cancer* 1992;70:1723-6.
- Colbert LH, Hartman TJ, Malila N, et al. Physical activity in relation to cancer of the colon and rectum in a cohort of male smokers. *Cancer Epidemiol Biomarkers Prev* 2001;10:265-8.
- Kushi L, Giovannucci E. Dietary fat and cancer. *Am J Med* 2002;113 Suppl 9B:63-70S.
- Slattery ML, Edwards S, Curtin K, et al. Physical activity and colorectal cancer. *Am J Epidemiol* 2003;158:214-24.
- Heavey PM, McKenna D, Rowland IR. Colorectal cancer and the relationship between genes and the environment. *Nutr Cancer* 2004;48:124-41.
- Potter JD. Nutrition and colorectal cancer. *Cancer Causes Control* 1996;7:127-46.
- Berlau J, Glej M, Pool-Zobel BL. Colon cancer risk factors from nutrition. *Anal Bioanal Chem* 2004;378:737-43.
- Raunio H, Husgafvel-Pursiainen K, Anttila S, Hietanen E, Hirvonen A, Pelkonen O. Diagnosis of polymorphisms in carcinogen-activating and inactivating enzymes and cancer susceptibility-a review. *Gene* 1995;159:113-21.
- Sachse C, Brockmoller J, Bauer S, Roots I. Functional significance of a C→A polymorphism in intron 1 of the cytochrome P450 CYP1A2 gene tested with caffeine. *Br J Clin Pharmacol* 1999;47:445-9.
- Tan W, Song N, Wang GQ, et al. Impact of genetic polymorphisms in cytochrome P450 2E1 and glutathione S-transferases M1, T1, and P1 on susceptibility to esophageal cancer among high-risk individuals in China. *Cancer Epidemiol Biomarkers Prev* 2000;9:551-6.
- Kiss I, Sandor J, Pajkos G, Bogner B, Hegedus G, Ember I. Colorectal cancer risk in relation to genetic polymorphism of cytochrome P450 1A1, 2E1, and glutathione-S-transferase M1 enzymes. *Anticancer Res* 2000;20:519-22.
- Martinez C, Garcia-Martin E, Ladero JM, et al. Association of CYP2C9 genotypes leading to high enzyme activity and colorectal cancer risk. *Carcinogenesis* 2001;22:1323-6.
- Sachse C, Smith G, Wilkie MJ, et al. A pharmacogenetic study to investigate the role of dietary carcinogens in the etiology of colorectal cancer. *Carcinogenesis* 2002;23:1839-49.
- de Jong MM, Nolte JM, te Meerman GJ, et al. Low-penetrance genes and their involvement in colorectal cancer susceptibility. *Cancer Epidemiol Biomarkers Prev* 2002;11:1332-52.
- Sachse C, Bhamra U, Smith G, et al. Polymorphisms in the cytochrome P450 CYP1A2 gene (CYP1A2) in colorectal cancer patients and controls: allele frequencies, linkage disequilibrium and influence on caffeine metabolism. *Br J Clin Pharmacol* 2003;55:68-76.
- Moonen H, Engels L, Kleinjans J, Kok T. The CYP1A2-164A→C polymorphism (CYP1A2*1F) is associated with the risk for colorectal adenomas in humans. *Cancer Lett* 2005;229:25-31.
- Landi S, Gemignani F, Moreno V, et al. A comprehensive analysis of phase I, phase II, metabolism gene polymorphisms and risk of colorectal cancer. *Pharmacogenomics* 2005;15:535-46.
- Rodriguez-Antona C, Ingelman-Sundberg M. Cytochrome P450 pharmacogenetics and cancer. *Oncogene* 2006;25:1679-91.
- Lang NP, Butler MA, Massengill J, et al. Rapid metabolic phenotypes for acetyltransferase and cytochrome P4501A2 and putative exposure to food-borne heterocyclic amines increase the risk for colorectal cancer or polyps. *Cancer Epidemiol Biomarkers Prev* 1994;3:675-82.
- Roberts-Thomson IC, Butler WJ, Ryan P. Meat, metabolic genotypes and risk for colorectal cancer. *Eur J Cancer Prev* 1999;8:207-11.
- Le Marchand L, Hankin JH, Wilkens LR, et al. Combined effects of well-done red meat, smoking, and rapid N-acetyltransferase 2 and CYP1A2 phenotypes in increasing colorectal cancer risk. *Cancer Epidemiol Biomarkers Prev* 2001;10:1259-66.
- Tranah GJ, Chan AT, Giovannucci E, Ma J, Fuchs C, Hunter DJ. Epoxide hydrolase and CYP2C9 polymorphisms, cigarette smoking, and risk of colorectal carcinoma in the Nurses' Health Study and the Physicians' Health Study. *Mol Carcinog* 2005;44:21-30.
- Bartsch H, Nair U, Risch A, Rojas M, Wikman H, Alexandrov K. Genetic polymorphism of CYP genes, alone or in combination, as a risk modifier of tobacco-related cancers. *Cancer Epidemiol Biomarkers Prev* 2000;9:3-28.
- Sole X, Guino E, Vallis J, Iñiesta R, Moreno V. SNPStats: a web tool for the analysis of association studies. *Bioinformatics* 2006;22:1928-9.
- Brennan P. Gene-environment interaction and aetiology of cancer: what does it mean and how can we measure it? *Carcinogenesis* 2002;23:381-7.
- Bingham SA, Day NE, Luben R, et al. Dietary fibre in food and protection against colorectal cancer in the European Prospective Investigation into Cancer and Nutrition (EPIC): an observational study. *Lancet* 2003;361:1496-501.
- Peters U, Sinha R, Chatterjee N, et al. Dietary fibre and colorectal adenoma in a colorectal cancer early detection programme. *Lancet* 2003;361:1491-5.
- Sanjoaquin MA, Appleby PN, Thorogood M, Mann JI, Key TJ. Nutrition, lifestyle and colorectal cancer incidence: a prospective investigation of 10998 vegetarians and non-vegetarians in the United Kingdom. *Br J Cancer* 2004;90:118-21.
- Soerjomataram I, de Vries E, Pukkala E, Coebergh JW. Excess of cancers in Europe: A study of eleven major cancers amenable to lifestyle change. *Int J Cancer* 2007;120:1336-43.
- Alvarez-Leon EE, Roman-Vinas B, Serra-Majem L. Dairy products and health: a review of the epidemiological evidence. *Br J Nutr* 2006;96 Suppl 1:594-9.
- Grau MV, Baron JA, Sandler RS, et al. Prolonged effect of calcium supplementation on risk of colorectal adenomas in a randomized trial. *J Natl Cancer Inst* 2007;99:129-36.
- Larsson SC, Wolk A. Meat consumption and risk of colorectal cancer: a meta-analysis of prospective studies. *Int J Cancer* 2006;119:2657-64.
- Tyagi P. Current data with HMG-CoA reductase inhibitors (statins) for colorectal cancer prevention. *Clin Colorectal Cancer* 2005;5:245-6.
- Hashibe M, Boffetta P, Janout V, et al. Esophageal cancer in Central and Eastern Europe: Tobacco and alcohol. *Int J Cancer* 2007;120:1518-22.
- Crispo A, Brennan P, Jockel KH, et al. The cumulative risk of lung cancer among current, ex- and never-smokers in European men. *Br J Cancer* 2004;91:1280-6.
- Bae SY, Choi SK, Kim KR, et al. Effects of genetic polymorphisms of MDR1, FMO3 and CYP1A2 on susceptibility to colorectal cancer in Koreans. *Cancer Sci* 2006;97:774-9.
- Ferrante JM, Ohman-Strickland P, Hudson SV, Hahn KA, Scott JG, Crabtree BF. Colorectal cancer screening among obese versus non-obese patients in primary care practices. *Cancer Detect Prev* 2006;30:459-65.
- McCarver DG, Byun R, Hines RN, Hichme M, Wegenek W. A genetic polymorphism in the regulatory sequences of human CYP2E1: association with increased chlorzoxazone hydroxylation in the presence of obesity and ethanol intake. *Toxicol Appl Pharmacol* 1998;152:276-81.
- Murray M. Altered CYP expression and function in response to dietary factors: potential roles in disease pathogenesis. *Curr Drug Metab* 2006;7:67-81.
- Le Marchand L, Wilkinson GR, Wilkens LR. Genetic and dietary predictors of CYP2E1 activity: a phenotyping study in Hawaii Japanese using chlorzoxazone. *Cancer Epidemiol Biomarkers Prev* 1999;8:495-500.
- Aklilu E, Oscarson M, Hildestrand M, Leidvik B, Otter C, Ingelman-Sundberg M. Functional analysis of six different polymorphic CYP1B1 enzyme variants found in an Ethiopian population. *Mol Pharmacol* 2002;61:586-94.
- Thier R, Bruning T, Roos PH, et al. Markers of genetic susceptibility in human environmental hygiene and toxicology: the role of selected CYP, NAT and GST genes. *Int J Hyg Environ Health* 2003;206:149-71.
- Wenzlaff AS, Cote ML, Bock CH, et al. CYP1A1 and CYP1B1 polymorphisms and risk of lung cancer among never smokers: a population-based study. *Carcinogenesis* 2005;26:2207-12.
- Roos PH, Bolt HM. Cytochrome P450 interactions in human cancers: new aspects considering CYP1B1. *Expert Opin Drug Metab Toxicol* 2005;1:187-202.
- Xu J, Lowey J, Wiklund F, et al. The interaction of four genes in the inflammation pathway significantly predicts prostate cancer risk. *Cancer Epidemiol Biomarkers Prev* 2005;14:2563-8.
- Goodman JE, Mechanic LE, Luke BT, Ambis S, Chanock S, Harris CC. Exploring SNPSNP interactions and colon cancer risk using polymorphism interaction analysis. *Int J Cancer* 2006;118:1790-7.
- Goode EL, Potter JD, Bamlet WR, Rider DN, Bigler J. Inherited variation in carcinogen-metabolizing enzymes and risk of colorectal polyps. *Carcinogenesis* 2007;28:328-41.
- Mononen N, Seppala EH, Duggal P, et al. Profiling genetic variation along the androgen biosynthesis and metabolism pathways implicates several single nucleotide polymorphisms and their combinations as prostate cancer risk factors. *Cancer Res* 2006;66:743-7.

II.2.1. Supplementary Data

II.2.1.1. Supplementary Table S1 : TaqMan-based genotyping analysis of the selected SNPs

Gene	Mutation tested	Primer/probe Name	Primer and probe sequences 5'→3' (strand)
CYP1A2	c.-163C>A	F	GAGAGCCAGCGTTCATGTTG (-)
		R	CAGAAAGACTAAGCTCCATCTACCAT (-)
		VIC	CGTCCTGTGCCAC (-)
		FAM	CGTCCTGGGCCAC (-)
CYP1A2	c.1548T>C (p.Asn516Asn)	F	GCCCGCTGTGAACATGTC (-)
		R	TCCCTGGCCTCAGAATGGT (-)
		VIC	TGTCTTCTTCAGTTGATGG (-)
		FAM	TGTCTTCTCAATTGATGG (-)
CYP2E1	c.-1293G>C	F	GCCAACGCCCTTCTTG (+)
		R	TCATTGGTTGTGCTGCACCTAA (+)
		VIC	CACTGCAGCTCTCCT (+)
		FAM	CACTGCACCTCTCCT (+)
CYP2E1	c.-1053C>T	F	CCACAAGTGATTTGGCTGGATTG (-)
		R	GTTTTTCATTCTGTCTTCTAACTGGCAAT (-)
		VIC	ATAGGTTGCAATTTTGTACTT (-)
		FAM	TAGGTTGCAATTTTATACTT (-)
CYP1B1	c.1294G>C (p.Val432Leu)	F	TTTGTCAACCAGTGGTCTGTGAAT (+)
		R	ACTCTGCTGGTCAGGTCCTT (+)
		VIC	CATGACCCACTGAAGTG (+)
		FAM	ATGACCCAGTGAAGTG (+)
CYP2C9	c.430C>T (p.Arg144Cys)	F	TCAGCAATGGAAAGAAATGGAAGGA (-)
		R	TGGTTTTTCTCAACTCCTCCACAAG (-)
		VIC	CTTGAACACGGTCCTC (-)
		FAM	CTTGAACACAGTCCTC (-)

Supplementary Table S1: TaqMan-based genotyping analysis of the selected SNPs.

II.2.1.2. Supplementary data S2 : Analysis of interaction between allelic variants of CYP genes and environmental covariates

Environmental factor	SNP : CYP1A2 g.-163A>C			SNP : CYP1A2 c.1548T>C			SNP : CYP2E1 g.-1293G>C			SNP : CYP2E1 g.-1053C>T			SNP : CYP1B1 c.1294C>G			SNP : CYP2C9 c.430C>T			
	A/A	A/C	C/C	T/T	T/C	C/C	G/G	G/C	C/C	C/C	C/T	T/T	C/C	C/G	G/G	C/C	C/T	T/T	
Tobacco	Current smokers	1.00	0.98 (0.81-1.19)	1.08 (0.76-1.54)	1.00	0.95 (0.78-1.16)	0.94 (0.70-1.27)	1.00	0.89 (0.62-1.27)	2.16 (0.18-25.42)	1.00	0.87 (0.61-1.24)	5.75 (0.67-49.30)	1.00	1.05 (0.85-1.29)	1.26 (0.96-1.66)	1.00	0.95 (0.77-1.18)	1.40 (0.71-2.73)
	Non Smokers	0.76 (0.48-1.20)	0.73 (0.46-1.16)	0.80 (0.23-2.77)	0.64 (0.39-1.06)	0.80 (0.51-1.27)	0.72 (0.29-1.81)	0.81 (0.58-1.12)	0.20 (0.05-0.92)	---	0.81 (0.58-1.12)	0.21 (0.05-0.92)	---	1.10 (0.63-1.90)	0.57 (0.36-0.92)	1.08 (0.53-2.20)	0.77 (0.53-1.11)	0.69 (0.37-1.28)	0.00
Physical Activity	0 to 3 hours/week	1.00	0.82 (0.60-1.13)	0.80 (0.46-1.41)	1.00	0.79 (0.57-1.10)	0.90 (0.56-1.46)	1.00	0.78 (0.45-1.36)	---	1.00	0.78 (0.45-1.36)	---	1.00	0.92 (0.66-1.30)	1.32 (0.83-2.10)	1.00	0.87 (0.62-1.23)	1.75 (0.42-7.28)
	>3 hours/week	0.46 (0.35-0.60)	0.47 (0.36-0.63)	0.58 (0.37-0.92)	0.45 (0.34-0.61)	0.48 (0.36-0.64)	0.45 (0.30-0.67)	0.52 (0.43-0.63)	0.42 (0.27-0.65)	1.46 (0.12-17.34)	0.52 (0.43-0.63)	0.41 (0.26-0.64)	3.86 (0.45-33.31)	0.51 (0.36-0.71)	0.51 (0.38-0.70)	0.61 (0.42-0.88)	0.50 (0.40-0.63)	0.50 (0.37-0.67)	0.81 (0.37-1.75)
Alcohol	<2 times/week	1.00	0.48 (0.29-0.79)	0.55 (0.25-1.24)	1.00	0.60 (0.35-1.00)	0.38 (0.19-0.76)	1.00	0.75 (0.29-1.92)	---	1.00	0.67 (0.26-1.67)	---	6.13 (0.72-52.62)	0	1.36 (0.66-2.79)	1.00	0.89 (0.51-1.57)	1.05 (0.28-3.86)
	≥2 times/week	0.32 (0.22-0.46)	0.34 (0.23-0.50)	0.37 (0.22-0.61)	0.31 (0.20-0.48)	0.32 (0.21-0.49)	0.34 (0.21-0.56)	0.48 (0.36-0.63)	0.39 (0.25-0.60)	0.39 (0.02-6.36)	0.48 (0.36-0.63)	0.39 (0.25-0.60)	1.30 (0.13-13.12)	---	0	0.66 (0.41-1.07)	0.47 (0.35-0.63)	0.45 (0.32-0.63)	0.63 (0.28-1.41)
Red Meat	≤4 times/week	1.00	0.97 (0.81-1.17)	1.08 (0.76-1.52)	1.00	0.96 (0.79-1.16)	0.93 (0.69-1.24)	1.00	0.81 (0.58-1.15)	2.29 (0.20-26.88)	1.00	0.80 (0.57-1.13)	6.13 (0.72-52.62)	1.00	1.00 (0.81-1.22)	1.21 (0.93-1.59)	1.00	0.98 (0.79-1.20)	1.38 (0.71-2.69)
	>5 times/week	3.16 (1.38-7.22)	3.46 (1.52-7.87)	1.41 (0.23-8.75)	2.91 (1.13-7.53)	3.80 (1.68-8.55)	1.59 (0.43-5.90)	3.02 (1.70-5.35)	3.73 (0.38-37.02)	---	3.03 (1.71-5.37)	3.74 (0.38-37.12)	---	3.62 (1.29-10.13)	3.53 (1.61-7.76)	2.00 (0.58-6.86)	3.86 (2.00-7.48)	1.60 (0.55-4.61)	---
Poultry - White Meat	≤4 times/week	1.00	0.98 (0.81-1.18)	1.06 (0.75-1.49)	1.00	0.97 (0.80-1.17)	0.95 (0.71-1.27)	1.00	0.80 (0.57-1.12)	2.18 (0.18-25.64)	1.00	0.79 (0.56-1.10)	5.79 (0.67-49.78)	1.00	0.99 (0.81-1.21)	1.17 (0.89-1.52)	1.00	0.96 (0.78-1.17)	1.44 (0.74-2.83)
	≥5 times/week	1.00 (0.31-3.24)	0.70 (0.28-1.73)	0.95 (0.13-7.01)	0.91 (0.25-3.28)	0.97 (0.38-2.47)	0.39 (0.08-2.00)	0.80 (0.41-1.57)	---	---	0.80 (0.41-1.58)	---	---	0.79 (0.26-2.45)	0.45 (0.16-1.30)	6.42 (0.72-56.95)	0.75 (0.32-1.76)	1.09 (0.35-3.37)	0.00
Cold cuts	≤4 times/week	1.00	0.83 (0.65-1.06)	0.85 (0.54-1.36)	1.00	0.82 (0.64-1.05)	0.74 (0.50-1.09)	1.00	0.80 (0.50-1.27)	0.88 (0.05-14.22)	1.00	0.78 (0.49-1.23)	4.77 (0.54-41.83)	1.00	1.14 (0.88-1.48)	1.32 (0.94-1.85)	1.00	1.07 (0.82-1.41)	1.24 (0.55-2.79)
	≥5 times/week	1.09 (0.84-1.40)	1.39 (1.06-1.83)	1.49 (0.90-2.47)	1.05 (0.79-1.38)	1.30 (0.99-1.71)	1.34 (0.89-2.03)	1.34 (1.11-1.62)	1.19 (0.72-1.96)	---	1.35 (1.11-1.63)	1.20 (0.73-1.97)	---	1.71 (1.24-2.37)	1.33 (1.01-1.76)	1.84 (1.24-2.74)	1.46 (1.18-1.80)	1.15 (0.85-1.56)	2.66 (0.79-8.95)
Fish	≤4 times/week	1.00	0.98 (0.81-1.17)	1.04 (0.74-1.46)	1.00	0.98 (0.81-1.18)	0.94 (0.70-1.25)	1.00	0.81 (0.58-1.15)	2.19 (0.19-25.77)	1.00	0.81 (0.58-1.14)	5.84 (0.68-50.16)	1.00	1.00 (0.82-1.23)	1.24 (0.95-1.62)	1.00	0.94 (0.77-1.16)	1.29 (0.66-2.51)
	≥5 times/week	0.67 (0.30-1.47)	0.95 (0.34-2.63)	0.00 (0.35-1.93)	0.82 (0.29-1.86)	0.73 (0.29-1.86)	0.00	0.69 (0.36-1.32)	1.07 (0.14-7.90)	---	0.72 (0.38-1.40)	0.70 (0.11-4.33)	---	1.13 (0.46-2.80)	0.68 (0.28-1.68)	0.00	0.62 (0.29-1.33)	0.85 (0.28-2.56)	---
Cooking	Roasting or frying with fats	1.00	1.11 (0.88-1.39)	1.33 (0.88-1.99)	1.00	1.09 (0.86-1.39)	1.19 (0.85-1.69)	1.00	0.77 (0.51-1.18)	---	1.00	0.77 (0.51-1.16)	---	1.00	0.91 (0.71-1.16)	0.97 (0.70-1.35)	1.00	0.93 (0.72-1.20)	0.87 (0.38-1.99)
	Roasting or frying without fats or grilling	1.33 (0.96-1.83)	1.06 (0.75-1.50)	1.22 (0.56-2.68)	1.45 (1.02-2.05)	1.06 (0.75-1.49)	0.92 (0.48-1.76)	1.08 (0.85-1.37)	1.26 (0.63-2.52)	---	1.08 (0.85-1.37)	1.26 (0.63-2.53)	---	0.86 (0.57-1.28)	1.11 (0.79-1.56)	1.42 (0.84-2.40)	1.16 (0.90-1.51)	0.84 (0.53-1.30)	6.07 (0.62-59.08)
	Steaming	1.53 (0.89-2.62)	0.96 (0.52-1.76)	0.36 (0.04-3.32)	1.84 (1.00-3.38)	0.91 (0.51-1.63)	0.55 (0.14-2.22)	1.09 (0.73-1.64)	0.79 (0.07-8.92)	1.55 (0.09-26.56)	1.12 (0.74-1.67)	0.50 (0.05-4.85)	1.55 (0.09-26.56)	1.08 (0.54-2.15)	0.82 (0.46-1.44)	2.41 (0.85-6.82)	1.16 (0.74-1.82)	0.76 (0.32-1.82)	2.38 (0.22-26.45)
Fruits	≤ once/day	1.00	1.01 (0.78-1.31)	1.29 (0.81-2.04)	1.00	0.94 (0.72-1.22)	0.92 (0.62-1.38)	1.00	0.96 (0.62-1.50)	---	1.00	0.92 (0.59-1.43)	---	1.00	0.96 (0.73-1.27)	1.36 (0.94-1.97)	1.00	0.89 (0.67-1.18)	1.02 (0.36-2.92)

Environmental factor	SNP : CYP1A2 g.-163A>C			SNP : CYP1A2 c.1548T>C			SNP : CYP2E1 g.-1293G>C			SNP : CYP2E1 g.-1053C>T			SNP : CYP1B1 c.1294C>G			SNP : CYP2C9 c.430C>T			
	A/A	A/C	C/C	T/T	T/C	C/C	G/G	G/C	C/C	C/C	C/T	T/T	C/C	C/G	G/G	C/C	C/T	T/T	
Vegetables	≥2 times/day	0.84 (0.65-1.07)	0.78 (0.60-1.01)	0.69 (0.41-1.14)	0.76 (0.57-0.99)	0.75 (0.57-0.98)	0.71 (0.47-1.07)	0.80 (0.66-0.96)	0.52 (0.31-0.88)	1.18 (0.07-20.19)	0.79 (0.66-0.95)	0.52 (0.31-0.88)	3.35 (0.36-31.35)	0.77 (0.56-1.05)	0.82 (0.62-1.08)	0.84 (0.58-1.23)	0.74 (0.60-0.91)	0.77 (0.57-1.03)	1.29 (0.55-3.02)
	≤ once/day	1.00	1.09 (0.85-1.40)	1.14 (0.73-1.79)	1.00	1.03 (0.80-1.33)	1.05 (0.71-1.54)	1.00	0.97 (0.63-1.51)	---	1.00	0.94 (0.61-1.45)	---	1.00	0.94 (0.71-1.23)	1.46 (1.02-2.10)	1.00	1.03 (0.78-1.37)	1.48 (0.57-3.80)
	≥2 times/day	0.99 (0.77-1.27)	0.86 (0.66-1.11)	0.96 (0.57-1.61)	0.97 (0.74-1.27)	0.87 (0.67-1.14)	0.81 (0.53-1.22)	0.92 (0.76-1.10)	0.58 (0.34-1.00)	0.00	0.91 (0.75-1.09)	0.58 (0.34-0.99)	2.06 (0.21-19.98)	0.88 (0.64-1.21)	0.94 (0.71-1.24)	0.90 (0.62-1.31)	0.92 (0.75-1.13)	0.82 (0.61-1.10)	1.16 (0.46-2.94)
Pastries	< once/week	1.00	0.98 (0.76-1.26)	1.19 (0.74-1.91)	1.00	1.05 (0.81-1.37)	1.06 (0.72-1.57)	1.00	0.86 (0.54-1.36)	0.81 (0.05-12.99)	1.00	0.83 (0.53-1.31)	2.24 (0.23-21.75)	1.00	0.98 (0.74-1.29)	1.05 (0.73-1.51)	1.00	0.93 (0.71-1.23)	1.11 (0.40-3.05)
	≥ once/week	1.11 (0.86-1.42)	1.12 (0.86-1.45)	1.06 (0.65-1.73)	1.20 (0.92-1.58)	1.11 (0.85-1.46)	1.00 (0.66-1.53)	1.10 (0.92-1.33)	0.87 (0.53-1.44)	---	1.10 (0.92-1.32)	0.87 (0.53-1.44)	---	1.01 (0.74-1.39)	1.05 (0.80-1.39)	1.48 (1.01-2.16)	1.08 (0.88-1.33)	1.06 (0.78-1.43)	1.68 (0.70-4.06)
Dairy Products	≤ once/day	1.00	0.99 (0.71-1.36)	0.83 (0.45-1.51)	1.00	1.05 (0.76-1.46)	0.82 (0.49-1.38)	1.00	0.87 (0.50-1.51)	1.03 (0.06-17.61)	1.00	0.84 (0.48-1.45)	1.03 (0.06-17.49)	1.00	0.92 (0.65-1.30)	1.56 (0.96-2.53)	1.00	0.56 (0.39-0.79)	4.04 (0.86-18.85)
	≥2 times/day	0.68 (0.52-0.89)	0.68 (0.52-0.89)	0.81 (0.52-1.26)	0.72 (0.54-0.96)	0.68 (0.51-0.91)	0.73 (0.50-1.07)	0.70 (0.58-0.86)	0.55 (0.35-0.86)	---	0.70 (0.57-0.85)	0.54 (0.35-0.84)	---	0.70 (0.50-0.98)	0.71 (0.52-0.97)	0.78 (0.54-1.14)	0.59 (0.47-0.73)	0.73 (0.55-0.97)	0.48 (0.20-1.13)
Bread	White bread	1.00	1.02 (0.83-1.25)	1.05 (0.72-1.53)	1.00	1.01 (0.82-1.24)	1.03 (0.75-1.42)	1.00	0.76 (0.52-1.09)	1.99 (0.17-23.30)	1.00	0.74 (0.51-1.07)	4.69 (0.53-41.56)	1.00	0.91 (0.73-1.13)	1.07 (0.80-1.43)	1.00	0.98 (0.78-1.23)	1.43 (0.71-2.91)
	Wholemeal bread	0.65 (0.47-0.89)	0.52 (0.36-0.76)	0.83 (0.38-1.79)	0.68 (0.47-0.96)	0.58 (0.41-0.82)	0.46 (0.23-0.91)	0.58 (0.46-0.74)	0.75 (0.33-1.72)	---	0.58 (0.45-0.73)	0.75 (0.33-1.72)	---	0.39 (0.25-0.61)	0.63 (0.45-0.88)	0.85 (0.51-1.42)	0.61 (0.47-0.81)	0.58 (0.38-0.88)	0.35 (0.04-3.41)
Body Mass index	BMI ≤ 24.9	1.00	1.14 (0.87-1.49)	1.01 (0.61-1.70)	1.00	1.09 (0.83-1.44)	0.89 (0.58-1.35)	1.00	0.86 (0.53-1.38)	---	1.00	0.82 (0.52-1.32)	---	1.00	1.09 (0.81-1.45)	1.57 (1.07-2.32)	1.00	0.86 (0.64-1.17)	1.16 (0.42-3.20)
	BMI 25 – 29.9	1.23 (0.94-1.61)	1.14 (0.86-1.51)	1.35 (0.78-2.35)	1.21 (0.90-1.63)	1.10 (0.82-1.47)	1.26 (0.79-2.03)	1.14 (0.93-1.40)	0.85 (0.48-1.50)	0.85 (0.05-13.75)	1.13 (0.92-1.38)	0.85 (0.48-1.50)	3.83 (0.42-35.10)	1.45 (1.03-2.03)	1.15 (0.85-1.56)	1.34 (0.89-2.01)	1.09 (0.87-1.36)	1.10 (0.80-1.51)	1.16 (0.42-3.19)
	BMI > 30	1.29 (0.89-1.88)	1.07 (0.72-1.58)	1.29 (0.60-2.75)	1.20 (0.79-1.81)	1.12 (0.75-1.66)	1.18 (0.63-2.21)	1.12 (0.85-1.47)	1.02 (0.40-2.59)	---	1.12 (0.85-1.47)	1.01 (0.40-2.58)	---	1.08 (0.68-1.74)	1.37 (0.93-2.02)	1.33 (0.72-2.47)	1.07 (0.79-1.45)	1.05 (0.65-1.71)	4.57 (0.50-42.02)
Diabetes	No	1.00	0.99 (0.82-1.19)	1.04 (0.73-1.49)	1.00	0.99 (0.81-1.20)	0.93 (0.69-1.26)	1.00	0.76 (0.54-1.09)	1.35 (0.08-22.85)	1.00	0.75 (0.53-1.07)	5.42 (0.62-47.81)	1.00	0.97 (0.79-1.19)	1.28 (0.98-1.69)	1.00	1.00 (0.81-1.23)	1.31 (0.67-2.57)
	Yes	2.27 (1.39-3.71)	1.96 (1.14-3.38)	2.34 (0.87-6.28)	2.20 (1.30-3.73)	2.07 (1.20-3.56)	2.10 (0.93-4.72)	2.00 (1.40-2.85)	4.85 (1.02-23.05)	---	2.02 (1.41-2.88)	4.87 (1.02-23.12)	---	2.08 (1.11-3.89)	2.68 (1.64-4.40)	1.39 (0.64-3.03)	2.50 (1.66-3.77)	1.39 (0.74-2.61)	---
Dyslipidemia	No	1.00	1.06 (0.87-1.31)	1.07 (0.73-1.58)	1.00	1.04 (0.84-1.28)	1.03 (0.74-1.42)	1.00	0.82 (0.56-1.20)	1.21 (0.07-20.88)	1.00	0.80 (0.55-1.17)	3.44 (0.34-34.37)	1.00	1.05 (0.84-1.31)	1.38 (1.02-1.86)	1.00	0.92 (0.73-1.17)	1.32 (0.62-2.80)
	Yes	0.84 (0.62-1.14)	0.62 (0.45-0.85)	0.91 (0.47-1.78)	0.86 (0.62-1.19)	0.65 (0.47-0.90)	0.66 (0.38-1.12)	0.71 (0.57-0.89)	0.62 (0.30-1.25)	---	0.71 (0.57-0.88)	0.62 (0.30-1.25)	---	0.89 (0.62-1.29)	0.72 (0.52-0.99)	0.73 (0.44-1.19)	0.68 (0.53-0.87)	0.75 (0.51-1.10)	1.44 (0.33-6.27)

Supplementary data S2: Analysis of interaction between allelic variants of CYP genes and environmental covariates. In order to assess possible gene-environment interactions modifying CRC risk, interactions were tested between the six single allelic variants CYP1A2 c.-163A>C, CYP1A2 c.1548T>C, CYP2E1 g.-1293G>C, CYP2E1 g.-1053C>T, CYP1B1 c.1294C>G, and CYP2C9 c.430C>T, and every environmental factor of the questionnaire on life habits (Figure 1), i.e., tobacco, physical activity, alcohol, red meat, poultry and/or white meat, cold cuts, fish, cooking method, fruits, vegetables, pastries, dairy products, and bread. Interactions were also tested between the six allelic variants and diabetes and dyslipidemia status. ORs were estimated, together with 95% CI, for every possible pair of SNP and environmental factor. For each pair of genetic and environmental factor, the interaction was tested in three different ways here we have just represent the influence of the genetic and environmental factor were tested in the same time.

II.2.1.3. Supplementary Table S3 : Analysis of association to CRC risk of multiple CYP SNPs combinations interacting with the body mass

A. Multiple-SNPs combination and Body Mass Index cross-classification interaction table (n=2088, adjusted by age and sex)

Combination	Number of individuals exhibiting the combination	BMI ≤ 24.9	BMI 25 – 29.9	BMI ≥ 30
		OR (95% CI)	OR (95% CI)	OR (95% CI)
ATGCCC	662 (31.7%)	1.00	1.38 (0.91 - 2.09)	1.35 (0.86 - 2.14)
ATGCGC	452 (21.7%)	1.35 (1.01 - 1.80)	1.22 (0.85 - 1.77)	1.23 (0.79 - 1.90)
CCGCCC	276 (13.2%)	0.92 (0.64 - 1.32)	1.33 (0.87 - 2.04)	0.97 (0.56 - 1.69)
CCGCGC	210 (10.1%)	1.00 (0.69 - 1.47)	1.03 (0.60 - 1.76)	1.08 (0.56 - 2.08)
ATGCGT	95 (4.5%)	0.52 (0.29 - 0.94)	1.17 (0.57 - 2.40)	6.54 (2.39 - 17.93)

Interaction p-value: 0.51

B. Multiple-SNPs combinations within Body Mass Index (n=2088, adjusted by age and sex)

Combination	Number of individuals exhibiting the combination	BMI ≤ 24.9	BMI 25 – 29.9	BMI ≥ 30
		OR (95% CI)	OR (95% CI)	OR (95% CI)
ATGCCC	0.3171	1.00	1.00	1.00
ATGCGC	0.2167	1.35 (1.01 - 1.80)	0.89 (0.64 - 1.24)	0.91 (0.53 - 1.54)
CCGCCC	0.1323	0.92 (0.64 - 1.32)	0.97 (0.64 - 1.47)	0.72 (0.39 - 1.33)
CCGCGC	0.1007	1.00 (0.69 - 1.47)	0.75 (0.48 - 1.17)	0.80 (0.43 - 1.50)
ATGCGT	0.0454	0.52 (0.29 - 0.94)	0.85 (0.44 - 1.63)	4.83 (2.07 - 11.29)

C. Body Mass Index within multiple-SNPs combinations (n=2088, adjusted by age and sex)

Combination	Number of individuals exhibiting the combination	BMI ≤ 24.9	BMI 25 – 29.9	BMI ≥ 30
		OR (95% CI)	OR (95% CI)	OR (95% CI)
ATGCCC	0.3171	1.00	1.38 (0.91 - 2.09)	1.35 (0.86 - 2.14)
ATGCGC	0.2167	1.00	0.91 (0.65 - 1.27)	0.91 (0.59 - 1.41)
CCGCCC	0.1323	1.00	1.45 (0.93 - 2.26)	1.06 (0.59 - 1.90)
CCGCGC	0.1007	1.00	1.02 (0.55 - 1.89)	1.08 (0.52 - 2.26)
ATGCGT	0.0454	1.00	2.24 (0.94 - 5.31)	12.55 (6.28 - 25.09)

Supplementary Table S3: Analysis of association to CRC risk of multiple CYP SNPs combinations interacting with the body mass (BMI= Body Mass Index). Allelic variants composing the combinations of six SNPs are ordered as follows: 1st CYP1A2 c.-163A>C, 2nd CYP1A2 c.1548T>C, 3rd CYP2E1 g.-1293G>C, 4th CYP2E1 g.-1053C>T, 5th CYP1B1 c.1294C>G, and 6th CYP2C9 c.430C>T. Only the five more frequent combinations (>4%) are represented. A., Cross-classification using a common reference category for both interacting variables. The group of individuals with a BMI ≤ 24.9 and exhibiting the most frequent combination of SNPs is selected as the reference category. ORs are estimated, together with 95% CI, for all other combinations distributed between the groups of individuals with BMI ≤ 24.9 (underweighted and normal weighted), 25 ≤ BMI ≤ 29.9 (overweighted), and BMI ≥ 30 (obese), according to a log-additive model. Significant associations to modification of CRC risk are typed in bold characters. B., The influence of SNP combination on modification of CRC risk is estimated by ORs calculation within each group of normal weighted, overweighted, and obese individuals. C., Body mass nested within SNPs combinations: the influence of the body mass is estimated by ORs calculation within each combination of SNPs.

II.3. Etude de 52 SNP de faible prédisposition au cancer colorectal sporadique (Article 2)

Ce projet s'inscrit dans la continuité de l'article précédent, mais, cette fois, l'analyse porte sur un nombre bien supérieur de SNP. En effet, cette étude a été réalisée sur un ensemble de 52 SNP, répartis dans 35 gènes déjà montrés dans la littérature comme étant de probables facteurs de prédisposition : *ALOX5*, *ALOX12*, *IL6*, *IL8*, *PLA2G2A*, *PDL2*, *PTGS1*, *PTGS2*, *PPARG*, *CYP1A2*, *CYP2E1*, *CYP1B1*, *CYP2C9*, *EPHX1*, *GSTA1*, *GSTM1*, *GSTM3*, *GSTP1*, *GSTT1*, *NQO1*, *SULT1A2*, *UGT1A1*, *UGT1A6*, *GH1*, *IGF1*, *IGFBP3*, *IRS1*, *VDR*, *CBS*, *MTHFD1*, *MTHFR*, *MTR*, *MTRR*, *TYMS*, *OGG1*. Ces gènes appartiennent tous aux voies citées précédemment comme intervenant dans la carcinogenèse colorectale (Chapitre II : II.1.1. Les polymorphismes choisis). Parmi celles-ci, on retient en particulier les voies de l'inflammation, de détoxification des xénobiotiques, de l'insuline, des monocarbonés et de réparation de l'ADN, dont le rôle carcinogénétique a été brièvement résumé au début de ce chapitre (Chapitre II : II.1.1. Les polymorphismes choisis).

Sur ces 52 polymorphismes, 46 sont des SNP qui ont été génotypés par *5'-nuclease assay* et six autres sont des polymorphismes complexes étudiés grâce à des PCR réalisées avec des amorces marquées par un fluorochrome. Sur ces 52 SNP, les 6 polymorphismes des cytochromes P450 étudiés dans l'article 1 ont également été analysés pour mettre en avant une possible interaction de ces SNP avec d'autres variants de la voie de détoxification ou avec des variants d'autres voies métaboliques.

L'association entre les génotypes de chaque SNP et le CCR sporadique a été analysée en suivant une régression logistique inconditionnelle ajustée sur le sexe et l'âge des participants (âge du diagnostic pour les patients et âge du prélèvement sanguin pour les témoins) et une régression logistique conditionnelle appliquée aux paires de cas et témoins appariés par sexe et par âge. Afin d'éliminer les faux-positifs, le calcul de la q-value a été réalisé par la méthode des FDR (*false discovery rate*) proposée par Storey *et al.* (2003). Chaque SNP ayant été trouvé associé au CCR dans la première partie de cette étude a été analysé avec les autres SNP également associés afin de faire ressortir éventuellement une combinaison de SNP en forte association avec le cancer.

L'analyse indépendante des SNP montre une association significative pour six SNP appartenant aux gènes, *PTGS1*, *IL8*, *MTHFR*, *PLA2G2A*, *PPARG* et *GSTM1*. Pour les trois premiers gènes (*PTGS1*, *IL8* et *MTHFR*), le SNP correspondant a été trouvé prédisposant au CCR sporadique, alors que pour les trois autres gènes (*PLA2G2A*, *PPARG* et *GSTM1*) le polymorphisme correspondant a été trouvé protecteur. Pour *GSTM1*, qui est un polymorphisme complexe, ce résultat est surprenant, car d'un point de vue biologique il est difficilement explicable que l'absence d'un allèle de cette

enzyme protège du CCR alors que l'absence des deux allèles n'aurait aucun effet sur la pathologie. C'est pourquoi ce polymorphisme a été exclu dans la suite des analyses et n'a pas été pris en compte dans les tests d'association de combinaisons alléliques et génotypiques.

L'analyse des combinaisons génotypiques a montré une association positive pour les génotypes trouvés prédisposants au CCR au travers des analyses indépendantes de chaque SNP. Par comparaison à toutes les autres combinaisons génotypiques de notre population, cette combinaison comporte les allèles minoritaires des gènes *PTGS1*, *IL8*, *MTHFR* (à l'état homozygote ou hétérozygote) et les allèles majoritaires des gènes *PLA2G2A* et *PPARG* (à l'état homozygote). Il est bien entendu que ce résultat est à prendre avec précaution, car il ne porte que sur une petite fraction d'individus. Il faudrait encore augmenter la taille des cohortes pour s'assurer d'une réelle association. Une analyse de cette combinaison en relation avec les facteurs environnementaux à notre disposition a été réalisée, mais aucun d'entre eux n'a montré une association significative.

ARTICLE 2

Low-penetrance alleles predisposing to sporadic colorectal cancers: a French case-controlled genetic association study.

KÜRY S^{*}, BUECHER B^{*}, ROBIOU-DU-PONT S^{*}, SCOUL C, COLMAN H, LE NEEL T, LE HOUÉROU C, FAROUX R, OLLIVERY J, LAFRAISE B, CHUPIN LD
SEBILLE V, BEZIEAU S

BMC Cancer **8** : 326 (2008)

* Contribution équivalente au travail

Research article

Open Access

Low-penetrance alleles predisposing to sporadic colorectal cancers: a French case-controlled genetic association study

Sébastien Küry*^{†1}, Bruno Buecher^{†2}, Sébastien Robiou-du-Pont^{†3}, Catherine Scoul³, Hélène Colman³, Tanguy Le Neel⁴, Claire Le Houérou⁵, Roger Faroux⁶, Jean Ollivry⁷, Bernard Lafraise⁸, Louis-Dominique Chupin⁹, Véronique Sébille¹⁰ and Stéphane Bézieau^{1,3}

Address: ¹Service de Génétique Médicale, Pôle de Biologie, CHU de Nantes, 9 quai Moncousu, Nantes, 44093 cedex 1, France, ²Service de Génétique Oncologique, unité de Génétique Constitutionnelle, Institut Curie, 26 rue d'Ulm, Paris, 75248 cedex 05, France, ³Laboratoire Biométadys, Université de Nantes, Faculté de Médecine, 1 rue Gaston Veil, Nantes, 44035, France, ⁴Biofortis, Nantes, France, ⁵Service d'Hématologie, Pôle de Cancérologie, CHU de Nantes, 9 quai Moncousu, Nantes, 44093 cedex 1, France, ⁶Service d'Hépatogastroentérologie, CHD de La Roche-sur-Yon, Les Oudairies, La Roche-sur-Yon, 85925 cedex 9, France, ⁷Association des Gastroentérologues de Vendée, Challans, France, ⁸Centre d'Examen de Santé de Saint-Nazaire, France, ⁹Centre d'Examen de Santé de La Roche-Sur-Yon, France and ¹⁰Laboratoire de Biomathématiques – Biostatistique, Université de Nantes, Faculté de Pharmacie, 1 rue Gaston Veil, Nantes, 44035, France

Email: Sébastien Küry* - sebastien.kury@chu-nantes.fr; Bruno Buecher - bruno.buecher@curie.net; Sébastien Robiou-du-Pont - sebastien.robious-du-pont@etu.univ-nantes.fr; Catherine Scoul - Catherine.Scoul@univ-nantes.fr; Hélène Colman - Helene.Colman@univ-nantes.fr; Tanguy Le Neel - tleneel@sante.univ-nantes.fr; Claire Le Houérou - claire.lehouerou@chu-nantes.fr; Roger Faroux - roger.faroux@chd-vendee.fr; Jean Ollivry - dr.ollivry@wanadoo.fr; Bernard Lafraise - b.lafraise@wanadoo.fr; Louis-Dominique Chupin - dominique.chupin@cpam-laroche.cnams.fr; Véronique Sébille - veronique.sebille@univ-nantes.fr; Stéphane Bézieau - stephane.bezieau@chu-nantes.fr

* Corresponding author †Equal contributors

Published: 7 November 2008

Received: 19 April 2008

BMC Cancer 2008, 8:326 doi:10.1186/1471-2407-8-326

Accepted: 7 November 2008

This article is available from: <http://www.biomedcentral.com/1471-2407/8/326>

© 2008 Küry et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: Sporadic colorectal cancers (CRC) are multifactorial diseases resulting from the combined effects of numerous genetic, environmental and behavioral risk factors. Genetic association studies have suggested low-penetrance alleles of extremely varied genes to be involved in susceptibility to CRC in Caucasian populations.

Methods: Through a large genetic association study based on 1023 patients with sporadic CRC and 1121 controls, we tested a panel of these low-penetrance alleles to find out whether they could determine "genotypic profiles" at risk for CRC among individuals of the French population. We examined 52 polymorphisms of 35 genes – drawn from inflammation, xenobiotic detoxification, one-carbon, insulin signaling, and DNA repair pathways – for their possible contribution to colorectal carcinogenesis. The risk of cancer associated with these polymorphisms was assessed by calculation of odds ratios (OR) using multivariate analyses and logistic regression.

Results: Whereas all these polymorphisms had previously been found to be associated with CRC risk, especially in Caucasian populations, we were able to replicate the association for only five of them. Three SNPs were shown to increase CRC risk: *PTGS1* c.639C>A (p.Gly213Gly), *IL8* c.-352T>A, and *MTHFR* c.1286A>C (p.Ala429Glu). On the contrary, two other SNPs, *PLA2G2A* c.435+230C>T and *PPARG* c.1431C>T (p.His477His), were associated with a decrease in CRC risk.

Further analyses highlighted genotypic combinations having a greater predisposing effect on CRC (OR 1.97, 95%CI 1.31–2.97, $p = 0.0009$) than the allelic variants that were examined separately.

Conclusion: The identification of CRC-predisposing combinations, composed of alleles *PTGS1* c.639A, *PLA2G2A* c.435+230C, *PPARG* c.1431C, *IL8* c.-352A, and *MTHFR* c.1286C, highlights the importance of inflammatory processes in susceptibility to sporadic CRC, as well as a possible crosstalk between inflammation and one-carbon pathways.

Background

As in most Western countries, colorectal cancer (CRC) is a major public health issue in France, where it is the second most common cause of death from cancer among adults [1]. Unfortunately, CRC diagnosis is often made at too late a stage and this induces a dismal prognosis, emphasizing the need for prevention and early diagnostic tools [2]. The development of such tools is, nonetheless, highly dependent on the form of cancer being screened. Thus, reliable genetic tests are already available to detect phenotypically well-characterized familial forms of CRC associated with high-penetrance alleles of genes, including APC in familial adenomatous polyposis (FAP), DNA mismatch repair genes in Lynch syndrome or hereditary non-polyposis colorectal cancer (HNPCC), and MUTYH in MUTYH-associated polyposis (MAP) [3]. However, familial diseases together probably make up only a small percentage of all CRCs [2], and the development of comparable screening tests appears more arduous and remote for sporadic cases, which account for the large majority of CRCs.

In contrast with their major role in familial CRCs, heritable factors are not the only components of the etiology of sporadic CRCs. Susceptibility to sporadic CRCs is multifactorial and derives from multiple interactive combinations of numerous low-penetrance alleles and relevant environmental or behavioral risk factors [2,4]. Independently, each low-penetrance allele contributes modestly to the increase in CRC risk, but its interactions with other susceptibility alleles and environmental factors can lead to a substantial increase in CRC risk, especially when exposed to certain dietary and lifestyle habits [4-6]. The number of interactions is all the higher because the susceptibility genes can be involved in many different biological pathways, which explains the extremely variable phenotype encountered in sporadic CRCs.

Despite much criticism for their non-reproducibility and weak statistical power [6], genetic association studies have been widely used to decipher the mechanisms of cancer susceptibility. Besides, their quality has improved greatly over the past few years [7]. They recently started to produce very valuable results, as illustrated by the identification of several susceptibility loci for colorectal cancers [8-12]. Great expectations can now be held about the results

and positive consequences on medical oncology provided by such studies. Beyond the search for susceptibility genes, a global effort is currently being made in the field of sporadic cancers, in order to determine which combinations of genetic variants, i.e., which genetic backgrounds, present at non rare frequencies in the general population, are likely to confer an increase in cancer risk, either alone, or by interacting with usual environmental factors [4,5,13,14].

In order to be part of this effort, we have conducted a case-controlled genetic association study based on a large French population sample. This one already turned out to be very useful by contributing to the finding of the new CRC susceptibility locus at chromosome 8q24.21 made by Zanke et al [9]. Yet, in the present study, our purpose was somehow more modest as we did not attempt to identify new susceptibility loci or variants by a pangenomic approach. Through an exploratory study, we tried to find out whether combinations of variants already known for their possible involvement in carcinogenesis – especially in various Caucasian populations – could determine 'genotypic profiles' at risk for CRC among individuals of our French study population. Thus, we focused on 52 allelic variants of 35 candidate genes, selected through a review of the literature on CRC susceptibility, and drawn from five biological pathways relating to inflammation, xenobiotics detoxification, one-carbon, insulin signaling, and DNA repair. Here, we report the results of our investigation on the risk for CRC associated with all 52 allelic variants, analyzed singly or in combinations.

Methods

Experimental design and study population

From December 2002 to March 2006, two groups of 1023 patients (632 males and 391 females; mean \pm SD age at diagnosis 65.7 \pm 10.1 years) and 1121 controls (609 males and 512 females; mean \pm SD age at inclusion 61.9 \pm 10. years), all of Caucasian origin, were recruited within the Pays de la Loire region of France.

Details of the study population characteristics have been described elsewhere [15]. In short, all cases were patients recruited in regional hospitals and clinics, with a personal history of colorectal cancer diagnosed at an age of \geq 40 years old. Any patient suspected of having a familial form

of colorectal cancer (either Familial Adenomatous Polyposis, Hereditary Non Polyposis Colorectal Cancer or MUTYH-Associated Polyposis) was excluded from the study. Age- and sex-matched controls were recruited at two regional Health Examination Centres; all were ≥ 40 years old and had no family history of colorectal cancer or polyps. Two venous blood samples were collected from each participant in an anonymous manner, after they had legally provided a written informed consent. The study was approved by both the local ethics committee CCPPRB (Consultative Committee for the Protection of Person in Biomedical Research) and by the national French ethics committee CNIL (National Commission for Data protection and the Liberties). Each participant also answered the same one-page standardized questionnaire pertaining to life and food habit information. Endoscopy and histology reports were obtained for each patient.

Selection of the low-penetrance genes and allelic variants

A review of the literature focused on susceptibility to sporadic colorectal cancer highlighted five biological pathways frequently cited for their role in colorectal carcinogenesis: inflammation, one-carbon, signaling insulin, xenobiotics detoxification, and DNA repair. In these pathways, we selected 35 genes reported as known or possible factors predisposing to CRC (*ALOX5*, *ALOX12*, *IL6*, *IL8*, *PLA2G2A*, *PDL2*, *PTGS1*, *PTGS2*, *PPARG*, *CYP1A2*, *CYP2E1*, *CYP1B1*, *CYP2C9*, *EPHX1*, *GSTA1*, *GSTM1*, *GSTM3*, *GSTP1*, *GSTT1*, *NQO1*, *SULT1A2*, *UGT1A1*, *UGT1A6*, *GH1*, *IGF1*, *IGFBP3*, *IRS1*, *VDR*, *CBS*, *MTHFD1*, *MTHFR*, *MTR*, *MTRR*, *TYMS*, *OGG1*), and chose to investigate 52 polymorphisms -46 single nucleotide polymorphisms (SNPs) and 6 complex polymorphisms - in these 35 genes (Table 1). Among these 52 polymorphisms, 38 polymorphisms were exclusively selected through our review of the literature, because of their possible involvement in the predisposition to CRC [16-30]; most of these polymorphisms reportedly modified the *in vitro* activity of the corresponding gene. It should be noted that 6 of these polymorphisms - found in *CYP1A2*, *CYP2E1*, *CYP1B1* and *CYP2C9*-, which were tested previously in our study population [15], are still cited in the present study, given that they were included in the combined analyses described below. The remaining 14 SNPs, located in the candidate genes *PLA2G2A*, *PTGS1*, *PTGS2*, and *PPARG*, were first found by a dHPLC screening in two samples of 50 patients and 50 controls randomly chosen within the whole study population. In fact, we chose to select these 14 SNPs, because of their uneven distribution between the two samples. With the exception of the 4 SNPs in *PLA2G2A*, the other ten SNPs were also found by our bibliographical survey on susceptibility to CRC [5,17,31-34].

As regards the nomenclature that we used for the SNPs, we referred to the databases dbNSP, SNPPER and SNP500 Cancer for the ID numbers, and we followed the Human Variation Society (HGVS) mutation nomenclature for the description of nucleotide and proteic sequence variants [35]. Thereby, we corrected wrong notations abundantly used in literature for some of the selected SNPs (e.g., *PPARG* C161T instead of c.1431C>T, or *MTHFR* A1298C instead of c.1286A>C).

Genotype analysis

Every study participant was genotyped for the 52 polymorphisms selected. Genotypes were determined using high-throughput TaqMan allelic discrimination tests for the 46 SNPs, and fluorescent multiplex PCRs for the 6 complex polymorphisms [for details, see Additional files 1 and 2].

Statistical analysis

Single-SNP association study

All polymorphisms were tested for Hardy-Weinberg equilibrium by a χ^2 test in both patients and controls. An appraisal was made on the association between allelic variants and CRC risk, by calculation the odds ratio (OR) of cancer cases together with 95 percent confidence intervals (CI). Association tests consisted of comparing genotype heterogeneity between the groups of patients and controls. For each allelic variant, homozygotes for the major allele were used as the reference group and compared to the other two groups of heterozygotes and homozygotes for the minor allele considered either separately (codominant model), or gathered in a same group (dominant model). We used logistic regression models for assessing single SNP effects adjusted for gender and age covariates. To estimate the risk for CRC, we first performed multivariate analyses based on unconditional logistic regression, adjusting for sex and age at date of reference (date of diagnosis for patients, and date of blood sample collection for controls). To detect a potential overestimation of the odds ratio by unconditional logistic regression, we then employed conditional logistic regression on 811 age- and sex-matched pairs of individuals (75.6% of the whole study population, including 2*324 women, and 2*487 men). In each case, the significance of ORs was assessed by calculation of p-values derived from likelihood-ratio tests.

Multiple testing of 52 polymorphisms implies inevitably a certain rate of false positive results. A correction is therefore required to lower as much as possible the rate of false positives, even though the need to adjust multiple comparisons has been questioned by Rothman [36]. Yet, classical methods of correction such as Bonferroni's adjustment are too drastic for genetic association studies and lead to the rejection of many results, including true

Table 1: Statistical analysis of association between CRC risk and polymorphisms selected for the study (n = 2144; OR calculations were adjusted according to age and sex).

Gene Name	Nucleotidic change **	Proteic change **	SNP ID **	Ref ***	Controls			Cases			+/- vs. +/+		-/- vs. +/+		+/- and -/- vs. +/+			
					+/+	+/-	-/-	+/+	+/-	-/-	OR (95%CI)	p	OR (95% CI)	p	OR (95% CI)	p		
Inflammation	ALOX5	c.760G>A	p.Glu254Lys	rs2228065	16	1121	0	0	1023	0	0	Rare		Rare		Rare		
	ALOX12	c.782G>A	p.Gln261Arg	rs1126667	16	361	549	211	357	498	168	0.93 (0.76-1.12)	0.37	0.82 (0.63-1.06)	0.09	0.90 (0.75-1.08)	0.24	
		IL6	c.-239G>C		rs1800795	17	435	504	182	363	489	171	1.12 (0.92-1.35)	0.11	1.13 (0.87-1.46)	0.35	1.12 (0.94-1.34)	0.22
		IL8	c.-352T>A		rs4073	17	375	516	230	307	511	205	1.25 (1.03-1.53)	0.02	1.12 (0.88-1.44)	0.49	1.21 (1.01-1.46)	0.04
		PLA2G2A	c.-859C>G		rs11573156	d	738	337	46	627	349	47	1.20 (1.00-1.45)	0.05	1.16 (0.75-1.78)	0.39	1.20 (1.00-1.43)	0.05
		PLA2G2A	c.132C>T	p.Tyr44Tyr	rs4744	d	742	333	46	635	342	46	1.17 (0.97-1.42)	0.05	1.13 (0.73-1.73)	0.47	1.17 (0.98-1.40)	0.09
		PLA2G2A	c.185+88G>A		rs2236772	d	1062	59	0	952	69	2	1.30 (0.90-1.88)	0.14	Rare		1.35 (0.94-1.94)	0.11
		PLA2G2A	c.435+230C>T		rs11677	d	829	276	16	789	226	8	0.82 (0.67-1.01)	0.07	0.50 (0.21-1.19)	0.13	0.80 (0.66-0.98)	0.03
		PLD2	c.1731C>T	p.Thr577Ile	rs1052748	18	316	559	246	280	520	223	1.05 (0.86-1.29)	0.63	1.00 (0.78-1.27)	0.85	1.04 (0.85-1.26)	0.82
		PTGS1	c.22C>T	p.Tip8Arg	rs1236913	d,32,34	976	139	6	879	143	1	1.19 (0.92-1.54)	0.30	0.18 (0.02-1.54)	0.08	1.15 (0.89-1.48)	0.29
		PTGS1	c.50C>T	p.Pro17Leu	rs3842787	d,32,34	970	147	4	874	142	7	1.06 (0.82-1.37)	0.58	2.17 (0.62-7.61)	0.28	1.09 (0.85-1.40)	0.51
		PTGS1	c.639C>A	p.Gly213Gly	rs5788	d,32	858	243	20	747	254	22	1.24 (1.01-1.53)	0.03	1.26 (0.67-2.36)	0.45	1.24 (1.02-1.52)	0.03
		PTGS1	c.123G>A	p.Gln41Gln	rs3842788	d,32	1074	47	0	978	45	0	1.10 (0.72-1.68)	0.81	Rare		1.10 (0.72-1.68)	0.81
		PTGS2	c.-646C>T		rs20420	d,16	1121	0	0	1023	0	0	Rare		Rare		Rare	
		PTGS2	c.306G>C	p.Val102Val	rs5277	d,31	783	312	26	707	285	31	1.00 (0.82-1.21)	0.90	1.25 (0.73-2.16)	0.30	1.02 (0.84-1.23)	0.85
		PTGS2	c.1815+427T>C		rs5275	d,31	494	499	128	433	476	114	1.08 (0.90-1.30)	0.36	1.01 (0.76-1.35)	0.91	1.07 (0.90-1.27)	0.47
		PTGS2	c.1815+1912A>G		rs4648298	d,31	1062	59	0	959	64	0	1.31 (0.90-1.89)	0.32	Rare		1.31 (0.90-1.89)	0.32
		PPARG	c.36C>G	p.Pro12Ala	rs1801282	17	896	212	13	822	194	7	1.02 (0.81-1.27)	0.98	0.47 (0.19-1.20)	0.25	0.98 (0.79-1.22)	0.85
		PPARG	c.1431C>T	p.His477His	rs3856806	33	875	217	29	803	211	9	1.07 (0.86-1.33)	0.59	0.30 (0.14-0.65)	0.003	0.97 (0.79-1.20)	0.8
	Xenobiotics detoxification	CYP1A2	c.-163A>C		rs762551	19-21	554	482	85	520	423	80	0.97 (0.81-1.16)	0.46	1.08 (0.77-1.51)	0.99	0.99 (0.83-1.17)	0.87
CYP1A2		c.1548T>C	p.Asn516Asn	rs2470890	20,21	455	525	141	433	467	123	0.97 (0.81-1.17)	0.47	0.97 (0.73-1.28)	0.53	0.97 (0.81-1.16)	0.94	
CYP2E1		c.-1293G>C		rs3813867	20	1032	88	1	954	67	2	0.81 (0.58-1.13)	0.25	2.15 (0.18-25.3)	0.52	0.82 (0.59-1.15)	0.37	
CYP2E1		c.-1053C>T		rs2031920	20	1030	90	1	950	67	6	0.79 (0.57-1.11)	0.20	5.74 (0.67-49.2)	0.09	0.85 (0.61-1.18)	0.33	
CYP1B1		c.1294C>G	p.Leu432Val	rs1056836	19-21	370	577	174	322	510	191	0.98 (0.81-1.20)	0.87	1.22 (0.94-1.58)	0.07	1.04 (0.86-1.25)	0.68	
CYP2C9		c.430C>T	p.Arg144Cys	rs1799853	19-21	823	281	17	751	251	21	0.96 (0.79-1.18)	0.83	1.34 (0.69-2.59)	0.36	0.98 (0.81-1.20)	0.88	
EPHX1		c.337T>C	p.Tyr113His	rs1051740	19-21	564	469	88	525	409	89	0.92 (0.77-1.11)	0.47	1.12 (0.81-1.54)	0.61	0.95 (0.80-1.13)	0.58	
GSTA1		c.-4605G>A		rs3957356	22	372	549	200	333	517	173	1.07 (0.88-1.30)	0.60	1.03 (0.79-1.33)	0.79	1.06 (0.88-1.28)	0.53	

Table 1: Statistical analysis of association between CRC risk and polymorphisms selected for the study (n = 2144; OR calculations were adjusted according to age and sex).

	GSTM1	null		19-21	161	392	568	174	305	544	0.72 (0.55-0.94)	0.01	0.86 (0.67-1.10)	0.33	0.80 (0.63-1.02)	0.07
	GSTM3	c.468+21delAGG		19,20	788	308	25	679	316	27	1.19 (0.98-1.44)	0.07	1.09 (0.62-1.92)	0.42	1.18 (0.98-1.42)	0.08
	GSTP1	c.313A>G	p.Ile105Val	rs1695	19-21	511	486	124	465	447	1.04 (0.86-1.24)	0.90	0.96 (0.72-1.29)	0.91	1.02 (0.86-1.21)	0.82
	GSTP1	c.341C>T	p.Ala114Val	rs1138272	19-21	966	146	9	882	137	1.00 (0.77-1.29)	0.83	0.44 (0.13-1.47)	0.22	0.96 (0.75-1.24)	0.78
	GSTT1	null		19-21	433	483	205	411	429	183	0.95 (0.78-1.15)	0.49	0.97 (0.75-1.23)	0.62	0.95 (0.80-1.14)	0.58
	NQO1	c.415C>T	p.Arg139Trp	rs4986998	20,21	1056	65	0	961	62	1.06 (0.74-1.54)	0.80	Rare		1.06 (0.74-1.54)	0.80
	NQO1	c.559C>T	p.Pro187Ser	rs1800566	19,21	719	351	51	633	343	1.09 (0.91-1.32)	0.26	1.08 (0.71-1.64)	0.83	1.09 (0.91-1.31)	0.34
	SULT1A2	c.714A>C	p.Asn235Thr	rs1059491	21	535	458	128	481	437	1.09 (0.90-1.31)	0.52	0.92 (0.69-1.23)	0.53	1.05 (0.88-1.25)	0.58
	UGT1A1	c.-3279T>G		rs4124874	23	359	546	216	308	515	1.09 (0.90-1.33)	0.34	1.08 (0.84-1.39)	0.54	1.09 (0.90-1.31)	0.37
	UGT1A1	c.-3156G>A		rs10929302	23	535	495	91	474	458	1.03 (0.86-1.24)	0.65	1.11 (0.80-1.53)	0.45	1.04 (0.88-1.24)	0.63
	UGT1A6	c.541A>G	p.Thr181Ala	rs2070959	22	508	504	109	470	443	0.94 (0.78-1.13)	0.97	1.07 (0.80-1.45)	0.56	0.96 (0.81-1.15)	0.67
	UGT1A6	c.552A>C	p.Arg184Ser	rs1105879	22	460	529	132	438	454	1.03 (0.74-1.07)	0.26	1.03 (0.77-1.36)	0.77	0.92 (0.77-1.09)	0.34
Insulin	GHI	c.456+90T>A		rs2665802	24	326	561	234	316	527	0.96 (0.79-1.17)	0.75	0.80 (0.62-1.03)	0.07	0.91 (0.76-1.10)	0.36
	IGFI	c.-1006CA(19)			25	764	152	205	689	159	1.12 (0.87-1.44)	0.24	0.95 (0.75-1.20)	0.63	1.02 (0.85-1.23)	0.81
	IGFBP3	c.-336A>C		rs2854744	25	287	571	263	259	502	0.99 (0.80-1.22)	0.80	1.13 (0.88-1.44)	0.42	1.03 (0.85-1.26)	0.76
	IRSI	c.2911G>A	p.Gly971Arg	rs1801278	25	956	158	7	865	152	1.01 (0.79-1.29)	0.62	0.86 (0.28-2.66)	0.92	1.00 (0.79-1.28)	0.97
	VDR	c.1024+283G>A		rs1544410	25	423	520	178	372	493	1.11 (0.92-1.35)	0.43	1.06 (0.81-1.37)	0.94	1.10 (0.92-1.31)	0.31
One-carbon	CBS	c.844ins68			26,27	928	185	8	829	185	1.12 (0.89-1.41)	0.32	1.20 (0.44-3.28)	0.82	1.12 (0.90-1.41)	0.31
	MTHFD1	c.1958G>A	p.Arg653Gln	rs2236225	28	339	557	225	322	494	0.91 (0.75-1.12)	0.49	0.99 (0.77-1.27)	0.80	0.94 (0.78-1.13)	0.48
	MTHFR	c.665C>T	p.Ala222Val	rs1801133	26,27	457	515	149	435	452	0.93 (0.77-1.12)	0.38	0.95 (0.73-1.25)	0.76	0.93 (0.78-1.11)	0.44
	MTHFR	c.1286A>C	p.Ala429Glu	rs1801131	26,27	577	443	101	484	432	1.20 (1.00-1.45)	0.04	1.26 (0.93-1.70)	0.12	1.21 (1.02-1.44)	0.03
	MTR	c.2756A>G	p.Asp919Gly	rs1805087	26,27	742	335	44	706	289	0.90 (0.74-1.09)	0.31	0.64 (0.39-1.06)	0.10	0.87 (0.72-1.05)	0.14
	MTRR	c.66A>G	p.Ile22Met	rs1801394	26,27	322	568	231	291	515	1.01 (0.83-1.24)	0.97	1.01 (0.79-1.30)	0.75	1.01 (0.84-1.23)	0.89
	TYMS	c.943+447del TTTAAAG			27	524	482	115	456	456	1.13 (0.94-1.36)	0.36	1.04 (0.77-1.39)	0.52	1.11 (0.94-1.33)	0.22
DNA repair*	OGG1	c.977C>G	p.Cys326Ser	rs1052133	29	668	402	51	651	321	0.84 (0.69-1.01)	0.05	1.04 (0.69-1.57)	0.90	0.86 (0.72-1.03)	0.09

* SNP OGG1 c.977C>G was chosen to be analyzed together with 6 monoallelic germline mutations of the MUTYH gene – also belonging to the base excision repair system like OGG1 – which were studied previously for their predisposing effect on sporadic CRC (Küry et al., Genet Test; ref [30]).

** Nomenclature referring to the databases dbNSP, SNPPER and SNP500Cancer.

*** Ref. numbers correspond to bibliographical references, and the "d" mark corresponds to SNPs identified by dHPLC screening in 50 patients with CRC and 50 controls from the study population.

associations. We therefore followed the approach proposed by Storey et al., and we estimated the false discovery rate (FDR), more appropriate in large sets of hypotheses than multiple testing procedures [37]. For this estimation, we calculated the q-value, setting the threshold at 0.05. We considered findings that met a false discovery rate (FDR) criterion of 20% to be robust.

Multiple-SNP association study

Modeling strategy for analysis of genotypic combinations

According to the results of single polymorphisms analyses, we tested all possible combinations of N polymorphisms composed by the N allelic variants found to be significantly associated with CRC risk, and by polymorphisms associated with marginally or non-significant ORs having a p value comprised between 0.05 and 0.25. As a comparison, we also analyzed a hundred of combinations of N polymorphisms randomly selected from any of the 52 ones included in our study. For each polymorphism, the genotype was coded 0 if the rare variant was absent, 1 if present at heterozygous state, 2 if present at homozygous state. Because most of the homozygotes for the minor allele of the polymorphisms tested were often extremely rare, we assumed in our model that the rare effect allele was dominant. In other terms, if a polymorphism predisposed to CRC, genotypes 1 and 2 were considered predisposing to CRC; on the contrary, genotypes 1 and 2 were considered protective if the effect of the minor variant of the polymorphism was rather protective. For each polymorphism included in a combination, we therefore tested two hypotheses: 1. either we considered that the minor allele was predisposing to CRC and therefore the CRC-predisposing genotypes were 1 and 2, whereas genotype 0 was considered protective; 2. or we considered that the minor allele was associated with a decreased CRC risk, and therefore 0 was considered to be the predisposing genotype, whereas genotypes 1 and 2 were considered protective. Thus, for every combination of N SNPs tested, we also tested every possible combination of x protective and N-x predisposing minor alleles (with x ranging from 0 to N), each combination tested representing a different model.

For each test of a combination or model, we compared three groups. The first group was composed by individuals exhibiting genotype 0 for the N SNPs, i.e., carrying the frequent allele at homozygous state for the N SNPs; this group, theoretically the most frequent group in the study cohort, was set as a reference, and the relating combination of genotypes was considered as the 'reference pattern' exhibiting a null or very weak effect on CRC risk. The second group was composed of individuals exhibiting patterns of genotypic combinations composed of N predisposing genotypes (predisposing patterns). The third

group comprised the individuals exhibiting all other possible combinations of genotypes (mixed patterns).

In the same way as for single SNP association study, heterogeneity between the groups of patients and controls was translated into colorectal cancer risk associated with genotypic patterns, assessed by OR calculation together with 95 percent confidence intervals. We first used unconditional logistic regression, adjusting for sex and age, and we compared the results observed to those obtained by conditional logistic regression analysis restricted to 811 age- and sex-matched pairs of individuals. We eventually assessed the robustness of the associations observed by applying the false discovery rate (FDR) method, setting the threshold at 0.05.

Validation of the logistic regression model

In order to validate the logistic regression model, we applied a standard Monte Carlo bootstrap (random resampling with replacement from the original dataset) procedure based on 1000 replicates. We calculated the 5th and 95th percentiles from the resulting distributions to determine the lower and upper limits of the confidence interval.

Alternatively, we examined the internal consistency of the results obtained for the model. Analyses of genotypic combinations were replicated on stratifications of the study population determined according to age, geographical origin, random selection, or inference of population structure [Additional file 1].

Results

The allele frequencies we found for each of the 52 polymorphisms tested were consistent with those reported in literature and in dbSNP <http://www.ncbi.nlm.nih.gov/projects/SNP/> for Caucasian populations. Two of the 52 polymorphisms analyzed turned out to be monomorphic. The 50 others were all at Hardy-Weinberg equilibrium. Results of the unconditional logistic regression analyses for CRC association with polymorphisms considered independently are described in Table 1. Alternative results of conditional logistic regression restricted to 811 age- and sex-matched pairs of individuals are reported in Additional file 3.

Six associations between CRC risk and allelic variants were determined by both unconditional and conditional logistic regression analyses. For SNPs *PTGS1* c.639C>A (p.Gly213Gly), *IL8* c.-352T>A, and *MTHFR* c.1286A>C (p.Ala429Glu), minor alleles appeared associated with an increase in CRC, whereas for SNPs *PLA2G2A* c.435+230C>T, *PPARG* c.1431C>T (p.His477His), they were associated with a decrease in CRC risk. Application of the false discovery rate to the results obtained by

unconditional logistic regression analysis showed a value of about 0.5 for these five findings, suggesting that the associations were not very robust. Yet, on the contrary, *q*-values determined from conditional logistic regression analysis were found between 0.07 and 0.26 for the same findings, rather suggesting quite robust associations [Additional file 3]. Our analyses also showed a protective effect of the variant *GSTM1* null carried at the heterozygous state on CRC risk. However, this observation is not relevant at the biological level since the loss of activity of the *GSTM1* enzyme, or *GSTM1* deficiency, is associated with a deletion carried on both alleles. In fact, complementary analyses in our study population revealed that individuals carrying two deleted alleles did not have a significantly different risk of CRC compared to carriers of one or zero deleted alleles (OR 1.08, 95% CI 0.91–1.21, *p* = 0.4), which is consistent with two meta-analyses performed on this polymorphism [19,38]. Therefore, the CRC-predisposing effect observed for *GSTM1* null allele is to consider with great caution, all the more because the *q*-value of 0.4 calculated for this finding would rather suggest a false positive result. Two additional CRC-predisposing effects were observed for SNPs *PLA2G2A* c.-859C>G and *CYP1B1* c.1294C>G by conditional logistic regression analysis [Additional file 3], but were not found when using unconditional method applied to the whole study population; they were therefore considered as false positives.

Since five SNPs were undoubtedly found associated with a modification of CRC risk in single-SNP analyses, we focused our further analyses on combinations of five allelic variants, as described in Methods. Among all the genotypic combinations tested, only one showed a statistically significant association with an increased risk of CRC, which appeared robust according to the calculation of FDR (*q*-value < 0.1). This combination was composed by the very same five SNPs that had been found independently associated with CRC risk from single-SNP analyses. Additional file 4 uses the example of this precise set of five SNPs to illustrate the different models of genotypic combinations tested for every set of five polymorphisms – picked from the 52 polymorphisms of the study – which we tested. The most predisposing patterns for this combination, presented in Table 2, combine genotypes 1 and 2 for *PTGS1* c.639C>A, *IL8* c.-352T>A, and *MTHFR* c.1286A>C, and 0 for *PLA2G2A* c.435+230C>T and *PPARG* c.1431C>T. These predisposing patterns appeared to be associated with a highly significant increase in colorectal cancer, either compared to the reference pattern alone (OR 2.65; 95% CI 1.58–4.42 with *p* = 0.0005), or compared to reference and mixed patterns gathered (OR 1.97; 95% CI 1.31–2.97 with *p* = 0.0009).

Validation of the model by bootstrapping and examination of internal consistency strengthened the above results. By bootstrapping, we calculated a mean ± SD OR of 1.86 ± 0.41 (95%CI 1.23–2.85, *p* = 0.003). Examination of internal consistency confirmed a trend in an increased CRC risk associated with the predisposing combination patterns, regardless of the mode of stratification used [Additional file 5]. No effect of gender or anatomical sub-location was noted. Most of the associations observed were statistically significant, but certain stratifications determined according to geographical origin and/or age displayed a greater effect regarding the predisposing combination patterns (Table 3). Among the 2144 individuals composing the whole study population, the strongest association was found in individuals of ≤ 67 years of age and originating from the French *département* of the Vendée.

Discussion

In this study, we have used a candidate gene approach to examine the associations between colorectal cancer risk and 52 allelic variants distributed in 35 genes drawn from pathways of inflammation, metabolism of xenobiotics detoxification, one-carbon, insulin signaling, and DNA repair. To our knowledge, we are describing here within one of the most comprehensive investigations on populations of this kind, covering more than 1000 patients with sporadic colorectal cancer, and 1000 controls, i.e., in the range of 500–2000 case-control pairs defined by Brennan to detect the statistically significant effect of polymorphisms [39]. Obviously, the list of polymorphisms which we designed here is non-comprehensive, and it must be seen as a panel test, a first attempt in the search for CRC-predisposing 'genetic profiles'.

With the exception of the 4 SNPs in *PLA2G2A* that we had selected by dHPLC, all the allelic variants chosen for the present study had previously been found to be associated with a modification of CRC risk in at least one study. However, we have been able to replicate the association with CRC risk for only 5 of the 52 polymorphisms tested (Table 1). By independent single-SNP analyses, three SNPs were shown to increase CRC risk: *PTGS1* c.639C>A, *IL8* c.-352T>A, and *MTHFR* c.1286A>C. Two other SNPs, *PLA2G2A* c.435+230C>T and *PPARG* c.1431C>T, were on the contrary associated with a decrease in CRC risk. Combinations of the CRC-predisposing alleles relating to these five variants determine 'genotypic profiles' at significantly higher risk of CRC (OR 1.97, 95% CI 1.31–2.97). Among the individuals exhibiting these profiles, younger individuals (≤ 67 years) from the Vendée seem to be even more predisposed to CRC (OR 6.36, 95% CI 1.44–28.09). However, the size of the population sample analyzed here is too small to draw definitive conclusions on any age- or geographical-effect. In the same way, the consistency of

Table 2: Analysis of association between combinations of genotypes and risk for colorectal cancer, in the entire study population.

A. Encoding of theoretical combinations of genotypes					
Patterns of genotypic combinations					Effect on colorectal cancer
PTGS1 c.639C>A	PLA2G2A c.435+230C>T	PPARG c.1431C>T	IL8 c.-352T>A	MTHFR c.1286A>C	
0	0	0	0	0	Null or very weak (reference pattern)
0	1 or 2	1 or 2	0	0	Protective (protective patterns) *
1 or 2	0	0	1 or 2	1 or 2	Predisposing (predisposing patterns) Average (mixed patterns)
Other genotypes					
B. Analysis of observed combination of genotypes association with colorectal cancer (n = 2144, adjusted by sex and age)					
Patterns of genotypic combinations	Controls	Patients	OR (95% CI)	P-value**	
Reference pattern	95 (8.5%)	63 (6.2%)	1.00		
Protective patterns	7 (0.6%)	4 (0.4%)	0.86 (0.24–3.06)	0.8180	
Mixed patterns	978 (87.2%)	890 (87.0%)	1.37 (0.98–1.90)	0.0602	
Predisposing patterns	41 (3.7%)	66 (6.4%)	2.65 (1.58–4.42)	0.0005	
Reference and mixed patterns	1080 (96.3%)	957 (93.5%)	1.00		
Predisposing patterns	41 (3.7%)	66 (6.5%)	1.97 (1.31–2.97)	0.0009	

* These patterns were inferred from the identification of the predisposing patterns.

** Observed associations in bold met a false discovery rate criterion of < 0.1.

the risk profile was lost in some of the study population subgroups used for the analyses reported in Additional file 5.b, certainly because of their relatively small size. Caution is therefore required, when considering the true significance of this risk profile, even though bootstrapping results tend to show its robustness [Additional file 5.a]. It is noteworthy that a sixth polymorphism, GSTM1 null allele, showed a protective effect in independent analyses of allelic variants, but the significance of its association with CRC risk remained dubious, and was not confirmed by multiple-SNP analyses, contrary to the five other polymorphisms.

To investigate further the CRC-predisposing combinations that emerged from our analyses, we tested their possible interactions with the environmental and lifestyle risk factors reported in our study questionnaire (physical activity, cooking methods, and consumption of alcohol, tobacco, red meat, cold cuts, white meat and poultry bread, dairy products, fish, fruits, pastries, or vegetables), by use of SNPStats [40]. Indeed, in the same study popu-

lation, we had previously observed an interaction between allelic combinations of CYP genes and consumption of red meat which leads to a strong increase in CRC risk [15]. In the present study, we did not observe any comparable gene-environment interaction that could remain statistically significant throughout multiple adjustment and/or test for internal consistency in patient and control sub-groups (data not shown). Given that four of the SNPs composing the CRC-predisposing combinations are related to inflammation, the most relevant 'environmental' factor to be tested here would have actually been NSAIDs treatment. But, since this item did not figure in our questionnaire on life-habits [15] – we assumed that it would have introduced a bias in the design of the control group-, we were unfortunately unable to analyze its interaction with the CRC-predisposing genotypic combinations of the five SNPs mentioned above.

The respective biological impact of the five variants *PTGS1* c.639C>A, *IL8* c.-352T>A, *MTHFR* c.1286A>C, *PLA2G2A* c.435+230C>T, and *PPARG* c.1431C>T provide some

Table 3: Replications of genotypic combinations analyses on stratifications of the patients and controls groups determined according to geographical origin and/or age.

Genotypic combinations patterns	Individuals originating from Vendée				Individuals ≤ 67 years				Individuals from Vendée ≤ 67 years			
	Controls (n = 189)	Cases (n = 420)	OR (95% CI)	p*	Controls (n = 768)	Cases (n = 529)	OR (95% CI)	p*	Controls (n = 134)	Cases (n = 232)	OR (95% CI)	p*
Reference pattern	14	29	1.00		65	31	1.00		12	16	1.00	
Mixed patterns	171	359	0.92 (0.47–1.83)		676	460	1.42 (0.91–2.21)		120	195	1.15 (0.52–2.56)	
Predisposing patterns	4	32	3.80 (1.10–13.19)	0.011	27	38	3.00 (1.56–5.77)	0.003	2	21	7.22 (1.38–37.80)	0.009
Reference and mixed patterns	185	388	1.00		741	491	1.00		132	211	1.00	
Predisposing patterns	4	32	4.14 (1.42–12.08)	0.002	27	38	2.17 (1.31–3.61)	0.003	2	21	6.36 (1.44–28.09)	0.002

OR calculations were all adjusted by age and sex.

clues to the predisposition to CRC associated with some of the genetic combinations that they compose. Thus, *PLA2G2A* c.435+230C>T and *PTGS1* c.639C>A, belong to genes which encode proteins following each other in the enzymatic cascade of the arachidonic acid pathway. *PLA2G2A* catalyzes the hydrolysis of membrane phospholipids, thereby releasing unsaturated fatty acids, including arachidonic acid [41]. The latter becomes the substrate of *PTGS1*, which in turn catalyzes the formation of prostaglandin H₂ (PGH₂), a precursor for a number of inflammatory molecules – eicosanoids – that promote colorectal carcinogenesis [42]. Until now, out of the potential key players in the CRC process, more attention has been given to another element of the arachidonic acid pathway, *PTGS2*, rather than to *PTGS1* or *PLA2G2A* [31,43]. The interest for *PTGS2* in CRC risk stems from its expression induced by pathophysiological conditions such as tumorigenesis or inflammatory situations [16]. However, we found no significant modification of CRC risk associated with any of the four *PTGS2* SNPs we tested (Table 1). In contrast, *PTGS1* is constitutively expressed and its implication in CRC risk has not been investigated very much, since its carcinogenic role through induction of *PTGS2* has been suggested only recently [16,34]. Thus, to our knowledge, the synonymous polymorphism c.639C>A (p.Gly213Gly) that we found as having a predisposing effect on CRC (OR 1.24) at the heterozygous (CA) or homozygous state (AA), had not been examined within this context to date and its precise impact on *PTGS1* activity would require further functional studies. In the same way, the four *PLA2G2A* SNPs we studied have not yet been tested within the CRC risk context. Therefore, there is no possible point of comparison for the protective effect relating to CRC that we found for the allele c.435+230T carried at homozygous state (OR 0.80, 95%CI 0.66–0.98). Because of its localization within the Mom-1 (Min modifier) locus, and its promotion of tumors in APC^{Min} mice, it had first been suggested that *PLA2G2A* may represent a tumor suppressor gene involved in familial forms of CRC [44]. As a result, such a role of *PLA2G2A* in FAP genesis had been ruled out in humans [41,45]. However, the contribution of *PLA2G2A* to sporadic CRC predisposition cannot be excluded, and it has been suggested that the increase in *PLA2G2A* expression could cause the accumulation of arachidonic acid, a molecule likely to have pro-apoptotic properties [41]. A hypothesis for the protective role of variant c.435+230T might therefore be its enhancement of *PLA2G2A* expression.

PPARG c.1431C>T (aka C161T) and *IL8* c.-352T>A, also relate to inflammation or immune response. *PPAR*_γ (peroxisome proliferator-activated receptor gamma) is a nuclear receptor tightly linked to the arachidonic acid pathway, in that it is activated by various eicosanoids [46].

Essential to adipocyte differentiation and to regulation of lipid metabolism, *PPAR*_γ is thought to have overall tumor suppressive properties, exerted notably in CRC [46-48]. A protective effect on CRC risk had been reported for the rare allele of variant *PPARG* c.36C>G (p.Pro12Ala) [49,50], but we were not able to reproduce this observation in our study population, perhaps because the reference studies were designed on smaller populations – about 200 case-control pairs-. On the other hand, we noted a significantly decreased risk of CRC associated with the rare allele of variant c.1431C>T (p.His477His) at the homozygous state (OR 0.30, 95%CI 0.14–0.65). These results are consistent with those reported in an earlier study on colorectal adenoma [48], but they are inconsistent with the contrary observations made by another team investigating CRC [33]. Given the controversial effect assigned to *PPARG* c.1431C>T, functional assays would be required to understand the exact role of this little-investigated polymorphism.

As regards *IL8*, it encodes a pro-inflammatory chemokine, released by infiltrating lymphocytes, in response to exposure of the colonic epithelium to toxic and pathologic challenges [50]. The allelic variant *IL8* c.-352T>A has been found to be associated with CRC risk, even though the effect attributed to rare allele A goes from null [51], to predisposing or even protective [17,52]. As regards our study, we found that genotypes c.352AA or c.352TA were associated with an elevated risk of CRC compared to genotype c.352TT (OR 1.21, 95%CI 1.01–1.46). The debate on the true carcinogenic role of variant c.-352T>A probably comes from its controversial impact on *IL8* transcription, sometimes described as an enhancement [51], sometimes as a downregulation [52]. As *IL8* contributes to chronic inflammation, it would make sense that its overexpression would increase the risk for CRC, and therefore the allele c.-352A may rather enhance the transcriptional activity of the gene.

The fifth gene highlighted by our analyses, *MTHFR* (5,10-methylenetetrahydrofolate reductase), is part of the one-carbon metabolic pathway, involved in both DNA methylation and DNA synthesis. The *MTHFR* enzyme synthesizes 5-methyltetrahydrofolate, the primary circulatory form of folate, which represents a fundamental methyl donor in cellular metabolism [26]. The influence of *MTHFR* activity on folate status could be important in CRC neoplasia, since folate deficiency could cause DNA hypomethylation, and/or induce uracil misincorporation during DNA synthesis, which would lead to mutations and chromosomal damage [26,27,53]. Two polymorphisms c.665C>T (also reported as C677T) and c.1286A>C (aka c.1298A>C), have been widely tested for modification of CRC risk, because of the reduction in *MTHFR* activity induced by their minor allele. In the present study, we found no effect for c.665C>T, whereas we observed

an elevated risk of CRC associated with genotypes c.1286CC or c.1286AC compared to genotype c.1286AA (OR 1.21, 95%CI 1.02–1.44). At first sight, these results appear difficult to compare with the greatly inconsistent or even conflicting earlier results, and reviewed in two recent comprehensive meta-analyses [54,55]. However, this discrepancy in the overall results could be explained by the variation of the polymorphisms effect according to folate status. Mu et al. suggested indeed that genotypes reducing MTHFR activity – here c.1286CC or c.1286AC – would favor cancer risk when dietary folate levels are low, by inducing a DNA hypomethylation causing DNA damage and mutations [56]. When folate levels are adequate, the reduced MTHFR activity induced by the same genotypes would lead to a great pool of methylenetetrahydrofolate available for DNA synthesis, and therefore prevent cancer by diminishing uracil misincorporation. Moreover, it has been suggested that, according to its administration prior or further to the existence of preneoplastic lesions, folate would rather prevent or increase tumor development, respectively [53].

All these biological data considered, our work underlines the important contribution of inflammatory processes to CRC susceptibility in our study population, and it points to a possible crosstalk between inflammation and one-carbon pathways. The four allelic variants of genes *PTGS1*, *PLA2G2A*, *PPARG*, and *IL8* might favor inflammatory processes, whereas the *MTHFR* allelic variant could induce a DNA hypomethylation altering the expression of the putative tumor suppressor gene *PPARG*. Two recent studies on diabetes and cardiovascular diseases have suggested a common pathobiological mechanism between the inflammation process and genotype TT of *MTHFR* c.665C>T (aka C677T), i.e., a genotype inducing a lower MTHFR activity like genotype *MTHFR* c.1286CC [57,58]. However, our results do not enable to conclude to an interaction between the five allelic variants. Indeed, the odd ratios associated with the observed CRC-predisposing combinations are certainly not strong enough, and our method is not the most appropriate one for an investigation on gene-gene interactions.

Conclusion

Combinations of alleles *PTGS1* c.639A, *PLA2G2A* c.435+230C, *PPARG* c.1431C, *IL8* c.-352A, and *MTHFR* c.1286C determine 'genotypic profiles' at significantly higher risk of CRC (OR 1.97, 95% CI 1.31–2.97), in a small percentage of the French population. The identification of these CRC-predisposing combinations highlights the importance of inflammatory processes in susceptibility to sporadic CRC, as well as a possible crosstalk between inflammation and one-carbon pathways.

To find their significance, our results would obviously need to be reproduced. Indeed, replication represents a

key step for establishing the credibility of any genotype-phenotype association [59]. Only a validation in other Caucasian populations would therefore enable to consider the five above variants and their related genotypic combinations as relevant risk markers for sporadic CRC. Moreover, even though the size of our population sample is quite large, it is not large enough to draw definitive conclusions on the statistical significance of the polymorphisms-disease associations observed. In order to reach such conclusions, our results should be combined to a meta-analysis, since this tool is particularly robust to estimate population-wide effect of genetic risk factors in diseases [60]. In any case, our observations would warrant investigations on possible interactions between NSAIDs and combinations of polymorphisms from the five genes we have highlighted here, which our questionnaire unfortunately did not enable us to achieve. The results of our study also call for more comprehensive investigations into polymorphisms of inflammation genes – especially those belonging to the arachidonic acid pathway-, as candidate risk factors for CRC. In fact, the CRC-predisposing combinations we identified account for only 5% of our study population, which means that many more predisposing combinations, involving different and maybe common polymorphisms, are still to be found. Very large genetic association studies on candidate genes of inflammation may make it possible to find the most frequent predisposing combinations, and thus make the first step towards drawing up recommendations for clinical practice guidelines regarding sporadic CRC.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

SK, BB, and SRdP equally contributed to this work. SK participated in the design of the study, carried out statistical analyses, contributed to genotyping assays, and drafted the manuscript; BB conceived the study, participated in its design, and coordinated the recruitment of patients and controls; SRdP participated in statistical analyses and genotyping assays, and helped to draft the manuscript; CS did samples management and contributed to genotyping assays; HC contributed to molecular genetic analyses; TLN coordinated data storage and management; CLH coordinated the recruitment of patients and controls; RF and JO contributed to the collection of patients samples; BL and LDC contributed to the collection of controls samples; VS participated in statistical analyses; SB conceived the study, participated in its design and coordination and helped to draft the manuscript. All authors read and approved the final manuscript.

Additional material

Additional file 1

Supplemental materials and methods.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2407-8-326-S1.rtf>]

Additional file 2

List of primers and probes used for the study.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2407-8-326-S2.xls>]

Additional file 3

Statistical analysis of association between CRC risk and polymorphisms selected for the study using conditional logistic regression (number of age- and sex-matched pairs = 811, i.e., 1622 individuals).

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2407-8-326-S3.xls>]

Additional file 4

Test of the different possible models of genotypic combinations related to a given set of five polymorphisms picked among the 52 polymorphisms analysed in the study.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2407-8-326-S4.xls>]

Additional file 5

Estimation by use of bootstrapping of the external validity of the model showing association between CRC risk and genotypes combinations.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2407-8-326-S5.xls>]

Acknowledgements

We are very grateful to all the patients, control individuals, and physicians who participated in this work and made it possible. This work was funded by a regional Hospital Clinical Research Program (PHRC) and supported by the Regional Council of Pays de la Loire, the Groupement des Entreprises Françaises dans la Lutte contre le Cancer (GEFLUC), the Association Anne de Bretagne Génétique and the Ligue Régionale Contre le Cancer (LRCC).

References

- Remontet L, Esteve J, Bouvier AM, Grosclaude P, Launoy G, Menegoz F, Exbrayat C, Tretare B, Carli PM, Guizard AV, Troussard X, Bercelli P, Colonna M, Halna JM, Hedelin G, Mace-Lesech J, Peng J, Buemi A, Velten M, Jouglu E, Arveux P, Le Bodic L, Michel E, Sauvage M, Schwartz C, Faivre J: **Cancer incidence and mortality in France over the period 1978-2000.** *Rev Epidemiol Sante Publique* 2003, **51(1 Pt 1)**:3-30.
- de la Chapelle A: **Genetic predisposition to colorectal cancer.** *Nat Rev Cancer* 2004, **4(10)**:769-780.
- Davidson NO: **Genetic testing in colorectal cancer: who, when, how and why.** *Keio J Med* 2007, **56(1)**:14-20.
- Hunter DJ, Riboli E, Haiman CA, Albanes D, Altshuler D, Chanock SJ, Haynes RB, Henderson BE, Kaaks R, Stram DO, Thomas G, Thun MJ, Blanche H, Buring JE, Burt NP, Calle EE, Cann H, Canzian F, Chen YC, Colditz GA, Cox DG, Dunning AM, Feigelson HS, Freedman ML, Gaziano JM, Giovannucci E, Hankinson SE, Hirschhorn JN, Hoover RN, Key T, et al: **A candidate gene approach to searching for low-penetrance breast and prostate cancer genes.** *Nat Rev Cancer* 2005, **5(12)**:977-985.
- Goodman JE, Mechanic LE, Luke BT, Ambis S, Chanock S, Harris CC: **Exploring SNP-SNP interactions and colon cancer risk using polymorphism interaction analysis.** *Int J Cancer* 2006, **118(7)**:1790-1797.
- Tabor HK, Risch NJ, Myers RM: **Candidate-gene approaches for studying complex genetic traits: practical considerations.** *Nat Rev Genet* 2002, **3(5)**:391-397.
- Naccarati A, Pardini B, Hemminki K, Vodicka P: **Sporadic colorectal cancer and individual susceptibility: a review of the association studies investigating the role of DNA repair genetic polymorphisms.** *Mutat Res* 2007, **635(2-3)**:118-145.
- Tomlinson I, Webb E, Carvajal-Carmona L, Broderick P, Kemp Z, Spain S, Penegar S, Chandler I, Gorman M, Wood W, Barclay E, Lubbe S, Martin L, Sellick G, Jaeger E, Hubner R, Wild R, Rowan A, Fielding S, Howarth K, Silver A, Atkin W, Muir K, Logan R, Kerr D, Johnstone E, Sieber O, Gray R, Thomas H, Peto J, et al: **A genome-wide association scan of tag SNPs identifies a susceptibility variant for colorectal cancer at 8q24.21.** *Nat Genet* 2007, **39(8)**:984-988.
- Zanke BW, Greenwood CM, Rangrej J, Kustra R, Tenesa A, Farrington SM, Prendergast J, Olschwang S, Chiang T, Crowdy E, Ferretti V, Laflamme P, Sundararajan S, Roumy S, Olivier JF, Robidoux F, Sladek R, Montpetit A, Campbell P, Bézieau S, O'Shea AM, Zogopoulos G, Cotterchio M, Newcomb P, McLaughlin J, Younghusband B, Green R, Green J, Porteous ME, Campbell H, et al: **Genome-wide association scan identifies a colorectal cancer susceptibility locus on chromosome 8q24.** *Nat Genet* 2007, **39(8)**:989-994.
- Picelli S, Vandrovčova J, Jones S, Djureinovic T, Skoglund J, Zhou XL, Velculescu VE, Vogelstein B, Lindblom A: **Genome-wide linkage scan for colorectal cancer susceptibility genes supports linkage to chromosome 3q.** *BMC Cancer* 2008, **8(1)**:87.
- Tenesa A, Farrington SM, Prendergast JG, Porteous ME, Walker M, Haq N, Barnetson RA, Theodoratou E, Cetnarskyj R, Cartwright N, Semple C, Clark AJ, Reid FJ, Smith LA, Kavoussanakis K, Koessler T, Pharoah PD, Buch S, Schafmayer C, Tepel J, Schreiber S, Volzke H, Schmidt CO, Hampe J, Chang-Claude J, Hoffmeister M, Brenner H, Wilkerson S, Canzian F, Capella G, et al: **Genome-wide association scan identifies a colorectal cancer susceptibility locus on 11q23 and replicates risk loci at 8q24 and 18q21.** *Nat Genet* 2008, **40(5)**:631-637.
- Tomlinson IP, Webb E, Carvajal-Carmona L, Broderick P, Howarth K, Pittman AM, Spain S, Lubbe S, Walther A, Sullivan K, Jaeger E, Fielding S, Rowan A, Vijaykrishnan J, Domingo E, Chandler I, Kemp Z, Qureshi M, Farrington SM, Tenesa A, Prendergast JG, Barnetson RA, Penegar S, Barclay E, Wood W, Martin L, Gorman M, Thomas H, Peto J, Bishop DT, et al: **A genome-wide association study identifies colorectal cancer susceptibility loci on chromosomes 10p14 and 8q23.3.** *Nat Genet* 2008, **40(5)**:623-630.
- Onay VU, Briollais L, Knight JA, Shi E, Wang Y, Wells S, Li H, Rajendram I, Andrusis IL, Ozelik H: **SNP-SNP interactions in breast cancer susceptibility.** *BMC Cancer* 2006, **6**:114.
- Wu X, Gu J, Grossman HB, Amos CI, Etzel C, Huang M, Zhang Q, Millikan RE, Lerner S, Dinney CP, Spitz MR: **Bladder cancer predisposition: a multigenic approach to DNA-repair and cell-cycle-control genes.** *Am J Hum Genet* 2006, **78(3)**:464-479.
- Küry S, Buecher B, Robiou-du-Pont S, Scoul C, Sebille V, Colman H, Le Houérou C, Le Neel T, Bourdon J, Faroux R, Ollivry J, Lafraisse B, Chupin LD, Bézieau S: **Combinations of cytochrome P450 gene polymorphisms enhancing the risk for sporadic colorectal cancer related to red meat consumption.** *Cancer Epidemiol Biomarkers Prev* 2007, **16(7)**:1460-1467.
- Goodman JE, Bowman ED, Chanock SJ, Alberg AJ, Harris CC: **Ara-chidonate lipoygenase (ALOX) and cyclooxygenase (COX) polymorphisms and colon cancer risk.** *Carcinogenesis* 2004, **25(12)**:2467-2472.
- Landi S, Moreno V, Gioia-Patricola L, Guino E, Navarro M, de Oca J, Capella G, Canzian F: **Association of common polymorphisms in inflammatory genes interleukin (IL)6, IL8, tumor necrosis factor alpha, NFKB1, and peroxisome proliferator-activated receptor gamma with colorectal cancer.** *Cancer Res* 2003, **63(13)**:3560-3566.
- Yamada Y, Hamajima N, Kato T, Iwata H, Yamamura Y, Shinoda M, Suyama M, Mitsudomi T, Tajima K, Kusakabe S, Yoshida H, Banno Y, Akao Y, Tanaka M, Nozawa Y: **Association of a polymorphism of**

- the phospholipase D2 gene with the prevalence of colorectal cancer.** *J Mol Med* 2003, **81(2)**:126-131.
19. de Jong MM, Nolte IM, te Meerman GJ, Graaf WT van der, de Vries EG, Sijmons RH, Hofstra RM, Kleibeuker JH: **Low-penetrance genes and their involvement in colorectal cancer susceptibility.** *Cancer Epidemiol Biomarkers Prev* 2002, **11(11)**:1332-1352.
 20. Landi S, Gemignani F, Moreno V, Gioia-Patricola L, Chabrier A, Guino E, Navarro M, de Oca J, Capella G, Canzian F: **A comprehensive analysis of phase I and phase II metabolism gene polymorphisms and risk of colorectal cancer.** *Pharmacogenet Genomics* 2005, **15(8)**:535-546.
 21. Sachse C, Smith G, Wilkie MJ, Barrett JH, Waxman R, Sullivan F, Forman D, Bishop DT, Wolf CR: **A pharmacogenetic study to investigate the role of dietary carcinogens in the etiology of colorectal cancer.** *Carcinogenesis* 2002, **23(11)**:1839-1849.
 22. Logt EM van der, Bergevoet SM, Roelofs HM, van Hooijdonk Z, Te Morsche RH, Wobbes T, de Kok JB, Nagengast FM, Peters WH: **Genetic polymorphisms in UDP-glucuronosyltransferases and glutathione S-transferases and colorectal cancer risk.** *Carcinogenesis* 2004, **25(12)**:2407-2415.
 23. Innocenti F, Grimsley C, Das S, Ramirez J, Cheng C, Kuttub-Boulos H, Ratain MJ, Di Rienzo A: **Haplotype structure of the UDP-glucuronosyltransferase IAI promoter in different ethnic groups.** *Pharmacogenetics* 2002, **12(9)**:725-733.
 24. Le Marchand L, Donlon T, Seifried A, Kaaks R, Rinaldi S, Wilkens LR: **Association of a common polymorphism in the human GHI gene with colorectal neoplasia.** *J Natl Cancer Inst* 2002, **94(6)**:454-460.
 25. Slattery ML, Samowitz W, Hoffman M, Ma KN, Levin TR, Neuhausen S: **Aspirin, NSAIDs, and colorectal cancer: possible involvement in an insulin-related pathway.** *Cancer Epidemiol Biomarkers Prev* 2004, **13(4)**:538-545.
 26. Le Marchand L, Donlon T, Hankin JH, Kolonel LN, Wilkens LR, Seifried A: **B-vitamin intake, metabolic genes, and colorectal cancer risk (United States).** *Cancer Causes Control* 2002, **13(3)**:239-248.
 27. Sharp L, Little J: **Polymorphisms in genes involved in folate metabolism and colorectal neoplasia: a HuGE review.** *Am J Epidemiol* 2004, **159(5)**:423-443.
 28. Chen J, Kyte C, Valcin M, Chan W, Wetmur JG, Selhub J, Hunter DJ, Ma J: **Polymorphisms in the one-carbon metabolic pathway, plasma folate levels and colorectal cancer in a prospective study.** *Int J Cancer* 2004, **110(4)**:617-620.
 29. Hansen R, Saebo M, Skjelbred CF, Nexø BA, Hagen PC, Bock G, Bowtitz Lothe IM, Johnson E, Aase S, Hanstean IL, Vogel U, Kure EH: **GPX Pro198Leu and OGG1 Ser326Cys polymorphisms and risk of development of colorectal adenomas and colorectal cancer.** *Cancer Lett* 2005, **229(1)**:85-91.
 30. Kiry S, Buecher B, Robiou-du-Pont S, Scoull C, Colman H, Lelièvre B, Olschwang S, Le Houérou C, Le Neel T, Faroux R, Ollivry J, Lafraise B, Chupin LD, Béziau S: **The thorough screening of the MUTYH gene in a large French cohort of sporadic colorectal cancers.** *Genet Test* 2007, **11(4)**:373-9.
 31. Cox DG, Pontes C, Guino E, Navarro M, Osorio A, Canzian F, Moreno V: **Polymorphisms in prostaglandin synthase 2/cyclooxygenase 2 (PTGS2/COX2) and risk of colorectal cancer.** *Br J Cancer* 2004, **91(2)**:339-343.
 32. Halushka MK, Walker LP, Halushka PV: **Genetic variation in cyclooxygenase 1: effects on response to aspirin.** *Clin Pharmacol Ther* 2003, **73(1)**:122-130.
 33. Jiang J, Gajalakshmi V, Wang J, Kuriki K, Suzuki S, Nakamura S, Akasaka S, Ishikawa H, Tokudome S: **Influence of the C161T but not Pro12Aa polymorphism in the peroxisome proliferator-activated receptor-gamma on colorectal cancer in an Indian population.** *Cancer Sci* 2005, **96(8)**:507-512.
 34. Ulrich CM, Bigler J, Sparks R, Whitton J, Sibert JG, Goode EL, Yasui Y, Potter JD: **Polymorphisms in PTGS1 (= COX-1) and risk of colorectal polyps.** *Cancer Epidemiol Biomarkers Prev* 2004, **13(5)**:889-893.
 35. den Dunnen JT, Antonarakis SE: **Nomenclature for the description of human sequence variations.** *Hum Genet* 2001, **109(1)**:121-124.
 36. Rothman KJ: **No adjustments are needed for multiple comparisons.** *Epidemiology* 1990, **1(1)**:43-46.
 37. Storey JD, Tibshirani R: **Statistical significance for genomewide studies.** *Proc Natl Acad Sci USA* 2003, **100(16)**:9440-9445.
 38. Ye Z, Parry JM: **A meta-analysis of 20 case-control studies of the glutathione S-transferase M1 (GSTM1) status and colorectal cancer risk.** *Med Sci Monit* 2003, **9(10)**:SR83-91.
 39. Brennan P: **Gene-environment interaction and aetiology of cancer: what does it mean and how can we measure it?** *Carcinogenesis* 2002, **23(3)**:381-387.
 40. Sole X, Guino E, Valls J, Iniesta R, Moreno V: **SNPStats: a web tool for the analysis of association studies.** *Bioinformatics* 2006, **22(15)**:1928-1929.
 41. Diaz BL, Arm JP: **Phospholipase A(2).** *Prostaglandins Leukot Essent Fatty Acids* 2003, **69(2-3)**:87-97.
 42. Steele VE, Hawk ET, Viner JL, Lubet RA: **Mechanisms and applications of non-steroidal anti-inflammatory drugs in the chemoprevention of cancer.** *Mutat Res* 2003, **523-524**:137-144.
 43. Tan W, Wu J, Zhang X, Guo Y, Liu J, Sun T, Zhang B, Zhao D, Yang M, Yu D, Lin D: **Associations of functional polymorphisms in cyclooxygenase-2 and platelet 12-lipoxygenase with risk of occurrence and advanced disease status of colorectal cancer.** *Carcinogenesis* 2007, **28(6)**:1197-1201.
 44. MacPhee M, Chepenik KP, Liddell RA, Nelson KK, Siracusa LD, Buchberg AM: **The secretory phospholipase A2 gene is a candidate for the Mom1 locus, a major modifier of ApcMin-induced intestinal neoplasia.** *Cell* 1995, **81(6)**:957-966.
 45. Nakanishi M, Rosenberg DW: **Roles of cPLA2alpha and arachidonic acid in cancer.** *Biochim Biophys Acta* 2006, **1761(11)**:1335-1343.
 46. Auwerx J: **Nuclear receptors. I. PPAR gamma in the gastrointestinal tract: gain or pain?** *Am J Physiol Gastrointest Liver Physiol* 2002, **282(4)**:G581-585.
 47. Michalik L, Desvergne B, Wahli W: **Peroxisome-proliferator-activated receptors and cancers: complex stories.** *Nat Rev Cancer* 2004, **4(1)**:61-70.
 48. Siezen CL, van Leeuwen AI, Kram NR, Luken ME, van Kranen HJ, Kampman E: **Colorectal adenoma risk is modified by the interplay between polymorphisms in arachidonic acid pathway genes and fish consumption.** *Carcinogenesis* 2005, **26(2)**:449-457.
 49. Gong Z, Xie D, Deng Z, Bostick RM, Muga SJ, Hurley TG, Hebert JR: **The PPAR(gamma) Pro12Aa polymorphism and risk for incident sporadic colorectal adenomas.** *Carcinogenesis* 2005, **26(3)**:579-585.
 50. Theodoropoulos G, Papaconstantinou I, Felekouras E, Nikiteas N, Karakitsos P, Panoussopoulos D, Lazaris A, Patsouris E, Bramis J, Gazouli M: **Relation between common polymorphisms in genes related to inflammatory response and colorectal cancer.** *World J Gastroenterol* 2006, **12(31)**:5037-5043.
 51. Gunter MJ, Canzian F, Landi S, Chanock SJ, Sinha R, Rothman N: **Inflammation-related gene polymorphisms and colorectal adenoma.** *Cancer Epidemiol Biomarkers Prev* 2006, **15(6)**:1126-1131.
 52. Lee WP, Tai DI, Lan KH, Li AF, Hsu HC, Lin EJ, Lin YP, Sheu ML, Li CP, Chang FY, Chao Y, Yen SH, Lee SD: **The -251T allele of the interleukin-8 promoter is associated with increased risk of gastric carcinoma featuring diffuse-type histopathology in Chinese population.** *Clin Cancer Res* 2005, **11(18)**:6431-6441.
 53. Ulrich CM, Potter JD: **Folate and cancer - timing is everything.** *Jama* 2007, **297(21)**:2408-2409.
 54. Huang Y, Han S, Li Y, Mao Y, Xie Y: **Different roles of MTHFR C677T and A1298C polymorphisms in colorectal adenoma and colorectal cancer: a meta-analysis.** *J Hum Genet* 2007, **52(1)**:73-85.
 55. Hubner RA, Houlston RS: **MTHFR C677T and colorectal cancer risk: A meta-analysis of 25 populations.** *Int J Cancer* 2007, **120(5)**:1027-1035.
 56. Mu LN, Cao W, Zhang ZF, Cai L, Jiang QW, You NC, Goldstein BY, Wei GR, Chen CW, Lu QY, Zhou XF, Ding BG, Chang J, Yu SZ: **Methylenetetrahydrofolate reductase (MTHFR) C677T and A1298C polymorphisms and the risk of primary hepatocellular carcinoma (HCC) in a Chinese population.** *Cancer Causes Control* 2007, **18(6)**:665-675.
 57. Araki A, Hosoi T, Orimo H, Ito H: **Association of plasma homocysteine with serum interleukin-6 and C-peptide levels in patients with type 2 diabetes.** *Metabolism* 2005, **54(6)**:809-814.
 58. Dedoussis GV, Panagiotakos DB, Pitsavos C, Chrysohooou C, Skoumas J, Choumerianou D, Stefanadis C: **An association between the methylenetetrahydrofolate reductase (MTHFR) C677T mutation and inflammation markers related to cardiovascular disease.** *Int J Cardiol* 2005, **100(3)**:409-414.

BMC Cancer 2008, 8:326

<http://www.biomedcentral.com/1471-2407/8/326>

59. Chanock SJ, Manolio T, Boehnke M, Boerwinkle E, Hunter DJ, Thomas G, Hirschhorn JN, Abecasis G, Altshuler D, Bailey-Wilson JE, Brooks LD, Cardon LR, Daly M, Donnelly P, Fraumeni JF Jr, Freimer NB, Gerhard DS, Gunter C, Guttmacher AE, Guyer MS, Harris EL, Hoh J, Hoover R, Kong CA, Merikangas KR, Morton CC, Palmer LJ, Phimister EG, Rice JP, Roberts J, et al.: **Replicating genotype-phenotype associations.** *Nature* 2007, **447(7145)**:655-660.
60. Ioannidis JP, Ntzani EE, Trikalinos TA, Contopoulos-Ioannidis DG: **Replication validity of genetic association studies.** *Nat Genet* 2001, **29(3)**:306-309.

Pre-publication history

The pre-publication history for this paper can be accessed here:

<http://www.biomedcentral.com/1471-2407/8/326/prepub>

Publish with **Bio Med Central** and every scientist can read your work free of charge

BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime.

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp



II.3.1. Supplementary data

II.3.1.1. Additional file 1: Supplemental materials and methods.

Genotype analysis

Determination of genotypes by 5' exonuclease assay

Genomic DNA was extracted from collected blood samples using the Nucleon BACC2 kit (GE Healthcare, Uppsala, Sweden). DNA concentrations were calculated using Picogreen technology (Molecular Probes, Eugene, OR) and diluted to 10 ng/μl in 96-well format following a customized automated procedure developed by Microlab Star (Hamilton Robotics, Reno, NV).

Every study participant was genotyped for 46 SNPs of 29 selected genes (Table 1; and Supplemental Table 1). All these SNPs were selected for their relevance to sporadic CRC, according to a review of the literature performed through Pubmed. Genotypes were determined using high-throughput TaqMan allelic discrimination tests. Primers and dye-labeled MGB-NFQ probes were designed and synthesized by Applied Biosystems (Foster City, CA; Supplemental Table 1). Reactions were set up in a 384-well plate in a 6 μl final volume, including: 2.25 μl of 2X Universal Master Mix, 0.11 μl of 40X Assay Mix, 1.64 μl of water, and 20 ng of genomic DNA (10 ng/μl). Reaction plates were thermocycled in an i-cycler (Bio-Rad Laboratories, Hercules, CA): an initial 10 minutes denaturation at 95°C was followed by 45 cycles including a denaturation step at 92°C during 15 s and an annealing/extension step at 60°C during 1 min. After amplification, end-point fluorescence readings were conducted on an Applied Biosystems ABI 7900HT sequence detection system. Genotypes were assigned using the allelic discrimination software SDS v2.1 (Applied Biosystems, Foster City, CA).

Determination of genotypes by fluorescent multiplex PCR

A first multiplex PCR was used to coamplify four complex polymorphisms: a deletion of 3 bp in *GSTM3*, a deletion of 6 bp in *TYMS*, a CA repeat in *IGF1*, and an insertion of 68 bp in *CBS* (primers sequences are detailed in Supplemental Table 1); reverse primers of the four primer sets used for the multiplex PCR were labelled at their 5' end with FAM. PCR were performed in a final volume of 30 μl including components all provided by the same manufacturer (Eurogentec, Angers, France): 3 μl of 10X mix PCR buffer, 4 mM MgCl₂, 200 μM of dNTPs, 0.2 μM of each primer, 1.5 U of HotGoldstar Taq Polymerase and 60 ng of genomic DNA. PCR were performed on a MJ Research PTC-200 (GMI, Minnesota, USA) and consisted of an initial denaturation step at 94°C for 10 mins, followed by 30 cycles of 94°C for 30 s, 64°C for 15 s, and 72°C for 20s, and ended by a 10 mins extension at 72°C. Electrophoresis was performed on a 3130 capillary sequencer, and data were analyzed using the Genescan 3.1 and genotyper 2.5.2 softwares (Applied biosystems, Foster City, USA).

In a second PCR, primers sets were designed to perform a quantitative multiplex PCR of short fluorescent fragments (QMPSF), in order to test whole gene deletion polymorphism of *GSTM1* and *GSTT1* (Supplemental Table 1). Two primers sets were designed within genes *GSTM1* and *GSTT1* - adapting primers designed by Arand *et al.* (1996)-, and a third set of primers was designed to coamplify an additional fragment corresponding to the exon 14 of the *MLH1* gene, used as a control as defined by Charbonnier *et al.* (2000). A universal extension (5'-CGTTAGATAG-3') was added at the 5' end of each forward primer, and another universal extension (5'-GATAGGGTTAG-3') was added to each reverse primer. PCR were performed in a final volume of 30 µl including components all provided by the same manufacturer (Eurogentec, Angers, France): 3 µl of 10X mix PCR buffer, 3mM MgCl₂, 200 µM of dNTPs, 0.24 µM of each primer, 1.5 U of HotGoldstar Taq Polymerase and 100 ng of genomic DNA. PCR were performed on a MJ Research PTC-200 (GMI, Minnesota, USA) and consisted of an initial denaturation step at 94°C for 10 mins, followed by 22 cycles of 94°C for 1 min, 60°C for 1 min, and 72°C for 1 min, and ended by a 10 mins extension at 72°C. Electrophoresis was performed on a 3130 capillary sequencer, and data were analyzed using the Genescan 3.1 and genotyper 2.5.2 softwares (Applied biosystems, Foster City, USA). Peak heights of the amplicon obtained from each patient were normalized with the reference gene (*MLH1* gene) and then compared with those generated from a normal control. A two-fold reduction in the height of a peak indicated a heterozygous deletion of the gene *GSTM1* or *GSTT1*.

Inference of population structure from genetic data

Stratifications of the cohorts were searched using Structure 2.1 (DAVIES *et al.* 2002), according to the genotype data get for the 52 allelic variants studied. Monomorphic polymorphisms were excluded from the analyses and, following the authors' recommendations to avoid false stratifications, polymorphisms physically too close each others were not included in the analyses. As a consequence, only one allelic variant was analyzed per gene, which led to a global analysis performed on 35 polymorphisms out of 52 (*ALOX5* c.760G>A, *PLA2G2A* c.435+230C>T, *PDL2* c.1731C>T, *PTGS1* c.639C>A, *PTGS2* c.-646C>T, *PPARG* c.36C>G, *ALOX12* c.782G>A, *CYP1A2* c.-163A>C, *CYP2E1* c.-1053C>T, *CYP1B1* c.1294C>G, *CYP2C9* c.430C>T, *EPHX1* c.337T>C, *GSTA1* c.-4605G>A, *GSTM1* null, *GSTM3* c.468+21delAGG, *GSTP1* c.313A>G, *GSTT1* null, *NQO1* c.559C>T, *SULT1A2* c.714A>C, *UGT1A1* c.-3156G>A, *UGT1A6* c.541A>G, *GH1* c.456+90T>A, *IGF1* c.-1006CA(19), *IGFBP3* c.-336A>C, *IL6* c.-237G>C, *IL8* c.-352T>A, *IRS1* c.2911G>A, *VDR* c.1024+283G>A, *CBS* c.844ins68, *MTHFD1* c.1958G>A, *MTHFR* c.1286A>C, *MTR* c.2756A>G, *MTRR* c.66A>G, *TYMS* c.943+447delTTAAAG, *OGG1* c.977C>G).

We used both admixture and no admixture models, assuming a number of population K=2 or K=3, and running 100,000 iterations, following a burn-in period of 100,000 iterations.

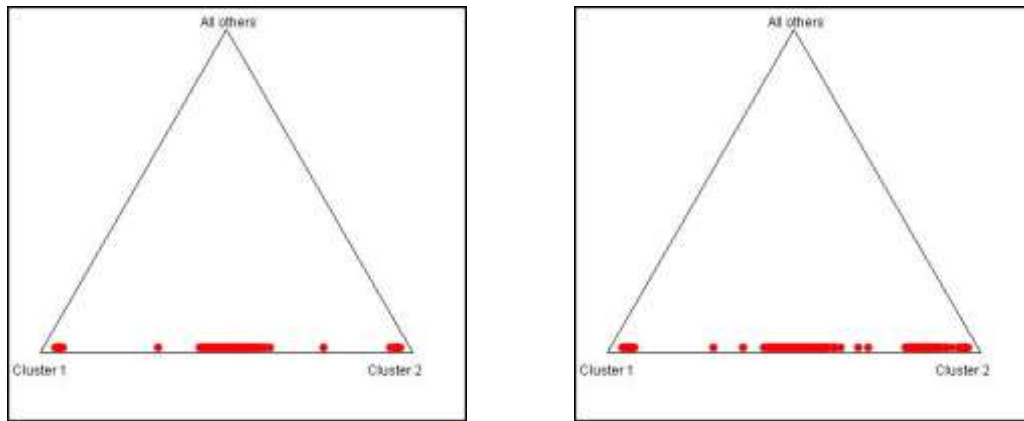


Figure a. Structure analyses assuming a number of population $K=2$, and running 100,000 iterations, following a burn-in period of 100,000 iterations. On the left, we used an admixture model. On the right, we used a no admixture model.

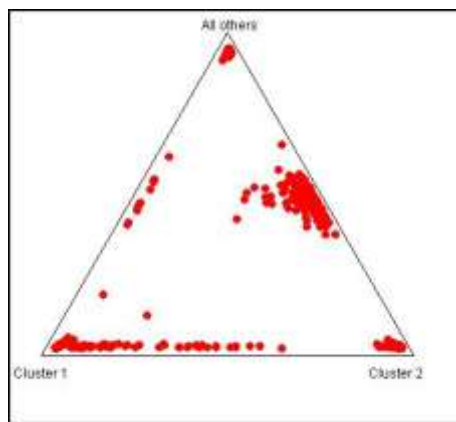


Figure b. Structure analyses using a no admixture model, assuming a number of population $K=3$, and running 100,000 iterations, following a burn-in period of 100,000 iterations.

At least three stratifications were predicted by the software. Analyses of the main effects of the 52 polymorphisms were done, using the three populations as confounder effects. No difference was observed compared to analyses adjusted according to sex and age; the six polymorphisms-CRC associations remained unchanged.

Since the present analysis of stratification was merely exploratory and not confirmatory, these results were not taken into account in the analysis of the main effects. Obviously, the number of polymorphisms included in the present analyses remained too weak to enable a robust stratification, since such analyses would rather require hundreds or even thousands of polymorphisms. However, these stratifications were used as a basis for further analyses on analyses of CRC risk associated with genotypic combinations, i.e., for the examination of internal consistency for multiple-SNP analyses (Supplemental Table 2). It appeared that the CRC risk observed did not significantly differ from a stratification to another, indicating that the genetic differences between the stratifications determined by Structure were whether artificial, or did not rely on the five SNPs composing the genotypic combinations studied.

Statistical analysis

Analysis of gene-environment interactions

Analyses of interactions between allelic variants and environmental factors were performed with SNPStats, which proposes statistical models for calculation of risk associated with one pair of interacting variables. For each polymorphism found to be associated independently with modification of CRC risk in univariate analyses, we investigated all the possible interactions with the environmental co-variables of the questionnaire on life habits completed by the study participants. In addition, we calculated 95% CI ORs of cancer cases to determine whether any “polymorphism-environmental factor” pair could modify the risk of CRC in our groups. A likelihood ratio test was used to investigate interaction among variables of interest. The same analyses were performed to assess possible interactions between environmental factors and genotypic combinations of these polymorphisms.

We assumed that there was an interaction between a genetic factor –either a polymorphism or a genotypic combination- and an environmental factor when the effect on CRC risk was two-sided, which means that both factors modified each other’s intrinsic effect on CRC risk. For the environmental factor, we compared the associated CRC risk calculated within stratifications of the study population determined according to the genotype with the associated risk calculated in the whole study population, which were reported elsewhere (10). Conversely, for the genotype related to a polymorphism or a genotypic combination, we compared the associated CRC risk calculated within stratifications of the study population determined according to the environmental factor to the associated risk calculated in the whole study population and described in Table 2.

For each positive gene-environment interaction we adjusted the p-values obtained for multiple testing, taking into account the twelve environmental risk factors of our questionnaire. Thus, we rejected any results with a p-value above 0.001. We also tested the consistency of the observations in sub-groups of the study population determined according to age, geographical origin, random selection, or inference of population structure (Supplemental Data 1).

II.3.1.2. Additional file 2: List of primers and probes used for the study.

Metabolic or regulation pathway	Gene	Polymorphism tested	Primer/probe Name	Primer and probe sequences 5'>3' (strand)	Design strand	
Acid arachidonic pathway	<i>ALOX5</i>	c.760G>A	F R VIC FAM	TGGTGCAACCCCTGTGTT CAGGCTGCACTCTACCATCTC CTGCCAAGAAGC CTGCCGAGAAAGC	Forward	
	<i>ALOX12</i>	c.782G>A	F R VIC FAM	CTGCCCTCCAGGCTAGTG GAAGGAGAGGTACTGGAAGTTCTTT AGTTGAGCCTGAAGCT TTGAGCCGGAAGCT	Reverse	
	<i>COX1</i>	c.123G>A	F R VIC FAM	CCTCCATTTCTACCCACAGT GCCGAAGCGGACACAGAT TCCATGCCAACACCAG CATGCCAGCACCAG	Forward	
	<i>COX1</i>	c.639C>A	F R VIC FAM	CACCTCACCCACCAGTCTCTCAA GCCTTGGTGAAGCCAGGA AACTTCTGAAAAGATG AACTTCTGGCAAGATG	Forward	
	<i>COX1</i>	c.22C>T	F R VIC FAM	GCTCAGCCCTCATCTCTCT GGAGCAGGAGCAGGAACAG CAAGAACCGGAGCAAG CAAGAACCAGAGCAAG	Reverse	
	<i>COX1</i>	c.50C>T	F R VIC FAM	TCTGCAGGAGTCTCTTGCT GGTCCGCGACAGGA CTGCTCCGCCGCT CTGCTCTGCCGCT	Forward	
	<i>COX2</i>	c.-646C>T	F R VIC FAM	ACATGTCAGCCTTTCTAACCTTACTC GTCTTAGAGGGTCCAGGAAGTC CGTCCGGACAGACT ACGTCGGAACAGACT	Reverse	
	<i>COX2</i>	c.306G>C	F R VIC FAM	TCCCTTCTCGAATGCAATTATGA GCTTGGAAATATGTTTTAGATTAGGCTTACAGTATTA TTACATGTCAACACATAAC ACATGTCAAGACATAAC	Reverse	
	<i>COX2</i>	c.1815+1912A>G	F R VIC FAM	ACTGACATTTAATGGTACTGTATATTACTTAATTTATTGAAGATTATT GTGACATAACAAAAAAATCAATGATTGTAGGC CACTATGGTTATAAACTGTG CACTATGGTTATAAAGCTGTG	Forward	
	<i>COX2</i>	c.1815+427T>C	F R VIC FAM	GCATCTTCCATGATGCATTAGAAGTAAC GGTAATGTCTAATTTAAATATTCATTTAATAATGCACCTGATACC AAGTACTTTGGTTATTTTT ACTTTGGTCATTTT	Forward	
	<i>PLA2G2A</i>	c.-859C>G	F R VIC FAM	CAGCCTTGTCCTCACCTA CAGGCCGTCTGTGTTGTTCTG CTTAAATAGTGCTCCCTC CCTTAAATAGTCTCCTCCTC	Reverse	
	<i>PLA2G2A</i>	c.132C>T	F R VIC FAM	AAAGGAAGCCGCACTCAGTTAT CCACCCACGCCACAGT AGCCGTAGAAGCC CAGCCATAGAAGCC	Reverse	
	<i>PLA2G2A</i>	c.435+230C>T	F R VIC FAM	CAACTCCGTGCTTAAACCAAGAAG GGCAATGCATACACACACATAT CCCCCGAGTACAG CCCCCAGAGTACAG	Reverse	
	<i>PLA2G2A</i>	c.185+88G>A	F R VIC FAM	CTGGGAGCTGCAAAAGACAGT CCAGCCAGGCCATCCT CCTGATGGGCACAGAG CTGATGGGCGCAGAG	Forward	
	<i>PLD2</i>	c.1731C>T	F R VIC FAM	ACCACCAAGGCCAAGTACAAG TGGTAGACTTGGGAAGCAGGTA CTCCACATACCCC ACTCCCATATACCCC	Forward	
	<i>PPARG</i>	c.36C>G	F R VIC FAM	GTTATGGGTGAAACTCTGGGAGATT GCAGACAGTGTACAGTGAAGGAAT CTCCTATTGACCCAGAAAG CTATTGACGCAGAAAG	Forward	
	<i>PPARG</i>	c.1431C>T	F R VIC FAM	CTCCAGAAAATGACAGCCTCAGA GTCTGTCTCCGTCTTCTTGATCAC CTGCAGTGTCCG CTGCACATGTTCCG	Reverse	
	Xenobiotics detoxification pathway	<i>CYP1A2</i>	c.-163A>C §	F R VIC FAM	GAGAGCCAGCGTTTCATGTTG CAGAAAGACTAAGCTCCATCTACCAT CGTCTGTGCCAC CGTCTGGGCCAC	Reverse
		<i>CYP1A2</i>	c.1548T>C §	F R VIC FAM	GCCCCGTGTGAACATGTC TCCCTGGCCTCAGAATGGT TGTCTTCTCAGTTGATGG TGTCTTCTCAATTGATGG	Reverse
		<i>CYP2E1</i>	c.-1293G>C §	F R VIC FAM	GCCAAAGCCCCCTTTG TCATTGGTTGTGCTGCACCTAA CACTGCAGCTCTCT CACTGCACCTCTCT	Reverse
<i>CYP2E1</i>		c.-1053C>T §	F R VIC FAM	CCACAAGTGATTTGGCTGGATTG GTTTTTCATTCTGTCTTCTAACTGGCAAT ATAGGTTGCAATTTTGTACTT TAGGTTGCAATTTTATACTT	Reverse	
<i>CYP1B1</i>		c.1294G>C §	F R VIC FAM	TTTGTCAACCAGTGGTCTGTGAAT ACTCTGCTGGTCAAGTCTCT CATGACCCACTGAAGTG ATGACCCAGTGAAGTG	Forward	

Metabolic or regulation pathway	Gene	Polymorphism tested	Primer/probe Name	Primer and probe sequences 5'>3' (strand)	Design strand	
Metabolic or regulation pathway	<i>CYP2C9</i>	c.430C>T [§]	F R VIC FAM	TCAGCAATGGAAAGAAATGGAAGGA TGTTTTTCTCAACTCTCCACAAG CTTGAACACGGTCTCT CTTGAACACAGTCTCT	Reverse	
	<i>EPHX1</i>	c.337T>C	F R VIC FAM	ND * ND * ND * ND *	Reverse	
	<i>GSTP1</i>	c.313A>G	F R VIC FAM	CCTGGTGGACATGGTGAATGAC TGGTGCAGATGCTCACATAGTTG CTGCAAAATACATCTCC CTGCAAAATACGCTCTCC	Forward	
	<i>GSTP1</i>	c.341C>T	F R VIC FAM	AGTAGGATGATACATGGTGGTGTCT GGCAGTGCCTTACATAGTCAT CTTGCCCGCTCTCT CTTGCCCACTCTCT	Reverse	
	<i>GSTA1</i>	c.-4605G>A	F R VIC FAM	GGAGTGGCTTTTCCCTAACTTGAC CTGTACCCTCTGGCT AGTGGGAGGGAACCTA AGTGGGAGAGAACTA	Forward	
	<i>SULT1A2</i>	c.714A>C	F R VIC FAM	GACTGTGGACCTCATGTTGA GACGGTGGTGTAGTTGGTCAT AGATGAAGAAGAACCCT AGATGAAGAAGACCCT	Forward	
	<i>NQO1</i>	c.559C>T	F R VIC FAM	TGCATTCTGTGGCTTCCAAGT TGGAGTGTGCCAATGCTATATG TCAGTTGAGGTTCTAAG TCAGTTGAGATTCTAAG	Reverse	
	<i>NQO1</i>	c.415C>T	F R VIC FAM	CGCTGCCATGTATGACAAAGGA GCTGTACAGCATTGAGAACCAT CCACTACCCGGAAGG CCACTACCCGGAAGG	Reverse	
	<i>UGT1A1</i>	c.-3156G>A	F R VIC FAM	TCCAGAATGGCTAGAGGTAAGAG CTTGCTCTCAAACCTCTGGGATAGA CCTGTCCAAGCTCA ACCTGTCTAAGCTCA	Reverse	
	<i>UGT1A1</i>	c.-3279G>T	F R VIC FAM	GTAAGCAGCAATGAACAGTCA TGCTGTTCCCAACTCTCTTT TTGTTCACTGAACTC TTTGTCAAACCTGAACTC	Reverse	
	<i>UGT1A6</i>	c.541A>G	F R VIC FAM	CCTACCATCTGTGTACTCTTCAG TGTAGCACCTGGGAATGTAGGA CTGGAGCATACATTC TGGAGCATGCATTC	Forward	
	<i>UGT1A6</i>	c.552A>C	F R VIC FAM	CCGTGTCCCTGGAGCAT CACCTGGGAATGTAGGACACA TTCAGCAGAAGCCAG CAGCAGCAGCCAG	Forward	
	Insulin signaling pathway	<i>IGFBP3</i>	c.-336A>C	F R VIC FAM	GCACACCTGGTTCITGTAGAC GCGTGCAGCTCAGACT CCTCGTGCTCAGGCC TCGTGCGCACGCC	Reverse
		<i>GH1</i>	c.456+90T>A	F R VIC FAM	ACTGACTTTGAGAGCTGTGTTAGAG GAGAAAGCCCTGGAGATTAC TCTTTTAGCAGTCAGGCC TTTTTAGCAGCAGGCC	Forward
		<i>IRS1</i>	c.2911G>A	F R VIC FAM	GGTCGAGATGGCAGACT CCGGTAGGCTGCAAAAT CTAGCAGCCCTGGGAG CTAGCAGCCCGGGAG	Reverse
		<i>IL8</i>	c.-352T>A	F R VIC FAM	GTTCTAAACCTGCCACTCTAGTAC CAAACCTGAGTATCACACTTCTTA AAGCATACAATTGATAATT AGCATACATTTGATAATT	Forward
		<i>IL6</i>	c.-237G>C	F R VIC FAM	GACGACCTAAGCTGCACCTTTTC GGGCTGATTGGAAACCTTATAAGATTG CCTTTAGCATCGCAAGAC CTTTAGCATGGCAAGAC	Reverse
		<i>VDR</i>	c.33061G>A	F R VIC FAM	AGTGTGCAGGCGATTGTA AGCCAGTTCACGCAAGAG CTGCGATTCCCA CCTGCACATTCCCA	Forward
		Monocarboxylate pathway	<i>MTHFD1</i>	c.1958G>A	F R VIC FAM	CATCGCACATGGCAATTCCT TCTGGGCCAACCAAGCTTGAG CAGACCGGATCGCA CAGACCAGATCGCA
	<i>MTHFR</i>		c.665C>T	F R VIC FAM	GCACCTGAAGGAGAAGGTGTCT CCTCAAAGAAAAGCTGCGTGATG ATGAAATCGGCTCCCGC ATGAAATCGACTCCCGC	Reverse
<i>MTHFR</i>	c.1286A>C		F R VIC FAM	GGAGGAGCTGCTGAAGATGTG TGGTTCTCCCGAGAGGTAAGA ACCAGTGAAGAAAGTGT CAGTGAAGCAAGTGT	Forward	
<i>MR</i>	c.2756A>G		F R VIC FAM	GGATGAATACTTTGAGGAAATCATGGAAGA TGTTTCTACCACTTACCTTGAGAGACT ATTAGACAGGACCAATTATG ACAGGGCCATTATG	Forward	
<i>MTRR</i>	c.66A>G		F R VIC FAM	AGCAGGGACAGGCAAAGG AAGATCTGCAGAAAATCCATGTACCA TTGCTCACATATTTCTT CTCACACATTTCTT	Reverse	

Metabolic or regulation pathway	Gene	Polymorphism tested	Primer/probe Name	Primer and probe sequences 5'>3' (strand)	Design strand
Base excision repair (DNA repair)	<i>MUTYH</i>	c.494A>G ^{§§}	F	CCACAGGAGGTGAATCAACTCT	Forward
			R	CCTTACCTCCGAGCTCCCT	
			VIC	CTGGGCTACTATTCT	
	<i>MUTYH</i>	c.1145G>A ^{§§}	FAM	TGGGCTGCTATTCT	Forward
			F	GTAACAAGAGAGAATGGAGGAATCG	
			R	GACGGAACTCCACAGT	
	<i>MUTYH</i>	c.1435G>T ^{§§}	VIC	CCTCTCAGGTCTGCTG	Forward
			FAM	CCTCTCAGATCTGCTG	
			F	GCCTCACCTCCCTGCT	
	<i>MUTYH</i>	c.694G>T ^{§§}	R	GCTGTTGGCCCTGATACACA	Forward
			VIC	CTTGCTAGGTTTTCC	
			FAM	CTTGCTAGTTTTCC	
	<i>MUTYH</i>	c.1269G>A ^{§§}	F	GGTGGATGGCAACGTAGCA	Forward
			R	GGGTGCTGTGGGATCAG	
			VIC	CTGTGCCGTGCCGA	
<i>MUTYH</i>	c.1269G>A ^{§§}	FAM	CTGTGCCGTTTTCCGA	Forward	
		F	TGCAAGAACTACAGCGTTGG		
		R	CTATCCGCTGCTCACTTACCT		
<i>OGG1</i>	C.977C>G	VIC	ACCTCCGACCTT	Forward	
		FAM	ACCTCCGACCTT		
		F	ND **		
			R	ND **	
			VIC	ND **	
			FAM	ND **	

Additional file 2a: Primers used for the study 5' exonuclease assays (TaqMan probes). [§] SNP already tested previously (KÜRY *et al.* 2007b); ^{§§} mutation already tested previously (KÜRY *et al.* 2007a); * sequence not provided by the manufacturer : cf. Drug Metabolism Genotyping Assays C___14938_30; ** sequence not provided by the manufacturer : cf. TaqMan SNP Genotyping Assay C___3095552_1_.

Metabolic or regulation pathway	Gene	Polymorphism tested	Localization	Primer sequence 5'>3'	Design strand	PCR product length	
Xenobiotics	<i>GSTM3</i>	3 bp deletion	Intron 6	ACCTGGACAACGAAACAATC	Forward	168 bp	Multiplex PCR 1
				FAM --AATCCACAAAGGTGAGCTAGAAG	Reverse		
Monocarbonate	<i>TYMS</i>	6 bp deletion	3' UTR	CAATGCTGAGGTTATGAACAAAG	Forward	386 bp	Multiplex PCR 1
				FAM --GCATCCAAACCAAGAATACAGC	Reverse		
Insulin signaling	<i>IGF1</i>	CA repeat	5'UTR	CCAGTTGAGGGATTGAATGAC	Forward	241 bp	Multiplex PCR 1
				FAM --CCACCACTCTGGGAGAAGG	Reverse		
Monocarbonate	<i>CBS</i>	68 bp insertion	Exon 8	GGTTTCTCATCCTGCCTCTG	Forward	266 bp	Multiplex PCR 1
				FAM --AAGTCGTAGCCGATCCCTTC	Reverse		
Xenobiotics	<i>GSTM1</i>	whole gene deletion	Exons 6-7	cgtagatagGAATCCCTGAAAAGCTAAAGC	Forward	240 bp	Multiplex PCR 2
				FAM --gatagggttagTCACCGGATCATGCCAGCA	Reverse		
Xenobiotics	<i>GSTT1</i>	whole gene deletion	Exons 4-5	cgtagatagTTCCTTACTGGTCTCACATCTC	Forward	480 bp	Multiplex PCR 2
				FAM --gatagggttagTCACCGGATCATGCCAGCA	Reverse		
Base excision repair	<i>MLH1</i>	68 bp insertion	Exon 14	cgtagatagATTTTTGTTTTGCAGTTCTCC	Forward	181 bp	Multiplex PCR 2
				FAM --gatagggttagTAGTAGCTCTGCTTGTTCACACA	Reverse		

Additional file 2b: Primers used for genotyping of complex polymorphisms by quantitative multiplex PCR of short fluorescent fragments (QMPSF).

II.3.1.3. Additional file 3: Statistical analysis of association between CRC risk and polymorphisms selected for the study using conditional logistic regression (number of age- and sex-matched pairs = 811, i.e., 1622 individuals).

	Gene Name	Nucleotide change **	Protein change **	SNP ID **	Ref ***	Controls			Cases			+/- vs. +/+			-/- vs. +/-			+/- and -/- vs. +/-		
						+/+	+/-	-/-	+/+	+/-	-/-	OR (95%CI)	p	q	OR (95% CI)	p	q	OR (95% CI)	p	q
Inflammation	ALOX5	c.760G>A	p.Glu254Lys	rs2228065	16	811 (100%)	0 (0%)	0 (0%)	811 (100%)	0 (0%)	0 (0%)	Rare			Rare			Rare		
	ALOX12	c.782G>A	p.Gln261Arg	rs1126667	16	257 (31.7%)	395 (48.7%)	159 (19.6%)	281 (34.6%)	400 (49.3%)	130 (16%)	0.93 (0.74-1.15)	0.49	0.88	0.75 (0.56-0.99)	0.047	0.40	0.87 (0.71-1.08)	0.21	0.66
	IL6	c.-239G>C		rs1800795	17	315 (38.8%)	372 (45.9%)	124 (15.3%)	284 (35%)	388 (47.8%)	139 (17.1%)	1.16 (0.93-1.43)	0.18	0.63	1.24 (0.93-1.66)	0.14	0.61	1.18 (0.96-1.44)	0.11	0.59
	IL8	c.-352T>A		rs4073	17	275 (33.9%)	368 (45.4%)	168 (20.7%)	220 (27.1%)	417 (51.4%)	174 (21.4%)	1.42 (1.13-1.78)	0.002	0.10	1.29 (0.98-1.71)	0.07	0.46	1.38 (1.11-1.70)	0.003	0.11
	PLA2G2A	c.-859C>G		rs11573156	d	537 (66.2%)	244 (30.1%)	30 (3.7%)	498 (61.4%)	280 (34.5%)	33 (4.1%)	1.24 (1.00-1.53)	0.05	0.40	1.19 (0.71-1.97)	0.51	0.89	1.23 (1.01-1.51)	0.04	0.40
	PLA2G2A	c.132C>T	p.Tyr44Tyr	rs4744	d	541 (66.7%)	240 (29.6%)	30 (3.7%)	503 (62%)	275 (33.9%)	33 (4.1%)	1.23 (1.00-1.52)	0.05	0.40	1.18 (0.71-1.97)	0.52	0.89	1.23 (1.00-1.50)	0.05	0.40
	PLA2G2A	c.185+88G>A		rs2236772	d	761 (93.8%)	50 (6.2%)	0 (0%)	747 (92.1%)	62 (7.6%)	2 (0.2%)	1.26 (0.86-1.86)	0.23	0.66	Rare			1.30 (0.89-1.91)	0.17	0.63
	PLA2G2A	c.435+230C>T		rs11677	d	520 (64.1%)	275 (33.9%)	16 (2%)	587 (72.4%)	216 (26.6%)	8 (1%)	0.70 (0.56-0.86)	0.001	0.07	0.44 (0.19-1.04)	0.06	0.43	0.68 (0.55-0.84)	0.0003	0.04
	PLD2	c.1731C>T	p.Thr577Ile	rs1052748	18	223 (27.5%)	403 (49.7%)	185 (22.8%)	220 (27.1%)	416 (51.3%)	175 (21.6%)	1.05 (0.83-1.32)	0.70	0.97	0.96 (0.73-1.27)	0.77	0.97	1.02 (0.82-1.27)	0.87	0.98
	PTGS1	c.22C>T	p.Trp8Arg	rs1236913	d,32,34	708 (87.3%)	98 (12.1%)	5 (0.6%)	689 (85%)	122 (15%)	0 (0%)	1.28 (0.96-1.70)	0.09	0.52	Rare			1.22 (0.92-1.61)	0.17	0.63
	PTGS1	c.50C>T	p.Pro17Leu	rs3842787	d,32,34	700 (86.3%)	109 (13.4%)	2 (0.2%)	695 (85.7%)	109 (13.4%)	7 (0.9%)	1.01 (0.76-1.34)	0.96	0.10	3.53 (0.73-17.01)	0.09	0.52	1.05 (0.80-1.39)	0.72	0.97
	PTGS1	c.639C>A	p.Gly213Gly	rs5788	d,32	618 (76.2%)	177 (21.8%)	16 (2%)	577 (71.2%)	216 (26.6%)	18 (2.2%)	1.31 (1.04-1.64)	0.02	0.26	1.20 (0.61-2.39)	0.59	0.89	1.30 (1.04-1.62)	0.02	0.26
	PTGS1	c.123G>A	p.Gln41Gln	rs3842788	d,32	778 (95.9%)	33 (4.1%)	0 (0%)	772 (95.2%)	39 (4.8%)	0 (0%)	1.19 (0.74-1.91)	0.47	0.88	Rare			Rare		
	PTGS2	c.-646C>T		rs20420	d,16	811 (100%)	0 (0%)	0 (0%)	811 (100%)	0 (0%)	0 (0%)	Rare			Rare			Rare		
	PTGS2	c.306G>C	p.Val102Val	rs5277	d,31	563 (69.4%)	227 (28%)	21 (2.6%)	560 (69%)	227 (28%)	24 (3%)	1.01 (0.81-1.25)	0.96	0.99	1.15 (0.63-2.09)	0.65	0.94	1.02 (0.82-1.26)	0.87	0.99
	PTGS2	c.1815+427T>C		rs5275	d,31	770 (94.9%)	41 (5.1%)	0 (0%)	757 (93.3%)	54 (6.7%)	0 (0%)	1.34 (0.88-2.04)	0.17	0.63	Rare			Rare		
	PTGS2	c.1815+1912A>G		rs4648298	d,31	361 (44.5%)	354 (43.6%)	96 (11.8%)	340 (41.9%)	386 (47.6%)	85 (10.5%)	1.16 (0.94-1.42)	0.16	0.63	0.94 (0.68-1.30)	0.71	0.97	1.11 (0.91-1.35)	0.29	0.77
	PPARG	c.36C>G	p.Pro12Ala	rs1801282	17	643 (79.3%)	155 (19.1%)	13 (1.6%)	633 (78%)	172 (21.2%)	6 (0.7%)	1.13 (0.88-1.44)	0.33	0.79	0.47 (0.18-1.24)	0.12	0.60	1.08 (0.85-1.37)	0.54	0.88
PPARG	c.1431C>T	p.His477His	rs3856806	33	623 (76.8%)	164 (20.2%)	24 (3%)	611 (75.3%)	192 (23.7%)	8 (1%)	1.19 (0.94-1.51)	0.14	0.61	0.34 (0.15-0.76)	0.01	0.24	1.08 (0.86-1.36)	0.48	0.88	
Xenobiotics detoxification	CYP1A2	c.-163A>C		rs762551	19,20,21	405 (49.9%)	346 (42.7%)	60 (7.4%)	398 (49.1%)	345 (42.5%)	68 (8.4%)	1.01 (0.83-1.24)	0.89	0.99	1.15 (0.79-1.68)	0.45	0.88	1.04 (0.85-1.26)	0.73	0.96
	CYP1A2	c.1548T>C	p.Asn516Asn	rs2470890	20,21	335 (41.3%)	381 (47%)	95 (11.7%)	331 (40.8%)	380 (46.9%)	100 (12.3%)	1.01 (0.82-1.24)	0.93	0.99	1.07 (0.77-1.47)	0.70	0.97	1.02 (0.84-1.24)	0.84	0.99
	CYP2E1	c.-1293G>C		rs3813867	20	747 (92.1%)	63 (7.8%)	1 (0.1%)	753 (92.8%)	57 (7%)	1 (0.1%)	0.90 (0.62-1.30)	0.57	0.89	0.99 (0.06-15.89)	1.00	1.00	0.90 (0.62-1.30)	0.57	0.89
	CYP2E1	c.-1053C>T		rs2031920	20	745 (91.9%)	65 (8%)	1 (0.1%)	750 (92.5%)	57 (7%)	4 (0.5%)	0.87 (0.60-1.26)	0.46	0.88	3.97 (0.44-35.60)	0.18	0.63	0.92 (0.64-1.32)	0.64	0.94
	CYP1B1	c.1294C>G	p.Leu432Val	rs1056836	19,20,21	272 (33.5%)	426 (52.5%)	113 (13.9%)	255 (31.4%)	403 (49.7%)	153 (18.9%)	1.01 (0.81-1.26)	0.94	0.99	1.44 (1.07-1.94)	0.02	0.26	1.10 (0.89-1.35)	0.37	0.79

Gene Name	Nucleotidic change **	Proteic change **	SNP ID **	Ref ***	Controls			Cases			+/- vs. +/-			-/- vs. +/-			+/- and -/- vs. +/-			
					+/+	+/-	-/-	+/+	+/-	-/-	OR (95%CI)	p	q	OR (95% CI)	p	q	OR (95% CI)	p	q	
<i>CYP2C9</i>	c.430C>T	p.Arg144Cys	rs1799853	19,20,21	594 (73.2%)	202 (24.9%)	15 (1.8%)	598 (73.7%)	196 (24.2%)	17 (2.1%)	0.96 (0.77-1.21)	0.75	0.97	1.13 (0.56-2.28)	0.74	0.97	0.98 (0.78-1.22)	0.82	0.99	
<i>EPHX1</i>	c.337T>C	p.Tyr113His	rs1051740	19,20,21	405 (49.9%)	345 (42.5%)	61 (7.5%)	422 (52%)	314 (38.7%)	75 (9.2%)	0.87 (0.71-1.07)	0.20	0.64	1.18 (0.82-1.70)	0.37	0.79	0.92 (0.76-1.12)	0.40	0.80	
<i>GSTA1</i>	c.-4605G>A		rs3957356	22	264 (32.5%)	398 (49.1%)	149 (18.4%)	265 (32.7%)	408 (50.3%)	138 (17%)	1.02 (0.82-1.27)	0.85	0.99	0.92 (0.69-1.23)	0.58	0.89	0.99 (0.81-1.22)	0.96	0.99	
<i>GSTM1</i>	null			19,20,21	406 (50.1%)	289 (35.6%)	116 (14.3%)	432 (53.3%)	245 (30.2%)	134 (16.5%)	0.80 (0.64-0.99)	0.04	0.40	1.09 (0.82-1.44)	0.57	0.89	0.88 (0.72-1.07)	0.20	0.64	
<i>GSTM3</i>	c.468+21delAGG			19,2	561 (69.2%)	227 (28%)	23 (2.8%)	541 (66.7%)	249 (30.7%)	21 (2.6%)	1.14 (0.92-1.41)	0.24	0.66	0.95 (0.52-1.73)	0.86	0.99	1.12 (0.91-1.38)	0.29	0.77	
<i>GSTP1</i>	c.313A>G	p.Ile105Val	rs1695	19,20,21	367 (45.2%)	343 (42.3%)	101 (12.4%)	368 (45.4%)	355 (43.8%)	88 (10.8%)	1.03 (0.84-1.27)	0.76	0.97	0.87 (0.63-1.20)	0.39	0.80	1.00 (0.82-1.21)	0.96	0.99	
<i>GSTP1</i>	c.341C>T	p.Ala114Val	rs1138272	19,20,21	698 (86.1%)	106 (13.1%)	7 (0.9%)	706 (87%)	101 (12.4%)	4 (0.5%)	0.94 (0.70-1.26)	0.69	0.97	0.56 (0.16-1.94)	0.36	0.80	0.92 (0.69-1.22)	0.56	0.89	
<i>GSTT1</i>	null			19,20,21	308 (38%)	353 (43.5%)	150 (18.5%)	320 (39.5%)	339 (41.8%)	152 (18.7%)	0.92 (0.74-1.15)	0.48	0.88	0.98 (0.74-1.28)	0.86	0.99	0.94 (0.77-1.15)	0.54	0.89	
<i>NQO1</i>	c.415C>T	p.Arg139Trp	rs4986998	20,21	782 (96.4%)	29 (3.6%)	0 (0%)	792 (97.7%)	19 (2.3%)	0 (0%)	0.65 (0.36-1.16)	0.14	0.61	Rare			Rare			
<i>NQO1</i>	c.559C>T	p.Pro187Ser	rs1800566	19,21	523 (64.5%)	251 (30.9%)	37 (4.6%)	505 (62.3%)	270 (33.3%)	36 (4.4%)	1.11 (0.90-1.38)	0.32	0.79	1.01 (0.63-1.62)	0.97	1.00	1.10 (0.90-1.35)	0.35	0.80	
<i>SULT1A2</i>	c.714A>C	p.Asn235Thr	rs1059491	21	382 (47.1%)	344 (42.4%)	85 (10.5%)	372 (45.9%)	356 (43.9%)	83 (10.2%)	1.06 (0.87-1.31)	0.56	0.89	1.00 (0.72-1.40)	0.99	1.00	1.05 (0.86-1.28)	0.62	0.93	
<i>UGT1A1</i>	c.-3279T>G		rs4124874	23	384 (47.4%)	351 (43.3%)	76 (9.4%)	375 (46.2%)	364 (44.9%)	72 (8.9%)	1.06 (0.87-1.30)	0.56	0.89	0.97 (0.68-1.38)	0.87	0.99	1.05 (0.86-1.27)	0.65	0.94	
<i>UGT1A1</i>	c.-3156G>A		rs10929302	23	263 (32.4%)	382 (47.1%)	166 (20.5%)	244 (30.1%)	407 (50.2%)	160 (19.7%)	1.15 (0.92-1.44)	0.22	0.66	1.04 (0.79-1.37)	0.79	0.99	1.12 (0.90-1.38)	0.31	0.79	
<i>UGT1A6</i>	c.541A>G	p.Thr181Ala	rs2070959	22	371 (45.8%)	359 (44.3%)	81 (10%)	375 (46.2%)	346 (42.7%)	90 (11.1%)	0.95 (0.78-1.17)	0.65	0.94	1.10 (0.79-1.53)	0.58	0.89	0.98 (0.81-1.19)	0.84	0.99	
<i>UGT1A6</i>	c.552A>C	p.Arg184Ser	rs1105879	22	337 (41.5%)	375 (46.2%)	99 (12.2%)	348 (42.9%)	359 (44.3%)	104 (12.8%)	0.93 (0.75-1.14)	0.48	0.88	1.02 (0.74-1.39)	0.91	0.99	0.95 (0.78-1.15)	0.58	0.89	
<i>Insulin</i>	<i>GH1</i>	c.456+90T>A	rs2665802	24	242 (29.8%)	398 (49.1%)	171 (21.1%)	245 (30.2%)	417 (51.4%)	149 (18.4%)	1.03 (0.83-1.30)	0.76	0.97	0.86 (0.65-1.14)	0.30	0.78	0.98 (0.79-1.22)	0.87	0.98	
	<i>IGF1</i>	c.-1006CA(19)		25	550 (67.8%)	109 (13.4%)	152 (18.7%)	548 (67.6%)	120 (14.8%)	143 (17.6%)	1.10 (0.83-1.47)	0.49	0.88	0.94 (0.73-1.22)	0.66	0.94	1.01 (0.82-1.25)	0.92	0.99	
	<i>IGFBP3</i>	c.-336A>C	rs2854744	25	189 (23.3%)	414 (51%)	208 (25.6%)	222 (27.4%)	390 (48.1%)	199 (24.5%)	0.80 (0.63-1.02)	0.07	0.46	0.81 (0.62-1.07)	0.14	0.61	0.81 (0.64-1.01)	0.06	0.43	
	<i>IRS1</i>	c.2911G>A	p.Gly971Arg	rs1801278	25	687 (84.7%)	118 (14.6%)	6 (0.7%)	687 (84.7%)	122 (15%)	2 (0.2%)	1.03 (0.79-1.36)	0.81	0.99	0.33 (0.07-1.65)	0.16	0.63	1.00 (0.76-1.31)	1.00	1.00
	<i>VDR</i>	c.1024+283G>A		rs1544410	25	313 (38.6%)	376 (46.4%)	122 (15%)	288 (35.5%)	394 (48.6%)	129 (15.9%)	1.14 (0.92-1.41)	0.23	0.66	1.15 (0.86-1.54)	0.36	0.79	1.14 (0.93-1.40)	0.20	0.64
<i>One-carbon</i>	<i>CBS</i>	c.844ins68		26,27	675 (83.2%)	129 (15.9%)	7 (0.9%)	662 (81.6%)	143 (17.6%)	6 (0.7%)	1.13 (0.87-1.47)	0.36	0.79	0.87 (0.29-2.61)	0.81	0.99	1.12 (0.86-1.44)	0.40	0.80	
	<i>MTHFD1</i>	c.1958G>A	p.Arg653Gln	rs2236225	28	249 (30.7%)	400 (49.3%)	162 (20%)	249 (30.7%)	398 (49.1%)	164 (20.2%)	1.00 (0.80-1.24)	0.96	0.99	1.01 (0.77-1.34)	0.93	0.99	1.00 (0.81-1.23)	1.00	1.00
	<i>MTHFR</i>	c.665C>T	p.Ala222Val	rs1801133	26,27	327 (40.3%)	372 (45.9%)	112 (13.8%)	347 (42.8%)	360 (44.4%)	104 (12.8%)	0.91 (0.74-1.12)	0.39	0.80	0.88 (0.64-1.19)	0.39	0.80	0.90 (0.74-1.10)	0.31	0.783
	<i>MTHFR</i>	c.1286A>C	p.Ala429Glu	rs1801131	26,27	425 (52.4%)	317 (39.1%)	69 (8.5%)	368 (45.4%)	354 (43.6%)	89 (11%)	1.29 (1.05-1.58)	0.02	0.26	1.49 (1.06-2.10)	0.02	0.26	1.33 (1.09-1.61)	0.005	0.14
	<i>MTR</i>	c.2756A>G	p.Asp919Gly	rs1805087	26,27	537 (66.2%)	242 (29.8%)	32 (4%)	554 (68.3%)	234 (28.9%)	23 (2.8%)	0.94 (0.76-1.16)	0.56	0.89	0.70 (0.40-1.21)	0.19	0.64	0.91 (0.74-1.12)	0.37	0.79
	<i>MTRR</i>	c.66A>G	p.Ile22Met	rs1801394	26,27	230 (28.4%)	403 (49.7%)	178 (21.9%)	228 (28.1%)	415 (51.2%)	168 (20.7%)	1.04 (0.83-1.31)	0.74	0.97	0.95 (0.72-1.26)	0.73	0.97	1.01 (0.82-1.26)	0.91	0.99
	<i>TYMS</i>	c.943+447del TTAAAG			27	394 (48.6%)	337 (41.5%)	80 (9.9%)	359 (44.3%)	363 (44.8%)	89 (11%)	1.18 (0.96-1.45)	0.11	0.59	1.22 (0.87-1.71)	0.24	0.66	1.19 (0.98-1.45)	0.08	0.50

Gene Name	Nucleotidic change **	Proteic change **	SNP ID **	Ref ***	Controls			Cases			+/- vs. +/+			-/- vs. +/-			+/- and -/- vs. +/+		
					+/+	+/-	-/-	+/+	+/-	-/-	OR (95%CI)	p	q	OR (95% CI)	p	q	OR (95% CI)	p	q
* OGG1	c.977C>G	p.Cys326Ser	rs1052133	29	489 (60.3%)	289 (35.6%)	33 (4.1%)	512 (63.1%)	256 (31.6%)	43 (5.3%)	0.85 (0.69-1.04)	0.12	0.59	1.24 (0.78-1.99)	0.36	0.79	0.89 (0.73-1.08)	0.24	0.66

Additional file 3 : Statistical analysis of association between CRC risk and polymorphisms selected for the study using conditional logistic regression (number of age- and sex-matched pairs = 811, i.e., 1622 individuals) ; p-values <0.05 and q-values<0.30 are written in bold ; * DNA repair pathway ; SNP OGG1 c.977C>G was chosen to be analyzed together with 6 monoallelic germline mutations of the MUTYH gene -also belonging to the base excision repair system like OGG1- which were studied previously for their predisposing effect on sporadic CRC (KURY *et al.* 2007a) ; ** Nomenclature referring to the databases dbNSP, SNPPER and SNP500Cancer ; * Ref: numbers correspond to bibliographical references, and the “d” mark corresponds to SNPs identified by dHPLC screening in 50 patients with CRC and 50 controls from the study population.**

II.3.1.4. Additional file 4: Test of the different possible models of genotypic combinations related to a given set of five polymorphisms picked among the 52 polymorphisms analyzed in the study

II.3.1.5. Additional file 5: Estimation by use of bootstrapping of the external validity of the model showing association between CRC risk and genotypes combinations

Le format et la taille de ces fichiers ne me permettaient pas de les inclure dans le présent travail.

II.4. Etude du gène *MUTYH* (Article 3)

Le troisième article présenté traite de travaux un peu en marge de notre étude principale d'analyse de SNP des gènes de prédisposition au cancer colorectal. Ces travaux sont, en effet, focalisés sur des mutations dans le gène *MUTYH* responsable d'une forme familiale de cancer colorectal la MAP (*MUTYH-Associated Polyposis*). Cette dernière se caractérise par une polypose colique modérée (<1000 polypes) et par un âge de survenue plus tardif (environ 50 ans) que dans les polyposes classiques. Ce sont Al-Tassan *et al.* (2002) qui ont découvert que des mutations bialléliques de *MUTYH* étaient responsables de la MAP alors qu'ils examinaient des familles présentant au niveau somatique des transversions C:G→A:T dans le gène *APC*, signes d'une anomalie dans les gènes de réparation de l'ADN, notamment les gènes du système BER.

En 2007, Cheadle et Sampson ont recensé 30 mutations pouvant engendrer des protéines tronquées, parmi lesquelles 11 mutations non-sens, 9 petites insertions et délétions et 10 variants dans les sites d'épissages. En outre, 52 mutations faux-sens et insertions ou délétions dans le cadre de lecture ont été recensées, dont les deux mutations Y165C (p.Tyr165Cys) et G382D (p.Gly382Asp) trouvées par Al-Tassan comme causales dans les formes MAP de cancer colorectal familial. À elles deux, ces mutations représentent environ 80% de toutes les mutations reportées de *MUTYH* dans la plupart des populations étudiées (CHEADLE and SAMPSON 2007). De manière plus confidentielle, d'autres mutations ont été trouvées associées à la MAP : par exemple, 1395delGGA chez des patients italiens, 1186–1187insGG chez des patients portugais, p.Pro391Leu chez des patients allemands, p.Tyr90X chez des patients pakistanais (JASS 2008).

Lorsque l'étude de ce gène a été réalisée, très peu d'informations étaient disponibles sur l'effet des mutations hétérozygotes dans la population générale. Nous n'avions à notre disposition que quelques études qui se contredisaient sur l'effet de ces mutations monoalléliques. C'est dans le but d'essayer de donner une réponse à cette interrogation que nous avons réalisé cette étude.

Cette étude, comprenant initialement 1024 cas et 1121 témoins, nous a permis d'exclure un patient, car celui-ci était hétérozygote composite pour les mutations Y165C et G382D, ce qui en faisait un patient atteint de MAP. C'est pourquoi, dans les deux articles précédents, la cohorte « patients » ne comprend que 1023 cas conformément à l'objectif primaire de notre étude qui était de ne travailler que sur des cas sporadiques de cancers colorectaux.

Dans un premier temps, nous avons réalisé un séquençage de 100 patients de notre cohorte, afin de déterminer les différents polymorphismes présents dans notre population. Ainsi, nous avons pu mettre en évidence 13 variants dont la plupart étaient connus. L'analyse bioinformatique a

montré que 10 de ces 13 variants n'avaient aucun effet sur la pathologie colorectale. Deux autres variants correspondaient aux mutations c.494A>G (p.Tyr165Cys) et c.1145G>A (p.Gly382Asp) responsables de la majorité des formes familiales de MAP. Le dernier variant, c.1435G>T (p.Val479Phe) était prédit comme étant délétère par les différents programmes bioinformatiques utilisés (SIFT, Tcoffee).

Ces trois variants délétères ou prédits comme tels, c.494A>G, c.1145G>A et c.1435G>T, ont donc, dans un deuxième temps, été étudiés sur la totalité de nos cohortes « patients » et « témoins » afin de connaître leur association éventuelle avec un risque de cancer colorectal sporadique. Les résultats obtenus montrent que les individus de notre population ne sont porteurs de ces mutations qu'à l'état hétérozygote et que le risque de CCR a tendance à être augmenté chez les individus hétérozygotes mais pas de manière significative. La stratification de la population en fonction de l'âge montre un effet plus important et significatif de ces variants chez les individus de plus de 55 ans. Bien sûr ces résultats sont à considérer avec précaution, car ils ne portent que sur un très faible nombre d'individus. Cinq autres mutations, non retrouvées par séquençage, mais qui ont déjà été montrées associées au risque de cancer colorectal chez des populations caucasiennes (BAI *et al.* 2005; ZHOU *et al.* 2005), ont également été analysées sur la totalité de nos cohortes. Aucun de ces variants n'a cependant été retrouvé dans notre population ce qui explique qu'aucun résultat ne soit présentement reporté.

Une troisième partie de l'article montre qu'aucune mutation délétère n'a été retrouvée par séquençage des exons des autres gènes (*OGG1*, *NUDT1*) de la voie de réparation de l'ADN par excision de base (BER). Par cette analyse, nous aurions pu mettre en évidence une probable association entre des mutations des gènes d'une même voie donnant lieu à une prédisposition pour la pathologie.

ARTICLE 3

The thorough screening of the MUTYH gene in a large French cohort of sporadic colorectal cancers.

KÜRY S^{*}, BUECHER B^{*}, ROBIOU-DU-PONT S^{*}, SCOUL C, COLMAN H,
LELIÈVRE B, OLSCHWANG S, LE HOUÉROU C, LE NEEL T, FAROUX R,
OLLIVERY J, LAFRAISE B, CHUPIN LD, BEZIEAU S

Genet Test **11** : 373-379 (2007)

* Contribution équivalente au travail

GENETIC TESTING

Volume 11, Number 4, 2007

© Mary Ann Liebert, Inc.

DOI: 10.1089/gte.2007.0029

The Thorough Screening of the *MUTYH* Gene in a Large French Cohort of Sporadic Colorectal Cancers

SÉBASTIEN KÜRY,^{1,2*} BRUNO BUECHER,^{2,3,*} SÉBASTIEN ROBIYOU-DU-PONT,^{1,2,*} CATHERINE SCOUL,^{1,2} HÉLÈNE COLMAN,^{1,2} BÉNÉDICTE LELIÈVRE,^{1,2} SYLVIANE OLSCHWANG,⁴ CLAIRE LE HOUÉROU,² TANGUY LE NEEL,^{1,5} ROGER FAROUX,⁶ JEAN OLLIVRY,⁷ BERNARD LAFRAISE,⁸ LOUIS-DOMINIQUE CHUPIN,⁹ and STÉPHANE BÉZIEAU^{1,2}

ABSTRACT

The *MUTYH* gene encodes a key glycosylase of the base-excision repair system that is involved in maintaining genomic DNA stability against oxidative damage. Biallelic germline *MUTYH* mutations have been proved to greatly predispose to non-familial adenomatous polyposis (FAP) and non-hereditary non-polyposis colorectal cancer (HNPCC) familial recessive forms of colorectal cancer with multiple adenomas. To date, there is still much debate over the impact of monoallelic germline *MUTYH* mutations on colorectal carcinogenesis. To evaluate their role in the susceptibility to sporadic colon and rectum cancers, we screened 1024 French sporadic colorectal cancer cases and 1121 French healthy controls for Caucasian *MUTYH*-associated polyposis mutations, including already known mutations p.Gly382Asp and p.Tyr165Cys, and new mutation p.Val479Phe. We observed a nonstatistically significant association between these *MUTYH* mutations at a heterozygous state and an increase in colorectal cancer risk (odds ratio [OR] 1.26, 95% confidence interval [CI] 0.70–2.27). As a result, we conclude that heterozygous *MUTYH* mutations do not play a major role in sporadic colorectal carcinogenesis although a modest effect on this process cannot be ruled out.

INTRODUCTION

AMONG THE THOUSANDS of new cases of colorectal cancer (CRC) diagnosed every year in Western countries, approximately 20% are inherited (Houlston *et al.*, 1992; Croitoru *et al.*, 2004). Two autosomal dominant forms have been identified: familial adenomatous polyposis (FAP) associated with germline mutations in the adenomatous polyposis coli (*APC*) tumor suppressor gene, and hereditary non-polyposis colorectal cancer (HNPCC) due to mutations in the DNA mismatch repair genes (e.g. *MLH1*, *MSH2*); both syndromes account for

3%–5% of CRC cases. Recently, Al-Tassan *et al.* (2002) suggested the existence of a third familial form of CRC, *MUTYH*-associated polyposis (MAP), linked to the base-excision repair (BER) gene *MUTYH* (MIM 604933). MAP, which represents 0.5% of unselected colorectal cancers (Fleischmann *et al.*, 2004), is clinically comparable to attenuated FAP but differs notably in its recessive mode of transmission and its later age of onset (Sampson *et al.*, 2005).

MUTYH participates with two other genes, *MTH1* (or *NUDT1*; MIM 600312) and *OGG1* (MIM 601982), in maintaining genome stability against oxidative damage to cellular

¹Laboratoire d'Etude de l'ADN, Faculté de Médecine de Nantes, 44035 France.

²Institut des Maladies de l'Appareil Digestif et CIC INSERM, Centre Hospitalier Universitaire de Nantes 44093, France.

³Service d'Hépatogastroentérologie, Centre Hospitalier Universitaire de Nantes 44093, France.

⁴Département d'Oncologie Génétique, Prévention et Dépistage, Institut Paoli-Calmettes, Marseille 13273, France.

⁵Biofortis, Nantes 44200, France.

⁶Service d'Hépatogastroentérologie, Centre Hospitalier Universitaire de la Roche-Sur-Yon, 85000 France.

⁷Association des Gastroentérologues de Vendée, 85300 Challans, France.

⁸Centre d'Examen de Santé de Saint-Nazaire 44616, France.

⁹Centre d'Examen de Santé de La Roche-Sur-Yon 85000, France.

*Sébastien Küry, Bruno Buecher, and Sébastien Robiou-du-Pont contributed equally to this work.

DNA. It encodes a DNA glycosylase that prevents a deleterious C:G to A:T transversion. This is achieved by removing adenines mispaired during replication with a 7,8-dihydro-8-oxoguanine (8-oxoG), the most stable mutagenic lesion induced by DNA oxidation. To date, at least 24 putative pathogenic mutations have been reported in *MUTYH*, including 12 missense or nonsense mutations, 6 splicing mutations, and 4 small and 1 gross deletions (Al-Tassan *et al.*, 2002; Jones *et al.*, 2002; Sampson *et al.*, 2003; Sieber *et al.*, 2003; Isidro *et al.*, 2004; Eliason *et al.*, 2005). Several studies have underlined an association between biallelic germline mutations of *MUTYH* and multiple colorectal adenomas and carcinoma (Al-Tassan *et al.*, 2002; Jones *et al.*, 2002; Sampson *et al.*, 2003; Sieber *et al.*, 2003; Isidro *et al.*, 2004; Eliason *et al.*, 2005; Aretz *et al.*, 2006).

The two most common nonconservative sequence mutations in Caucasian Europeans, p.Gly382Asp and p.Tyr165Cys, account for 80% of MAP associated alleles (Lipton and Tomlinson 2004), which represents a prevalence of approximately 0.5% in a general Caucasian population (Al-Tassan *et al.*, 2002; Sieber *et al.*, 2003). Other deleterious mutations have been found in only one population, such as p.Glu466X and p.Tyr90X, observed in Indian and Pakistani descent, respectively (Jones *et al.*, 2002; Sampson *et al.*, 2003). According to recent meta-analysis by Webb *et al.* (2006) no case of unaffected carrier of biallelic *MUTYH* mutations has been documented as yet. However, the question of a pathogenic effect of *MUTYH* monoallelic mutations has not yet been unravelled and is still the subject of much debate. Farrington *et al.* (2005) suggested a small dominant effect of *MUTYH* mutations contradicted by Webb *et al.* (2006) who rejected the hypothesis of an increased risk for CRC in monoallelic carriers on the basis of current data.

To address this important clinical issue and perhaps contribute to its elucidation, we conducted a large genetic association study based on two cohorts assembled with 1024 French patients with sporadic CRC and 1121 French healthy individuals respectively. We describe hereafter the results of our investigation on the predisposing role of *MUTYH* and other BER genes in sporadic CRC.

MATERIALS AND METHODS

Assembly of the cohorts and sample collection

From December 2002 to March 2006, we made up two cohorts of 1024 patients (633 males and 391 females; mean age at diagnosis 67 [40–99]) and 1121 controls (609 males and 512 females; mean age 62 [41–101]), all of Caucasian origin and recruited within the Pays de la Loire region of France. All cases are sporadic CRC diagnosed in one of the six public hospitals and five clinics collaborating in the project. Every patient with a personal history of colorectal cancer diagnosed at an age \geq 40 years, was eligible for the case cohort provided that his questioning by one of the study physician revealed no suspicion of major genetic predisposition (familial form of colorectal cancer); every patient aware of any member of his family affected with colorectal cancer or polyposis, among his first-degree relatives or even among more distant relatives, was automatically excluded from the study. Age- and gender-matched controls were included in two Health Examination Centers of the Pays

de la Loire region, and the recruitment of individuals \geq 70 years was completed in the departments of internal medicine and hepatogastroenterology at the University Hospital of Nantes. Control individuals were eligible for the study if they were \geq 40 years of age, and had no family history of colorectal cancer or polyps. Blood samples were collected in ethylenediaminetetraacetic acid (EDTA) tubes from participants after counseling and receipt of informed consent, and we used a Nucleon BACC2 kit (GE Healthcare, Uppsala, Sweden) for the extraction of genomic DNA. In the presence of the physician, each participant answered the same 1-page standardized questionnaire pertaining to family information, medical history, lifestyle, and dietary exposure by means of food-frequency questions validated by a nutritionist. For cases, endoscopy and histology reports were additionally requested from the study oncologist or hepatogastroenterologist by means of a multiple-choice questionnaire. All information regarding participants was rendered anonymous after collection of blood samples, as required by the local (CCPPRB) and national (CNIL) ethical committees that approved the study. All data provided by the questionnaires, as well as all the genetic data generated and described below, were pooled within a common electronic SQL database specifically set up for the need of the present study.

Detection screening of MUTYH variants

Taking into account the population-specific characteristics of the variants reported in *MUTYH* to date, we chose to perform a blind screening of the whole coding regions and splicing sites of the gene. For this purpose, we randomly selected 100 subjects within the patient cohort. From their genomic DNA, we polymerase-chain reaction (PCR)-amplified every exon and its flanking intronic region and sequenced the products on ABI 3730 using ABI Big Dye terminator V3.1 chemistry (Applied Biosystems, Foster City, CA).

The variants we observed that had already been reported in literature about MAP were kept for further testing of the whole cohorts as described below. Each novel variant we identified was analyzed with online bioinformatic software and also kept for further testing if predicted to be deleterious. Missense mutations were analyzed with SIFT Blink, Polyphen, and T-Coffee programs that provided us with information about the conservation degree across species and the putative pathogenic effect of the amino-acid residue concerned. Regarding variants found within the intronic boundaries of exons, we ran Genscan programs and used a splice-site prediction tool of the Berkeley Drosophila Genome Project to assess the potential effects on splicing events. URLs of the programs used are: <http://blocks.fhcr.org/sift/SIFT.html>; <http://genetics.bwh.harvard.edu/pph/>; and <http://archimede.cnrs-mrs.fr/Tcoffee>; <http://www.fruitfly.org>.

Analysis of MUTYH variants detected by screening

The entire patient and control cohorts (2145 study participants) were tested for pathogenic and potentially pathogenic *MUTYH* variants identified in the previous step. Genotypes were determined using high-throughput TaqMan allelic discrimination tests. Primers and dye-labeled MGB-NFQ probes, described in Table 1, were designed and synthesized by Applied Biosystems from sequences corresponding to Genbank ac-

SCREENING OF *MUTYH* IN FRENCH SPORADIC COLORECTAL CANCERS

375

TABLE 1. TAQMAN-BASED GENOTYPING ANALYSIS OF *MUTYH* MUTATIONS p.GLY382ASP, p.TYR165CYS, AND p.VAL479PHE

Mutation tested	Primer/probe name	Primer and probe sequences (5' to 3')
c.494A>G (p.Tyr165Cys)	F	CCACAGGAGGTGAATCAACTCT
	R	CCTTACCTTCCGAGCTCCCT
	VIC	CTGGGCTACTATTCT
	FAM	TGGGCTGCTATTCT
c.1145G>A (p.Gly382Asp)	F	GTAACAAGAGAGAATGGAGGGAATCG
	R	GACTGGGAACCTCCCACAGT
	VIC	CCTCTCAGGTCTGCTG
	FAM	CCTCTCAGATCTGCTG
c.1435G>T (p.Val479Phe)	F	GCCCTCACCTCCCTGTCT
	R	GCTGTTGGCCCTGATACACA
	VIC	CTTGTCTAGGTTTTCC
	FAM	CTTGTCTAGTTTTCC

cession NT_032977 and U63329.1. Reactions were set up in a 384-well plate in a 6 μ l final volume, including: 2.25 μ l of 2 \times Universal Master Mix, 0.11 μ l of 40 \times Assay Mix, 1.64 μ l of water, and 20 ng of genomic DNA (10 ng/ μ l). Reaction plates were thermocycled in an i-cycler (Bio-Rad Laboratories, Hercules, CA): an initial 10 min denaturation at 95°C was followed by 45 cycles including a denaturation step at 92°C for 15 sec and an annealing/extension step at 60°C for 1 min. After amplification, end point fluorescence reads were conducted on an Applied Biosystems ABI 7900HT sequence detection system. Genotypes were assigned using allelic discrimination software SDS v2.1 (Applied Biosystems). Every variant identified by 5'-exonuclease assay was subsequently confirmed by automated sequencing on ABI 3730 using ABI Big Dye terminator V3.1 chemistry (Applied Biosystems).

First statistical analyses of genotyping data were performed with SNPstats (Sole *et al.*, 2006) for the entire cohort. Then, in order to reproduce the investigation of an age-specific effect previously made by Farrington *et al.* (2005) and Webb *et al.* (2006), we analyzed partitions of the cohort determined the same way than these authors, i.e., according to early- (≤ 55 years) and late- (> 55 years) onset disease. CRC risks associated with genetic variations were estimated and translated to an odds ratio (OR) of cancer cases along with 95% confidence intervals.

Analysis of *NUDT1* and *OGG1* genes

To assess the hypothesis of an oligogenic phenomenon, every monoallelic carrier of either of the two mutations underwent a systematic sequencing of his entire *MUTYH*, *NUDT1* and *OGG1* coding regions, in order to find potential pathogenic heterozygous mutations. Sequences of the genomic contigs used as references correspond to Genbank accession numbers NT_022517 for *OGG1*, and NT_007819 for *NUDT1*.

RESULTS

Detection screening of *MUTYH* variants

Through our screening of the sample of 100 patients, we detected 13 variants. Five were nonconservative missense vari-

ants, including the 2 major Caucasian mutations c.494A>G (p.Tyr165Cys) in exon 7 and c.1145G>A (p.Gly382Asp) in exon 13, 2 referenced polymorphisms rs3219484 (c.64G>A; p.Val22Met) in exon 2 and rs3219489 (c.972G>C; p.Gln324His) in exon 12, and a novel variant c.1435G>T (p.Val479Phe) in exon 15. The 8 remaining variants identified were located in introns: only 3 variants were referenced in dbSNP as rs3219485 (c.157+30A>G) in intron 2, rs3219487 (c.462+35G>A) in intron 6 and rs3219488 (c.989-C>T) in intron 11; the 5 other variants were c.346+56G>A (intron 4), c.421-47C>T (intron 5), c.648+21C>A (intron 8), c.1437-17C>T (intron 14), and c.1481-83G>C (intron 15). All these variants were carried at the heterozygous state, except the nonconservative variant c.972G>C (p.Gln324His) and intronic variant c.462+35G>A found homozygous in 1 and 3 subjects, respectively.

In addition to the two obvious mutations, p.Tyr165Cys and p.Gly382Asp, we decided to test the variant c.1435G>T (p.Val479Phe) in the entire patient and control cohorts. The latter, found to be heterozygous in one patient of the cohort, was predicted to be malignant by the SIFT Blink program and benign by Polyphen, whereas T-coffee showed that the amino-acid residue at position 479 of *MUTYH* protein sequence was very much conserved between species. In addition, this mutation had been previously observed in a French MAP family (see Discussion).

The other two missense mutations, p.Val22Met and p.Gln324His, found in 6% and 20% of the patient sample, respectively, are already identified as polymorphisms (Al-Tassan *et al.*, 2002; Croitoru *et al.*, 2004). They concern amino-acid residues not well-conserved across species and are predicted to be benign by both SIFT Blink and Polyphen. For this reason, they were not kept for further analyses. None of the intronic variants found in the patient sample was predicted to produce an effect on splicing events. As a consequence, they were not submitted to further analyses either.

Analysis of *MUTYH* mutations p.Tyr165Cys, p.Gly382Asp, and p.Val479Phe

The three mutations p.Tyr165Cys, p.Gly382Asp, and p.Val479Phe were tested by 5'-exonuclease assay in the 2145

TABLE 2. STATISTICAL EVALUATION OF THE RELATIONSHIP BETWEEN DELETERIOUS *MUTYH* MUTATIONS DETECTED IN THE STUDY COHORTS AND SPORADIC COLORECTAL CANCER

Variants	Genotype	Entire cohort			≤55 years			>55 years		
		Controls	Cases		Controls	Cases		Controls	Cases	
		(n = 1121)	(n = 1023)	OR (95% CI)	(n = 344)	(n = 179)	OR (95% CI)	(n = 777)	(n = 844)	OR (95% CI)
G382D+	+/+	1101	1000	1.00	336	177	1.00	765	823	1.00
Y165C	+/-	20	23	1.27 (0.69–2.32)	8	2	0.47 (0.10–2.26)	12	21	1.63 (0.80–3.33)
				<i>p</i> = 0.44			<i>p</i> = 0.32			<i>p</i> = 0.18
G382D+	+/+	1100	999	1.00	335	177	1.00	765	822	1.00
Y165C+	+/-	21	24	1.26 (0.70–2.27)	9	2	0.42 (0.09–1.97)	12	22	1.71 (0.84–3.47)
V479F				<i>p</i> = 0.45			<i>p</i> = 0.23			<i>p</i> = 0.13

OR, odds ratio; CI, confidence interval; +/+ : non carriers; +/- : monoallelic carriers; G382D, mutation p.Gly382Asp Y165C, p.Tyr165Cys; V479F, p.Val479Phe.

The patient found compound heterozygote for mutations p.Gly382Asp and p.Tyr165Cys was excluded from the present analyses because he was considered as a familial case of attenuated polyposis.

SCREENING OF *MUTYH* IN FRENCH SPORADIC COLORECTAL CANCERS

377

study subjects included in patient and control cohorts. Among the patients, none was found homozygote for any of the three mutations. One patient (0.1%) was identified as a compound heterozygote for mutations p.Tyr165Cys, p.Gly382Asp. According to his genotype, this patient is by definition a familial case of *MUTYH*-associated polyposis; therefore, as our study deals with predisposition to sporadic forms of colorectal cancer, we chose to exclude him from the statistical analyses presented below. We also observed 24 monoallelic carriers: 20 (2%) for p.Gly382Asp, 3 (0.3%) for p.Tyr165Cys, and 1 (0.1%) for p.Val479Phe. By comparison, 16 (1.4%) controls were monoallelic carriers of p.Gly382Asp, 4 (0.4%) of p.Tyr165Cys, and 1 (0.1%) of p.Val479Phe, whereas we observed no biallelic carrier, neither homozygote nor compound heterozygote.

The 25 patients and 21 controls were screened by sequencing for other possible variants in the whole *MUTYH*, *NUDT1*, and *OGG1* coding and exon flanking regions. No known or potential deleterious variant emerged from this screening. It is worth noting however that a novel variant c.1502C>T (p.S501F) was observed in *MUTYH* Exon 16 of a control found monoallelic carrier of c.494A>G (p.Tyr165Cys). Variant p.S501F concerns an amino-acid residue not well conserved between species, although it is predicted to be possibly deleterious by Polyphen software. Regarding the other variants revealed by sequencing, they were the same as the one noted above according to the previous screening of 100 patients.

From genotyping results we estimated CRC risk associated with monoallelic mutations p.Tyr165Cys, p.Gly382Asp, and p.Val479Phe (Table 2). None of the three nonconservative mutations were in a Hardy-Weinberg equilibrium among affected and control subjects. We observed nonstatistically significant risks for all three heterozygous mutations: OR 1.45 (95% CI 0.75–2.79 ; $p = 0.27$) for p.Gly382Asp, OR 1.10 (95% CI 0.27–4.39 ; $p = 0.9$) for p.Tyr165Cys, and OR 1.09 (95% CI 0.07–17.53 ; $p = 0.95$) for p.Val479Phe. Combined ORs for monoallelic carriers remained nonstatistically significant: OR 1.27 (95% CI 0.69–2.32) for p.Tyr165Cys or p.Gly382Asp carriers, and OR 1.26 (95% CI 0.70–2.27) for p.Tyr165Cys, p.Gly382Asp or p.Val479Phe carriers. The increased tendency was strengthened for late-onset disease (>55 years) although the results remained inconclusive: OR 1.63 (95% CI 0.80–3.33) for p.Tyr165Cys or p.Gly382Asp monoallelic carriers, and OR 1.71 (95% CI 0.84–3.47) for carriers of any of the three heterozygous mutations. On the contrary, the risk for CRC was found decreased for individuals ≤ 55 years, although it was again in a nonsignificant manner: OR 0.47 (0.10–2.26) for p.Tyr165Cys or p.Gly382Asp carriers. Table 3 shows that in-

dividuals above 55 years old are at significantly higher risk of CRC compared to younger individuals of the cohort, whenever they carry monoallelic mutation p.Tyr165Cys or p.Gly382Asp or not; yet, this age effect is very strengthened in the group of monoallelic carriers compared to the group of noncarriers of the mutations.

Furthermore, we found no obvious association between the tumor location and heterozygous mutation carrier status for p.Tyr165Cys, p.Gly382Asp, or p.Val479Phe. In addition, the percentage of adenocarcinomas with synchronic polyps did not significantly diverge between the whole patient cohort and the group of patients heterozygotes for one of the three mutations.

DISCUSSION

The objective of the present study was to examine the impact of monoallelic *MUTYH* mutations on the susceptibility to sporadic CRC through a large population-based study. With this in mind, we screened two cohorts of 1024 patients with sporadic CRC and 1121 controls for sequence variants in *MUTYH*. In order to balance the cost of the experiments and the efficiency of our screening, we chose a two-steps strategy, i.e., after detection of *MUTYH* variants in a random sample of 100 patients chosen in the case cohort, three of these variants were tested in the whole patient and control cohorts to assess a possible increase in CRC risk as suggested by prior studies (Farrington *et al.*, 2005; Jenkins *et al.*, 2006).

We are aware that the first screening step may not have enabled the detection of all the variants present in the entire cohorts. Certain extremely rare variants, even pathogenic ones, may have slipped through the screening, whereas others such as c.494A>G (p.Tyr165Cys) or c.1435G>T (p.Val479Phe) would not have been revealed by sequencing the totally different random samples of patients. However, we assume that the mutations we may have overlooked are very minor MAP alleles that have a very small incidence on CRC and would not radically change any statistical calculations of associated risks. Moreover, in order to overlook as few mutations as possible, we also analyzed in the entire cohorts five additional variants that had not been detected through sequencing of the patient sample: p.Gly232Phe and p.Arg227Trp, two minor MAP mutations producing deleterious effects upon MutYH functions (Bai *et al.*, 2005), and p.Arg423Gln, p.Arg423Pro and p.Arg423Arg described in Swedish sporadic CRC patients (Zhou *et al.*, 2005). We did not provide any detailed analysis of these variants in the above sections, since nothing was found within the cohorts.

TABLE 3. ANALYSIS OF THE INTERACTION BETWEEN AGE AND CARRIER STATUS OF MUTATIONS p.Gly382Asp AND p.Tyr165Cys THROUGH ITS IMPACT ON SPORADIC COLORECTAL CANCER RISK

Genotype	Age	Controls	Cases	OR (95% CI)	p value
+/+	<55	336	177	1.00	8.48×10^{-12}
	>55	765	823	2.04 (1.66–2.51)	
+/-	<55	8	2	1.00	0.015
	>55	12	21	7.00 (1.27–38.48)	

OR, odds ratio; CI, confidence interval; +/+, non carriers; +/-, monoallelic carriers of mutation p.Gly382Asp or p.Tyr165Cys.

The second step of our screening did not include the intronic variants detected in patient sample and the missense variant c.1502C>T (p.S501F) found later in a control individual heterozygous for c.494A>G (p.Tyr165Cys). The demonstration of their possible deleterious effect on *MUTYH* expression or function would indeed require further *in vitro* investigations. By contrast, the selection of mutations p.Tyr165Cys and p.Gly382Asp was obvious because they are the main MAP-associated alleles in Caucasians (Lipton and Tomlinson, 2004) and they have been proved to induce a total loss of function of MutYH protein (Wooden *et al.*, 2004). The novel variant p.Val479Phe was selected on the basis of its existence at a homozygous state in an APC mutation-negative proband exhibiting multiple adenomas. This case, diagnosed in a 42-year-old in the Hospital of Nantes within a familial polyposis context with no evidence of vertical transmission, exactly fitted the clinical indications for *MUTYH* genetic analysis advocated by Sampson *et al.* (2003). In addition, as no other possibly defective mutation was found in *MUTYH*, we assume that p.Val479Phe is a new MAP-associated mutation.

The results we obtained by screening p.Tyr165Cys and p.Gly382Asp in the entire cohorts are consistent with previous reported works summarized in Webb's meta-analysis (Webb *et al.*, 2006). Indeed, monoallelic carriers of the two mutations account for 2.2% of the patients and 1.8% of the controls, these values being just slightly higher than those reported by Zhou *et al.* (2005) for a smaller cohort of sporadic CRC cases. In addition, we found no homozygote and the number of compound heterozygotes was proportionally smaller (0.1% of the patients and 0% of the controls) than in most studies. Genotyping results get for p.Val479Phe did not change the pattern as the variant was observed at heterozygous state in only one patient and one control. As a matter of fact, these data are quite logical, in that our study patient cohort was assembled with sporadic CRC, and by definition, we had carefully excluded all dubious familial forms of CRC. As biallelic germline mutations have been shown to predispose to a multiple-adenoma phenotype (Halford *et al.*, 2003; Gismondi *et al.*, 2004), the presence of many of these among our study patients would have questioned the very method used for the inclusion of the participants. For this reason, we will exclude the only patient we found to have compound heterozygote for p.Tyr165Cys and p.Gly382Asp from our future research on genetic susceptibility to sporadic CRC.

Sequencing of the 25 patient and 21 control monoallelic carriers of p.Tyr165Cys, p.Gly382Asp or p.Val479Phe showed that no other biallelic carrier of *MUTYH* deleterious mutations was present in the cohort. Disregarding possible and very rare MAP mutations undetected by our screening, we analyzed the three deleterious mutations found in the cohort to estimate the global sporadic CRC risk associated with monoallelic *MUTYH* mutations. Our results appear consistent with previous and recent findings (Webb *et al.*, 2006). Combined ORs for monoallelic mutations carriers show a slight increase in CRC risk which remains statistically insignificant nevertheless, whether considering the entire cohort or heterozygotes >55 years old (i.e., late-onset disease; see Table 2). As a matter of fact, the sporadic CRC risk attributable to monoallelic germline *MUTYH* mutations appears as modest

as the one reported in literature for polymorphisms found, for example, in genes of xenobiotics metabolizing enzymes (reported ORs range from approximately 1.1 to 2; de Jong *et al.*, 2002; Landi *et al.*, 2005). It is worth noting, however, that we observed the same interaction between age and monoallelic germline mutations as described previously by Farrington *et al.* (2005); it appears that in our cohort, the predisposing effect of age on CRC risk is enhanced by the genotype (Table 3). Yet the 95% confidence interval associated with our results is too broad, because of too small a number of monoallelic carriers present in our cohort. It would therefore not be careful to draw definitive conclusions and further analyses on larger cohorts would be required to confirm this interaction phenomenon.

Our observations do not contradict any of the recent comparable studies. None of them could either confirm the dominant effect hypothesized for monoallelic *MUTYH* mutations (Croitoru *et al.*, 2004; Fleischmann *et al.*, 2004), or provide evidence for their strong predisposing impact on colorectal carcinogenesis. We could thus speculate that heterozygous *MUTYH* mutations participate in colorectal carcinogenesis through an oligogenic mechanism involving other BER genes. In the present study, complementary mutational analyses of *NUDT1* and *OGG1* in p.Tyr165Cys, p.Gly382Asp or p.Val479Phe monoallelic carriers were negative; this indicates that no study CRC case is apparently due to a digenic mechanism comparable to that suggested for example by Farrington's cases (Farrington *et al.*, 2005). Yet, we cannot totally exclude that our patients carry mutations in other BER genes, like *NTHL1*, *NEIL1*, *NEIL2*, *MPG*, *TDG*, *UNG*, or *SMUG*, presumed to be little involved in CRC development (Broderick *et al.*, 2006). Further experiments would be required to address this question.

In conclusion, the predisposing effect of one of the three *MUTYH* monoallelic deleterious mutations observed in the present sporadic CRC population is relatively modest. The cohorts used for this study were not designed to investigate familial forms of CRC, but they are precisely adapted to examine minor susceptibility factors contributing to a low/moderate-penetrance modification for colorectal cancer development. As suggested by previous authors, pooling of different study data could bring more power to statistical analyses (Ioannidis *et al.*, 2006). Nevertheless, future genetic studies on *MUTYH* should also take into account the crosstalk between DNA repair systems (Kovtun and McMurray, 2007) and consider interactions of *MUTYH* germline mutations with BER and non-BER glycosylase variants.

ACKNOWLEDGMENTS

We are very grateful to all the patients, control individuals, and physicians who participated in this work and made it possible. This work would have not been possible without its funding by a regional Hospital Clinical Research Program (PHRC) and the support of the Groupement des Entreprises Françaises dans la Lutte contre le Cancer (GEFLUC), the association Anne de Bretagne Génétique, the Conseil Régional des Pays de la Loire and the Ligue Régionale Contre le Cancer (LRCC).

REFERENCES

- Al-Tassan N, Chmiel NH, Maynard J, Fleming N, Livingston AL, Williams GT, Hodges AK, Davies DR, David SS, Sampson JR, Cheadle JP (2002) Inherited variants of MYH associated with somatic G:C → T:A mutations in colorectal tumors. *Nat Genet* 30:227–232.
- Arzt S, Uhlhaas S, Goergens H, Siberg K, Vogel M, Pagenstecher C, Mangold E, Caspari R, Propping P, Friedl W (2006) *MUTYH*-associated polyposis: 70 of 71 patients with biallelic mutations present with an attenuated or atypical phenotype. *Int J Cancer* 119:807–814.
- Bai H, Jones S, Guan X, Wilson TM, Sampson JR, Cheadle JP, Lu AL (2005) Functional characterization of two human MutY homolog (hMYH) missense mutations (R227W and V232F) that lie within the putative hMSH6 binding domain and are associated with hMYH polyposis. *Nucleic Acids Res* 33:597–604.
- Broderick P, Bagratuni T, Vijayakrishnan J, Lubbe S, Chandler I, Houlston RS (2006) Evaluation of NTHL1, NEIL1, NEIL2, MPG, TDG, UNG and SMUG1 genes in familial colorectal cancer predisposition. *BMC Cancer* 6:243.
- Croitoru ME, Cleary SP, Di Nicola N, Manno M, Selander T, Aronson M, Redston M, Cotterchio M, Knight J, Gryfe R, Gallinger S (2004) Association between biallelic and monoallelic germline MYH gene mutations and colorectal cancer risk. *J Natl Cancer Inst* 96:1631–1634.
- de Jong MM, Nolte IM, te Meerman GJ, van der Graaf WT, de Vries EG, Sijmons RH, Hofstra RM, Kleibeuker JH (2002) Low-penetrance genes and their involvement in colorectal cancer susceptibility. *Cancer Epidemiol Biomarkers Prev* 11:1332–1352.
- Eliason K, Hendrickson BC, Judkins T, Norton M, Leclair B, Lyon E, Ward B, Noll W, Scholl T (2005) The potential for increased clinical sensitivity in genetic testing for polyposis colorectal cancer through the analysis of MYH mutations in North American patients. *J Med Genet* 42:95–96.
- Farrington SM, Tenesa A, Barnetson R, Wiltshire A, Prendergast J, Porteous M, Campbell H, Dunlop MG (2005) Germline susceptibility to colorectal cancer due to base-excision repair gene defects. *Am J Hum Genet* 77:112–119.
- Fleischmann C, Peto J, Cheadle J, Shah B, Sampson J, Houlston RS (2004) Comprehensive analysis of the contribution of germline MYH variation to early-onset colorectal cancer. *Int J Cancer* 109:554–558.
- Gismondi V, Meta M, Bonelli L, Radice P, Sala P, Bertario L, Viel A, Fornasari M, Arrigoni A, Gentile M, Ponz de Leon M, Anselmi L, Mareni C, Bruzzi P, Varesco L (2004) Prevalence of the Y165C, G382D and 1395delGGA germline mutations of the MYH gene in Italian patients with adenomatous polyposis coli and colorectal adenomas. *Int J Cancer* 109:680–684.
- Halford SE, Rowan AJ, Lipton L, Sieber OM, Pack K, Thomas HJ, Hodgson SV, Bodmer WF, Tomlinson IP (2003) Germline mutations but not somatic changes at the MYH locus contribute to the pathogenesis of unselected colorectal cancers. *Am J Pathol* 162:1545–1548.
- Houlston RS, Collins A, Slack J, Morton NE (1992) Dominant genes for colorectal cancer are not rare. *Ann Hum Genet* 56:99–103.
- Ioannidis JP, Gwinn M, Little J, Higgins JP, Bernstein JL, Boffetta P, Bondy M, Bray MS, Brenchley PE, Buffler PA, Casas JP, Chokkalingam A, Danesh J, Smith GD, Dolan S, Duncan R, Gruis NA, Hartge P, Hashibe M, Hunter DJ, Jarvelin MR, Malmer B, Maraganore DM, Newton-Bishop JA, O'Brien TR, Petersen G, Riboli E, Salanti G, Seminara D, Smeeth L, Taioli E, Timpson N, Uitterlinden AG, Vineis P, Wareham N, Winn DM, Zimmern R, Khoury MJ; Human Genome Epidemiology Network and the Network of Investigator Networks (2006) A road map for efficient and reliable human genome epidemiology. *Nat Genet* 38:3–5.
- Isidro G, Laranjeira F, Pires A, Leite J, Regateiro F, Castro e Sousa F, Soares J, Castro C, Giria J, Brito MJ, Medeira A, Teixeira R, Morna H, Gaspar I, Marinho C, Jorge R, Brehm A, Ramos JS, Boavida MG (2004) Germline *MUTYH* (MYH) mutations in Portuguese individuals with multiple colorectal adenomas. *Hum Mutat* 24:353–354.
- Jenkins MA, Croitoru ME, Monga N, Cleary SP, Cotterchio M, Hopper JL, Gallinger S (2006) Risk of colorectal cancer in monoallelic and biallelic carriers of MYH mutations: a population-based case-family study. *Cancer Epidemiol Biomarkers Prev* 15:312–314.
- Jones S, Emmerson P, Maynard J, Best JM, Jordan S, Williams GT, Sampson JR, Cheadle JP (2002) Biallelic germline mutations in MYH predispose to multiple colorectal adenoma and somatic G:C → T:A mutations. *Hum Mol Genet* 11:2961–2967.
- Kovtun IV, McMurray CT (2007) Crosstalk of DNA glycosylases with pathways other than base excision repair. *DNA Repair (Amst)* 6:517–529.
- Landi S, Gemignani F, Moreno V, Gioia-Patricola L, Chabrier A, Guino E, Navarro M, de Oca J, Capella G, Canzian F (2005) A comprehensive analysis of phase I and phase II metabolism gene polymorphisms and risk of colorectal cancer. *Pharmacogenet Genomics* 15:535–546.
- Lipton L, Tomlinson I (2004) The multiple colorectal adenoma phenotype and MYH, a base excision repair gene. *Clin Gastroenterol Hepatol* 2:633–638.
- Sampson JR, Dolwani S, Jones S, Eccles D, Ellis A, Evans DG, Frayling I, Jordan S, Maher ER, Mak T, Maynard J, Pigatto F, Shaw J, Cheadle JP (2003) Autosomal recessive colorectal adenomatous polyposis due to inherited mutations of MYH. *Lancet* 362:39–41.
- Sampson JR, Jones S, Dolwani S, Cheadle JP (2005) *MUTYH* (MYH) and colorectal cancer. *Biochem Soc Trans* 33:679–683.
- Sieber OM, Lipton L, Crabtree M, Heinemann K, Fidalgo P, Phillips RK, Bisgaard ML, Orntoft TF, Aaltonen LA, Hodgson SV, Thomas HJ, Tomlinson IP (2003) Multiple colorectal adenomas, classic adenomatous polyposis, and germ-line mutations in MYH. *N Engl J Med* 348:791–799.
- Sole X, Guino E, Valls J, Iniesta R, Moreno V (2006) SNPStats: a web tool for the analysis of association studies. *Bioinformatics* 22:1928–1929.
- Webb EL, Rudd MF, Houlston RS (2006) Colorectal cancer risk in monoallelic carriers of MYH variants. *Am J Hum Genet* 79:768–771; author reply 771–762.
- Wooden SH, Bassett HM, Wood TG, McCullough AK (2004) Identification of critical residues required for the mutation avoidance function of human MutY (hMYH) and implications in colorectal cancer. *Cancer Lett* 205:89–95.
- Zhou XL, Djureinovic T, Werelius B, Lindmark G, Sun XF, Lindblom A (2005) Germline mutations in the MYH gene in Swedish familial and sporadic colorectal cancer. *Genet Test* 9:147–151.

Address reprint requests to:

Sébastien Kury
Laboratoire d'Etude de l'ADN
Faculté de Médecine de Nantes
1 rue Gaston Veil
44035 Nantes Cedex
France

E-mail: Sebastien.Kury@univ-nantes.fr

III. Approche *Genome-Wide Association*

III.1. Introduction

La deuxième partie des présents travaux est basée sur une étude d'association réalisée à l'échelle du génome (*Genome-Wide Association* ou GWA). Notre étude se distingue toutefois des études d'association classiques, en ce sens où nous avons hybridé non pas des individus isolés, mais des pools d'ADN, ce qui nous a permis de diminuer le coût de nos analyses, tout en gardant une puissance statistique satisfaisante.

Cette méthode appelée SNP-MaP est basée sur la constitution de groupes ou pools d'ADN. Chaque pool a été conçu de manière à faire éventuellement ressortir des SNP spécifiques d'une certaine frange de notre population d'étude. Ainsi, avons-nous réalisé deux groupes de pools différents, le premier constitué à partir des catégories d'âge des individus et le second basé sur l'origine géographique des individus. Dans le premier groupe, notre but initial était de mettre en avant des marqueurs différents entre les individus développant un cancer colorectal sporadique avant l'âge de 60 ans et ceux le développant le plus tard possible. Ces deux sous-groupes ont été analysés contre les témoins les plus âgés de notre cohorte, afin de s'affranchir le plus possible d'un développement tardif d'un CCR sporadique. Le second groupe a été réalisé en se basant sur les origines géographiques des individus de notre cohorte. Comme notre recrutement a été réalisé dans la région des Pays de la Loire, tous les départements la composant sont représentés (Loire Atlantique, Vendée, Maine et Loire, Mayenne et Sarthe). La participation a toutefois été nettement plus forte dans les centres hospitaliers de la Vendée. De plus, il a été montré que les habitants originaires de ce département partagent un fond génétique commun important. On peut avancer l'hypothèse que si les individus ont un fond génétique commun, les marqueurs d'une pathologie seront plus facilement identifiables du fait d'un « bruit de fond » moins important.

Après de multiples analyses bioinformatiques, les données croisées de ces deux groupes nous ont permis de mettre en évidence, par cette approche, sept SNP appartenant aux gènes *SPAM1*, *DDX58*, *COL4A2*, *MTMR10*, *CDH13*, *UTP6*, *ZNF543* associés au CCR sporadique. De plus, nous avons validé la prédisposition du locus 8q24.21 déjà montré associé à la pathologie. Un génotypage individuel de tous les individus de notre cohorte nous a permis d'éliminer les SNP des gènes *SPAM1*, *DDX58*, *COL4A2*, *MTMR10*, *ZNF543* dont l'association n'était pas suffisamment robuste. Les deux SNP restants (rs2306907 du gène *CDH13* et rs3760454 du gène *UTP6*) ont subi une réplique sur une population caucasienne différente afin de valider leur association. En parallèle, une étude de l'expression de l'ARN messager (ARNm) des gènes *CDH13* et *UTP6* a été réalisée ce qui nous a permis

de mettre en évidence une augmentation de l'expression de l'ARNm du gène *CDH13*. Cette augmentation est soumise au génotype du SNP rs2306907 de *CDH13*. En effet, nous avons pu observer une diminution de l'expression de *CDH13* lorsque l'allèle variant du SNP rs2306907 était présent, nous laissant imaginer une probable implication de ce SNP dans la carcinogenèse colorectale sporadique.

III.2. Matériels et Méthodes

Les différentes étapes réalisées lors de ces travaux sont synthétisées sous forme d'un organigramme disponible en Annexe 3.

III.2.1. Constitution des cohortes et préparation des ADN

Le recrutement des cohortes, soutenu par un PHRC (Programme Hospitalier de Recherche Clinique) régional, s'est déroulé dans la région des Pays de la Loire de décembre 2002 à mars 2006. Il a permis d'inclure 1121 contrôles indemnes de toutes pathologies colorectales (609 hommes et 512 femmes avec un âge moyen à l'inclusion de 62 ans [41-101]) et 1023 patients atteints de cancers colorectaux sporadiques (633 hommes et 391 femmes avec un âge moyen au diagnostic de 67 ans [40-99]). Tous les patients dont les apparentés de premier degré ont présenté des pathologies colorectales ont été exclus de l'étude. Chaque individu ayant accepté de participer au projet s'est vu prélever deux tubes de 5 ml de sang périphérique et a dû répondre, en présence du médecin recruteur, à un questionnaire portant sur ses antécédents médicaux et sur ses habitudes de vie (Annexe 1 et Annexe 2).

Les ADN génomiques de tous les participants ont été extraits à partir des prélèvements sanguins, grâce au kit Nucleon® BACC2 (GE Healthcare, Uppsala, Suède). Ces ADN ont ensuite été dosés en utilisant la molécule fluorescente Picogreen® (Molecular Probs, Eugene, OR), puis ont été mis en plaque 96 puits à une concentration de 10 ng/μl et de 50 ng/μl, par une méthode automatisée développée sur Microlab Star® (Hamilton Robotics, Bonaduz, Suisse).

De plus amples informations sur le recrutement, les questionnaires et la préparation des échantillons sont disponibles dans la publication de Küry *et al.* (2007b) (Article 1).

III.2.2. La méthode SNP-MaP

Le principe de la méthode SNP-MaP (*SNP-Microarray and Pooling*) décrite par Butcher *et al.* (2005) consiste, dans un premier temps, à réaliser des « pools » d'ADN génomiques. La réalisation d'un pool revient à mélanger dans un même tube et en quantité équivalente un nombre défini d'ADN génomiques provenant d'individus différents. Ces « pools » sont ensuite hybridés sur des puces de génotypage ; dans notre cas ce sont des puces « GenChip® Human Mapping 250K-Nsp Array » de la société Affymetrix (Santa Clara, CA). Cette technique de « poolage » d'ADN suivie d'une hybridation permet une économie budgétaire substantielle, en ce sens où tous les ADN de notre projet peuvent théoriquement être hybridés sur quelques puces seulement (30 dans notre étude), au lieu de

milliers, si nous avons dû hybrider chacun des 2144 ADN génomiques de notre étude individuellement sur des puces indépendantes.

III.2.2.1. Sélection des individus intégrés dans les pools d'ADN

III.2.2.1.1. Groupe d'individus classés par catégories d'âge (projet 1) : Groupe « catégorie d'âge »

Le critère d'âge d'apparition du cancer permet de faire une relativement bonne distinction entre les cancers colorectaux familiaux et les cancers sporadiques. Les premiers, dans la majorité des cas, surviennent à un âge plus précoce que les cancers sporadiques, dont la probabilité de survenue augmente avec les années. Pour notre étude, nous avons fixé à 60 ans l'âge limite permettant de séparer les cas sporadiques des cas familiaux ; cette limite d'âge n'est pas reconnue de façon unanime et varie selon les sources de 50 à 60 ans, mais nous nous sommes appuyés sur la tendance actuelle qui tend à déplacer cette limite vers la borne supérieure de l'intervalle (60 ans).

Partant du principe que, dans les formes sporadiques, la précocité devait être influencée par des facteurs génétiques partagés par les cas les plus jeunes, nous avons décidé d'opposer les patients les plus âgés (> 60 ans, les plus à même de présenter de véritables formes sporadiques) aux patients les plus jeunes (<60 ans), en n'occultant pas le fait que ces derniers puissent présenter non pas une forme précoce de cancer sporadique, mais plutôt une forme familiale non découverte. Dans notre cohorte, nous avons identifié 243 patients de moins de 60 ans (moyenne=54 ans [40-60] dont 58% d'hommes) et, afin d'harmoniser la taille des effectifs, nous avons sélectionné les 288 patients les plus âgés (moyenne=76 ans [69-99] dont 59,4% d'hommes). De la même manière, nous avons choisi nos témoins dans la population la plus âgée de notre cohorte, afin de nous affranchir le plus possible d'un développement tardif de cancer colorectal sporadique. Nous avons ainsi sélectionné les 288 témoins les plus âgés de notre cohorte (moyenne=74 ans [68-101] dont 52,7% d'hommes).

En résumé, dans cette première étape, nous avons donc déterminé trois groupes d'individus classés par âge : un groupe de 243 patients de moins de 60 ans, un groupe de 288 patients de plus de 60 ans et un groupe de témoins de plus de 60 ans.

A partir de ces différents effectifs, nous avons divisé chacun des trois groupes d'individus en sous-groupes (ou « pools ») de tailles équivalentes : trois pools de patients de moins de 60 ans, trois pools de patients de plus de 60 ans et trois pools de témoins ont ainsi été créés. Pour chaque groupe (patients de moins de 60 ans, patients de plus de 60 ans, témoins), un quatrième pool a été réalisé en mélangeant certains des ADN des trois pools précédemment réalisés (Tableau 23).

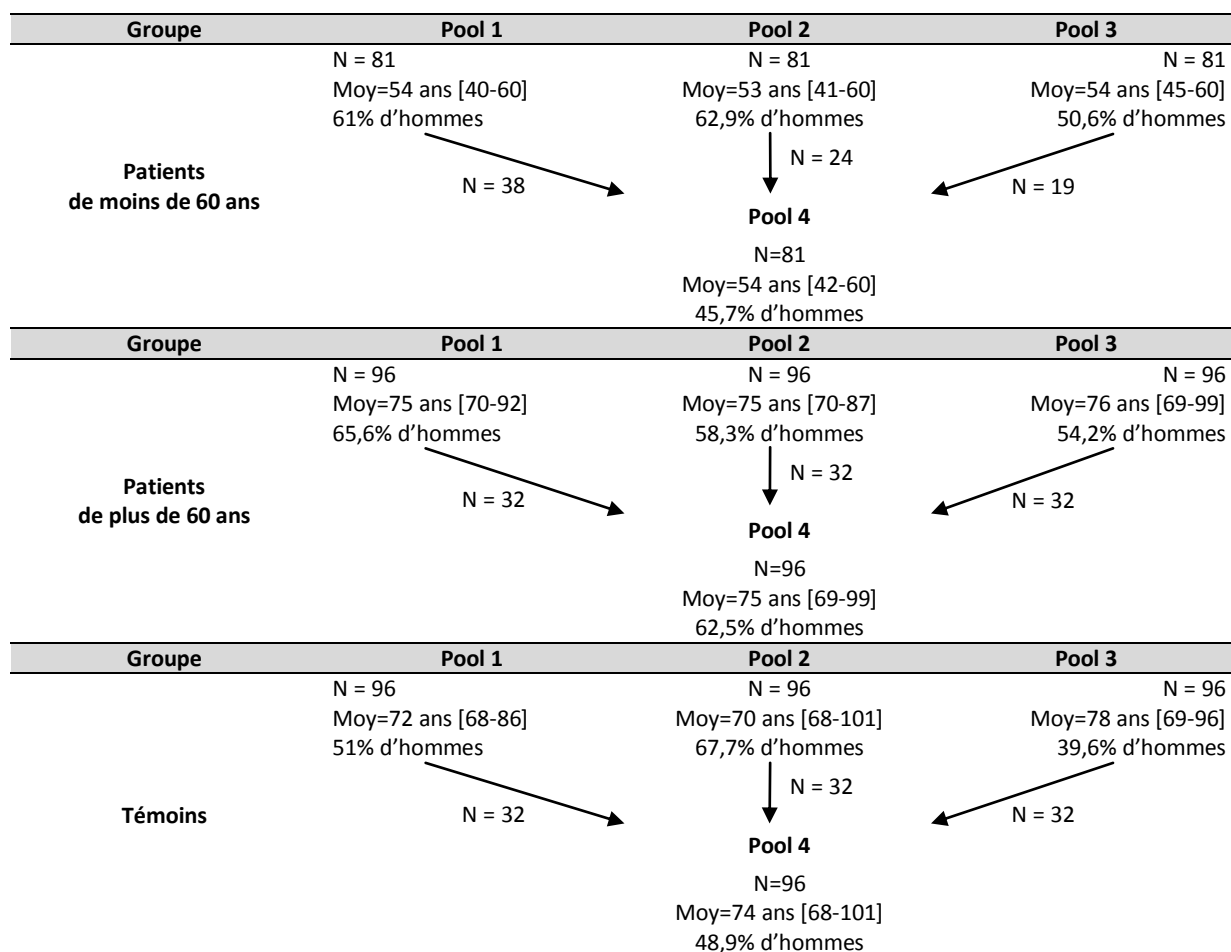


Tableau 23 : Composition des sous-groupes « catégorie d'âge ». Chaque groupe (patients de moins de 60 ans, patients de plus de 60 ans, témoins) est divisé en sous-groupes ou pools (notés pools 1, 2, 3). Un quatrième pool est réalisé (pool 4) en mélangeant certains des ADN des trois pools déjà réalisés. (N, effectifs du pool ; Moy, moyenne d'âge des individus constituant le pool avec entre crochets les bornes de l'effectif ; →, indique le nombre d'ADN du pool 1, 2 ou 3 qui vont constituer le pool 4).

III.2.2.1.2. Groupe d'individus classés par localisation géographique (projet 2) : Groupe « Vendéens »

De la même manière que pour le groupe « catégorie d'âge », nous avons recherché dans nos cohortes des individus provenant d'un même département de France. Notre choix s'est porté sur la Vendée, car, d'une part, nos cohortes sont constituées d'une quantité non négligeable d'individus provenant de ce département, et d'autre part, ce département a été décrit comme ayant eu un très faible flux migratoire durant son histoire. Cela confère aux individus issus de cette zone géographique un fond génétique commun plus important que chez des individus provenant d'autres régions et/ou départements qui ont subi plus de mixité géographique et/ou ethnique.

Nous avons donc recherché tous les individus de notre cohorte habitant ce département et possédant quatre grands-parents provenant également de cette région. Nous avons déterminé que notre cohorte possédait 427 patients (moyenne d'âge = 65 ans [40-98] dont 63% d'hommes) et 191 témoins (moyenne d'âge = 60 ans [41-101] dont 47,6% d'hommes) d'origine Vendéenne. Patients et

témoins devaient être nés et avoir leurs quatre grands-parents également nés dans cette région. Deux sous-groupes ou « pools » de 215 et 212 patients et un « pool » de 191 témoins ont été réalisés (Tableau 24).

Groupe	Pool 1	Pool 2
Patients Vendéens	N = 215 Moy=65 ans [40-87] 69,7% d'hommes	N = 212 Moy=66 ans [41-98] 56,1% d'hommes
Groupe	Pool 1	
Témoins Vendéens	N = 191 Moy=60 ans [41-101] 47,6% d'hommes	

Tableau 24 : Composition des sous-groupes « Vendéens ». Nous avons sélectionné les individus provenant d'un même département, la Vendée. Le groupe de patients a été divisé en deux sous-groupes (notés pool 1 et 2) tandis que le groupe témoin n'a pas été divisé et est donc représenté par un pool unique (pool1) (N, effectif du pool ; Moy, moyenne d'âge des individus constituant le pool avec entre crochets les bornes de l'effectif).

Il est à noter que les pools vendéens ont été constitués pour partie des individus déjà inclus dans les sous-groupes « catégorie d'âge ». Le Tableau 25 montre la provenance et la proportion des individus constituant ces pools.

Sous-groupes catégorie d'âge de provenance		Pool 1 Vendéens	Pool 2 Vendéens
Individus également inclus dans les pools « catégorie d'âge » (Tableau 23)	Pool 1 Moins de 60 ans	N=46 (21,4%) dont 19 appartiennent également au pool 4 moins de 60 ans	
	Pool 2 Moins de 60 ans	N=7 (3,3%)	N=36 (17%) dont 14 appartiennent également au pool 4 moins de 60 ans
	Pool 3 Moins de 60 ans		N=26 (12,3%) dont 10 appartiennent également au pool 4 moins de 60 ans
	Pool 1 Plus de 60 ans	N=36 (16,7%) dont 6 appartiennent également au pool 4 plus de 60 ans	
	Pool 2 Plus de 60 ans	N=11 (5,1%) dont 8 appartiennent également au pool 4 plus de 60 ans	N=23 (10,8%) dont 7 appartiennent également au pool 4 plus de 60 ans
	Pool 3 Plus de 60 ans		N=38 (17,9%) dont 12 appartiennent également au pool 4 plus de 60 ans
Nouveaux ADN		N=115 (53,5%)	N=89 (42%)
Sous-groupes catégorie d'âge de provenance		Pool Témoins Vendéens	
Individus également inclus dans les pools « catégorie d'âge » (Tableau 23)	Pool 1 Témoins Catégorie d'âge	N=5 (2,6%) dont 1 appartient également au pool 4 Témoins Catégorie d'âge	
	Pool 2 Témoins Catégorie d'âge	N=26 (13,6%) dont 12 appartiennent également au pool 4 Témoins Catégorie d'âge	
	Pool 3 Témoins Catégorie d'âge	N=6 (3,1%) dont 1 appartient également au pool 4 Témoins Catégorie d'âge	
	Nouveaux ADN	N=154 (80,7%)	

Tableau 25 : Composition des pools vendéens avec les effectifs de provenance des ADN. Chaque pool « vendéen » est composé pour partie d'ADN constituant également des pools « catégorie d'âge » (N, nombre d'ADN entrant déjà dans la composition des sous-groupes « catégorie d'âge » qui vont entrer dans les sous-groupes « vendéens »).

III.2.2.2. Construction des pools d'ADN pour l'analyse SNP-MaP

Pour réaliser ces pools d'ADN, nous avons utilisé des ADN disponibles dans notre DNAtèque, ajustés à une concentration de 50 ng/μl. Afin que tous les individus sélectionnés pour construire un pool donné contribuent équitablement à ce pool, nous avons mélangé des volumes équivalents (10 μl) d'ADN équidosés. Le mélange d'ADN obtenu a été dosé en triplicata au NanoDrop® ND-1000 Spectrophotometer (NanoDrop Technologies), avant d'être ajusté à 50 ng/μl. Avant cette opération, chaque ADN a subi un test d'amplification, afin d'éliminer les ADN de moindre qualité. Ce test a consisté en la réalisation d'une PCR quantitative en temps réel en utilisant une sonde Taqman de *5'-nuclease assay* (Applied Biosystems, Foster City, CA) permettant le test d'un polymorphisme mono-morphique étudié précédemment au laboratoire (rs2228065) (SHAM *et al.* 2002) ; le résultat d'amplification a été lu sur un lecteur ABI PRISM 7900HT (Applied Biosystems).

III.2.2.3. Hybridation sur puce

La préparation et l'hybridation des échantillons (ici, un échantillon correspond à un pool d'ADN) ainsi, que les réactifs utilisés, ont été réalisés comme recommandé dans le document édité par Affymetrix (Santa Clara, CA), « GeneChip® Mapping 500K Assay Manual »*.

La société Affymetrix a développé deux puces de génotypage (« GenChip Human Mapping 250K-Nsp Array » et « GenChip Human Mapping 250K-Sty Array ») afin d'analyser environ 500000 SNP. Dans notre étude, nous avons fait le choix de n'hybrider que la puce Nsp permettant l'analyse d'environ 250000 SNP et par conséquent de ne pas hybrider la puce Sty. Nous avons, en effet, jugé que la couverture de la puce Nsp et la densité des SNP présents sur celle-ci étaient suffisantes pour notre étude.

III.2.3. Analyses statistiques

III.2.3.1. Calculs des fréquences alléliques

Les pools d'ADN 1, 2 et 3 du projet 1 (« catégorie d'âge ») ont été réalisés en duplicata alors que les pools 4 n'ont été hybridés qu'une fois, ce qui correspond à 21 puces hybridées pour ce projet. Les pools d'ADN du projet 2 (Vendéens) ont été hybridés 3 fois, ce qui donne un total de 9 puces hybridées.

Comme l'expliquent Butcher *et al.* (2005) dans leur publication, les données générées par les puces doivent être analysées d'une manière différente de celle employée communément pour des puces réalisées avec des échantillons individuels. Dans le cas de la méthode SNP-MaP les génotypes

* https://www.affymetrix.com/support/downloads/manuals/500k_assay_manual.pdf

généérés par le logiciel GTYPE (Affymetrix) ne nous ont pas été utiles. En effet, le résultat donné par ce logiciel est un génotype moyen de l'ensemble des génotypes des individus constituant le pool, ce qui ne reflète pas fidèlement la réalité. Dans le cas d'un pool d'ADN, il est préférable de calculer les fréquences alléliques de chaque SNP. Le calcul de ces fréquences est réalisé pour chaque puce à partir des données brutes de fluorescence émise par les sondes correspondant aux SNP (Liu *et al.* 2003). Nous avons donc calculé chacune des fréquences alléliques en suivant la méthode préconisée par Liu *et al.* (2003) (Encadré 1 : Méthode de Liu *et al.* (2003) permettant le calcul de la fréquence allélique à partir de la fluorescence des puces à ADN.).

Calcul de la fréquence allélique à partir de la fluorescence

Les puces à ADN utilisent une unité appelée quartet pour définir le génotype d'un SNP. Le quartet inclut une paire de sondes pour l'allèle A et une paire de sondes pour le second allèle B. Une paire de sondes comprend une sonde sans mésappariement pour le polymorphisme (sonde *match*) et une sonde avec mésappariement pour le polymorphisme (sonde *mismatch*). Afin d'améliorer la sélectivité, le génotype est défini par plusieurs quartets qui possèdent une position différente pour le polymorphisme sur la séquence de la sonde de 25 nucléotides. Ces positions sont en -4, -2, -1, 0, +1, +2, +4, où 0 est le centre de la séquence nucléotidique de la sonde. Tous ces quartets sont présents pour le brin sens et anti-sens de la séquence d'ADN. Cela permet d'avoir au maximum 56 sondes pour déterminer le génotype d'un SNP, mais pour la puce utilisée ici, au maximum 24 sondes peuvent être choisies parmi les 56 pour déterminer le génotype du SNP (Figure 42).

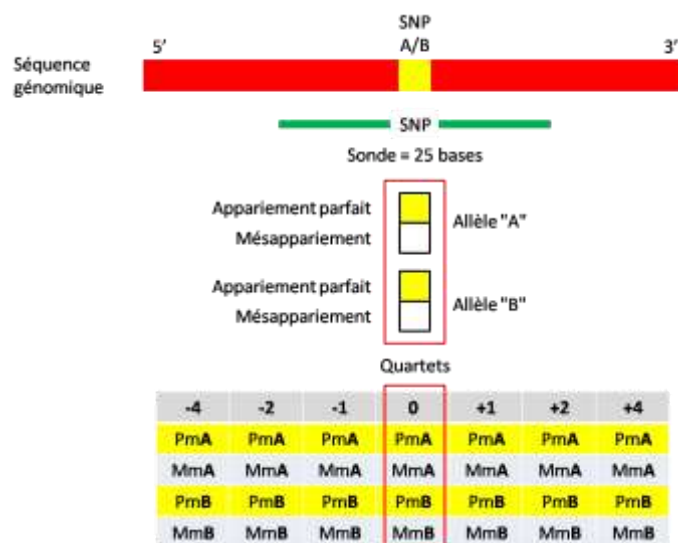


Figure 42 : Les différentes sondes de la puce permettant de calculer la fréquence allélique. Pour chaque SNP, la puce contient au maximum 24 sondes différentes réparties sur le brin sens et anti-sens. Pour une position donnée du polymorphisme sur la sonde, il y aura une sonde sans mésappariement (Pm, *perfect match allele*) pour le polymorphisme et une sonde avec mésappariement (Mm, *mismatch*) pour le polymorphisme et cela pour l'allèle A et l'allèle B, tout en sachant qu'il y a sept positions possibles du polymorphisme sur la sonde.

Afin de calculer la fréquence allélique appelée RAS (*Relative Allele Signal*) de chaque SNP (appelé *i*), il faut réaliser indépendamment pour les deux brins d'ADN différents calculs. Ci-dessous sont explicités les calculs pour le brin sens (noté *(s)*) :

- Il faut calculer la moyenne des intensités des sondes sans mésappariement des deux allèles pour tous les quartets que l'on va noter, $PM_i^{(sA)}$ pour l'allèle A et $PM_i^{(sB)}$ pour l'allèle B.
- Puis la moyenne des intensités pour toutes les sondes avec mésappariement des allèles A et B ($MM_i^{(s)}$)
- De ces différents calculs, la valeur pour $A_i^{(s)}$ pour l'allèle A et $B_i^{(s)}$ pour l'allèle B sont calculées de la sorte :

$$A_i^{(s)} = \max(\text{PM}_i^{(SA)} - \text{MM}_i^{(s)}; 0) \text{ et } B_i^{(s)} = \max(\text{PM}_i^{(SB)} - \text{MM}_i^{(s)}; 0)$$

Ainsi, la valeur retenue pour $A_i^{(s)}$ et $B_i^{(s)}$ sera la plus grande entre 0 et la différence des intensités des sondes avec et sans mésappariement pour les deux allèles.

- Puis, pour le brin considéré, nous allons calculer le signal relatif à l'allèle $s_i^{(s)}$:

$$s_i^{(s)} = A_i^{(s)} / (A_i^{(s)} + B_i^{(s)})$$

Une fois les calculs de signal relatif réalisés pour le brin sens ($s_i^{(s)}$) et anti-sens ($s_i^{(t)}$), une moyenne de ces signaux est réalisée afin d'obtenir la fréquence allélique du SNP.

$$\text{RAS} = (s_i^{(s)} + s_i^{(t)}) / 2$$

Encadré 1 : Méthode de Liu *et al.* (2003) permettant le calcul de la fréquence allélique à partir de la fluorescence des puces à ADN.

A la suite de cette opération, nous avons vérifié pour chaque SNP présent sur la puce 250K Nsp que l'allèle considéré comme minoritaire par Affymetrix était bien celui dont la fréquence était minoritaire dans la population caucasienne. Cette donnée étant fournie par Affymetrix, nous avons pu identifier les SNP de la puce dont l'allèle, décrit comme minoritaire par Affymetrix, avait une fréquence supérieure à 0,5. Ces allèles correspondaient en réalité à des allèles ancestraux, majoritaires, mal annotés par Affymetrix. Nous avons alors corrigé toutes ces « erreurs » d'annotations, afin d'uniformiser les données de fréquences alléliques que nous avons calculées pour l'ensemble des SNP de la puce 250K Nsp. Par cette procédure, nous avons pu exprimer tous les SNP de la puce en fonction de la fréquence de leur allèle minoritaire. Une fois cette opération réalisée pour chaque puce, une moyenne élaguée (moyenne excluant les valeurs extrêmes) des fréquences alléliques de chaque SNP a été calculée pour chaque groupe, à partir des valeurs obtenues pour chaque sous-groupe. Cette étape est illustrée par les Tableau 26 à Tableau 29, qui présentent des exemples de calculs de moyenne élaguée pour deux SNP du chromosome 22. Chaque tableau montre le calcul pour les différentes analyses réalisées.

Méthode de calcul des fréquences moyennes d'allèles minoritaires pour les groupes : « patients de moins de 60 ans », « patients de plus de 60 ans » et témoins													
dbSNP	ch	position	Freq	Pool 1	Duplicata Pool 1	Pool 2	Duplicata Pool 2	Pool 3	Duplicata Pool 3	Pool 4	Moyenne	Moyenne élaguée	MAF
Patients Moins de 60 ans - Groupe catégorie d'âge													
				0,853	0,914	0,862	0,853	0,830	0,890	0,894	0,867	0,870	0,130
Patients de plus de 60 ans - Groupe catégorie d'âge													
rs5746679	22	15460378	0,890	0,910	0,885	0,914	0,918	0,890	0,896	0,922	0,905	0,906	0,094
Témoins - Groupe catégorie d'âge													
				0,865	0,878	0,840	0,877	0,874	0,839	0,902	0,862	0,867	0,133

Tableau 26 : Pour ces deux analyses, la fréquence moyenne élaguée pour l'allèle minoritaire est calculée à partir de celle obtenue pour la catégorie « patients » et comparée à la fréquence moyenne calculée à partir des puces témoins. Les données qui ne rentrent pas en compte dans la moyenne élaguée sont surlignées en noir. Pour ce SNP, il est nécessaire de convertir la moyenne obtenue, car elle ne correspond pas à celle de l'allèle minoritaire, la conversion se fait de la sorte : $\text{MAF} = 1 - \text{moyenne élaguée}$. (dbSNP, numéro de référence du SNP ; ch, chromosome sur lequel se trouve le SNP ; Position, position du SNP sur le chromosome ; Freq, fréquence du SNP dans la population caucasienne dans la base SNP ; Pool 1 à 3 et les duplicata, correspondent aux fréquences calculées à partir des données brutes des puces ; MAF (*Minor allele frequency*), fréquence de l'allèle minoritaire).

Méthode de calcul des fréquences moyennes d'allèles minoritaires pour les groupes : « patients catégorie d'âge » et témoins														
dbSNP	ch	position	Freq	Pool 1	Duplicata Pool 1	Pool 2	Duplicata Pool 2	Pool 3	Duplicata Pool 3	Pool 4	Moyenne	Moyenne élaguée	MAF	
Patients Moins de 60 ans - Groupe catégorie d'âge														
rs5746679	22	15460378	0,890	0,853	0,914	0,862	0,853	0,830	0,890	0,894	0,888	0,890	0,110	
				Patients de plus de 60 ans - Groupe catégorie d'âge										
				0,910	0,885	0,914	0,918	0,890	0,896	0,922				
				Témoins - Groupe catégorie d'âge										
				0,865	0,878	0,840	0,877	0,874	0,839	0,902	0,862	0,867	0,133	

Tableau 27 : Pour cette analyse, la fréquence moyenne élaguée pour l'allèle minoritaire est calculée à partir de celles obtenues pour les catégories « patients » de moins et de plus de 60 ans et comparée à la fréquence moyenne calculée à partir des puces témoins. Les données qui ne rentrent pas en compte dans la moyenne élaguée sont surlignées en noir. Pour ce SNP, il est nécessaire de convertir la moyenne obtenue, car elle ne correspond pas à celle de l'allèle minoritaire, la conversion se fait de la sorte : $MAF=1-moyenne\ élaguée$. (dbSNP, numéro de référence du SNP ; ch, chromosome sur lequel se trouve le SNP ; Position, position du SNP sur le chromosome ; Freq, fréquence du SNP dans la population caucasienne dans la base SNP ; Pool 1 à 3 et les duplicata, correspondent aux fréquences calculées à partir des données brutes des puces ; MAF (*Minor allele frequency*), fréquence de l'allèle minoritaire).

Méthode de calcul des fréquences moyennes d'allèles minoritaires pour les groupes : « patients vendéens pool1 », « patients vendéens pool2 » et témoins														
dbSNP	ch	position	Freq	Puce 1	Puce 2	Puce3	Moyenne	Moyenne élaguée	MAF					
Patients Pool 1 - Groupe vendéens														
rs175148	22	15690488	0,420	0,421	0,407	0,374	0,400	0,407	0,407					
				Patients Pool 2 - Groupe vendéens										
				0,390	0,329	0,375	0,364	0,375		0,375				
				Pool Témoins - Groupe vendéens										
				0,349	0,292	0,330	0,324	0,330	0,330					

Tableau 28 : Pour ces deux analyses, la fréquence moyenne élaguée pour l'allèle minoritaire est calculée à partir de celle obtenue pour le pool « patient » correspondant et comparée à la fréquence moyenne calculée à partir des puces témoins. Les données qui ne rentrent pas en compte dans la moyenne élaguée sont surlignées en noir. Pour ce SNP, il n'est pas nécessaire de convertir la moyenne obtenue, car elle correspond à celle de l'allèle minoritaire. (dbSNP, numéro de référence du SNP ; ch, chromosome sur lequel se trouve le SNP ; Position, position du SNP sur le chromosome ; Freq, fréquence du SNP dans la population caucasienne dans la base SNP ; Pool 1 à 3 et les duplicata, correspondent aux fréquences calculées à partir des données brutes des puces ; MAF (*Minor allele frequency*), fréquence de l'allèle minoritaire).

Méthode de calcul des fréquences moyennes d'allèles minoritaires pour les groupes : « patients vendéens » et témoins														
dbSNP	ch	position	Freq	Puce 1	Puce 2	Puce3	Moyenne	Moyenne élaguée	MAF					
Patients Pool 1 - Groupe vendéens														
rs175148	22	15690488	0,420	0,421	0,407	0,374	0,382	0,387	0,387					
				Patients Pool 2 - Groupe vendéens										
				0,390	0,329	0,375								
				Pool Témoins - Groupe vendéens										
				0,349	0,292	0,330	0,324	0,330	0,330					

Tableau 29 : Pour cette analyse, la fréquence moyenne élaguée pour l'allèle minoritaire est calculée à partir de celles obtenues pour les deux pools vendéens et comparée à la fréquence moyenne calculée à partir des puces témoins. Les données qui ne rentrent pas en compte dans la moyenne élaguée sont surlignées en noir. Pour ce SNP, il n'est pas nécessaire de convertir la moyenne obtenue, car elle correspond à celle de l'allèle minoritaire. (dbSNP, numéro de référence du SNP ; ch, chromosome sur lequel se trouve le SNP ; Position, position du SNP sur le chromosome ; Freq, fréquence du SNP dans la population caucasienne dans la base SNP ; Pool 1 à 3 et les duplicata, correspondent aux fréquences calculées à partir des données brutes des puces ; MAF (*Minor allele frequency*), fréquence de l'allèle minoritaire).

Les SNP présentant une fréquence allélique pour l'allèle minoritaire (MAF, *minor allele frequency*) inférieure à 5% ont été exclus de l'analyse, car les puces n'assuraient pas une grande

fiabilité pour ces données (DOCHERTY *et al.* 2007). Nous avons également exclu tous les SNP pour lesquels nous avons obtenu moins de deux données de fréquence pour les différentes analyses réalisées (« Patients catégorie d'âge », « Patients moins de 60 ans », « Patients plus de 60 ans », « Patients vendéens », « Patients vendéens pool1 », « Patients vendéens pool2 ») ; le but était ici de ne pas prendre en compte des SNP avec des fréquences alléliques aberrantes.

III.2.3.2. Calculs des OR et p-value

A partir des fréquences alléliques calculées et connaissant le nombre d'individus composant chaque pool ainsi que le nombre de réplifications d'hybridation réalisées, nous avons pu calculer un nombre théorique d'allèles minoritaires et majoritaires pour les patients et les témoins pour chaque SNP de la puce Affymetrix 250K Nsp (Figure 43).

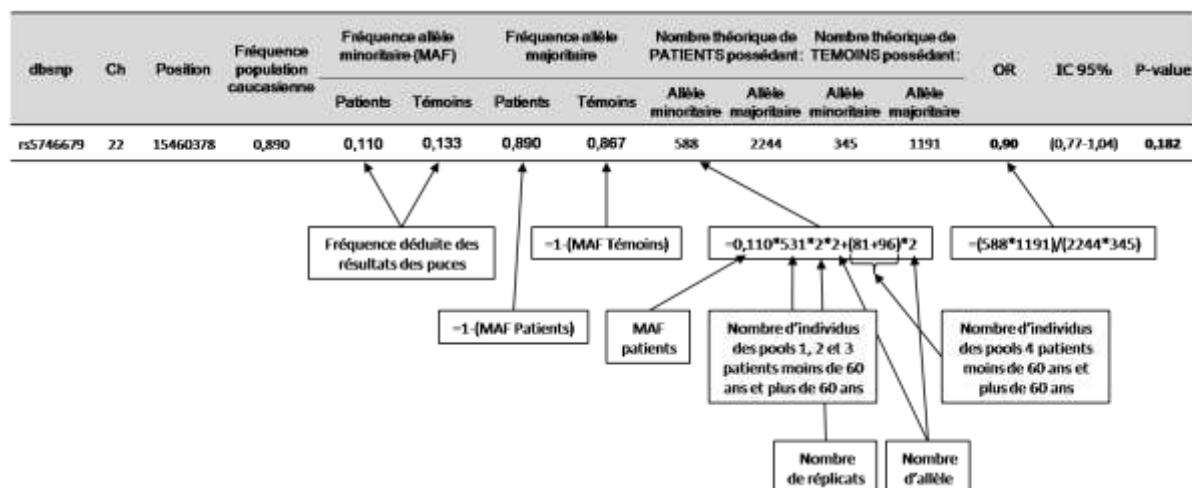


Figure 43 : Exemple de calcul de l'OR d'un SNP à partir des fréquences alléliques obtenues par l'analyse des puces. (dbSNP, numéro de référence du SNP ; ch, chromosome sur lequel se trouve le SNP ; Position, position du SNP sur le chromosome ; MAF (*Minor allele frequency*), fréquence de l'allèle minoritaire ; IC, Intervalle de confiance ; OR, Odds Ratio).

A partir de ces répartitions alléliques théoriques entre patients et témoins, nous avons réalisé pour tous les polymorphismes et toutes les analyses un test pour l'équilibre d'Hardy-Weinberg, par un test de χ^2 . Une évaluation de l'association de l'allèle variant et du risque de développer un cancer colorectal a été déterminée en calculant l'Odds Ratio (OR) de chaque polymorphisme avec un intervalle de confiance (IC) fixé à 95%. La formule permettant le calcul de l'Odds Ratio est la suivante :

$$OR = \frac{(\text{patients allèle minoritaire} \times \text{témoins allèle majoritaire})}{(\text{patients allèle majoritaire} \times \text{témoins allèle minoritaire})}$$

Ce test permet de déterminer l'association à la pathologie de l'allèle variant du polymorphisme en comparant les allélotypes du groupe « patient » et du groupe « témoin ». Si l'OR

est supérieur à 1, l'allèle variant est prédisposant ; s'il est inférieur à 1, il est protecteur. Cette association est considérée comme significative si, et seulement si, la p-value* est inférieure à 0,05. La p-value représente en quelque sorte, un indicateur de significativité pour un test d'association entre un SNP et le risque de développer un cancer colorectal. Plus la p-value est faible, et plus l'association observée est significative.

III.2.4. Choix des SNP d'intérêt

III.2.4.1. Graphique : $-\log_{10}(p\text{-value}) = f(\text{position chromosomique})$

Pour chaque SNP, nous avons reporté sur un graphique l'indice de significativité de son test d'association au risque de prédisposition aux cancers colorectaux. Pour des raisons pratiques, nous avons reporté non pas la valeur de la p-value correspondant au test d'association, mais la valeur de « $-\log_{10}(p\text{-value})$ ». Plus cette valeur est élevée, plus l'association observée est significative, qu'elle soit protectrice ou prédisposante. Sur un même graphique, nous avons reporté les indices de significativité des associations de tous les SNP d'un même chromosome (HUNTER *et al.* 2007; MARCHINI *et al.* 2007; MOFFATT *et al.* 2007; THOMAS *et al.* 2008).

III.2.4.2. Analyses bioinformatiques

Pour chaque chromosome, un seuil de significativité a été fixé pour la valeur « $-\log_{10}(p\text{-value})$ », au-delà de laquelle toutes les associations entre les SNP correspondants et le risque de cancer colorectal ont été considérées comme très probablement significatives, ce qui revient à sélectionner les meilleurs SNP. Tous les SNP présentant une valeur supérieure à ce seuil ont été sélectionnés pour nos analyses ultérieures.

Nous avons ensuite déterminé la position de chacun de ces SNP sélectionnés par rapport aux gènes présents dans la même région chromosomique en interrogeant les bases de données publiques (dbSNP, Ensembl) et le logiciel d'analyse des puces Affymetrix (GTYPE).

Nous avons ainsi retenu, pour la suite de nos analyses, les premiers gènes rencontrés en amont et en aval des SNP sélectionnés, dans une région de 2Mb au maximum centrée sur ce polymorphisme (Figure 44a). Lorsque deux SNP en association avec le CCR se trouvaient à moins de 1Mb l'un de l'autre, nous avons retenu tous les gènes situés dans l'intervalle entre ces polymorphismes, ainsi que le premier gène rencontré à moins de 1 Mb en amont et en aval de cet intervalle (Figure 44b).

* P-value : Probabilité d'obtenir le résultat au hasard sous l'hypothèse nulle. Par exemple, une p-value égale à 1% correspond à une probabilité de 1% d'obtenir la même valeur par hasard.

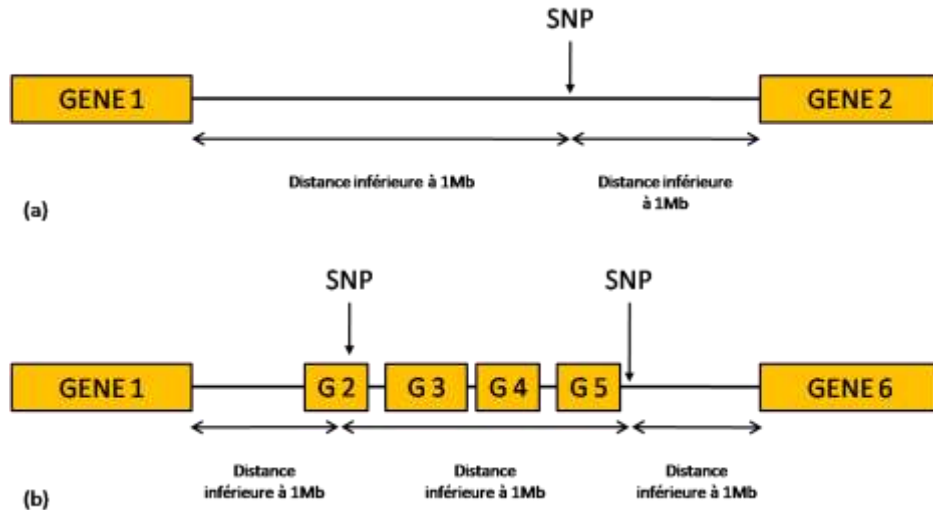


Figure 44 : (a) Mode de sélection des gènes dans le cas d'un SNP intergénique. Ici, nous aurions sélectionné les gènes 1 et 2, car ils sont les premiers gènes rencontrés à moins de 1Mb du polymorphisme à analyser. (b) Mode de sélection des gènes dans le cas de deux SNP distants de moins de 1Mb. Dans ce cas, un des SNP est intragénique (dans le gène 2) et le deuxième est situé en aval à moins de 1Mb. Les gènes sélectionnés sont donc les gènes 2, 3, 4, 5 et 6 situés dans l'intervalle entre les deux SNP. Le gène 1 n'est pas pris en compte dans cette analyse, car le SNP le plus proche trouvé comme associé au cancer appartient à un gène.

Dans certains cas, ces SNP ont été mis en évidence par la même analyse (exemple : analyse du groupe d'individus classés par catégorie d'âge) ; les SNP ainsi identifiés, distants de moins de 1Mb, étaient en déséquilibre de liaison, très probablement liés à la mutation causale ou au polymorphisme associé au CCR (Figure 45). Dans d'autres cas, les différents SNP identifiés pour un même gène ont été mis en évidence séparément par des analyses portant sur des groupes différents (exemple : l'analyse du groupe des individus classés par âge a permis d'identifier le SNP1, tandis que l'analyse des individus vendéens a fait ressortir le SNP2).

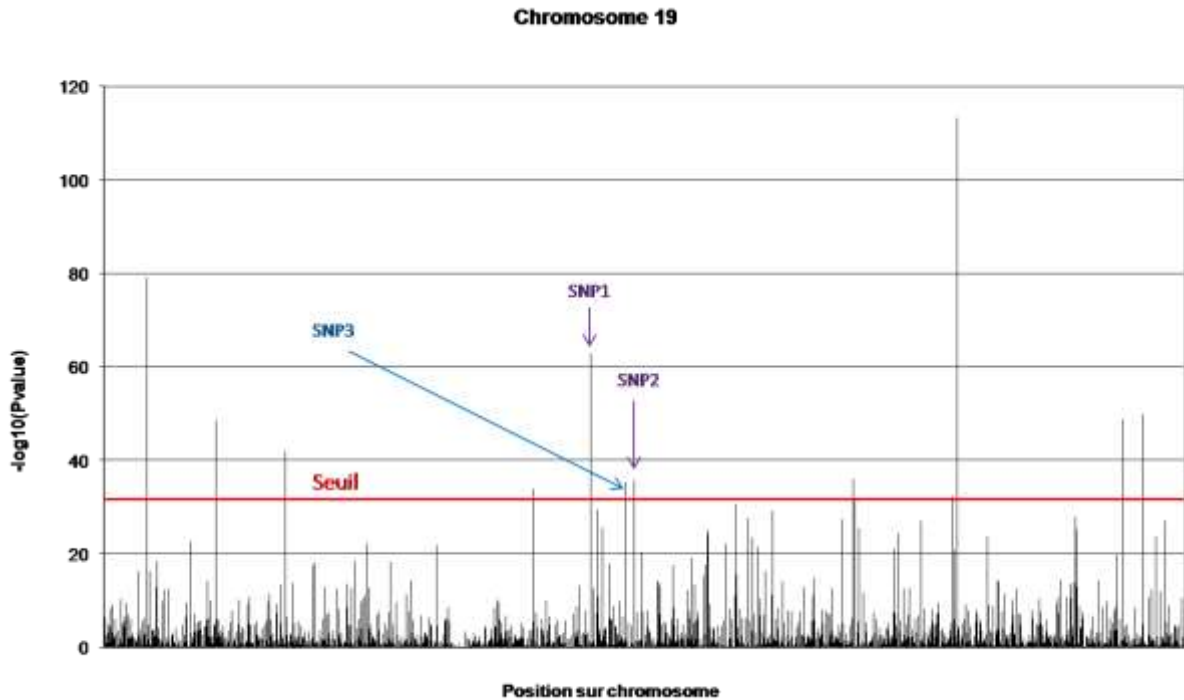


Figure 45 : Représentation du graphique $-\log_{10}(p\text{-value})=f(\text{position chromosomique})$ pour le chromosome 19 et pour l'analyse « individus classés par catégorie d'âge ». Sur ce graphique, une région de moins de 1MB délimitée par les SNP 1 et 2 contient également un SNP en association avec le CCR, le SNP3. Comme tous les SNP dépassant le seuil de significativité sont pris en compte, le SNP3 sera analysé de la même manière qu'un SNP isolé et pourra faire ressortir des gènes déjà sélectionnés pour leur proximité aux SNP1 et SNP2.

Chaque gène retenu pour sa proximité à un SNP associé significativement à un risque de cancer colorectal a fait l'objet d'une recherche bibliographique dans PubMed*. Nous avons ainsi réalisé un filtre de sélection pour déterminer les gènes candidats connus pour leur rôle dans la carcinogénèse colorectale ou étudiés dans le cadre de la prédisposition au cancer (colorectal ou autre). Pour cette recherche, nous avons soumis le mot clé « cancer » au moteur de recherche de PubMed, en limitant notre requête au titre et/ou au résumé de l'article. Par ce criblage large, nous avons pu retenir les gènes présélectionnés pour leur position chromosomique ayant déjà fait l'objet d'une publication scientifique contenant le mot cancer dans son titre et/ou son résumé. Toutes les publications prises en compte pour cette analyse sont antérieures au mois de février 2008.

III.2.4.3. Critères de sélection d'un SNP à analyser en détail

Nous avons réalisé une « méta-analyse » des six analyses indépendantes et nous avons croisé leurs données. Pour classer par ordre d'intérêt les associations au cancer colorectal correspondant à chacun des SNP présélectionnés, nous avons attribué un score à chacune de ces associations et, par voie de conséquence, à chacun des SNP en tenant compte des cinq critères détaillés ci-après.

* <http://www.ncbi.nlm.nih.gov/pubmed/>

III.2.4.3.1. Nombre de fois où le SNP apparaît dans les différentes analyses des projets 1 et 2

Pour ce critère, nous avons compté le nombre de fois où nous avons retrouvé le SNP associé significativement au CCR dans les six analyses indépendantes (1 point par apparition). Le fait qu'un SNP ressorte associé au CCR dans les deux projets (catégorie d'âge et vendéens) apportait naturellement plus de crédit à ce candidat. Un SNP n'apparaissant que dans un seul projet ne pouvait être pris en compte pour les analyses ultérieures. Ainsi, dans notre système d'analyse, un polymorphisme qui ressort dans les deux projets est un meilleur candidat et plus le nombre d'analyses dans lequel il ressort est élevé, plus son score augmente. Nous avons fixé à trois le nombre d'apparitions minimales d'un SNP à sélectionner (deux apparitions dans un projet et une dans le second). Cette méthode nous permet d'augmenter nos chances de sélectionner le meilleur marqueur polymorphique possible d'un gène de prédisposition.

III.2.4.3.2. Concordance des OR calculés pour le SNP entre les projets 1 et 2

Si, d'un projet à l'autre, les OR calculés sont concordants et signent tous, soit un effet protecteur, soit un effet prédisposant du SNP, six points sont attribués au SNP. Dans le cas où l'effet du SNP varie d'une analyse à l'autre (ex : protecteur pour un projet et prédisposant pour le second), seulement trois points sont attribués au SNP. Dans ce dernier cas, la prédisposition liée au SNP est vraisemblablement très faible, car trop variable.

III.2.4.3.3. Pertinence du gène candidat proche du SNP avec le cancer

Dans la carte du gène produite par le site « Entrez Gene » de NCBI* le résumé de la fonction du gène permet d'attribuer un score en fonction de la présence de mots-clés listés ci-dessous :

- Les termes dont la définition même renvoie à la notion de cancer, tels que « cancer » ou « tumeur » confèrent deux points,
- Les termes suggérant un lien potentiel, direct ou indirect, au cancer, tels que « cycle cellulaire »,..., ou « *RNA binding* » confèrent un point.

III.2.4.3.4. Pertinence bibliographique avec le cancer colorectal du gène proche du SNP

Pour préciser davantage encore la valeur du gène candidat dans le contexte thématique de notre étude, à savoir le cancer colorectal, nous avons choisi d'attribuer des points supplémentaires au gène proche d'un SNP trouvé associé, en fonction du nombre de publications en rapport avec le CCR dans lesquelles il a été étudié. Pour cela une recherche bibliographique dans la base de données

* <http://www.ncbi.nlm.nih.gov/sites/entrez>

PubMed a été réalisée. Les mots-clés utilisés étaient : le « nom du gène » et « cancer colorectal » ; ces mots devaient être présents dans le titre et/ou le résumé des publications scientifiques. Un nombre de publications supérieur ou égal à cinq rapportait deux points supplémentaires au gène et 1 point était donné pour un nombre de publications situé entre un et quatre.

III.2.4.3.5. Localisation précise du SNP par rapport au gène candidat

Grâce au logiciel SNPBrowser v.4.0 (Applied Biosystems), nous avons regardé si le SNP se situait dans le gène lui-même, était lié au gène ou en déséquilibre de liaison avec des SNP présents dans le gène (deux points) ou au contraire s'il n'était pas lié du tout au gène (0 point). De cette manière, nous avons assuré un bonus de points aux polymorphismes dont nous pensions qu'ils pourraient avoir un impact direct sur la fonction du gène. Bien sûr, par cette méthode nous n'avons pas valorisé les SNP situés dans une possible région de régulation par exemple.

Nous avons calculé le total de points obtenus pour chaque SNP, puis nous avons fixé une limite de points minimale nous garantissant théoriquement l'obtention d'un bon SNP candidat, afin de ne garder que les SNP dont les scores dépassaient cette limite.

III.2.5. 5'-nuclease assay

Pour valider les associations mises en évidence précédemment, ou pour valider sur notre cohorte des associations trouvées dans des publications (BRODERICK *et al.* 2007), les polymorphismes ont été analysés par la technique de 5'-nuclease assay sur la totalité de nos cohortes (1121 contrôles, 1023 patients). Les sondes utilisées font partie du panel de sondes MGB désignées par Applied Biosystems : la première sonde *reporter* est marquée par le fluorochrome VIC et la seconde par le fluorochrome FAM, alors que la sonde *quencher* n'est pas marquée par une molécule fluorescente.

dbSNP	Gène	Séquence encadrant le SNP
rs2189684	SPAM1	CTATCCTGCTGATCCCCACCTGCC[G/T]TTGGCTACAAATTACTTTTTCT
rs17289655	DDX58	ACGCAAAGTGTCAAAAACACAGGG[A/G]AAGTAACACCTAATTCTCCAGGGA
rs4773171	COL4A2	TTGTATCTAAAACACTGTTGGGGTG[A/G]ATAACTCTTAGAGGTTTGAACCCAG
rs2955788	MTMR10	GTGCAATGAAGATGCAAAGAGAAG[G/T]CCCCTAECTCTCGCTGCAGAGAGAG
rs2306907	CDH13	TTCCGTGCCTTTCTGTCTTGCCCG[A/G]TAGGCACCACAGTGATGCGGATGAC
rs3760454	UTP6	TCTTCTACTACTGTTCTTCTC[C/T]GGATCAGCTCCAAAAGATTAATTC
rs8099984	ZNF543	TCATAAGGGTTTTGTAATTTTCAGC[C/T]TATAGATTATGGACATATTTAAAA
rs4939827	SMAD7	TCACAGCCTCATCCAAAAGAGGAAA[C/T]AGGACCCAGAGCTCCCTCAGACTC

Tableau 30 : Séquences de référence pour l'analyse des polymorphismes d'intérêt.

Les réactions ont été réalisées en plaques 384 puits dans un volume final de 6 µl, incluant : 2,25 µl de Taqman® Genotyping Master Mix (2X), 0,11 µl de Taqman® Assay Mix (40X), 1,64 µl d'eau et 20ng d'ADN génomique à 10 ng/µl. Les plaques ainsi réalisées ont été passées dans un thermocycleur icycler (Bio-Rad laboratories, Hercules, CA) avec une dénaturation initiale de 10 minutes à 95°C, suivie par 45 cycles incluant une étape de dénaturation à 92°C pendant 15 secondes

et une étape d'hybridation/extension à 60°C de 1 minute. Après cette amplification, la fluorescence en point final a été lue par un appareil ABI 7900HT d'Applied Biosystems. Les génotypes ont été assignés en utilisant le logiciel de discrimination allélique SDS v2.3 (Applied Biosystems).

Tous les polymorphismes analysés ont été testés pour l'équilibre d'Hardy-Weinberg par un test de χ^2 entre les patients et les témoins. La prédisposition de chaque polymorphisme a été déterminée en calculant l'Odds Ratio de ses génotypes avec un intervalle de confiance fixé à 95%.

III.2.6. Etude de réplication

Les polymorphismes trouvés comme étant prédisposants, lors de l'analyse par *5'-nuclease assay* sur notre population caucasienne issue de France, ont été envoyés à des collaborateurs canadiens (équipe de Zanke *et al.*) pour réplication sur des populations différentes.

III.2.6.1. Population de réplication : Zanke *et al.* (population Arctic)

III.2.6.1.1. Recrutement

De manière résumée, cette population, constituée comme décrit dans l'article de Zanke *et al.* (2007), est issue de 3 recrutements distincts : un d'Ontario (Canada), un autre du Newfoundland (Canada), un troisième de Seattle (USA). Chacune des trois sources de recrutement présente ses propres spécificités, mais les patients recrutés sont tous atteints de CCR, caucasiens, et, en règle générale, ont entre 20 et 74 ans (pas de précision sur le type exact de cancer). Les témoins ont été recrutés dans la population générale par téléphone. Chaque personne ayant accepté de participer à l'étude a subi un prélèvement sanguin et a dû répondre à un questionnaire médical portant sur ses antécédents familiaux et à un questionnaire épidémiologique portant notamment sur ses habitudes alimentaires.

III.2.6.1.2. Génotypage

Le génotypage des individus a été effectué par hybridation sur le panel de puces « GenChip® Human Mapping 500K Array » (Affymetrix) en suivant le protocole « GeneChip® Mapping 500K Assay Manual »*.

III.2.6.1.3. Analyse statistique

Tous les polymorphismes analysés ont été testés pour l'équilibre d'Hardy-Weinberg par un test de χ^2 entre les patients et les témoins. La prédisposition de chaque polymorphisme a été déterminée en calculant l'OR de ses génotypes avec un intervalle de confiance fixé à 95%.

* https://www.affymetrix.com/support/downloads/manuals/500k_assay_manual.pdf

III.2.7. Etude comparative de l'expression des gènes d'intérêt entre les tissus sains et tumoraux

III.2.7.1. Principe

III.2.7.1.1. Recrutement des échantillons et extraction de l'ARN et de l'ADN

Afin de connaître l'effet des gènes d'intérêt dans le cancer colorectal, nous avons étudié leur expression dans le tissu sain et le tissu cancéreux de différents patients. Pour cela, nous avons fait appel à la tumorothèque de Nantes (CHU G. et R. Laënnec, Bd Jean Monod, St Herblain, 44093 Nantes cedex) qui a pu nous fournir les tissus sains et cancéreux pour 36 individus atteints de cancer colorectal. L'ARN et l'ADN de chaque tissu a été extrait par du TRI Reagent® Solution (Ambion) en utilisant les protocoles recommandés par le fabricant^{*†}.

Tous les ADN et ARN obtenus ont été dosés en triplicata au NanoDrop® ND-1000 Spectrophotometer (NanoDrop Technologies) puis ajustés à 10 ng/μl.

III.2.7.1.2. 5'-nuclease Assay

Afin de réaliser une corrélation entre le génotype et l'expression des gènes étudiés, le génotypage des prélèvements a été réalisé par *5'-nuclease assay* comme précédemment expliqué, en utilisant les mêmes sondes.

III.2.7.1.3. PCR quantitative en temps réel

L'analyse de l'expression des différents tissus a été réalisée par un kit, disponible chez Applied Biosystems, qui permet d'obtenir les résultats d'expression (C_T^{\ddagger}) en une seule étape, en partant de l'ARN (Taqman® RNA-to-CT™ 1-Step Kit). Les réactions ont été réalisées en plaques 384 puits, avec des mélanges réactionnels composés de la sorte : 0,25 μl de Taqman® RT Enzyme Mix (40X), 5 μl de Taqman® RT-PCR Mix (2X), 0,5 μl de Taqman® Gene Expression Assay (20X) et 4,25 μl d'ARN à 10 ng/μl. Pour une plus grande précision dans l'évaluation de l'expression des gènes d'intérêt, chaque échantillon a été analysé en duplicata. La fluorescence émise a été analysée en temps réel par un lecteur ABI 7900HT (Applied Biosystems), durant tout le programme qui comprend les étapes suivantes : une transcription reverse à 48°C pendant 15 minutes, puis une activation de l'ADN polymérase pendant 10 minutes à 95°C suivie de 40 cycles comprenant chacun une

* http://www.ambion.com/techlib/prot/bp_9738.pdf pour l'ADN

† <http://www.ambion.com/techlib/append/supp/tri.pdf> pour l'ARN

‡ C_T , *Cycle Threshold* pour « cycle du seuil », correspond au nombre de cycles à partir duquel le signal de fluorescence associé à l'amplification par PCR dépasse le seuil

dénaturation à 95°C pendant 15 secondes et une étape d'hybridation/extension à 60°C pendant une minute.

Plusieurs contrôles internes ont été amplifiés en même temps que les échantillons : un gène rapporteur, la β -Actine, a été analysé pour chaque échantillon et, sur chaque plaque, un ARN calibrateur a été déposé pour chaque sonde, à différentes concentrations, afin de pouvoir comparer différentes analyses entre elles (qu'elles proviennent de la même plaque ou de plaques différentes). Les données d'expression ont ensuite été analysées par le logiciel SDS v2.3 (Applied Biosystems) qui permet de déterminer le C_T de chaque échantillon.

III.2.7.1.4. Analyses des données d'expression

L'analyse des C_T obtenus a été faite selon la méthode $2^{-\Delta\Delta C_T}$ décrite par Livak et Schmittgen (2001). Cette méthode permet de comparer deux conditions entre elles, ici l'expression du gène dans les tissus cancéreux et dans les tissus sains.

Echantillon	C_T Actine	C_T Actine Moyen	C_T CDH13	C_T CDH13 Moyen	$2^{-\Delta\Delta C_T}$	$\frac{2^{-\Delta\Delta C_T \text{ Cancer}}}{2^{-\Delta\Delta C_T \text{ Normal}}}$
ARN calibrateur	20,91		30,05			
Duplicata ARN calibrateur	20,42	20,66	30,02	30,03		
Tissu Tumoral	18,50		26,07			
Duplicata Tissu tumoral	18,01	18,26	25,96	26,01		
Tissu Normal	19,17		27,46			
Duplicata Tissu normal	18,88	19,02	26,99	27,22		
					3,06	1,36
					2,25	

Figure 46 : Exemple de calcul du $2^{-\Delta\Delta C_T}$ pour un individu. Les données permettant de calculer les moyennes sont encadrées en bleu ; les données permettant le calcul des ΔC_T et donc du $2^{-\Delta\Delta C_T}$ sont encadrées en rouge. Les données permettant le calcul du rapport des $2^{-\Delta\Delta C_T}$ du tissu tumoral et du tissu sain sont encadrées en vert.

Dans un premier temps, nous avons calculé la moyenne entre les duplicata des C_T des tissus tumoraux, normaux et de l'ARN calibrateur. Nous avons ensuite calculé les valeurs ΔC_T de l'actine et du gène à étudier pour les tissus normaux et tumoraux. Le ΔC_T correspond à la variation du C_T relatif à l'actine ou au gène à étudier, entre l'ARN calibrateur et l'ARN tumoral ou entre l'ARN calibrateur et l'ARN extrait du tissu normal. Puis, nous avons calculé pour le tissu sain et le tissu tumoral la variation des valeurs ΔC_T , c'est-à-dire le $\Delta\Delta C_T$, entre le gène à étudier et l'actine. Les rapports des $\Delta\Delta C_T$ de tissus tumoraux et normaux nous ont donné le rapport d'expression du gène à étudier (dans notre exemple de la Figure 46 : *CDH13*) entre les tissus tumoraux et sains. Si ce rapport est supérieur à 1, le gène est surexprimé dans la tumeur, et s'il est inférieur à 1, il est sous-exprimé.

Une moyenne de ces rapports d'expression a été calculée afin de constater la variation d'expression du gène de manière générale entre tissu sain et tissu tumoral. D'autres moyennes de ces rapports d'expression ont également été réalisées en fonction du génotype des échantillons, afin de mettre en avant une éventuelle corrélation entre le génotype du SNP et l'expression du gène. Toutes les moyennes présentent un intervalle de confiance de 95%.

III.2.7.2. Etude des transcrits du gène d'intérêt

À partir des ARN extraits précédemment, des RT-PCR ont été effectuées afin de vérifier la présence ou l'absence de transcrits alternatifs chez les patients présentant un génotype variant. Pour cela, deux couples d'amorces ont été synthétisés : le premier couple (couple n°1) devait permettre de mettre en évidence une possible altération d'épissage au niveau du site donneur d'épissage dans l'intron 6. Le second couple (couple n°2) permettait l'amplification de l'intégralité de la partie codante du gène. La transcription reverse a été réalisée avec le kit Thermoscript™ (Invitrogen) en utilisant comme amorces des oligonucléotides dT et en suivant le protocole disponible sur internet*.

Couples	Séquence
Couple n°1	5'-TGTC AATGGCAA AACTCTCG-3' dans l'exon 5 5'-CAGTCCAGCCATATCTTGAGC-3' dans l'exon 7
Couple n°2	5'-CTCGAGCCCATACTTCAAGG-3' 5'-TGGAAACTTGGGAGTCAAGC-3'

Tableau 31 : Séquence des couples d'amorces permettant de vérifier la présence de transcrit alternatif.

La PCR a quant à elle été réalisée grâce à l'ADN Polymérase Taq Platinum™ (Invitrogen) et en suivant le protocole du fournisseur[†].

La PCR permettant l'amplification des exons met en jeu le programme suivant : 2 minutes de dénaturation initiale à 94°C, puis 40 cycles comprenant une dénaturation de 30 secondes à 94°C, puis une hybridation des amorces de 30 secondes à 57,5°C et une élongation d'une minute. Le programme se termine par une élongation finale de 10 minutes à 72°C. La PCR de la totalité de la partie codante repose sur un programme similaire, exception faite de la température d'hybridation qui est à 59°C et du temps d'élongation, allongé à 2 minutes 30 secondes compte tenu de la longueur attendue du produit PCR.

La révélation des amplifications PCR a été faite sur un gel d'agarose à 1,2% sur lequel ont été déposés 5 µl de produit de PCR additionnés à un tampon de charge. Des marqueurs de taille ont également été déposés en même temps que les échantillons, afin de déterminer la taille des produits PCR obtenus.

* http://www.invitrogen.com/etc/medialib/en/filelibrary/pdf.Par.63607.File.dat/thermoscriptsystem_man.pdf

† http://www.invitrogen.com/etc/medialib/en/filelibrary/pdf.Par.63607.File.dat/thermoscriptsystem_man.pdf

Un séquençage automatique a également été réalisé à partir des produits d'amplification de la PCR faite avec le couple n°1 d'amorces. Cette technique permet une révélation plus précise que les gels d'agarose, permettant la détection de réarrangements (délétions, insertions,...) qui seraient passés inaperçus par une révélation classique.

III.2.8. Analyse par le logiciel GenePool

Très récemment, nous avons eu connaissance de l'existence de ce logiciel qui utilise une approche basée sur l'analyse de clusters et sur le calcul de la force relative d'allèle, RAS (*Relative Allele Strength*). De manière très simplifiée, la valeur de RAS correspond au rapport d'intensités des fluorescences de l'allèle majoritaire A et de l'allèle minoritaire B d'un SNP.

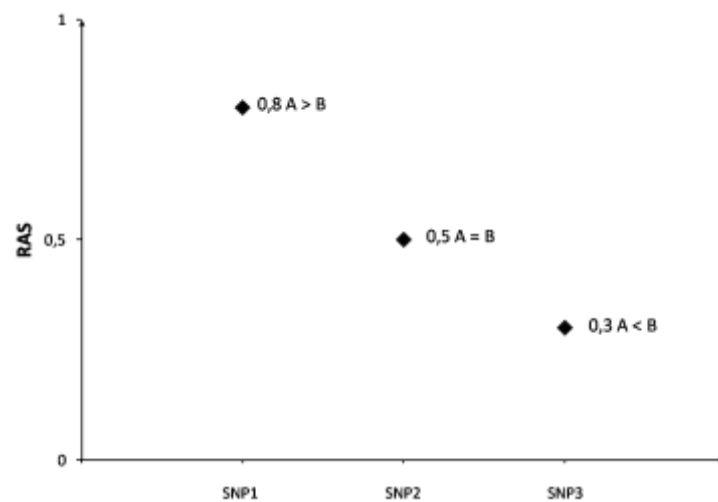


Figure 47 : Représentation simplifiée des résultats de l'analyse GenePool. Une valeur de RAS supérieure à 0,5 signifie que, dans le pool d'ADN testé, l'allèle majoritaire A est en excès par rapport à l'allèle B. Une valeur de RAS de 0,5 signe une population équivalente des deux allèles tandis qu'une valeur inférieure à 0,5 indique une prédominance de l'allèle B.

Pour chaque SNP de la puce Affymetrix, le logiciel GenePool va ainsi calculer des valeurs de RAS, qui correspondent aux quartets de sondes (Encadré 1), ces valeurs définissant un cluster (Figure 48). Le logiciel va répéter ce calcul pour chaque pool d'ADN inclus dans l'analyse. Chaque pool va alors être caractérisé par un cluster de valeur RAS. Une méthode mathématique telle que la distance silhouette (LOVMAR *et al.* 2005) va permettre de quantifier la séparation entre les pools correspondant à des pools patients et ceux correspondant à des pools témoins. Plus le score de silhouette sera proche de 1, meilleure sera la séparation entre les clusters de patients et de témoins. Plus ce score se rapprochera de 0, ou atteindra des valeurs négatives, moins la séparation sera bonne (Figure 48). Tous les SNP d'une même puce vont ainsi être classés par GenePool en fonction de leur score de silhouette, ceux ayant les meilleurs scores correspondant aux SNP présentant l'association la plus significative au risque de CCR. En d'autres termes, plus le score attribué par GenePool est élevé plus la probabilité que le SNP soit associé au CCR est forte.

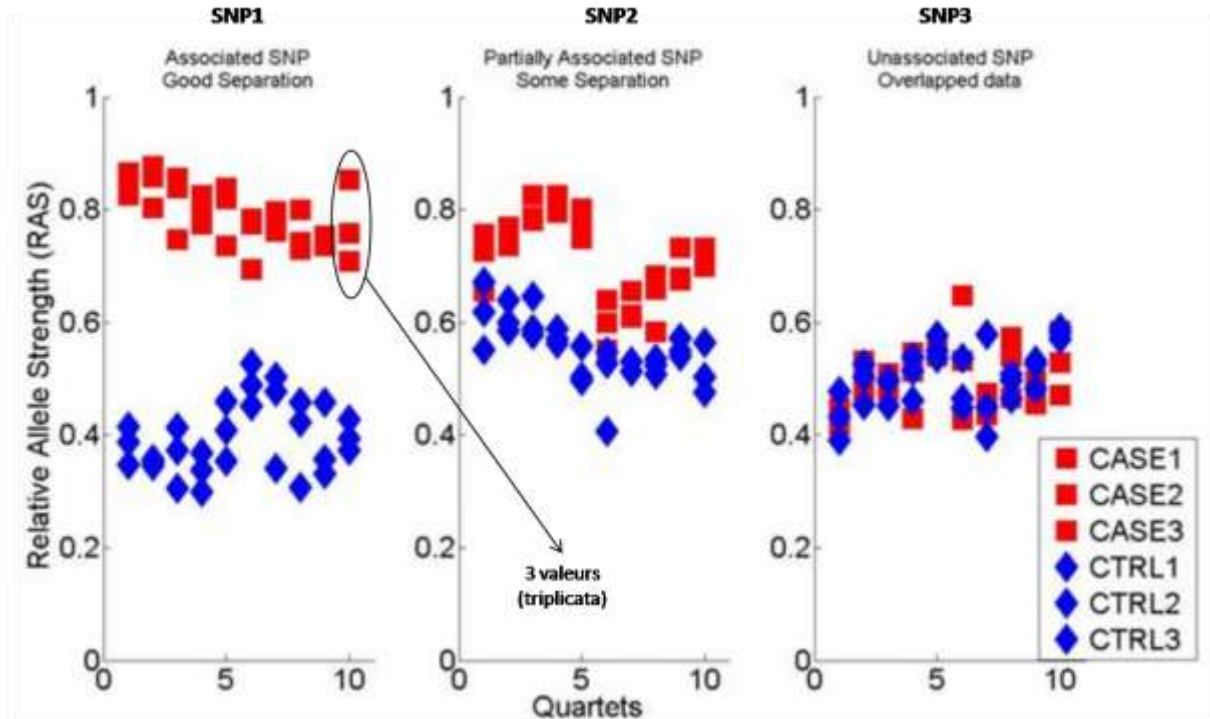


Figure 48 : Représentation des clusters de SNP pour trois répliqués. Le SNP1 présente une bonne séparation des clusters de RAS, ce qui va donner un score de silhouette élevé. Le SNP2 présente une séparation partielle des clusters ce qui donnera un score de silhouette moyen. Le SNP3 présente une mauvaise séparation entre les clusters ce qui donnera un score de silhouette faible (<http://genepool.tgen.org/docs/clustering/>).

III.3. Résultats

III.3.1. Méthode de validation des résultats

III.3.1.1. Validation de l'approche SNP-MaP par comparaison avec les données de la littérature

Tous les tests de significativité ont été calculés (OR) pour chaque SNP de l'ensemble des puces hybridées par la méthode SNP-MaP. Afin de valider notre approche, certains OR ont été comparés à ceux retrouvés dans la littérature par des études d'association du génome complet (GWA, *Genome-Wide association*). Les SNP pris en compte pour cette méta-analyse devaient être présents à la fois dans les articles publiés jusqu'en février 2009 (recherche bibliographique faite sur la base de données « PubMed » avec comme mots clés « *colorectal cancer Genome-Wide Association* ») et sur la puce Affymetrix « Genechip® Human Mapping 250K-Nsp ».

Les OR obtenus par la méthode SNP-MaP devaient se rapprocher de ceux obtenus dans la littérature.

Chromosome	SNP	Publication	OR des publications			OR calculé par SNP-MaP			
			OR pour allèle minoritaire	IC (95%)	p-value	OR pour allèle minoritaire	IC (95%)	p-value	
4	rs4148157	Campa <i>et al.</i> (2008)	0,87	(0,67-1,12)	0,297	1,11	(0,93-1,32)	0,242	
		Curtin <i>et al.</i> (2009)	1,11	(0,98-1,25)	0,093				
8	rs10505477	Ghousaini <i>et al.</i> (2008)	1,27	(1,19-1,33)	2,90E-08	1,32	(1,19-1,45)	2,83E-08	
		Zanke <i>et al.</i> (2007)	1,17	(1,12-1,23)	3,16E-11				
		Poynter <i>et al.</i> (2007)	1,04	(0,94-1,14)	0,457				
		Gruber <i>et al.</i> (2007)	1,23	(1,05-1,43)	8,00E-03				
			Schafmayer <i>et al.</i> (2009)	1,22	(1,14-1,32)	1,40E-07			
			Tomlinson <i>et al.</i> (2008)	1,12	(1,05-1,19)	3,73E-14			
		rs7816535	Schafmayer <i>et al.</i> (2009)	1,00	(0,91-1,11)	0,934	0,90	(0,80-0,99)	0,050
		rs11785277	Schafmayer <i>et al.</i> (2009)	0,95	(0,85-1,05)	0,317	1,20	(1,07-1,33)	7,80E-04
	rs7841264	Schafmayer <i>et al.</i> (2009)	0,90	(0,81-0,99)	0,033	0,76	(0,63-0,90)	2,40E-03	
	rs6470531	Schafmayer <i>et al.</i> (2009)	0,97	(0,90-1,05)	0,508	1,04	(0,94-1,15)	0,416	

Tableau 32 : Comparaison entre les OR publiés et les OR calculés dans notre étude par la méthode SNP-MaP à partir des puces pangénomiques (OR, *Odds Ratio* ; IC, Intervalle de confiance ; SNP, *Single Nucleotide Polymorphism*).

Dans la littérature, nous avons retrouvé six polymorphismes présents sur la puce Affymetrix « Genechip® Human Mapping 250K-Nsp » dont l'association à un risque de cancer colorectal a déjà été testée par un calcul d'OR. Sur ces six polymorphismes, seuls le rs10505477 et le rs7841264 ont reproduit, dans notre étude, une association significative au CCR (p-value<0.05) calculée auparavant dans d'autres études. Il apparaît donc, comme nous pouvons le supposer, que l'approche SNP-MaP aurait tendance à surexprimer la significativité des OR (p-value très faible) par rapport à ceux obtenus par génotypage individuel.

Pour trois autres SNP (rs4148157, rs7816535, rs6470531), l'association au CCR n'a pas été retrouvée significative dans notre approche SNP-MaP, ce qui correspond tout à fait aux résultats des équipes ayant publié leurs données. En revanche, nous avons retrouvé dans notre étude une association significative pour le rs11785277 qui n'avait pas été décrite auparavant dans d'autres études. Ce dernier résultat peut être considéré soit comme un faux-positif dû à notre approche, soit comme un phénomène populationnel. Il apparaît surtout qu'une véritable comparaison ne pourrait être efficace que pour une technologie de génotypage individuel sur la même population du fait des différences de recrutement.

III.3.1.2. Validation par génotypage individuel des associations mises en évidence par la méthode SNP-MaP

Dans des travaux antérieurs de notre équipe (KÜRY *et al.* 2008) (Article 2), nous avons génotypé différents polymorphismes de faible pénétrance appartenant à des gènes de prédisposition aux cancers colorectaux, afin de mettre en évidence leur prédisposition au CCR sporadique de façon isolée et/ou combinée à d'autres polymorphismes et/ou à des facteurs environnementaux. Deux de ces polymorphismes (rs1801282, rs1800566) sont présents sur la puce Affymetrix 250K Nsp.

Par ailleurs, nous disposons des génotypes de tous les individus de nos cohortes pour un troisième polymorphisme (rs10505477) trouvé comme associé aux cancers colorectaux et au cancer de la prostate dans de nombreuses publications (CURTIN *et al.* 2009; GHOUSSAINI *et al.* 2008; GRUBER *et al.* 2007; POYNTER *et al.* 2007; SCHAFMAYER *et al.* 2009; SCHUMACHER *et al.* 2007; THOMAS *et al.* 2008; TOMLINSON *et al.* 2008; ZANKE *et al.* 2007). Ces génotypes provenaient d'une collaboration entre l'équipe de B. Zanke et la nôtre ; notre cohorte avait en effet permis à l'équipe de Zanke de valider la région chromosomique 8q24.21 comme locus de susceptibilité au CCR. Toutefois, jusqu'à aujourd'hui, le mécanisme impliquant le locus 8q24.21 dans la carcinogénèse colorectale et prostatique reste encore non élucidé (SCHUMACHER *et al.* 2007; ZANKE *et al.* 2007). Ce SNP avait déjà été utilisé pour la première étape de validation décrite dans la section précédente (Chapitre II : III.3.1.1. Validation de l'approche SNP-MaP par comparaison avec les données de la littérature).

La comparaison des résultats issus du génotype individuel et de la méthode SNP-MaP donne des résultats tout à fait satisfaisants avec une bonne concordance telle que représentée dans le Tableau 33 :

SNP	Population totale de l'étude		Individus présents dans les 2 projets de l'étude (groupes « catégorie d'âge » et « Vendéens »)				Individus présents dans le groupe « Catégorie d'âge »				Individus présents dans le groupe « Vendéens »			
	Génotypage individuel		Génotypage individuel		Génotypage par SNP-MaP		Génotypage individuel		Génotypage par SNP-MaP		Génotypage individuel		Génotypage par SNP-MaP	
	OR (IC 95%)	p-value	OR (IC 95%)	p-value	OR (IC 95%)	p-value	OR (IC 95%)	p-value	OR (IC 95%)	p-value	OR (IC 95%)	p-value	OR (IC 95%)	p-value
rs1801282	0,953 (0,78-1,15)	0,630	0,914 (0,72-1,14)	0,436	0,894 (0,76-1,03)	0,146	0,838 (0,63-1,11)	0,220	0,998 (0,85-1,16)	0,977	1,055 (0,71-1,54)	0,784	0,823 (0,66-1,01)	0,067
Valeur absolue des différences entre OR d'un même projet				0,020				0,160				0,232		
rs1800566	1,073 (0,92-1,24)	0,352	0,927 (0,78-1,10)	0,388	0,980 (0,85-1,12)	0,774	0,935 (0,75-1,15)	0,536	1,005 (0,86-1,16)	0,945	0,910 (0,68-1,21)	0,524	0,912 (0,74-1,11)	0,379
Valeur absolue des différences entre OR d'un même projet				0,053				0,070				0,002		
rs10505477	1,134 (1,00-1,27)	0,039	1,116 (0,96-1,28)	0,130	1,316 (1,19-1,45)	2,83E-08	1,058 (0,88-1,26)	0,528	1,169 (1,03-1,32)	0,015	1,238 (0,97-1,57)	0,084	1,371 (1,19-1,57)	1,07E-05
Valeur absolue des différences entre OR d'un même projet				0,200				0,111				0,133		

Tableau 33 : Comparaison des résultats obtenus par le génotypage individuel et par la méthode SNP-MaP pour les polymorphismes rs1801282, rs1800566 et rs10505477. (OR : Odds Ratio, IC : Intervalle de confiance, SNP : Single Nucleotide Polymorphism). Ce tableau montre les OR calculés à partir des résultats de génotypage individuel et des résultats obtenus par la méthode SNP-MaP pour différents groupes d'individus constituant notre cohorte. Les OR de la population générale ont été indiqués comme référence et ont été calculés à partir des résultats de génotypage individuel des 1121 témoins et 1024 patients de notre cohorte. Les OR obtenus à partir de génotypage individuel et par la méthode SNP-MaP ont également été calculés pour l'ensemble des individus présents dans les groupes « catégorie d'âge » et « vendéens » soit au total 442 témoins et 646 patients. Considérant les différences théoriques entre les groupes d'individus classés par âge et issus de Vendée, un calcul d'OR a également été réalisé séparément pour chacun des deux groupes (288 témoins et 531 patients pour le groupe d'individus classé par âge ; 191 témoins et 427 patients « vendéens »).

Dans ce tableau, les différents OR obtenus par la méthode SNP-MaP ont été comparés à ceux obtenus par génotypage individuel pour les différents groupes d'individus analysés. Très peu de variations ont été observées entre les deux types d'analyses. La différence maximale entre les deux OR d'un même groupe d'individus est de 0,23 avec une différence moyenne entre les OR obtenus par génotypage individuel et ceux obtenus par l'approche SNP-MaP qui se situe à 0,11, ce qui renforce encore la pertinence de notre approche.

III.3.2. Résultats de génotypage par SNP-MaP

L'analyse des données des puces pangénomiques a permis de faire ressortir, grâce au graphique « $-\log_{10}(p\text{-value})=f(\text{position chromosomique})$ », plusieurs centaines de SNP, potentiellement associés aux CCR, qui dépassaient la valeur seuil de significativité que nous avons fixée. Ainsi, les différentes analyses du groupe « catégorie d'âge » ont fait ressortir une association avec le CCR pour 768 SNP. Cette association est également retrouvée pour 1050 SNP dans les analyses du groupe « vendéens » et l'analyse combinée de ces résultats a montré l'association hypothétique de 1586 SNP avec le CCR.

Chacun de ces SNP a été relié à un ou plusieurs gènes en suivant la méthode de sélection présentée dans la partie « Chapitre II : III.2.4.Choix des SNP d'intérêt ». Ainsi, ces polymorphismes ont permis de mettre en évidence 462 gènes pour les analyses des groupes « catégorie d'âge » et 742 gènes pour celles des groupes « vendéens ». Les gènes retenus ont été sélectionnés par une étude bibliographique en tenant compte de leur pertinence par rapport au cancer.

Afin de réduire le nombre de gènes à analyser, les deux listes de gènes obtenues ont été croisées pour ne garder que les gènes en commun aux deux groupes d'analyses (groupe « catégories d'âge » et groupe « vendéens »). À ce stade, 119 gènes ont été retrouvés comme intervenant potentiellement dans la carcinogénèse colorectale.

Le Tableau 34 liste les 119 gènes avec tous les polymorphismes qui ont permis de les faire ressortir. Ce tableau est simplifié, car les OR significatifs retrouvés dans chaque groupe sont représentés sous forme de croix.

Gènes	dbSNP	Groupe « Catégorie d'âge »			Groupe « Vendéens »			Liaison
		Patients - Témoins	Moins de 60 ans - Témoins	Plus de 60 ans - Témoins	Patients - Témoins	Pool 1 - Témoins	Pool 2 - Témoins	
<i>CYP2J2</i>	rs11572245	X	X					Intronique CYP2J2
	rs7537739				X	X		Liaison CYP2J2
<i>THEM4, S100A10</i>	rs2999525	X		X				Liaison THEM4, S100A10
	rs11803792		X		X	X	X	3'UTR PRUNE
	rs941933				X	X	X	Liaison CRNN
	rs9645406			X				Intronique THEM5
<i>CD1D, CD1A</i>	rs10796993		X		X	X	X	Intronique KIRREL
	rs393304		X					Liaison CDID
	rs1748366				X	X	X	Liaison CDID
	rs11265054				X	X		Liaison OR6K3
	rs2793846					X		0% de Liaison
<i>OR6K3, OR6K6, OR6N1, PYHIN1</i>	rs863327	X		X	X			Intronique SPTA1
	rs2793846			X		X		0% de Liaison
	rs1748366				X	X	X	5'UTR CD1D
	rs11265054				X	X		Liaison OR6K3, OR6K6
	rs2793846			X		X		0% de Liaison
<i>RTN4</i>	rs2165007	X	X	X	X	X	X	Intronique ACO13414.7-202
	rs1430195				X	X	X	Intronique EFEMP1
	rs4557020			X	X			Intronique SPTBN1
<i>LRRN1</i>	rs1682876					X	X	Liaison LRRN1
	rs4349487				X	X	X	Liaison LRRN1
	rs1829725			X				0% de Liaison
<i>LIMD1, CCR9, CXCR6, CCR1, CCR3, CCR2, CCR5, LTF</i>	rs267257	X	X	X				0% de Liaison
	rs939421	X	X	X				Intronique LIRC2
	rs267262				X		X	0% de Liaison
	rs6773430				X		X	0% de Liaison
<i>SLAH2, P2RY14</i>	rs6810143	X		X				Intronique TSC2D2
	rs3773621	X	X	X			X	Intronique P2RY14
	rs13066347				X		X	0% de Liaison
	rs6440770				X		X	0% de Liaison
<i>CENPC1</i>	rs7663519	X	X					0% de Liaison
	rs903508	X						0% de Liaison
	rs3775872	X	X	X				Intronique STAP1
	rs6821995		X					Intronique TMPRSS11A
	rs6552073				X	X		0% de Liaison
	rs10031769				X	X		Intronique TMPRSS11A
<i>GNRHR</i>	rs7663519	X	X					0% de Liaison
	rs6821995		X					Intronique TMPRSS11A
	rs6552073				X	X		0% de Liaison
	rs10031769				X	X		Intronique TMPRSS11F
<i>AFF1</i>	rs10516785	X		X				Liaison AFF1
	rs3755980		X		X		X	Intronique AFF1
<i>MAP1B</i>	rs1217782		X					0% de Liaison
	rs716327	X	X	X				Intronique FCHO2
	rs17242177						X	0% de Liaison
	rs397868				X		X	Liaison ZNF366
<i>TDRD6, PLA2G7, MEP1A</i>	rs6928546		X					Liaison TDRD6, PLA2G7
	rs476873	X	X					Liaison C6orf136
	rs1345062				X			Intronique RCAN2
	rs220687				X	X	X	Liaison GPR116
<i>TBX18, NT5E</i>	rs1407651	X		X				0% de Liaison
	rs9444325	X	X	X				0% de Liaison
	rs9450120				X	X	X	0% de Liaison
	rs1408274						X	0% de Liaison
<i>UTRN</i>	rs9376835	X	X	X				Intronique UTRN
	rs7750326				X	X	X	Intronique UTRN
	rs1104835	X		X	X		X	0% de Liaison
<i>GLI3</i>	rs10274027	X		X	X	X	X	0% de Liaison
	rs2163704	X	X	X				0% de Liaison
	rs17640127			X				Intronique GLI3
<i>SPAM1</i>	rs2189684	X	X	X	X			Liaison SPAM1
	rs7777181	X	X					0% de Liaison
<i>GIMAP7, GIMAP4</i>	rs13234689	X	X					Intronique NUB1
	rs4725919				X		X	Liaison GIMAP4
	rs323369	X		X				Liaison TEX15
<i>TEX15, WRN</i>	rs776385	X	X	X				0% de Liaison
	rs11784948					X		Liaison TMEM66
	rs12681336				X	X		0% de Liaison
	rs4733360	X		X	X			Intronique NRG1
<i>NRG1</i>	rs11784948					X		0% de Liaison
	rs12681336				X	X		0% de Liaison
<i>DEFB1</i>	rs2738159			X			X	0% de Liaison
<i>DEFA6</i>	rs6995735				X		X	0% de Liaison
<i>ANGPT1</i>	rs35143736		X					Intronique ANGPT1
	rs2029790				X	X	X	0% de Liaison
<i>TRPS1</i>	rs7825123	X						0% de Liaison
	rs7017226						X	0% de Liaison
	rs16886605				X		X	0% de Liaison

Gènes	dbSNP	Groupe « Catégorie d'âge »			Groupe « Vendéens »			Liaison
		Patients - Témoins	Moins de 60 ans - Témoins	Plus de 60 ans - Témoins	Patients - Témoins	Pool 1 - Témoins	Pool 2 - Témoins	
FAM84B	rs4291232	X	X					0% de Liaison
	rs16901065				X	X	X	0% de Liaison
	rs16902054				X	X	X	0% de Liaison
	rs961023				X	X	X	0% de Liaison
PTPRD	rs4742533	X	X					Intronique PTPRD
	rs10977409			X				Intronique PTPRD
	rs2151322					X		0% de Liaison
	rs10815658						X	0% de Liaison
	rs2031212				X		X	Intronique PTPRD
DDX58	rs17289655	X	X		X		X	Intronique DDX58
	rs1331225					X		0% de Liaison
	rs10814035					X		0% de Liaison
	rs1871159			X	X		X	0% de Liaison
NTRK2	rs7875895	X		X				0% de Liaison
	rs1443446	X	X					Intronique NTRK2
	rs3893398				X	X	X	0% de Liaison
	rs2769605				X		X	0% de Liaison
C9orf29, GNG10, C9orf84	rs7020283	X	X		X		X	Intronique PTGR1
	rs10817310	X	X				X	Intronique ROD1
	rs4978978				X			0% de Liaison
	rs315711				X	X	X	Intronique C9orf84
	rs10817041					X		Intronique SVEP1
UGCG, SUSD1	rs10817291					X		Liaison ROD1, SUSD1
	rs7020283	X	X		X		X	Intronique PTGR1
	rs10817310	X	X				X	Intronique ROD1
	rs10817041					X		Intronique SVEP1
	rs10817291					X		Liaison ROD1, SUSD1
DBC1	rs7870691	X	X					0% de Liaison
	rs7019950					X		0% de Liaison
	rs6478441	X			X	X		0% de Liaison
MMP26	rs10500622			X				0% de Liaison
	rs10742353				X		X	0% de Liaison
	rs12282135				X	X		0% de Liaison
	rs7934332				X	X		0% de Liaison
	rs11035102				X		X	0% de Liaison
LMO1	rs3911598	X		X			X	0% de Liaison
	rs7101471	X		X				Intronique STK33
	rs1026345				X	X		0% de Liaison
TMEM16C	rs12806221	X		X	X	X		0% de Liaison
	rs12274588				X	X	X	0% de Liaison
MGAT4C	rs2162673		X					Intronique MGAT4C
	rs11104501		X					0% de Liaison
	rs7297350				X	X		0% de Liaison
	rs17370612				X	X		Liaison MGAT4C
	rs11104677				X	X		0% de Liaison
	rs2405789				X		X	Liaison NTS, MGAT4C
SPATA13	rs7987854	X		X				0% de Liaison
	rs4769349	X	X	X				Liaison PARP4
	rs4644725				X	X	X	0% de Liaison
DNAJC15	rs11346		X	X				Exonique DNAJC15
	rs2281811				X		X	Liaison EPST11
	rs3764147				X	X		Exonique C13orf31
	rs9533433				X	X	X	Intronique DNAJC15
KLHL1	rs9541660	X	X					0% de Liaison
	rs1890045		X					Intronique KLHL1
	rs17086052				X	X		Intronique KLHL1
	rs9599643				X		X	0% de Liaison
SLITRK6	rs2479926	X		X	X		X	0% de Liaison
	rs12867065	X	X					0% de Liaison
	pos 85150654	X		X				0% de Liaison
	rs9524646						X	0% de Liaison
SLITRK5	rs4771285	X		X	X		X	0% de Liaison
	rs4619260	X	X	X				0% de Liaison
	rs373783	X		X		X		0% de Liaison
SLC10A2	rs1970270					X		0% de Liaison
	rs12874168	X	X	X				0% de Liaison
	rs7328579				X		X	0% de Liaison
LIG4, ABHD13, TNFSF13B	rs9558148				X		X	0% de Liaison
	rs1572870	X	X	X				0% de Liaison
	rs9521082	X	X					Intronique MYO16
	rs3905076				X	X		0% de Liaison
COL4A2	rs11839708				X	X		0% de Liaison
	rs6492261	X		X				Intronique COL4A2
	rs4773171	X	X	X	X	X	X	Intronique COL4A2
NOVA1	rs2289046				X		X	3'UTR IRS2
	rs12887210	X	X					0% de Liaison
	rs1956044	X	X	X				0% de Liaison
	rs8017114				X		X	0% de Liaison

Gènes	dbSNP	Groupe « Catégorie d'âge »			Groupe « Vendéens »			Liaison
		Patients - Témoins	Moins de 60 ans - Témoins	Plus de 60 ans - Témoins	Patients - Témoins	Pool 1 - Témoins	Pool 2 - Témoins	
	rs17111180				X	X		0% de Liaison
	rs1956073					X		0% de Liaison
<i>SIPA1L1</i>	rs36456	X	X	X				Liaison RGS6
	rs7156271		X		X	X		0% de Liaison
	rs2286136					X		Intronique <i>SIPA1L1</i>
	rs8040989	X	X	X				Liaison <i>MTMR10</i>
<i>MTMR10</i>	rs2955788	X	X		X	X	X	Intronique <i>MTMR10</i>
	rs1345919	X	X					Intronique <i>AVEN</i>
<i>SLC12A6</i>	rs12050753	X	X	X	X			Intronique <i>SLC12A6</i>
	rs9972505	X		X				Intronique <i>C15orf29</i>
	rs7174356	X		X				Intronique <i>AQR</i>
	rs9972505	X		X				Intronique <i>C15orf29</i>
<i>NOLA3, C15orf55, ACTC1</i>	rs7174356	X		X				Intronique <i>AQR</i>
	rs12050753	X	X	X	X			Intronique <i>SLC12A6</i>
	rs1562919				X	X	X	0% de Liaison
	rs1609750	X	X	X				0% de Liaison
<i>SLCO3A1</i>	rs4932599	X	X	X			X	Intronique <i>SLCO3A1</i>
	rs7169129	X	X	X		X	X	0% de Liaison
	rs1904392				X	X	X	0% de Liaison
	rs2418489	X	X	X				Intronique <i>ADAMTS17</i>
<i>ADAMTS17</i>	rs4965573				X		X	Intronique <i>ADAMTS17</i>
	rs17819964	X		X	X	X		Intronique <i>TCF12</i>
<i>TCF12</i>	rs8032600				X	X		0% de Liaison
	rs95529	X		X				0% de Liaison
<i>TPM1</i>	rs7163254	X	X		X	X		0% de Liaison
	rs1498128				X	X	X	0% de Liaison
	rs17185055				X	X	X	0% de Liaison
	rs7196399		X					0% de Liaison
<i>ZNF23, ZNF19, MARVELD3</i>	rs811053	X	X	X				0% de Liaison
	rs16970994				X	X		0% de Liaison
	rs12447045	X	X		X		X	0% de Liaison
	rs9921266	X						0% de Liaison
<i>WWOX</i>	rs7198580	X	X	X				0% de Liaison
	rs1876976		X				X	0% de Liaison
	rs8053895				X		X	0% de Liaison
	rs2306907	X		X			X	Intronique <i>CDH13</i>
<i>CDH13</i>	rs2306903				X	X		Intronique <i>CDH13</i>
	rs17699108				X		X	Intronique <i>CDH13</i>
	rs11644204				X	X		Intronique <i>CDH13</i>
	rs2932821				X		X	0% de Liaison
	rs244783						X	Intronique <i>WFDC1</i>
	rs2306907	X		X				Intronique <i>CDH13</i>
<i>HSBP1, OSGIN1</i>	rs1052716			X				Liaison <i>KIAA1609</i>
	rs2306905				X		X	0% de Liaison
	rs244783						X	Intronique <i>WFDC1</i>
	rs2306907	X		X				Intronique <i>CDH13</i>
<i>WFDC1</i>	rs1052716			X				Liaison <i>KIAA1609</i>
	rs244783						X	Intronique <i>WFDC1</i>
	rs3760454	X	X	X	X	X		Non synonymous coding <i>UTP6</i>
<i>CDK5R1</i>	rs16967358	X	X	X				Intronique <i>MYO1D</i>
	rs1019152				X		X	Liaison <i>CDK5R1</i>
	rs7212444				X	X	X	Intronique <i>ACCN1</i>
	rs179918	X	X	X				0% de Liaison
<i>SOX9</i>	rs17826646	X		X			X	0% de Liaison
	rs2429965				X		X	0% de Liaison
	rs7216312						X	0% de Liaison
	rs4794298	X		X				Intronique <i>CA10</i>
<i>CA10</i>	rs7213039				X	X	X	0% de Liaison
	rs203090		X			X		Intronique <i>CA10</i>
	rs16943975	X		X				0% de Liaison
<i>CDH2</i>	rs1615700						X	0% de Liaison
	rs2552488	X		X				0% de Liaison
<i>PIK3C3</i>	rs8087817	X	X	X	X	X		0% de Liaison
	rs4632208				X	X	X	0% de Liaison
	rs2703183				X	X	X	0% de Liaison
	rs1020050				X		X	0% de Liaison
	rs12968853	X	X					Intronique <i>KRTSP5</i>
<i>SETBP1</i>	rs11873246	X	X	X				Intronique <i>SETBP1</i>
	rs8098662	X	X					0% de Liaison
	rs9953694				X	X	X	Intronique <i>SETBP1</i>
	rs7235710		X					Intronique <i>PIGN</i>
<i>PIGN</i>	rs17068168					X	X	0% de Liaison
	rs17069472	X		X	X	X		Intronique <i>PIGN</i>
	rs12605064	X	X	X				0% de Liaison
<i>CDH19</i>	rs8098955				X	X		0% de Liaison
	rs9950015				X	X		0% de Liaison
	rs1667928	X	X	X				0% de Liaison
<i>MBP</i>	rs470216				X	X		Liaison <i>MBP</i>

Gènes	dbSNP	Groupe « Catégorie d'âge »			Groupe « Vendéens »			Liaison
		Patients - Témoins	Moins de 60 ans - Témoins	Plus de 60 ans - Témoins	Patients - Témoins	Pool 1 - Témoins	Pool 2 - Témoins	
	rs9956448				X	X		0% de Liaison
CASP14, NOTCH3, ABHD9, BRD4	rs12151210	X	X					0% de Liaison
	rs12982033	X	X		X		X	Intronique AP1M1
	rs7255326						X	0% de Liaison
UCA1	rs12151210	X	X					0% de Liaison
	rs12982033	X	X		X		X	Intronique AP1M1
	rs597012			X				Intronique CYP4F15
	rs2032882			X				Intronique RAB8A
	rs7255326						X	0% de Liaison
HPN, FXYP3, FXYP5	rs12976962			X				Liaison ZNF302
	rs9807905			X				0% de Liaison
	rs786506				X			0% de Liaison
	rs1830031				X	X	X	Intronique ZNF567
XRCC1	rs2854508	X			X			Intronique XRCC1
	rs10402361				X	X	X	Liaison CEACAM1, CEACAM8
	rs304720	X	X	X	X	X	X	Liaison IRGQ, ZNF576, ZNF428, CADM4
ETFB	rs273641	X	X	X				Liaison CD33, SIGLECP7
	rs36489	X		X				Liaison SIGLEC8
	rs2569436				X	X		Liaison SIGLEC9
	rs1868939					X		0% de Liaison
USP29, ZNF264, AURKC	rs892166	X	X		X		X	0% de Liaison
	rs8099984	X	X		X	X	X	Liaison ZNF543
	rs2304127				X	X	X	Intronique ZNF582
	rs10423167	X			X		X	Liaison ZNF549, ZNF550, ZNF416, ZIK1, ZNF530
	rs7256469					X		Intronique USP29
TGM3	rs6132407		X					0% de Liaison
	rs4814980				X	X		Intronique STK35
	rs2325971				X	X		Intronique VPS16
C20orf42	rs6139898			X				Intronique CLRS1
	rs2182712	X	X	X				0% de Liaison
	rs990928				X	X	X	0% de Liaison
	rs6038219				X	X	X	0% de Liaison
BMP2	rs2182712	X	X	X				0% de Liaison
	rs990928				X	X	X	Intronique BMP2
	rs6055172		X		X			0% de Liaison
	rs6038781				X		X	0% de Liaison
SNRBP2, PCSK2	rs3787509	X		X	X	X		Liaison SNRBP2
	rs1101128	X		X		X		0% de Liaison
	rs6131660				X	X		Intronique C20orf133
	rs11696277				X			Intronique PCSK2
	rs16999677				X	X		0% de Liaison
CDH4	rs11697815	X		X				0% de Liaison
	rs2865801				X		X	0% de Liaison
	rs4925290				X		X	Intronique CDH4
	rs6101205						X	0% de Liaison
USP25	rs991774	X	X			X		0% de Liaison
	rs2823607	X	X	X				Intronique C21orf34
	rs2823389			X				0% de Liaison
	rs2823411	X	X	X			X	0% de Liaison
	rs2823191					X		0% de Liaison
	rs2823298				X	X		Liaison CYCSP42
HMGN1, B3GALT5	rs2297256	X	X	X				3'UTR BRWD1 Liaison HMGN1
	rs1734860	X	X					Intronique IGSF5
	rs2836863		X					0% de Liaison
	rs2836687			X				Liaison ETS2
	rs1006601			X	X	X		0% de Liaison
	rs2836831				X	X	X	0% de Liaison
rs7283380				X	X	X	0% de Liaison	
ADRBK2	rs133164	X	X			X		Liaison IGLL3, LRP5L
	rs9613067		X					Intronique MYO18B
	rs485193			X				0% de Liaison
	rs19999		X				X	0% de Liaison
ARFGAP3	rs4822208	X	X					Liaison ARFGAP3
	rs5751297				X	X		Liaison SERHL
	rs7285376					X		0% de Liaison
BIK	rs4140554	X	X					Liaison PACSIN2
	rs1883264	X	X					Intronique BIK
	rs4822208	X	X					Liaison ARFAGAP3
	rs5751297				X	X		Liaison SERHL
	rs7285376					X		0% de Liaison

Tableau 34 : Tableau simplifié représentant tous les polymorphismes ayant permis de faire ressortir les gènes ayant un lien avec le cancer. (X représente la présence d'un OR significatif ; la colonne « liaison » indique la liaison ou non du SNP : « 0% de liaison » indique que le polymorphisme n'est lié à aucun gène selon SNPBrowser v4.0, et lorsqu'une liaison est repérée il est indiqué « liaison » suivie du nom du ou des gènes correspondant(s)).

Grâce à ce mode de sélection, nous avons pu déterminer sept gènes liés à plusieurs polymorphismes ayant un effet prédisposant ou protecteur significatif dans notre population d'étude. Le Tableau 35 fait la liste de tous les polymorphismes associés positivement au CCR ayant permis de sélectionner ces sept gènes. Le score attribué à ces différents SNP d'après les critères indiqués dans la section « Chapitre II : III.2.4.3.Critères de sélection d'un SNP à analyser en détail » a permis de sélectionner un SNP unique par gène candidat, c'est-à-dire le SNP ayant l'association *a priori* la plus significative au risque de CCR. Les sept SNP que nous avons sélectionnés par différentes étapes sont reportés dans le Tableau 36. Ce tableau indique la localisation chromosomique du SNP, son numéro dbSNP et le gène auquel il appartient, ou auquel il est lié. Il est nécessaire de vérifier par le génotypage de tous les individus présents dans nos cohortes si cette prédisposition est réelle ou due à un faux positif.

Gène	dbSNP	Groupe « Catégorie d'âge »						Groupe « Vendéens »						Liaison				
		Analyse des patients « Moins de 60 ans » et « Plus de 60 ans »				Analyse « Moins de 60 »		Analyse « Plus de 60 »		Analyse Patient		Analyse « Pool 1 »			Analyse « Pool 2 »			
		OR (IC 95%)		p-value		OR (IC 95%)		p-value		OR (IC 95%)		p-value			OR (IC 95%)		p-value	
		OR (IC 95%)	p-value	OR (IC 95%)	p-value	OR (IC 95%)	p-value	OR (IC 95%)	p-value	OR (IC 95%)	p-value	OR (IC 95%)	p-value		OR (IC 95%)	p-value		
SPAM1	rs2189684	2,74 (2,38-3,16)	6,50E-44	2,73 (2,31-3,21)	3,85E-33	2,29 (1,95-2,69)	2,87E-24	5,19 (4,37-6,14)	3,07E-80							Liaison SPAM1		
	rs17289655	1,75 (1,52-2,00)	2,13E-15	2,96 (2,52-3,45)	4,97E-42			3,82 (3,05-4,77)	5,60E-32					5,90 (4,69-7,41)	1,44E-52	Intronique DDX58		
DDX58	rs1331225									4,00 (3,23-4,95)	2,05E-37					0% de Liaison		
	rs10814035									0,33 (0,27-0,39)	3,49E-33					0% de Liaison		
	rs1871159					0,49 (0,41-0,57)	9,73E-18	0,40 (0,34-0,45)	3,67E-34					0,19 (0,15-0,23)	7,49E-61	0% de Liaison		
COL4A2	rs6492261	0,51 (0,45-0,58)	2,30E-25			0,39 (0,33-0,44)	7,24E-38									Intronique COL4A2		
	rs4773171	1,84 (1,62-2,09)	2,79E-21	1,88 (1,61-2,18)	1,29E-16	1,94 (1,67-2,23)	1,82E-19	0,42 (0,36-0,48)	7,59E-33	0,41 (0,34-0,48)	1,75E-26	0,44 (0,37-0,51)	5,45E-23			Intronique COL4A2		
	rs2289046							0,41 (0,35-0,47)	1,72E-33			0,39 (0,33-0,46)	2,73E-26			3'UTR IRS2		
MTMR10	rs8040989	0,36 (0,31-0,41)	4,56E-50	0,38 (0,32-0,44)	2,18E-30	0,35 (0,29-0,40)	2,45E-38									Liaison MTMR10		
	rs2955788	0,57 (0,49-0,65)	1,57E-16	0,43 (0,36-0,50)	1,63E-22			0,24 (0,20-0,28)	5,26E-65	0,26 (0,21-0,31)	2,21E-40	0,23 (0,18-0,27)	1,67E-46			Intronique MTMR10		
CDH13	rs2306907	1,60 (1,41-1,81)	1,45E-13			2,87 (2,47-3,32)	1,14E-44							2,22 (1,85-2,64)	2,76E-18	Intronique CDH13		
	rs2306903							0,12 (0,10-0,13)	9,61E-180	0,10 (0,08-0,12)	3,70E-138					Intronique CDH13		
	rs17699108							4,86 (4,10-5,75)	3,04E-75			7,43 (6,18-8,92)	7,04E-102			Intronique CDH13		
	rs11644204							0,42 (0,35-0,48)	5,69E-28	0,33 (0,27-0,39)	4,22E-36					Intronique CDH13		
	rs2932821							0,39 (0,34-0,45)	1,40E-37				0,39 (0,33-0,46)	6,06E-28		0% de Liaison		
	rs244783												0,36 (0,30-0,43)	2,60E-27		Intronique WFDC1		
CDK5R1	rs3760454	0,55 (0,48-0,61)	9,20E-21	0,57 (0,49-0,66)	2,93E-13	0,53 (0,46-0,61)	1,22E-17	0,46 (0,39-0,53)	4,71E-24	0,32 (0,26-0,37)	6,01E-42					Exonique UTP6		
	rs16967358	0,52 (0,45-0,60)	5,66E-19	0,54 (0,45-0,64)	9,03E-12	0,50 (0,42-0,59)	2,13E-15									Intronique MYO1D		
	rs1019152							2,74 (2,28-3,28)	1,12E-27			3,62 (2,98-4,40)	1,14E-38			Liaison CDK5R1		

Gène	dbSNP	Groupe « Catégorie d'âge »						Groupe « Vendéens »						Liaison		
		Analyse des patients « Moins de 60 ans » et « Plus de 60 ans »				Analyse « Moins de 60 »		Analyse « Plus de 60 »		Analyse Patient		Analyse « Pool 1 »			Analyse « Pool 2 »	
		OR (IC 95%)	p-value	OR (IC 95%)	p-value	OR (IC 95%)	p-value	OR (IC 95%)	p-value	OR (IC 95%)	p-value	OR (IC 95%)	p-value		OR (IC 95%)	p-value
	rs7212444								0,36	1,00E-44	0,38	2,25E-31	0,34	1,56E-38	Intronique ACCN1	
	rs892166	1,76	1,15E-15	2,54	5,11E-31				0,27	1,09E-50			0,33	5,55E-27	0% de Liaison	
		(1,53-2,02)		(2,16-2,97)					(0,23-0,32)				(0,27-0,40)			
	rs8099984	1,60	1,19E-13	2,16	7,77E-24				2,09	1,65E-24	1,93	8,70E-16	2,34	3,29E-25	Liaison ZNF543	
		(1,41-1,81)		(1,86-2,51)					(1,81-2,40)		(1,64-2,27)		(1,99-2,75)			
<i>USP29,</i> <i>ZNF264,</i> <i>AURKC</i>	rs2304127								0,29	1,44E-49	0,37	3,56E-25	0,22	1,14E-45	Intronique ZNF582	
									(0,24-0,34)		(0,30-0,44)		(0,18-0,27)			
	rs10423167	1,43	5,16E-08						2,31	7,79E-28			2,31	2,15E-22	Liaison ZNF549, ZNF550, ZNF416,	
		(1,25-1,63)							(1,98-2,67)				(1,94-2,72)			
	rs7256469										2,36	2,47E-25			ZIK1, Liaison USP29	
											(2,00-2,77)					

Tableau 35 : Représentation des sept gènes mis en évidence à la fois par l'analyse portant sur le groupe d'individus classés par âge et par celle portant sur le groupe d'individus vendéens. Tous les polymorphismes ayant permis de faire ressortir ces sept gènes d'intérêt figurent dans le tableau. Les OR significatifs de chaque SNP sont indiqués ainsi que les intervalles de confiance à 95%.

Localisation chromosomique	Gène d'intérêt	SNP associé	Localisation du SNP par rapport au gène d'intérêt
7q31.33	<i>SPAM1</i>	rs2189684	Intronique
9p21.1	<i>DDX58</i>	rs17289655	Intronique
13q34	<i>COL4A2</i>	rs4773171	Intronique
15q13.3	<i>MTMR10</i>	rs2955788	Intronique
16q23.3	<i>CDH13</i>	rs2306907	Intronique
17q11.2	<i>UTP6</i>	rs3760454	Exonique
19q13	<i>ZNF543</i>	rs8099984	Liaison à ZNF543

Tableau 36 : SNP trouvés associés au cancer colorectal dans notre étude par la méthode SNP-MaP.

III.3.3. Résultats de génotypage individuel

Le génotypage individuel de l'ensemble des patients et témoins de notre cohorte nous a permis de déterminer parmi les sept SNP du Tableau 36, ceux dont l'association trouvée au CCR n'était en réalité qu'un faux-positif dû à la méthode SNP-MaP. Le Tableau 37 compare les résultats obtenus par SNP-MaP et par génotypage individuel (par sonde Taqman) :

SNP	Résultats des génotypages obtenus à partir des puces sur tous les groupes d'individus		Résultats des génotypages individuels obtenus par sonde Taqman sur la population totale de l'étude	
	OR (IC 95%)	p-value	OR (IC 95%)	p-value
rs2189684 (SPAM1intron1A+14610C)	4,62 (4,07-5,22)	3,41E-129	1,00 (0,86-1,14)	9,70E-01
rs17289655 (DDX58intron9A+962G)	2,96 (2,59-3,37)	1,15E-59	1,09 (0,92-1,28)	3,06E-01
rs4773171 (COL4A2intron4A+9669G)	1,42 (1,28-1,56)	3,09E-12	0,99 (0,87-1,11)	8,20E-01
rs2955788 (MTMR10intron7C+5112A)	0,31 (0,27-0,34)	3,90E-92	0,94 (0,79-1,10)	4,45E-01
rs2306907 (CDH13intron6C+141467T)	1,54 (1,39-1,69)	5,70E-18	1,24 (1,08-1,42)	1,71E-03
rs3760454 (UTP6exon3A+29G)	0,43 (0,39-0,47)	5,39E-60	0,83 (0,73-0,93)	2,61E-03
rs8099984 (ZNF5433UTRT+1776C)	2,13 (1,93-2,34)	7,02E-53	0,90 (0,79-1,01)	8,81E-02

Tableau 37 : OR et p-value obtenus par génotypage individuel par sonde Taqman (1023 patients et 1121 témoins) et par génotypage selon la méthode SNP-MaP (646 patients et 442 témoins) pour les SNP d'intérêt (OR, Odds Ratio ; IC, intervalle de confiance).

La comparaison des deux méthodes de génotypage nous indique que, sur les sept SNP investigués, seulement deux (rs2306907, rs3760454) montrent une réelle association aux cancers colorectaux sporadiques dans notre population d'étude. En effet, le rs2306907 du gène *CDH13* possède un OR de 1,24 (IC 95% [1,08-1,42], p=0,002) et le rs3760454 de *UTP6* un OR de 0,83 (CI 95% [0,73-0,93], p=0,003). Ces résultats indiquent que la méthode SNP-MaP permet la détection de SNP de prédisposition, mais qu'il est nécessaire de vérifier par génotypage individuel les résultats obtenus pour s'assurer de la validité de l'association entre le polymorphisme et la pathologie.

III.3.4. Réplication sur d'autres populations du test d'association des SNP rs2306907 et rs3760454 associés dans notre population

Afin de valider la prédisposition de ces deux SNP au CCR, il était nécessaire de génotyper ces SNP sur au moins une autre population de témoins et de patients atteints de CCR. Pour cela nous avons fait appel à des équipes internationales ayant mis en place des études d'associations – et donc des populations – comparables à la nôtre.

Ainsi, nous avons confronté nos résultats aux génotypes obtenus pour ces deux mêmes SNP (rs2306907, rs3760454) sur la population de Zanke *et al.* (2007). Cette population est constituée de patients canadiens caucasiens atteints de CCR, sans distinction de forme pathologique (PAF, HNPCC, sporadique,...). Les résultats des tests d'association au CCR pour les SNP (rs2306907, rs3760454) sont indiqués dans le Tableau 38 et la Figure 49 ci-dessous :

SNP	Population	OR de l'allèle minoritaire		
		OR	IC 95%	P value
rs2306907 (CDH13intron6C+141467T)	Résultats sur toutes les populations	1,13	(1,02-1,25)	0,013
	Résultats population NANTES	1,24	(1,08-1,42)	0,002
	Résultats population ZANKE	1,02	(0,88-1,17)	0,818
rs3760454 (UTP6exon3A+29G)	Résultats sur toutes les populations	0,90	(0,82-0,98)	0,018
	Résultats population NANTES	0,83	(0,73-0,93)	0,003
	Résultats population ZANKE	0,98	(0,87-1,10)	0,775

Tableau 38 : OR calculés à partir du génotypage des rs2306907 et rs3760454 pour notre population d'étude (NANTES) et pour la population caucasienne canadienne (ARCTIC) de B. ZANKE.

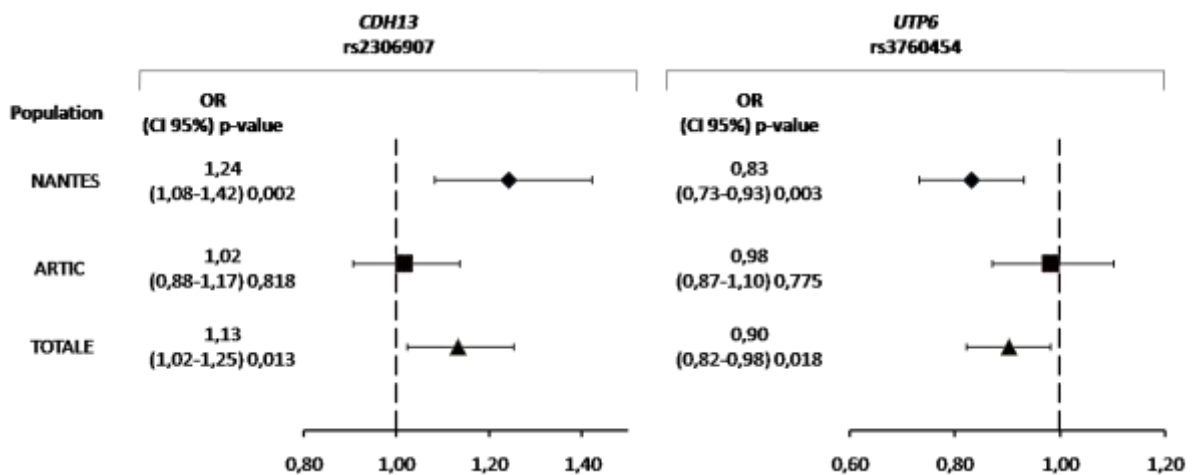


Figure 49 : OR associés aux allèles minoritaires des SNP (rs2306907, rs3760454) pour notre population (NANTES) et pour la population caucasienne canadienne ARTIC de l'équipe de B. ZANKE.

Le Tableau 38 et la Figure 49 montrent que la population canadienne ne réplique pas l'association significative mise en évidence dans notre cohorte. Dans cette dernière, l'OR calculé pour

le SNP rs2306907 est de 1,02 (IC 95% [0,88-1,17], p=0,818) tandis que celui calculé pour le rs3760454 est de 0,98 (IC 95% [0,87-1,10], p=0,775). Les associations au risque de CCR que nous avons trouvées par la méthode SNP-MaP et validées par génotypage individuel sur notre population ne sont pas retrouvées pour les deux SNP (rs2306907, rs3760454) dans la population canadienne.

III.3.5. Expression dans les tissus sain et tumoral des gènes candidats *CDH13* et *UTP6 (HCA66)*

Dans cette partie, nous avons dans un premier temps examiné l'expression des gènes dans les tissus coliques sain et tumoral de 36 patients. Nous avons croisé ces données avec les résultats de génotypage, afin de déterminer si le génotype des SNP rs2306907, rs3760454 intervenait dans l'expression de leur gène respectif. La localisation des SNP rs2306907 dans *CDH13* et rs3760454 dans *UTP6* pouvait en effet laisser augurer de leur impact sur l'expression de ces deux gènes.

Le SNP rs3760454 est localisé dans l'exon 3 du gène *UTP6*. Même si la présence du variant ne semble pas altérer la fonction de la protéine comme indiqué par les programme de prédiction Sift (NG and HENIKOFF 2006) et Tcoffee (NOTREDAME *et al.* 2000; POIROT *et al.* 2003), il est possible néanmoins qu'il ait un impact sur la conformation générale de la protéine et donc sur sa fonction.

Le SNP rs2306907 est situé dans le site accepteur d'épissage de l'intron 6, quatre nucléotides avant le début de l'exon 6. Ce nucléotide n'est pas très conservé entre les espèces. Une analyse par Alamut (Interactive Biosoftware, Rouen, France) utilisant les logiciels de prédiction de site d'épissage SpliceSiteFinder-like (ZHANG 1998), MaxEntScan (YEO and BURGE 2004) et GeneSplicer (PERTEA *et al.* 2001) montre que la présence du polymorphisme n'engendre très probablement pas d'épissage de transcrits alternatifs.

Génotype	CDH13 (rs2306907) c.782-4G>A N=36 (GG=20, GA=14, AA=2)		UTP6 (rs3760454) c.206A>G (p.Q69R) N=36 (AA=15, AG=16, GG=5)	
	$\frac{2^{-\Delta\Delta C_T} \text{ Cancer}}{2^{-\Delta\Delta C_T} \text{ Normal}}$	CI 95%	$\frac{2^{-\Delta\Delta C_T} \text{ Cancer}}{2^{-\Delta\Delta C_T} \text{ Normal}}$	CI 95%
Total	2,49	1,49	1,35	0,54
Homozygote Normal	3,57	1,85	1,44	0,51
Hétérozygote	1,19	0,48	1,30	0,58
Homozygote Variant	0,78	0,50	1,23	0,96
Hétérozygote +	1,14	0,45	1,28	0,49
Homozygote Variant				

Tableau 39 : Rapports $\frac{2^{-\Delta\Delta C_T} \text{ Cancer}}{2^{-\Delta\Delta C_T} \text{ Normal}}$ moyens pour chaque catégorie (N=effectif total, « génotype »=effectif du génotype,

$\frac{2^{-\Delta\Delta C_T} \text{ Cancer}}{2^{-\Delta\Delta C_T} \text{ Normal}}$, rapport des taux d'expression du gène entre le tissu tumoral et le tissu normal).

Les résultats de l'expression résumés dans le Tableau 39 montrent, pour le gène *UTP6*, que son expression est légèrement augmentée dans le tissu tumoral, mais pas de manière significative (rapport=1,35±0,54 avec un IC à 95%). La corrélation entre l'expression du gène *UTP6* et le génotype du SNP rs3760454 ne donne pas de résultats concluants. Pour le gène *CDH13*, le Tableau 39 montre que son expression est augmentée de presque 2,5 fois dans les tissus tumoraux par rapport aux tissus sains. La corrélation du génotype et de l'expression montre une diminution d'expression en présence de l'allèle variant A, la diminution étant accentuée lorsque l'allèle A est présent à l'état homozygote. Ce résultat suggère un effet possible du SNP rs2306907 sur le risque de CCR, ou d'un SNP en liaison avec celui-ci.

Une manière alternative d'examiner l'effet du génotype variant sur l'expression du gène *CDH13* est de comparer le génotype homozygote normal GG au génotype variant hétérozygote GA et/ou homozygote variant AA réunis. En faisant cette analyse, la diminution d'expression du gène est de 2,4 fois entre les individus possédant le génotype sauvage GG et ceux qui possèdent au moins un allèle variant, ce qui montre un effet de l'allèle variant sur l'expression du gène.

Ces résultats nous ont incités à examiner l'effet possible du SNP rs2306907 sur l'épissage de *CDH13*, afin d'essayer de corrélérer la diminution de l'expression du gène à une modification de l'épissage.

III.3.6. Recherche par RT-PCR d'épissage alternatif de *CDH13* induit par le SNP rs2306907

Afin d'identifier de possibles transcrits alternatifs dus à la présence de l'allèle variant du SNP rs2306907 de *CDH13*, nous avons réalisé des RT-PCR sur les exons 5 et 7 encadrant l'exon 6 qui devrait subir un épissage alternatif si le SNP avait un effet sur ce mécanisme. Une seconde RT-PCR a également été réalisée afin de permettre d'amplifier la totalité de la séquence codante de l'ARN. Le but était d'être sûr qu'il n'y avait pas plusieurs transcrits possibles dus au polymorphisme rs2306907. Les résultats sont les suivants :

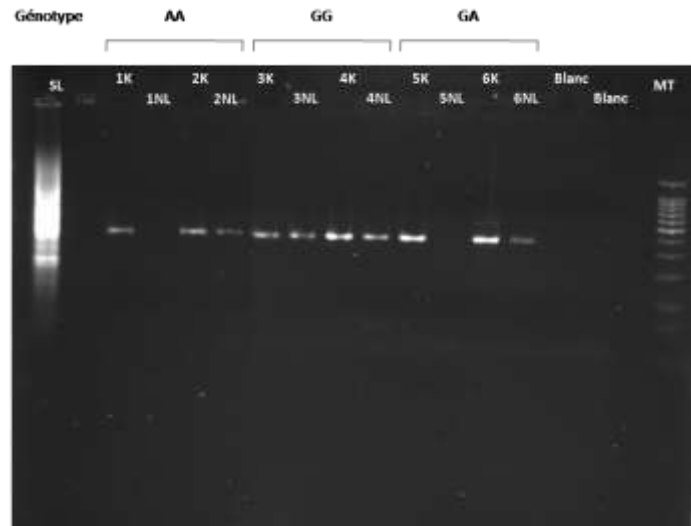


Figure 50 : Résultat de la PCR exons sur gel d'agarose à 1,2% révélée par du BET pour des individus homozygotes sauvages, hétérozygotes et homozygotes variants. Les numéros terminant par K correspondent aux tissus tumoraux, tandis que ceux terminant par NL correspondent aux tissus sains. Pour un même individu, les produits du tissu sain et du tissu tumoral ont été déposés dans des puits adjacents.

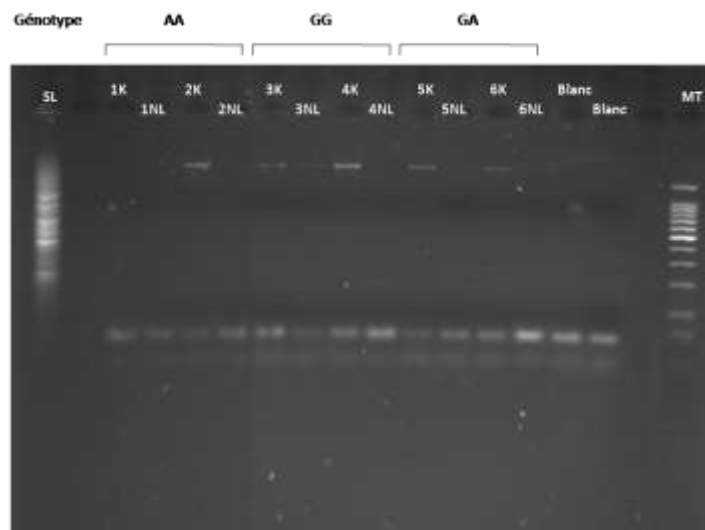


Figure 51 : Résultats de la PCR amplifiant la totalité de la séquence codante sur gel d'agarose à 1,2% révélée par du BET pour des individus homozygotes sauvages, hétérozygotes et homozygotes variants. Les numéros terminant par K correspondent aux tissus tumoraux tandis que ceux terminant par NL correspondent aux tissus sains. Pour un même individu, les produits du tissu sain et du tissu tumoral sont déposés l'un à côté de l'autre.

Ces gels ne permettent pas de distinguer la présence de transcrits alternatifs impliquant un épissage anormal de l'exon proche du SNP rs2306907 ou tout autre exon de *CDH13*.

Un séquençage des produits de la PCR des exons 5 à 7 a été réalisé afin de détecter un décalage de quelques nucléotides (insertion, délétion,...), non visible sur gel d'agarose et qui pourrait expliquer cette perte d'expression chez les individus porteurs de l'allèle variant. Ce séquençage n'a pas permis de mettre en évidence de tels remaniements.

III.3.7. Résultat des analyses de GenePool

Dans les analyses réalisées grâce au logiciel GenePool, nous avons choisi de ne garder que les 200 meilleurs SNP (TOP200) en appliquant la méthode mathématique « silhouette ». Ces analyses nous ont permis de faire ressortir 680 gènes différents qui, par cette méthode d'analyse, sont significativement associés au cancer colorectal (Tableau 40). Parmi ces 680 gènes, des gènes sont en communs avec ceux identifiés par la méthode SNP-MaP développée au laboratoire. En effet, si on considère la liste de 4560 gènes que nous avons générée par notre méthode, sans tenir compte de l'association « virtuelle » au cancer (données dont ne tient pas compte le logiciel GenePool), 185 sont identiques à ceux de l'analyse GenePool. En outre, en intégrant le lien putatif des gènes candidats au cancer, sur les 119 gènes associés aux cancers (Tableau 34) que nous avons trouvés parmi les 4560 gènes issus de notre méthode de sélection, 18 gènes sont également retrouvés par l'analyse GenePool (*CCR5*, *CCR3*, *CENPC1*, *TRPS1*, *FAM84B*, *PTPRD*, *MMP26*, *SPATA13*, *SLITRK5*, *SLCO3A1*, *ADAMTS17*, *WWOX*, *CDH13*, *CA10*, *CDH2*, *CDH19*, *BMP2*, *USP25*, ces gènes dans le Tableau 40 sont surlignés en noir et rouge).

Parmi ces 18 gènes, on retrouve notamment le gène *CDH13*, que nous avons identifié par notre méthode, puis validé par génotypage individuel, et qui est en association directe avec le cancer colorectal sporadique. Bien entendu, afin de pouvoir comparer les deux méthodes d'analyses, nous avons, pour l'analyse GenePool, recherché les gènes les plus proches des SNP les plus significatifs, de la même manière que pour notre analyse. Le fait de retrouver des gènes en commun par les deux analyses nous conforte dans la pertinence de notre approche.

Chromosome	Analyse de tous les individus		Analyse du groupe « Patients Moins de 60 ans » versus Témoins		Analyse du groupe « Patients Plus de 60 ans » versus Témoins		Analyse des groupe »Patients Moins et Plus de 60 ans » versus Témoins		Analyse du groupe « Patients Vendéens » versus Témoins Vendéens	
	dbSNP	Gène	dbSNP	Gène	dbSNP	Gène	dbSNP	Gène	dbSNP	Gène
1	rs11578279	CAMTA1	rs11578279	CAMTA1	rs6603832	AJAP1	rs11578279	CAMTA1	rs2493275	PRDM16
	rs4845882	MTHFR	rs3767555	EPHA8	rs4845882	MTHFR	rs4845882	MTHFR	rs2971426	SFPQ
	rs178134	RUNX3	rs855319	PGM1	rs10889185	NFIA	rs10889185	NFIA	rs10889008	PRKAA2
	rs12043383	RRAGC	rs924277	SLC22A15	rs2457831	INADL	rs2457831	INADL	rs6680953	KANK4
	rs2457831	INADL	rs319464	FAM46C ; GDAP2	rs1415974	SGIP1	rs1496024	CACHD1	rs10736412	LRRC7
	rs1496024	CACHD1	rs6677802	SPAG17	rs17130910	LRRC7	rs17130767	LRRC7	rs10493613	ELTD1
	rs12023616	LRRC7	rs6427545	LY9	rs2744901	PTGER3	rs12023616	LRRC7	rs1776308	LPHN2 ; ELTD1
	rs2744901	PTGER3	rs17838246	RABGAP1L	rs1698735	COL24A1	rs2744901	PTGER3	rs6677277	TTL7
	rs4650634	ELTD1	rs6664259	EEF1AL12	rs1725420	BARHL2	rs12126688	NEGR1 ; LRR1Q3	rs6693944	RAP1A
	rs924277	SLC22A15	rs2111932	DENND1B	rs7522388	FAM102B	rs924277	SLC22A15	rs6674938	ACP6
	rs7519748	ATP1A4	rs1166938	PLXNA2	rs17024828	EPS8L3	rs6677802	SPAG17	rs16841689	CTSK
	rs6427545	LY9	rs4920165	SLC35F3	rs924277	SLC22A15	rs7519748	ATP1A4	rs1935067	AIM2
	rs1166938	PLXNA2	rs4658661	WDR64	rs2881832	LMX1A	rs6427545	LY9	rs10494601	PLA2G4A
	rs1592073	FH	rs3863747	PLD5	rs16857962	FMO9P ; POGK	rs2881832	LMX1A	rs10800928	PRELP
	rs4658661	WDR64			rs10911794	HMCN1	rs6664259	EEF1AL12	rs2820879	ESRRG
	rs4369215	PLD5			rs10494661	RGS18	rs1166938	PLXNA2	rs4453078	OR2T3
	rs6587388	OR2T8			rs4951002	PCTK3	rs1592073	FH		
					rs11120512	DAF	rs4658661	WDR64		
					rs1592073	FH	rs4369215	PLD5		
							rs6587388	OR2T8		
2	rs7573823	SOX11	rs7573823	SOX11	rs7573823	SOX11	rs7573823	SOX11	rs11096709	FAM49A
	rs12710631	VSNL1 (un peu plus loin MYC-N)	rs12710631	VSNL1 (un peu plus loin MYC-N)	rs16858854	TRIB2	rs16858854	TRIB2	rs12710732	CISD1B
	rs1653721	KBTBD9	rs4525733	C2orf73	rs12710631	VSNL1 (un peu plus loin MYC-N)	rs12710631	VSNL1 (un peu plus loin MYC-N)	rs2300701	SRD5A2
	rs6743009	NCOA1	rs1446693	LRRTM4	rs6743009	NCOA1	rs6743009	NCOA1	rs1574417	CTNNA2
	rs17020366	CTNNA2	rs2945825	POLR1A	rs2707272	FAM82A1	rs11884752	LTBP1	rs17020366	CTNNA2 (assez éloigné)
	rs6727379	LONRF2	rs10174201	EDAR	rs12477458	CCDC85A	rs6734118	QPCT	rs6759953	SEPT10 ; EDAR
	rs17261964	MYO7B	rs17587043	SLC20A1	rs17050367	CCDC85A	rs4662880	H56ST1 ; RAB6C (éloigné)	rs11900983	BUB1
	rs4613217	RAB6C	rs11123212	ACTR3	rs719832	SPRED2 ; MEIS1	rs4613217	RAB6C	rs2713243	CLASP1
	rs7371606	CXCR4	rs4662880	RAB6C ; H56ST1	rs1434186	CTNNA2	rs7371606	CXCR4	rs12622810	CLASP1
	rs16852459	SCN7A ; XIRP2	rs4613217	RAB6C	rs4662880	H56ST1 ; RAB6C	rs16823807	FMNL2	rs17261964	MYO7B
	rs6718607	STK39	rs16823807	FMNL2	rs4613217	RAB6C	rs16843372	DAPL1	?	PLCH1
	rs10497418	OLA1	rs16843372	DAPL1	rs7371606	CXCR4	rs10497418	OLA1	rs16852459	SCN7A ; XIRP2
	rs12693471	TFPI	rs10931222	ZC3H15	rs10497418	OLA1	rs2628579	TMEFF2	rs873159	B3GALT1
	rs6434261	GULP1	rs12693471	TFPI	rs2663644	MAP2	rs2663644	MAP2	rs16854583	STK39
	rs12468359	CRYGD	rs10933628	AQP12B	rs2161765	EPHA4 ; PAX3	rs11686538	DOCK10	rs7603617	ZNF385B
	rs2663644	MAP2	rs3935095	PDCD1	rs7602069	DOCK10	rs7602069	DOCK10	rs2054122	PLCL1 ; SATB2
									rs295127	LOC26010
									rs12468359	CRYGD
									rs10804247	XRCC5
	3	rs7614915	GRM7	rs9869724	LSM3 ; SLC6A6	rs164730	GRM7	rs17011480	HMGB1L15	rs9869724
rs4234503		ATP2B2	rs13095794	GALNTL2	rs17011480	HMGB1L15	rs12495243	PDCD6IP	rs2157057	CCR3 ; CCR5
rs6550028		GADL1 ; TGFB2	rs17011480	HMGB1L15	rs12495243	PDCD6IP	rs17031353	PDCD6IP	rs4447759	PTPRG
rs6599159		ULK4	rs12495243	PDCD6IP	rs17031353	PDCD6IP	rs6599159	ULK4	rs1345180	WDR82P1
rs1374837		FAM19A1	rs191686	WNT5A	rs1256388	ULK4	rs1449865	GBE1	rs16852665	CCDC54 ; CBLB
rs1449865		GBE1	rs358917	ERC2 ; WNT5A	rs1374837	FAM19A1	rs2682386	SENP7	rs7429735	LSAMP ; IGSF11
rs2682386		SENP7	rs2034151	KBTBD8	rs6794581	CADM2	rs2712447	CBLB	rs4507250	TMEM108
rs2712447		CBLB	rs1449865	GBE1	rs2682386	SENP7	rs10512857	ACPP	rs16856591	PLOD2 ; SLC9A9

Chromosome	Analyse de tous les individus		Analyse du groupe « Patients Moins de 60 ans » versus Témoins		Analyse du groupe « Patients Plus de 60 ans » versus Témoins		Analyse des groupe »Patients Moins et Plus de 60 ans » versus Témoins		Analyse du groupe « Patients Vendéens » versus Témoins Vendéens		
	dbSNP	Gène	dbSNP	Gène	dbSNP	Gène	dbSNP	Gène	dbSNP	Gène	
4	rs4635724	NMD3	rs1869517	N/A	rs2712447	CBLB	rs6441441	N/A	rs6796427	COMMD2	
	rs4148594	ABCC5	rs4362744	HTR1F ; POU1F1	rs2626491	NAALADL2 ; TBL1XR1	rs4148594	ABCC5	rs6809795	NMD3 ; Sl (assez éloignés)	
			rs10512857	ACPP	rs17543305	KLHL6	rs4686605	FGF12	rs7615444	KCNMB2	
			rs4686605	FGF12	rs16861698	RTP1			rs16832543	SOX2	
									rs4855063	ATP11B	
									rs7622638	SST	
			rs3755849	CRMP1	?	PCDH7	?	PCDH7	rs10017988	CPEB2	
	rs4695152	GABRA2	rs11944501	AFAP	rs6448713	PCDH7	rs4695152	GABRA2	rs17082298	SCFD2	
	?	COX18	rs34836460	KCNIP4	rs17090207	LPHN3	rs2865707	PDHA2	rs7672620	CENPC1	
	rs1247611	IL8	rs7687409	PCDH7 (éloigné) ; STIM2	rs10518130	BTC	rs4240272	ANKRD50 ; SPRY1	rs4376173	ARD1B	
	rs2034002	SCARB2	rs4695152	GABRA2	rs2034002	SCARB2	rs2419537	PCDH10	rs10516727	WDFY3	
	rs6843152	AGPAT9	rs1350545	KDR	rs7664979	PDLIM5 ; PGDS	rs17504041	TLL1	rs17035260	PDGFC	
	rs7678230	PPP3CA	rs6843152	AGPAT9	rs2865707	PDHA2	rs11133061	ADAM29	rs11132004	ANXA10	
	rs4558930	NDST3	rs159478	PITX2	rs2419537	PCDH10	rs4862858	TRIML2	rs34442885	MFAP3L	
	rs2419537	PCDH10	rs4240272	ANKRD50 ; SPRY1	rs10519401	PCDH18			rs10029166	VEGFC	
rs17504041	TLL1	rs10001280	PCDH10 ; PCDH18 (très éloigné)	rs17504041	TLL1						
rs11133061	ADAM29	rs10434086	PCDH18	rs11133061	ADAM29						
rs4862858	TRIML2	rs3967768	NPY1R	rs10030071	AGA						
		?	MFAP3L ; CLCN3	rs2705994	AGA						
		rs552944	GALNTL6	rs2709334	FAM92A3						
		rs7435849	AGA (très éloigné) ; ODZ3 (très éloigné)								
		rs1467097	CLDN22 ; CDKN2AIP								
5	rs6881461	ANKH	rs3913498	TAS2R1	rs1876153	FAM173B ; TAS2R1	rs3913498	TAS2R1	rs11134011	IRX1	
	rs7708358	CDH18	rs10062100	DNAH5	rs2386870	CDH18	rs2386870	AGBL5	rs7720630	CDH18	
	rs7720630	CDH18	rs7720630	CDH18	rs17431811	CDH6 ; CDH9 (très éloigné)	rs10040617	UGT3A1	rs16889756	SUB1	
	rs163599	AGXT2	rs7732647	MAST4	rs10040617	UGT3A1	rs4314405	GHR	rs696758	C9	
	rs4314405	GHR	rs12652555	RIOK2	rs4314405	GHR	?	NR2F1 (désert génique)	rs1423416	C6	
	rs929820	PDE4D	rs32441	ST8SIA4	rs6865369	PIK3R1	rs6890798	AC113407.2	rs2648734	PELO	
	rs16872663	ANKDD1B	rs4958527	NMUR2	rs7703100	CARTPT	rs10519381	KCNN2	rs10042636	WDR41 ; OTP	
	rs9293749	SCAMP1	rs17069703	ODZ2	?	NR2F1 (désert génique)	rs2950955	SPOCK1	rs11738420	H2AFY	
	?	NR2F1 (désert génique)	rs12514656	PCDH24	rs6890798	CHD1			rs17561517	GABRG2	
	rs6890798	CHD1			rs12522800	PRR16			rs2125340	ERGIC1	
	rs2590424	FAM174A ; ST8SIA4			rs1075059	SLU7					
	rs2950955	SPOCK2			rs1816065	MAT2B					
	rs17069703	ODZ2									
	rs12514656	PCDH24									
	6	rs2328391	MBOAT1 ; E2F3	rs9378581	OFCC1 ; SLC35B3	rs862422	DNAH8	rs862422	DNAH8	rs9358363	CDKAL1
		rs16873629	CLIC5	rs862422	DNAH8	rs9380804	DNAH8	rs990189	COL21A1	rs11756275	VN1R6-4P
rs990189		COL21A1	rs1681947	PTP4A1 ; GLULD1	rs990189	COL21A1	rs1681947	PTP4A1 ; GLULD1	rs16873629	CLIC5	
rs2502481		COL12A1	rs9482736	POU3F2	rs12212516	RIMS1	rs2096182	KCNQ5	rs9453944	EGFL11	
rs184658		TRDN ; TCBA1	rs3923954	SOBP	rs2768940	MDN1	rs2768940	MDN1	rs2504668	IMP1	
rs229928		TBPL1	rs1935706	AL357514.19	rs17752168	KLHL32 ; POU3F2	rs229928	TBPL1	rs2608951	ARG1 ; CRSP3 ; AKAP7	
rs1416483		IYD	rs17065063	PDE7B	rs1884185	FBXL4	rs17077947	SASH1	rs6931460	PDE7B	
rs17081113		AKAP12	rs7774833	UST	rs3013685	MAN1A1 ; GJA1	rs17081113	AKAP12	rs1424935	ADAT2	
rs1103031		RPS6KA2			rs17637818	MOXD1			rs1416483	IYD	

Chromosome	Analyse de tous les individus		Analyse du groupe « Patients Moins de 60 ans » versus Témoins		Analyse du groupe « Patients Plus de 60 ans » versus Témoins		Analyse des groupe »Patients Moins et Plus de 60 ans » versus Témoins		Analyse du groupe « Patients Vendéens » versus Témoins Vendéens	
	dbSNP	Gène	dbSNP	Gène	dbSNP	Gène	dbSNP	Gène	dbSNP	Gène
					rs229928	TBPL1				
					rs17081113	AKAP12				
					rs6917661	CNKSR3				
7	rs12702976	NDUFA4	rs12702976	NDUFA4	rs10952098	ICA1	rs12702976	NDUFA4	rs6956986	SOSTDC1
	rs6967799	TWISTNB	rs978989	DGKB	rs6462372	AVL9	rs978989	DGKB	rs7796749	AGR3
	rs1637682	DPY19L1	rs2709738	SP8	rs1354536	PSMA2	rs6967799	TWISTNB	rs10272258	RAPGEF5
	rs7723	TMED4	rs1989838	IL6	rs7723	TMED4	rs2709738	SP8	rs7796681	TOMM7
	rs1997052	IGFBP3 ; IGFBP1	rs7723	TMED4	rs1997052	IGFBP3 ; IGFBP1	rs17154336	SKAP2	rs2462910	SKAP2 ; HOXA1
	?	COBL	rs4720545	IGFBP3 ; IGFBP1 ; TNS3	rs2960877	WBSCR17	rs6462351	PDE1C	rs1011605	NEUROD6 ; ADCYAP1R1
	rs7791296	VSTM2A	rs1997052	IGFBP3 ; IGFBP1	rs7785975	CCDC146	rs1637682	DPY19L1	rs215699	PDE1C ; LSM5
	rs2960877	WBSCR17	?	COBL	rs2269811	MTERF	rs7723	TMED4	rs6947141	COBL (assez éloigné)
	rs2269811	MTERF	rs1649705	COBL ; VSTM2A ;	rs10230949	SAMD9	rs10499637	IGFBP3 ; IGFBP1	rs17747038	IGFBP3 ; IGFBP1
	rs17251278	COL1A2	rs2960877	WBSCR17	rs6971944	SAMD9	rs1997052	IGFBP3 ; IGFBP1	rs1916886	PTPRZ1
	rs10499948	SMURF1	rs33995980	SEMA3D	rs17165914	BET1	?	COBL	rs1419523	GRM8
	rs3095038	TMEM168	rs1799053	TTC9C	rs10499948	SMURF1	rs2960877	WBSCR17	rs834777	MTPN
	rs2059320	DGKI	rs10499948	SMURF1	rs12703305	DPP6	rs33995980	SEMA3D	rs2555050	CHRM2
	rs6953539	CNTNAP2	rs4296966	LHFPL3			rs2269811	MTERF	rs2059320	DGKI
	rs10275919	RBM33	rs1023506	DNAJB9 ; IMMP2L			rs17165914	BET1	rs17162807	MOXD2
			rs4731707	KLF14			rs10499948	SMURF1	rs3817518	DPP6
			rs4445168	MKLN1			rs4731707	KLF14		
			rs6415372	PLXNA4 ; PODXL			rs13247935	DGKI		
			rs13247935	DGKI			rs2792448	NUB1		
			rs2792448	NUB1			rs10275919	RBM33		
			rs10275919	RBM33						
8	rs10503154	DLGAP2	rs2554707	CSMD1	rs2978340	CSMD1	rs10503154	DLGAP2	rs1038062	CSMD1
	rs1038062	CSMD1	rs2634432	PSD3	rs1038062	CSMD1	rs1038062	CSMD1	rs2010239	SLC7A2
	rs7832396	MFHAS1	rs16925736	CA8	rs2840489	MCPH1	rs2840489	MCPH1	rs12676858	TMEM66
	rs6996732	CSGALNACT1	rs622755	CHD7	rs13258090	CSGALNACT1	rs7464304	FAM110B	rs7004799	UNC5D ; ZNF703
	rs7464587	NEFM	rs7007978	IFITM8P	rs7464304	FAM110B	rs7007978	IFITM8P	rs4458918	RUNX1T1
	rs1914995	SNAI2	rs16935931	SULF1	rs2512057	LAPTM4B	rs16937936	TRPA1	rs9918880	TRPS1
	rs7842114	FAM150A	rs16937936	TRPA1	rs17732055	ZFAT	rs2512057	MED30	rs1485282	COLEC10
	rs11990963	PENK ; IMPAD1	rs2512057	LAPTM4B	rs2649108	KHDRBS3 ; FAM135B	rs10086780	TRHR	rs2469625	KCNQ3
	rs7464304	FAM110B	rs16898767	PABPC1			rs12542202	TMEM75	rs2976562	NDRG1 ; ST3GAL1
	rs16898767	PABPC1	rs10086780	TRHR						
	rs10086780	TRHR	rs2467614	FAM84B						
	rs9918880	TRPS1								
9	rs4742418	DMRT1	rs4742418	DMRT1	rs7045303	RFX3	rs4742418	DMRT1	rs11789377	C9orf92 ; BNC2
	rs7045303	RFX3	rs2180943	GLIS3	rs869625	BNC2	rs7045303	RFX3	rs10965969	RP11-315114.1
	rs869625	BNC2	rs10977614	PTPRD	rs7865016	DMRTA1	rs2180943	GLIS3	rs7048937	TJP2
	rs7865016	DMRTA1	rs10809216	PTPRD (très éloigné) ; TYRP1	rs17770335	ELAVL2 ; DMRTA1	rs869625	BNC2	rs17721472	RFK
	rs10969294	UNGO2	rs7865016	DMRTA1	rs10969294	UNGO2	rs7865016	DMRTA1	rs1556384	SHC3
	rs10781076	GDA	rs10969294	UNGO2	rs2026343	GAS1	rs10969294	LINGO2 ;	rs16915239	MUSK
	rs11143224	GDA	rs10781076	GDA	rs10991900	AUH	rs10991900	AUH	rs10817572	COL27A1
	rs10780262	CHCHD9 ; PSAT1	rs2296666	ECM2	rs10120783	COL27A1	rs16911681	C9orf3 ; FBP1	rs2012566	DEC1
	rs2026343	GAS1	rs6478108	TNFSF15	rs7042929	ASTN2	rs17829436	PPAPR3	rs16934099	ASTN2
	rs13285155	DAPK1					rs10979941	AKAP2		
	rs2296666	ECM2								
	rs10819931	ALDOB								

Chromosome	Analyse de tous les individus		Analyse du groupe « Patients Moins de 60 ans » versus Témoins		Analyse du groupe « Patients Plus de 60 ans » versus Témoins		Analyse de groupe « Patients Moins et Plus de 60 ans » versus Témoins		Analyse du groupe « Patients Vendéens » versus Témoins Vendéens	
	dbSNP	Gène	dbSNP	Gène	dbSNP	Gène	dbSNP	Gène	dbSNP	Gène
10	rs10979941	AKAP2								
	rs544276	LPAR1								
	rs16934099	ASTN2								
	rs2924269	AKR1CL2	rs7079674	AKR1CL2	?	CAMK1D	rs10828679	CACNB2	rs7914563	AKR1CL2
	rs9329334	CUGBP2	rs12265098	PRKCQ	rs10828679	CACNB2	rs2488576	ZNF438	rs2924269	AKR1CL2
	rs10828679	CACNB2	rs11009445	PLXDC2	rs11816461	MYO3A	rs11599920	CREM	rs10752275	CAMK1D
	rs11009445	PLXDC2	rs11015427	ANKRD26	rs3124176	LYZL2 ; MAP3K8	rs7900263	DRGX	rs10906425	PRPF18
	rs11599920	CREM	rs2025459	PARD3	rs2488576	ZNF438	rs6481166	SGMS1	rs10762634	PRKG1
	rs7900263	DRGX	rs6481166	SGMS1	rs17295689	ZEB1	rs12771335	MBL2	rs10994491	CDC2
	rs6481166	SGMS1	rs17529748	PRKG1	rs11599920	CREM	rs2133469	PCDH15	rs2256703	SFTPD
	rs12771335	MBL2	rs10762634	PRKG1	rs10508820	FZD8	rs1875160	C10orf11 ; ZNF503	rs12773130	GRID1
	rs2133469	PCDH15	rs17792619	CTNNA3	rs2133469	PCDH15	rs4570530	NRG3	rs11191995	ITPRIP
	rs10994491	CDC2	rs4570530	NRG3	rs12218465	SUPV3L1	rs2419871	PCDH10	rs1537685	ATRLN1
rs1227969	TSPAN15	rs963975	KCNK18	rs7089242	SORCS3 ; SORCS1	rs422221	RAB11FIP2	rs363226	SLC18A2	
rs4570530	NRG3	rs422221	RAB11FIP2	rs10732243	SORCS1	rs2981456	FGFR2	rs3750835	ATE1	
rs2532666	EMX2	rs2981456	FGFR2			rs1060561	DOCK1	rs11248499	GPR26	
rs1060561	DOCK1	rs1060561	DOCK1					rs2282439	DHX32	
11	rs10742372	MMP26	rs10742570	LRRC4C	rs11246002	RIC8A	rs10742372	MMP26	rs17588983	N/A
	rs7479001	SPON1	rs6591720	TTC28	rs10742372	MMP26	rs7479001	SPON1	rs4939197	OR10W1
	rs4519081	LRRC4C	rs880762	ODZ4	rs7479001	SPON1	rs4519081	LRRC4C	rs887647	DAGLA
	rs2276031	CAPN5	rs561832	ODZ4	rs1017513	NAV2	rs2280136	API5 ; LRRC4C (très éloigné)	rs17143592	FAM181B
	rs17817331	CCDC83	rs2186629	CHORDC1	rs11025736	NELL1	rs6591720	TTC9C	rs17207890	FAM76B
	rs2186629	CHORDC1	rs11021059	SESN3	rs6483940	GAS2 ; LUZP2	rs11227673	RHOD	rs17318118	PDGFD
	rs10895253	YAP1	rs11220658	CNTN5	rs2054582	LRRC4C ; RAG2 (très éloigné)	rs2276031	CAPN5	rs12280073	PDGFD
	rs11213582	RAD21 ; SLC30A8 ; EIF3H	rs12280073	PDGFD	rs6591720	TTC9C	rs2186629	CHORDC1	rs10891729	FAM55B
	rs2103173	PTS	rs7358466	ZC3H12C	rs2276031	CAPN5	rs624184	PIWIL4	rs1425206	ETS1
	rs702739	GRAMD1B	rs2103173	PTS	rs2186629	CHORDC1	rs10895253	YAP1		
			rs10750111	DSCAML1	rs4635043	CHORDC1	rs2103173	PTS		
					rs10895253	YAP1	rs10750111	DSCAML1		
					rs10502139	POU2AF1	rs702739	GRAMD1B		
12	rs11050047	FAR2	rs12815704	LMO3 ; MGST1	rs17526713	CCDC91	rs1471407	PPFIBP1	rs7296792	TSPAN9 ; PRMT8
	rs10878215	TBCID30 ; WIFI	rs7973993	CHST11	rs2408467	SLC38A1	rs11050047	FAR2	rs2969097	RIMKLB
	rs1688797	CNOT2	rs3924901	TMEM132C	rs3913067	LRIG3	rs17768388	PDZRN4	rs720052	LRIG3
	rs11612199	TMEM132C	rs10847672	SLC15A4	rs10878215	TBCID30 ; WIFI	rs3913067	LRIG3	rs10878215	TBCID30 ; WIFI
	rs10847672	SLC15A4			rs7137568	CNOT2	rs1688797	CNOT2	rs7308545	KCNC2
					rs7955537	SLC25A3	rs7137568	CNOT2	rs10862560	TMTC2
					rs7976158	RBM19	rs11057576	FAM101A	rs1565351	BTG1
13	rs3213617	KATNAL1	rs12872464	SGCG ; FGF9	rs927553	SPATA13	rs927553	SPATA13	rs2312302	MIPEP
	rs9596374	PDS5B	rs9551767	UBL3	rs7984339	ALOX5AP ; HSPH1	rs3213617	KATNAL1	rs17078075	PDS5B ; KL
	rs2875194	UFM1 ; FREM2	rs3213617	KATNAL1	rs2875194	UFM1 ; FREM2	rs1831549	C13orf33 ; ALOX5AP	rs17054640	SMAD9
	rs7999012	FAM124A	rs2875194	UFM1 ; FREM2	rs1170118	DGKH	rs17057315	UFM1	rs550910	SCEL
	rs7321909	ATP7B	rs9600149	KLF12	rs886347	PCDH8 ; LECT1	rs2875194	UFM1 ; FREM2	rs9301578	SLITRK5 ; MIRHG1 (très éloignés)

Chromosome	Analyse de tous les individus		Analyse du groupe « Patients Moins de 60 ans » versus Témoins		Analyse du groupe « Patients Plus de 60 ans » versus Témoins		Analyse des groupe »Patients Moins et Plus de 60 ans » versus Témoins		Analyse du groupe « Patients Vendéens » versus Témoins Vendéens	
	dbSNP	Gène	dbSNP	Gène	dbSNP	Gène	dbSNP	Gène	dbSNP	Gène
	rs973907	MIRHG1	rs9518615	FGF14	rs11843864	PCDH20 (éloigné); PCDH9 (très éloigné)	rs1160321	LHFP	rs973907	MIRHG1
	rs2761171	CLYBL	rs9586725	DAOA	rs7329163	GPC5	rs1170118	DGKH	rs9582434	NALCN
	rs9520140	ARGLU1			rs2761171	CLYBL	rs342774	SUGT1	rs2490535	NALCN
	rs472825	IRS2			rs9520140	ARGLU1	rs1436681	PCDH17		
					rs472825	IRS2	rs1414427	PCDH9; PCDH20		
						rs4614594	PCDH9; PCDH20			
						rs973907	MIRHG1			
						rs2761171	CLYBL			
14	rs8022457	PRKD1; FOXG1B	rs8022457	PRKD1; FOXG1B	rs10909625	GZMB	rs10909625	GZMB	rs1652947	EGLN3
	rs3813402	GCH1	rs3813402	GCH1	rs2296182	KTN1	rs8022457	PRKD1; FOXG1B	rs4981250	BAZ1A
	rs1152472	PELI2	rs1152472	PELI2	rs1940532	CKCNK10	rs1152472	PELI2	rs1777069	LRFN5
	rs7147705	NRXN3	rs8009268	FLRT2	rs12590101	EML5	rs8009268	FLRT2	rs3813402	GCH1
	rs12590101	EML5	rs12590101	EML5	rs917908	FBLN5	rs12590101	EML5	rs4390494	PRKCH
	rs917908	FBLN5	rs12434332	FOXN3	rs2148567	PRIMA1	rs917908	FBLN5	rs11622028	PLEKHG3
	rs17092772	TCL6	rs17841064	MARK3	rs17092772	TCL6	rs17092772	TCL6		
	rs8027939	GJD2	rs7175526	THSD4	rs8027939	GJD2	rs8027939	GJD2	rs16954432	RASGRP1
rs7175526	THSD4	rs17511735	SH3GL3	rs8042191	THBS1	rs7175526	THSD4	rs2433283	SLC27A2	
rs11632927	C15orf51; ADAMTS17	rs11638630	ZSCAN2	rs2056494	GATM	rs16949348	MCTP2	rs16966345	WDR72	
15					rs2414544	AQP9	rs11632927	C15orf51; ADAMTS17	rs12904969	SMAD3; SMAD6
					rs17695293	SLCO3A1			rs12439738	SLCO3A1
					rs12908968	CHD2				
					rs16949348	MCTP2				
					rs4965177	SPAT8; ARRDCA				
					rs1670247	SYNM				
					rs11632927	LYSM4; ADAMTS17				
	rs7199384	FAHD1	rs7187830	XYLT1	rs7206267	XYLT1	rs7206267	XYLT1	rs1006760	ABAT
	rs1902816	GPR139	rs1837024	CDH5	rs1478508	DYNLRB2	rs1902816	GPR139	rs16945621	ABCC12
	rs1837024	CDH5			rs7199903	FOXF1	rs1837024	CDH5	rs17772912	NUDT7; ADAMTS18
rs10514587	CDH13					rs7199903	FOXF1	rs2738622	WWOX	
								rs16958029	CDH13	
17	rs2269762	TNK1	rs2269762	TNK1	rs1391765	RNF222	rs2269762	TNK1	rs9788983	RPH3AL
	rs7221571	WSB1	rs7221571	WSB1	rs8073025	NMT1	rs7221571	WSB1	rs880335	USP43
	rs8073025	NMT1	rs8073025	NMT1	rs17220558	AKAP12	rs8073025	NMT1	rs4794990	CCL2
	rs1015759	UTP18; CA10	rs1495274	IGF2BP1			rs1015759	UTP18; CA10	rs16948381	MYST2
	rs7567	GAA	rs1015759	UTP18; CA10						
			rs7223292	BCAS3						
		rs4968433	TBX4							
18	rs9964988	ZNF519	rs383907	ZFP161	rs149424	CHST9	rs383907	ZFP161	rs579916	APCDD1
	rs16944213	CDH2	rs403814	L3MBTL4	rs7228141	ASXL3	rs9964988	ZNF519	rs2733119	DSC3
	rs1386369	CDH2	rs17621805	RNMT	rs533291	TXNDC10	rs1386369	CDH2	rs11872215	GALNT1
	rs2733119	DSC3	rs9964988	ZNF519	rs17082526	SOC6	rs2733119	DSC3	rs7239484	ATP8B1
	rs6507178	FHOD3	rs8082787	ZNF521	rs9956282	GALR1	rs6507178	FHOD3	rs589535	MCAR
	rs16950564	LIPG	rs16944213	CDH2			rs16950564	LIPG	rs8085480	CDH19
	rs12971007	CDH19	rs1386369	CDH2			rs12971007	CDH19	rs4294909	CDH19
	rs4294909	CDH19	rs16962939	FAM59A			rs12959657	NETO1	rs4799232	SALL3
	rs8099793	ZNF407	rs674372	FAM59A			rs17087084	NETO1		
	rs2849802	TSHZ1	rs9947251	SLC14A2			rs8099793	ZNF407		
			rs16950564	LIPG			rs2849802	TSHZ1		
			rs10432161	MAPK4						

Chromosome	Analyse de tous les individus		Analyse du groupe « Patients Moins de 60 ans » versus Témoins		Analyse du groupe « Patients Plus de 60 ans » versus Témoins		Analyse des groupe »Patients Moins et Plus de 60 ans » versus Témoins		Analyse du groupe « Patients Vendéens » versus Témoins Vendéens	
	dbSNP	Gène	dbSNP	Gène	dbSNP	Gène	dbSNP	Gène	dbSNP	Gène
19	rs1865098	<i>UQCRFS1</i>	rs12455922	<i>CCDC68</i>	rs7253544	<i>ZNF28</i>	rs1865098	<i>UQCRFS1</i>	rs8112148	<i>SAFB2</i>
	rs198963	<i>KLKP1</i>	rs198963	<i>KLKP1</i>					rs11084785	<i>ZNF599</i>
									rs198963	<i>KLKP1</i>
20	rs6516411	<i>PLCB1</i>	rs16979248	<i>RBBP9</i>	rs4141508	<i>SNAP25</i>	rs11087626	<i>ADRA1D</i>	rs6052946	<i>SLC23A2</i>
	rs6067742	<i>NFATC2</i>	rs6028939	<i>MAFB</i>	rs6067742	<i>NFATC2</i>	rs6100338	<i>MRPS16P</i>	rs6054260	<i>BMP2</i>
			rs743093	<i>ZFP64</i>					rs6104580	<i>SNAP25 ;MKKS</i>
			rs6100338	<i>MRPS16P</i>					rs4814886	<i>SLC24A3</i>
									rs292868	<i>HSPEP1</i>
									rs17217119	<i>CYP24A1</i>
21	rs6517676	<i>USP25</i>	rs6517676	<i>USP25</i>	rs17759053	<i>TUBAP ; MRPL39</i>	rs6517676	<i>USP25</i>	rs2831853	<i>HEMK2</i>
	rs2823642	<i>USP25</i>	?	<i>TUBAP</i>	rs876780	<i>SIM2</i>	rs41423945	<i>TUBAP</i>	rs9981186	<i>DSCAM</i>
	rs1738128	<i>HEMK2</i>	rs1738128	<i>HEMK2</i>	rs2040125	<i>TTC3</i>	rs1738128	<i>HEMK2</i>	rs8134657	<i>TMPRSS2</i>
	rs876780	<i>SIM2</i>	rs1026400	<i>ABCG1</i>						
22	rs5746636	<i>DGCR6</i>	rs181408	<i>BID</i>	rs5746636	<i>DGCR6</i>	rs5746636	<i>DGCR6</i>	rs5746636	<i>DGCR6</i>
	rs134176	<i>TTC28</i>	rs7287616	<i>IGLL1</i>	rs140022	<i>CACNG2</i>	rs134176	<i>TTC28</i>	rs1569785	<i>RABL4</i>
			rs134176	<i>TTC28</i>	rs6002045	<i>SLC25A17</i>	rs6002045	<i>SLC25A17</i>	rs7287281	<i>FAM19A5</i>
			rs10483170	<i>LARGE</i>	rs5769820	<i>BX324178.8</i>				
			rs6002045	<i>SLC25A17</i>						
X	rs10333	<i>TSPAN7</i>	rs1670436	<i>DIAPH2</i>	rs6639974	<i>PNPLA4</i>	rs1670436	<i>DIAPH2</i>	rs547617	<i>HTR2C</i>
	rs1670436	<i>DIAPH2</i>	rs3131425	<i>WDR40B</i>			rs16992781	<i>FMR1 ; SLITRK2 (éloigné)</i>	rs4830086	<i>TJAP1</i>
	rs5975728	<i>GPR112</i>								

Tableau 40 : Résultats des 200 meilleurs SNP obtenus pour cinq analyses différentes réalisées par le logiciel GenePool (PEARSON *et al.* 2007). Les analyses correspondent à : Tous les patients contre tous les témoins, les « patients de moins de 60 ans contre les témoins, les « patients de plus de 60 ans » contre les témoins, les « patients du groupe catégorie d'âge » contre les témoins et les « patients Vendéens » contre les témoins. Les gènes surlignés en noirs correspondent aux gènes en communs retrouvés dans notre liste de 119 gènes associés virtuellement au cancer par notre approche SNP-MaP. Les gènes surlignés en rouge correspondent aux cadhérines communes retrouvées dans notre liste de 119 gènes associés virtuellement au cancer par notre approche SNP-MaP. Les gènes surlignés en vert correspondent aux gènes de la famille des cadhérines communes retrouvés dans les 4560 gènes significatifs par notre approche SNP-MaP.

IV. Discussion

IV.1. Les différentes stratégies employées

IV.1.1. Choix de l'étude d'association cas-témoins

La présente étude a été construite autour de la thématique des cancers colorectaux sporadiques. Par conséquent, les formes familiales de CCR ont été exclues et il était impossible de recourir à des approches basées sur des analyses de liaison génétique. A l'entame du projet, en 2005, la stratégie qui nous a paru la plus appropriée reposait sur l'étude d'association génétique cas-témoins. Loin de faire l'unanimité au sein de la communauté scientifique, l'étude d'association est pourtant tout à fait adaptée à l'examen de polymorphismes génétiques de faible pénétrance (RISCH and MERIKANGAS 1996), c'est-à-dire très précisément le type de facteurs génétiques, qui, en conjonction avec des facteurs environnementaux très variés, contribuent à la prédisposition aux CCR sporadiques (DE LA CHAPELLE 2004). Dans le cadre d'une maladie multifactorielle, l'étude d'association présente le grand avantage de permettre l'étude simultanée de l'effet de facteurs environnementaux et génétiques.

Certes, l'étude d'association cas-témoins ne représente pas la panacée dans l'étude de maladies multifactorielles. Jusque très récemment encore, de nombreuses équipes rapportaient des résultats parfois totalement contradictoires pour des polymorphismes (SNP) qui, dans une étude, prédisposaient au CCR et, dans une autre, protégeaient du CCR. Ces discordances s'expliquent par une très grande faiblesse dans la désignation des études portant souvent sur des populations trop petites, trop hétérogènes sur le plan du phénotype ou des origines géographiques et ethniques (BRENNAN 2002). Pour contourner ce problème nous avons basé notre projet sur deux grandes populations de patients et de témoins, constituées chacune de plus de 1000 individus, appariés au plus juste par genre, âge et origine géographique.

Une fois le choix de la stratégie arrêtée sur l'étude d'association cas-témoins, deux approches étaient alors envisageables pour notre étude : l'approche « gènes candidats » et l'approche pangénomique.

IV.1.2. L'approche « gènes candidats »

Cette approche se focalise sur des variants fonctionnels déjà étudiés pour leur implication dans la carcinogenèse colorectale ou appartenant à une voie métabolique elle-même reliée à cette pathologie. Cette approche très appliquée n'a par conséquent pas pour objectif de rechercher de

nouvelles voies métaboliques ou de nouveaux gènes pouvant intervenir dans la prédisposition au CCR.

Le fait de n'étudier que des variants fonctionnels permet, lorsque l'association entre le variant et la pathologie est mise en évidence, d'établir une corrélation entre le phénotype et le génotype des patients. Il devient alors possible d'établir une classification de la pathologie en fonction de la présence ou de l'absence du variant. Le cas échéant, ce variant peut même être utilisé dans des tests diagnostiques. Par exemple, la mise en évidence d'un SNP associé au CCR qui affecte la fonction d'une enzyme va pouvoir ouvrir de nouvelles perspectives, soit d'un point de vue mécanistique, dans la compréhension de l'évolution de la pathologie, soit thérapeutique.

Le principal inconvénient de l'approche « gènes candidats » repose sur le fait que toutes les analyses sont basées sur les seules connaissances publiées. L'étude ainsi construite est alors très tributaire des travaux menés et rapportés par de tierces équipes. Elle dépend donc fortement de la qualité de ces travaux (NACCARATI *et al.* 2007). Cependant, de nombreuses études « gènes candidats » ont mis en avant des associations entre le CCR et un facteur génétique qui n'ont jamais été soumises à réplique sur d'autres populations (TABOR *et al.* 2002). La réplique est pourtant d'autant plus nécessaire qu'elle permet d'accroître la puissance statistique en augmentant le nombre d'individus analysés et, d'autre part, de s'affranchir de tous les phénomènes populationnels. Il a été montré que les populations, en fonction, par exemple, de leur ethnie ou de leur origine géographique, (RISCH 2000) n'ont pas la même fréquence allélique pour certains SNP. En effet, si un allèle « X » d'un SNP est montré minoritaire dans une population « A » et majoritaire dans une autre population appelée « B », la prédisposition due à cet allèle B sera donc différente dans les deux populations.

Le fait que l'approche « gènes candidats » soit orientée par les travaux antérieurs implique que le champ d'investigation devient très réduit lorsque la pathologie étudiée est nouvelle et que toutes les voies métaboliques en cause n'ont pas été identifiées. Dans ce cas, la portée de l'approche « gènes candidats » sera forcément plus réduite. Dans le même ordre d'idée, l'absence de publications des résultats négatifs (absence d'association entre un polymorphisme et le CCR) n'étant pas valorisée dans les revues scientifiques ceci constitue un biais dans une méta-analyse bibliographique.

Un autre inconvénient de cette approche « gènes candidats » est qu'elle est basée sur une analyse bibliographique réalisée à un instant *t*. Une veille bibliographique doit être faite, qui va périodiquement mettre à jour de nouveaux polymorphismes de prédisposition. Aussi, ce type d'approche oblige-t-il à réaliser des réajustements constants du plan des manipulations. L'intégration fréquente de nouvelles données à la base existante impose à l'expérimentateur une dynamique dans

l'analyse de ses résultats, qui se matérialise par de nombreuses analyses complémentaires susceptibles de totalement bouleverser les interprétations établies à un moment donné.

Ces inconvénients n'ont pas empêché la communauté scientifique d'utiliser cette approche pour la détection de facteurs de prédisposition. Il faut néanmoins, au minimum, que le projet soit construit sur de grandes populations, afin d'obtenir une bonne puissance statistique (BRENNAN 2002). Cette approche « gènes candidats » est en train de perdre du terrain face à l'analyse du génome entier, ou *Genome-Wide Association* (GWA), facilitée par l'apparition des puces à ADN de génotypage.

IV.1.3. L'approche pangénomique

Cette approche étendue au génome entier, ou approche pangénomique, permet d'analyser plusieurs milliers de SNP simultanément. Les puces à ADN permettent, à l'heure actuelle, d'analyser jusqu'à un million de SNP répartis sur la totalité du génome. L'obtention du génotype de tous ces SNP permet de s'affranchir de l'approche « gènes candidats » et de ne plus rester cantonné aux voies et aux gènes déjà identifiés dans la pathologie. Cette technologie est un grand pas en avant pour l'étude des maladies génétiques, car elle permet en une expérience unique d'obtenir les informations pour l'ensemble du génome et donc d'identifier de nouveaux loci de prédisposition.

Par opposition à l'approche « gènes candidats », tous les SNP sont étudiés sans distinction, qu'ils soient fonctionnels ou non : les SNP étudiés peuvent être introniques, exoniques, ou même appartenir à des régions chromosomiques dépourvues de gènes connus. Grâce au projet international HapMap (2003) qui vise à établir des cartes d'haplotype des différentes populations, un grand nombre de SNP ainsi analysés peuvent être reliés, par exemple, à un gène ou à une zone de régulation. Par cette approche, le SNP trouvé en association avec la pathologie ne sera pas forcément le SNP causal, mais simplement un marqueur génétique associé à la maladie. Pour un SNP qui n'est pas compris dans un gène, les données du projet HapMap (2003) nous indiquent les autres SNP en déséquilibre de liaison avec celui-ci, et, de proche en proche, nous permettent de remonter jusqu'au SNP et au gène causal.

En revanche, cette technologie est très onéreuse (HIRSCHHORN and DALY 2005). En effet, même si les prix ont baissé depuis l'apparition des toutes premières puces à ADN, le coût reste élevé pour un grand nombre de laboratoires. Si on compare ce coût à celui des méthodes de génotypage classiques, il n'est pas très élevé, compte tenu du fait que l'analyse sur puce SNP porte au minimum sur 10000 SNP. Afin de faire son choix dans la mise en place stratégique d'une étude, il est indispensable de savoir si l'on a besoin d'un aussi grand nombre d'informations, car le problème est ensuite de pouvoir gérer toutes ces données. En effet, le criblage sur puce SNP génère un très grand

nombre de données par individu. Par exemple, si pour des populations de 1000 patients et 1000 témoins, on analyse 250000 SNP (nombre de SNP que nous avons étudiés dans ce projet), le nombre de génotypes générés est de 50 millions. Un défi de taille est alors de stocker et d'analyser tous ces génotypes. A l'heure actuelle quelques logiciels, qui permettent d'effectuer des analyses sur un aussi grand nombre de données, ont été mis en ligne (GenePool (PEARSON *et al.* 2007), Packages SNP-MaP pour R (DAVIS *et al.* 2009)), mais ces logiciels ne correspondent pas toujours exactement aux attentes des expérimentateurs. Un gros travail reste nécessaire pour développer des outils adéquats pour les études d'association à partir de puces pangénomiques.

Par ailleurs, l'identification d'un SNP associé au CCR par cette approche ne garantit pas nécessairement l'identification d'un gène candidat voire d'une corrélation génotype/phénotype évidente (HIRSCHHORN and DALY 2005). La plupart des SNP trouvés associés sont *a priori* non causaux et reliés à aucun gène, ce qui complique l'interprétation. Seules ces approches sont cependant capables de mettre en évidence de tels polymorphismes, qui impliquent sûrement un mécanisme pathologique qui leur est propre, mais qui reste encore non élucidé (microARN, régulation à distance,...). Un des mécanismes probables à étudier repose sur la prise en compte de la conformation tridimensionnelle de l'ADN. En effet, il est possible que lors du repliement de l'ADN, le SNP intergénique mis en cause soit fortement rapproché du gène causal. Il peut ainsi exercer son action, alors que dans la structure d'ADN primaire ce SNP et ce gène sont trop éloignés pour que l'on puisse suspecter une quelconque association entre eux. A titre d'exemple, plusieurs SNP du locus 8q24.21 ont été retrouvés associés à différents cancers comme le cancer du côlon (ZANKE *et al.* 2007) et de la prostate (SCHUMACHER *et al.* 2007). Cependant, cette région ne contient pas de gène. Elle comprend uniquement un pseudogène (*POUF5FP1*) pouvant coder pour un facteur de transcription, situé à environ 0,01Mb du SNP d'intérêt, et un autre gène intervenant dans des cancers (*MYC*), lui-même situé à plus de 1Mb. A ce jour, le mystère reste entier quant à l'implication carcinogénétique de ce locus 8q24.21 et des SNP trouvés associés.

Le Tableau 41 résume les avantages et inconvénients des deux techniques.

Approche	Avantages	Inconvénients
« Gènes candidats »	<ul style="list-style-type: none"> Analyse des seuls gènes connus Analyse des SNP fonctionnels Coût modéré 	<ul style="list-style-type: none"> Seulement basée sur des connaissances publiées Non répliation des résultats Non publication des résultats non associés Analyse bibliographique à un instant <i>t</i>
Genome-Wide Association (GWA)	<ul style="list-style-type: none"> Analyse de plusieurs milliers de SNP Analyse étendue aux SNP appartenant aux régions intergénomiques (potentielles régions de régulation) Analyse simultanée de plusieurs voies métaboliques 	<ul style="list-style-type: none"> Coûteuse Analyses bioinformatiques complexes Analyse de SNP non fonctionnels

Tableau 41 : Avantages et inconvénients des approches « gènes candidats » et GWA.

IV.2. Approche « gènes candidats »

IV.2.1. Etude de six polymorphismes des cytochromes P450

(Article 1)

IV.2.1.1. Association entre les facteurs environnementaux et le risque de CCR sporadique

Avant d'explorer l'aspect génétique de la prédisposition aux cancers colorectaux sporadiques, nous avons cherché à savoir si les effets connus pour certains facteurs environnementaux étaient vérifiés dans notre population d'étude. Certains auteurs rapportaient notamment une augmentation du risque de CCR avec la consommation de viande rouge (SANDHU *et al.* 2001), ou de sucres hautement assimilables (GIOVANNUCCI 2001a). D'autres auteurs suggéraient, au contraire, une diminution du risque de CCR avec la consommation de produits laitiers (ALVAREZ-LEON *et al.* 2006) ou encore de fibres alimentaires (PETERS *et al.* 2003). Dans notre étude, nous avons confirmé la tendance décrite dans les études épidémiologiques pour la plupart des facteurs environnementaux intégrés dans notre questionnaire d'habitudes de vie. Parmi les résultats les plus marquants, nous avons notamment retrouvé un effet protecteur de la consommation de pain complet (OR=0,56), mais aussi un effet prédisposant pour une forte consommation de viande rouge (OR=3,41) et pour une absence d'activité physique (OR=1,82). Toutefois, certains résultats sont surprenants, tel que l'effet protecteur observé pour la consommation d'alcool ; dans notre population, en effet, les personnes qui ne consomment pas d'alcool semblent avoir un risque plus élevé de cancer colorectal que les consommateurs réguliers. Cette observation peut révéler un biais dans l'étude ; il est possible que les participants aient, lors de l'interrogatoire, minimisé leur consommation, ou bien, que certains patients aient diminué voire arrêté de consommer de l'alcool, du fait de leur maladie. Ce facteur psychologique directement lié à la maladie n'existe pas par définition chez les témoins sains, ce qui pourrait expliquer ces données discordantes par rapport à celles attendues.

Comme attendu d'après les études épidémiologiques de référence, entre autres de Franceschi (1999), Heavey *et al.* (2004), Park *et al.* (2005b) ou encore de Slattery *et al.* (2003), l'analyse combinée des facteurs environnementaux révèle que dans notre population d'étude, une consommation régulière de fruits, de légumes, de produits laitiers et de pain complet serait la meilleure des combinaisons qui protégerait du cancer colorectal sporadique.

IV.2.1.2. Association entre les facteurs environnementaux et les combinaisons génétiques des cytochromes P450

Les résultats que nous avons obtenus ne sont toutefois que tout à fait préliminaires. Ils correspondent à une étude plus exploratoire que confirmatoire. L'approche utilisée, à savoir l'étude combinée de plusieurs SNP, n'est qu'une première étape dans une démarche d'amélioration des connaissances du phénomène de prédisposition. Ces résultats mériteraient, dans un premier temps, d'être validés sur une population différente.

Les polymorphismes que nous avons choisis d'étudier, bien que sélectionnés d'après la littérature, ne constituent pas les seuls polymorphismes des cytochromes P450. Une étude plus exhaustive de la voie de détoxification des xénobiotiques serait indispensable pour déterminer un réel profil génétique de prédisposition dont l'effet sur le risque de CCR serait potentialisé par des interactions avec des facteurs environnementaux, tels que la consommation excessive de viande rouge. Etant donné que plusieurs enzymes CYP peuvent métaboliser un même substrat (NELSON *et al.* 2004), une voie parallèle peut prendre en charge le xénobiotique pour le fonctionnaliser ; l'effet observé lorsque le variant est présent sera nul, ou diminué. Ce phénomène permet d'expliquer la grande complexité qui existe dans les voies métaboliques et justifie la nécessité de faire une analyse la plus complète possible.

Dans notre étude, nous nous sommes focalisés sur les cytochromes P450 qui fonctionnalisent les amines hétérocycliques aromatiques et les hydrocarbures polycycliques aromatiques (RAUNIO *et al.* 1995). Cette analyse des polymorphismes des enzymes agissant sur le même substrat permet d'observer une voie complète de fonctionnalisation, en tenant compte de voies parallèles du métabolisme qui pourraient entrer en jeu. Cette analyse est nécessaire et rend caduques les analyses de SNP isolés, car la prise en charge d'un substrat par des voies parallèles peut modifier l'effet du polymorphisme vis-à-vis de la pathologie.

Comme Bartsch *et al.* (2000) l'ont fait dans leur étude, nous avons pu réaliser une combinaison d'allèles de SNP catalysant la même réaction afin de connaître le véritable impact de la présence des polymorphismes sur le CCR sporadique. Chacun des SNP analysés dans cette combinaison a, au préalable, été testé indépendamment sans qu'aucune association significative n'ait été mise en évidence. Aucune association significative n'est ressortie non plus pour l'analyse des combinaisons associant de deux à six SNP. Les résultats véritablement intéressants ne sont apparus qu'à partir du moment où nous avons intégré la composante environnementale à nos analyses génétiques. Ainsi, en testant les mêmes combinaisons de SNP conjointement avec les facteurs environnementaux relatifs à notre questionnaire, nous avons retrouvé une association

positive entre une combinaison de six allèles ATGCGT des polymorphismes *CYP1A2* c.-163A>C, *CYP1A2* c.1458T>C, *CYP2E1* g.-1293G>C, *CYP2E1* g.-1053C>T, *CYP1B1* c.1294C>G et *CYP2C9* c.430C>T et un risque accru de CCR, en interaction avec une forte consommation de viande rouge. En effet, nous avons vu dans le paragraphe « Chapitre II : IV.2.1.1.Association entre les facteurs environnementaux et le risque de CCR sporadique » que la consommation de viande rouge augmentait le risque de CCR d'environ 3,4 fois. Lorsque l'on étudie l'association des combinaisons avec le CCR sporadique en fonction de la consommation de viande rouge, on observe une augmentation du risque d'environ 43 fois pour la combinaison la plus prédisposante (ATGCGT) chez les individus qui consomment plus de cinq fois de la viande rouge par semaine par rapport aux individus qui mangent de la viande rouge moins de quatre fois par semaine. Cette combinaison potentialise l'effet carcinogène de la viande rouge, alors qu'elle ne montre aucun effet particulier sur le risque de CCR lorsqu'elle est analysée seule en dehors du contexte environnemental. En outre, cette même combinaison (ATGCGT) induit également un risque accru de CCR lorsqu'on l'analyse simultanément à la variable obésité. Cette observation illustre parfaitement l'intérêt d'étudier les combinaisons d'allèles, car ni l'obésité, ni la combinaison d'allèles ne montrent une association avec le CCR lorsqu'ils sont étudiés indépendamment, alors que l'étude simultanée des deux facteurs montre une corrélation positive. Ces deux exemples mettent en avant l'importance de l'étude concomitante des facteurs génétiques et environnementaux dans des pathologies telles que le cancer colorectal sporadique (GOODE *et al.* 2006; GOODMAN *et al.* 2006; MONONEN *et al.* 2006; XU *et al.* 2005). Comme de la Chapelle (2004) l'a montré sur les cancers colorectaux, la prédisposition aux maladies multifactorielles résulte de l'effet additif de multiples facteurs génétiques, dans un contexte environnemental à risque. Il apparaît donc logique d'étudier les interactions possibles entre les facteurs génétiques, intrinsèques, des patients et les facteurs environnementaux, extrinsèques, auxquels ils sont soumis.

Ces résultats encourageants sont à pondérer, car ils ne portent que sur un petit nombre d'individus (92) présentant la combinaison allélique la plus prédisposée à augmenter le risque de CCR, lorsqu'elle est associée à une forte consommation de viande rouge. Du fait de ce petit nombre d'individus, cette combinaison devrait être validée par réplication sur d'autres populations, en particulier des populations dont le régime carné tient une part prépondérante dans l'alimentation.

IV.2.2. Etude de 52 SNP de faible prédisposition au cancer colorectal sporadique (Article 2)

Cette étude basée sur l'analyse de 52 polymorphismes, répartis dans 35 gènes de prédisposition au cancer colorectal, avait pour objectif de répliquer sur une grande population (1023 patients et 1121 témoins) des associations au risque de CCR déjà mises en évidence par différents auteurs. Le choix des SNP à analyser était basé sur leur implication comme facteur protecteur ou prédisposant à la carcinogénèse colorectale. Cette réplification était nécessaire afin de déterminer l'effet de ces variants de faible pénétrance dans la population française, car bon nombre des résultats publiés portaient sur l'analyse d'un nombre limité de patients et de témoins, ce qui ne permettait pas d'avoir une puissance statistique suffisante pour détecter des variants de faible pénétrance (BRENNAN 2002). Pour preuve, sur les 52 SNP analysés sur notre population, seulement cinq SNP (*PTGS1* c.639C>A, *IL8* c.-352T>A, *MTHFR* c.1286A>C, *PLA2G2A* c.435+230C>T, *PPARG* c.1431C>T) montrent une réelle association avec le CCR sporadique. Cette non reproductibilité des résultats publiés met en avant les défauts de conception de certaines études d'association, car, théoriquement, si tous les variants avaient été véritablement liés au CCR, nous aurions dû obtenir des résultats significatifs pour la majorité d'entre eux. Si nous avons validé au moins 50% des associations décrites dans la littérature, nous aurions pu mettre cette différence sur le compte de faux-positifs ou de la variabilité ethnique entre les différentes populations étudiées même si les populations de référence étaient caucasiennes, elles n'étaient pas nécessairement françaises. Dans notre étude, seulement 10% des associations ont été validées, ce qui fait ressortir de gros problèmes dans la conception des études d'association. C'est pourquoi il est nécessaire de faire des réplifications sur différentes populations, ou de réaliser des méta-analyses de plusieurs études comparables, afin d'éprouver le facteur génétique testé et de valider son association avec la pathologie. Il en est de même pour les cinq polymorphismes trouvés associés au CCR sporadique dans notre population, car bien que notre population se caractérise par sa grande taille et par une très bonne homogénéité entre les échantillons de patients et de témoins, il serait important de vérifier ces associations, afin de valider définitivement ces polymorphismes comme des marqueurs de faible pénétrance du cancer colorectal sporadique.

Les cinq polymorphismes en question (c.639C>A du gène *PTGS1*, c.-352T>A du gène *IL8*, c.1286A>C du gène *MTHFR*, c.435+230C>T du gène *PLA2G2A*, c.1431C>T du gène *PPARG*) ont été trouvés associés au CCR sporadique dans notre population. Les polymorphismes *PTGS1* c.639C>A, *IL8* c.-352T>A et *MTHFR* c.1286A>C ont été montrés comme ayant un effet prédisposant vis-à-vis du CCR alors que les variants *PLA2G2A* c.435+230C>T et *PPARG* c.1431C>T sont protecteurs. Ces cinq SNP

ont, comme pour les polymorphismes des gènes *CYP* de la section précédente (Chapitre II : IV.2.1. Etude de six polymorphismes des cytochromes P450 (Article 1)), été analysés en combinaison et il s'est avéré que la combinaison des allèles prédisposants de ces polymorphismes montre un effet prédisposant accentué. Ce résultat doit cependant être nuancé, car seulement 41 individus de notre population portent la combinaison identifiée ; là encore, il serait intéressant de répliquer, voir de réaliser une méta-analyse, afin de détecter les vraies des fausses associations et de valider cette association.

Afin que les résultats faux-positifs soient éliminés de l'analyse nous avons effectué différentes corrections. La correction classique par le test de Bonferroni est trop draconienne pour les études d'association comme la nôtre, car il éliminerait beaucoup trop d'associations dont sans nul doute de vraies associations. C'est pourquoi, pour corriger nos données, nous avons utilisé d'autres tests tels que le test FDR (*False Discovery Rate*) proposé par Storey *et al.* (2003) qui permet le calcul de la q-value et l'élimination des résultats faux-positifs. Le second test utilisé pour valider nos résultats est le ré-échantillonnage aléatoire de nos données (*Monte Carlo Bootstrapping*) basé sur mille combinaisons aléatoirement choisies parmi les individus constituant notre population. Ces tests bien que nécessaires pour valider nos résultats ne peuvent remplacer une ou plusieurs réplique(s) sur d'autres populations caucasiennes atteintes de CCR et/ou une méta-analyse regroupant les SNP que nous trouvons significativement associés au CCR.

Parmi les cinq polymorphismes associés au CCR dans notre population, quatre (*PTGS1* c.639C>A, *IL8* c.-352T>A, *PLA2G2A* c.435+230C>T, *PPARG* c.1431C>T) appartiennent à des gènes intervenant dans la voie de l'inflammation et trois ont déjà été montrés comme modifiant le risque de cancer colorectal : *PTGS1* c.639C>A dans une étude de Halushka *et al.* (2003), *IL8* c.-352T>A dans une étude de Landi *et al.* (2003) et *PPARG* c.1431C>T dans une étude de Jiang *et al.* (2005). Le variant du gène *PLA2G2A*, qui n'a jamais été montré associé au risque de CCR contrairement aux trois autres SNP, a été sélectionné par dHPLC par l'étude de 50 patients et de 50 témoins. Les deux gènes *PLA2G2A* et *PTGS1* interviennent dans le métabolisme de l'acide arachidonique qui va, après transformation, aboutir aux prostaglandines. Celle-ci joue un rôle important dans la médiation de l'inflammation (NATHAN 2002; STEELE *et al.* 2003) menant les cellules vers une inflammation chronique lorsqu'elle est surexprimée (PRESCOTT and FITZPATRICK 2000). Elle joue également un rôle dans la carcinogenèse en favorisant le phénotype tumoral des cellules par l'augmentation de la transcription de VEGF (JAIN 2002). Le polymorphisme *PPARG* c.1431C>T a quant à lui, déjà été étudié pour son effet protecteur vis-à-vis du cancer colorectal. Ce gène, par son implication dans la différenciation des adipocytes et la régulation du métabolisme lipidique, est suspecté d'avoir des propriétés de suppresseur de tumeur, notamment dans les CCR (AUWERX 2002; MICHALIK *et al.* 2004). Enfin, le gène

IL8 code pour une chimiokine proinflammatoire ; celle-ci est relarguée par les lymphocytes infiltrant la tumeur. Les chimiokines, par leur action, vont favoriser la croissance tumorale et l'inflammation chronique (COUSSENS and WERB 2002).

Le cinquième SNP trouvé associé au CCR, *MTHFR* c.1286A>C, appartient à une voie différente de la voie de l'inflammation, à savoir la voie des radicaux monocarbonés. Cette dernière intervient dans la synthèse et la méthylation de l'ADN (HUBNER and HOULSTON 2009). Son métabolite de départ, le folate, est un facteur limitant de la synthèse de l'ADN, car il représente la seule source *de novo* de thymidine (CHOI and MASON 2002). Une diminution de l'apport en folate va entraîner des mauvaises incorporations d'uracile lors de la synthèse de l'ADN, ce qui va engendrer des mutations et des altérations chromosomiques pouvant mener jusqu'au cancer (BLOUNT *et al.* 1997). De la même manière la méthylation de l'ADN, connue comme étant un des processus de la carcinogenèse colorectale, peut être modifiée par cette voie (EDEN *et al.* 2003).

Il existerait peut-être une relation entre la voie de l'inflammation et des radicaux monocarbonés par l'intermédiaire de la méthylation du gène suppresseur de tumeur *PPARγ* par *MTHFR*. Deux études portant sur le diabète (ARAKI *et al.* 2005) et les maladies cardiovasculaires (DEDOUSSIS *et al.* 2005) confirment cette hypothèse, en suggérant une relation entre un polymorphisme de *MTHFR* et la voie de l'inflammation.

Ces résultats montrant une grande implication de la voie de l'inflammation dans la prédisposition au cancer colorectal sporadique, ainsi que les résultats obtenus par l'analyse croisée de la combinaison des cytochromes P450 et de certains facteurs environnementaux, tels que la viande rouge et l'obésité (Article 1), montrent l'importance d'étudier les associations qui peuvent exister entre les facteurs environnementaux et génétiques. En effet, au vu des résultats mis en avant dans l'analyse des polymorphismes des cytochromes P450 et leur association avec la consommation de viande rouge (Article 1), il aurait été intéressant de corréliser nos résultats sur la prédisposition de la voie de l'inflammation (Article 2) avec le facteur extrinsèque qui influence le plus cette voie, c'est-à-dire la prise d'anti-inflammatoires non stéroïdiens (AINS) (SHACTER and WEITZMAN 2002). Malheureusement, ce facteur ne faisait pas partie de notre questionnaire d'habitude de vie rempli par chaque participant à son entrée dans l'étude.

Pour éviter ces désagréments, il serait intéressant d'uniformiser les questionnaires de recrutement aux niveaux national et international pour les études d'association faisant appel à l'épidémiologie, afin que chaque collection d'ADN et chaque pathologie (dans un premier temps) aient un même questionnaire uniformisé regroupant toutes les informations nécessaires à l'étude.

Par cette uniformisation, les collections seraient d'autant plus riches et les études ne manqueraient plus d'aucune information, une fois le recrutement terminé, comme cela a pu être notre cas.

IV.2.3. Etude du gène *MUTYH* (Article 3)

L'étude des mutations du gène *MUTYH* sur notre cohorte peut paraître hors sujet dans le cadre d'une étude de formes sporadiques de cancers colorectaux. Ce gène, en effet, renvoie directement à la notion de formes familiale de CCR, puisque des mutations homozygotes ou hétérozygotes composites sont responsables d'une forme de polypose atténuée. Notre étude du gène *MUTYH* se situe à la limite entre les formes familiales et sporadiques, l'idée maîtresse étant de déterminer si des variants connus pour induire le phénotype MAP, quand ils sont présents sur les deux allèles de *MUTYH* ou présents conjointement avec un second variant, pouvaient accroître le risque de cancer colorectal dans la population générale lorsqu'ils sont à l'état hétérozygote. Compte tenu de notre recrutement (CCR sporadique), nous ne devions pas trouver de mutations connues de *MUTYH* à l'état homozygote chez les patients de notre population, ce qui aurait signé une MAP.

La question posée ici était de savoir si un variant monoallélique de *MUTYH* engendrait un risque de CCR qui suggérerait une prise en charge différente des cas patients, car différentes hypothèses sont avancées dans la littérature quant à l'implication de ces variants dans la prédisposition au cancer colorectal. Il convenait donc de déterminer sur notre population regroupant exclusivement des cancers colorectaux sporadiques l'effet de ces variants.

Dans notre projet, afin de ne pas négliger des variants autres que Y165C et G382D spécifiques de notre population, nous avons entrepris de réaliser un criblage par séquençage de la partie codante du gène *MUTYH* sur un panel de 100 patients atteints de CCR sporadique. Ce nombre de patients séquencés, choisi arbitrairement, pouvait nous empêcher de détecter certains variants très rares, mais il nous permettait tout de même de détecter un grand nombre de variants potentiellement causaux. Le séquençage nous a permis de mettre en évidence trois mutations à l'état hétérozygote, les deux mutations découvertes par Al-Tassan et une troisième (V489F) qui est potentiellement délétère d'après les analyses bioinformatiques.

L'analyse sur la totalité de la cohorte de ces trois mutations n'a fait ressortir que des patients porteurs des mutations à l'état hétérozygote. Un seul patient était un hétérozygote composite pour les mutations Y165C et G382D ce qui nous permis de l'exclure de notre cohorte. Les patients porteurs hétérozygotes de ces mutations ont montré une légère prédisposition non significative aux CCR (OR=1,2). Cette prédisposition n'est pas plus importante que celle observée pour un polymorphisme de prédisposition classique tel que les polymorphismes de détoxification des xénobiotiques (DE JONG *et al.* 2002; LANDI *et al.* 2003). Lorsque l'on stratifie nos individus en fonction de l'âge d'apparition du CCR, on trouve un risque augmenté pour les porteurs hétérozygotes de plus

de 55 ans, mais le nombre d'individus pris en compte pour cette analyse est vraiment trop peu conséquent pour conclure à un effet dans la population générale.

Cette étude de l'effet des mutations monoalléliques de *MUTYH* sur la population générale est soumise à débat. En effet, Balaguer *et al.* (2007) et Sampson *et al.* (2003) ont montré dans leurs publications respectives que le risque de CCR n'était pas augmenté pour les patients porteurs de mutations hétérozygotes de *MUTYH*. Au contraire, Farrington *et al.* (2005), Webb *et al.* (2006) et Fleischmann *et al.* (2004) ont montré un risque augmenté pour les porteurs hétérozygotes. Ces résultats ne sont toutefois pas forcément discordants, en ce sens où la population étudiée a toute son importance. Le risque semble différent en fonction du contexte familial dans lequel est faite l'observation d'une mutation monoallélique de *MUTYH*. En effet, Olschwang *et al.* (2007), dans une population composée d'une majorité de cas de cancers familiaux, proposent un mécanisme de transmission autosomique dominant pour les patients porteurs de mutations *MUTYH*.

Dans notre population composée exclusivement de cancers colorectaux sporadiques, le risque n'est pas augmenté de manière significative pour les porteurs hétérozygotes de mutations *MUTYH*. Ces résultats sont d'ailleurs confirmés dans une méta-analyse de Farrington *et al.* à venir dans laquelle nous avons collaboré. Cette absence d'augmentation de risque pour les porteurs monoalléliques de mutations *MUTYH* ne va pas à l'encontre des précédentes publications sur le sujet si l'on émet l'hypothèse d'une augmentation de risque pour ces individus seulement lorsqu'ils sont apparentés à des patients atteints de cancer colorectal. Au vu de ces résultats, un criblage systématique de *MUTYH* sur la population générale n'est pas nécessaire alors qu'une prise en charge et un suivi médical adapté devraient être mis en place pour les apparentés aux patients MAP, porteurs hétérozygotes des mutations. Il est possible que cette augmentation de risque soit due, soit à un mécanisme additif entre variants de *MUTYH* ou d'autres variants dans les gènes de la voie BER, soit à des variants dans le promoteur ou les parties régulatrices de *MUTYH* (OLSCHWANG *et al.* 2007).

C'est dans l'optique de mettre en avant un mécanisme oligogénique, que nous avons réalisé un séquençage des parties codantes des gènes *OGG1* et *NUDT1* chez les individus de notre population porteurs des mutations hétérozygotes de *MUTYH*. Cette analyse ne nous a pas permis de mettre en évidence des mutations potentiellement délétères dans ces gènes. Cependant, notre analyse n'était pas exhaustive et il se pouvait qu'une analyse des autres gènes de la voie BER puisse mettre en évidence un tel mécanisme.

Depuis la publication de cette étude, le séquençage des parties codantes de deux autres gènes (*TOE1* et *PRDX1*) de la voie BER a été réalisé, et comme pour les gènes *OGG1* et *NUDT1*, aucun

polymorphisme potentiellement délétère n'a été souligné. Cette analyse portait sur un nombre très restreint d'individus, ne permettant pas de conclure formellement dans l'implication de ces gènes.

IV.3. L'approche *Genome-Wide Association*

Au cours des trois dernières années, les études d'association pangénomique, ou GWAS (*Genome-Wide Association Studies*), sont devenues des standards dans l'investigation des maladies multifactorielles. Grâce à de très rapides avancées technologiques et à un impressionnant développement des outils de génotypage haut et très haut débit, les études pangénomiques sont devenues à la portée du plus grand nombre, permettant ainsi le génotypage simultané de plusieurs milliers de SNP sur de grandes cohortes de plusieurs centaines voire milliers d'individus. Ainsi, ces approches ont permis aux scientifiques de lier fortement des variants à de nombreuses maladies complexes, telles que la schizophrénie (SHIFMAN *et al.* 2008), le cancer de la prostate (DUGGAN *et al.* 2007), la maladie de Crohn (HAMPE *et al.* 2007). Ces quelques résultats sont encourageants, mais nécessitent un investissement financier très lourd qui ne peut pas être supporté par tous les laboratoires.

C'est pourquoi, dans ce travail de thèse, nous avons fait le choix d'employer une stratégie plus économique mais qui utilise tout de même la technologie des puces SNP pangénomiques, afin d'en garder les avantages (Chapitre II : IV.1. Les différentes stratégies employées). Cette approche est appelée SNP-MaP (*SNP Microarrays and pooling*).

IV.3.1. Approche par pool d'ADN

IV.3.1.1. Notion d'allélotype

Dans ce type d'analyse utilisant des pools d'ADN, une des caractéristiques principales est l'utilisation de l'allélotype et non du génotype. En effet, dans les approches GWA classiques, chaque ADN de patients est hybridé sur une puce. Le résultat obtenu est un profil génotypique personnalisé et propre à chaque individu. Avec ces données, tous les individus de l'étude d'association pourront être analysés de manière « classique » en calculant les OR par comparaison de la distribution des génotypes entre la population patiente et témoin. Par opposition, dans l'approche SNP-MaP, l'analyse « classique » ne pourra pas être faite, car le génotype déterminé par les logiciels de traitement des puces correspond à un génotype moyen de tous les ADN mélangés. Il est donc nécessaire de revenir aux données brutes de fluorescences des sondes de la puce afin de pouvoir calculer la fréquence allélique des différents pools hybridés (Encadré 1). Pour chaque SNP d'une puce, ce sont ces fréquences alléliques des patients et des témoins qui sont comparées. On ne peut pas déterminer le génotype des individus composant un pool, mais seulement l'allélotype du pool. L'obtention de l'allélotype n'empêche cependant pas l'utilisateur de calculer un OR en déterminant un nombre théorique de porteurs des allèles, en se basant sur le nombre d'individus présents dans le

pool. Cet OR compare alors la distribution des allèles entre les patients et les témoins, contrairement à la première approche classique (hybridation d'une puce par individu) qui compare la distribution des génotypes. Toutefois l'emploi de cette approche SNP-MaP nous enlève la possibilité d'effectuer des corrélations entre les génotypes et différents facteurs tels les habitudes de vie, les différents phénotypes, ou encore l'âge. Ces corrélations, comme nous avons pu le montrer dans les articles 1, 2 et 3, sont un plus dans l'analyse des maladies multifactorielles. En dehors de la méthode développée par notre équipe pour l'analyse des allélotypes de pools d'ADN, d'autres méthodes d'analyses existent désormais. Des outils libres d'accès apparus récemment permettent d'analyser spécifiquement les résultats d'étude d'association pangénomique utilisant des mélanges d'ADN. On peut citer par exemple le logiciel GenePool (PEARSON *et al.* 2007) qui compare, pour chaque SNP, les fréquences alléliques obtenues à partir des intensités de fluorescence entre les patients et les témoins (Chapitre II : III.2.8. Analyse par le logiciel GenePool). Différentes manières de calculer cette différence sont possibles avec le logiciel, mais dans tous les cas, nous obtenons un classement des SNP par significativité croissante. D'autres logiciels sont également disponibles, comme le package SNP-MaP (DAVIS *et al.* 2009) pour R (R Development Core Team, 2008 ; www.r-project.org) ou encore dChip (BEROUKHIM *et al.* 2006).

IV.3.1.2. Constitution des pools

Afin de pouvoir détecter des variations significatives, la constitution des pools est un paramètre très important aussi bien du point de vue technique que théorique, même si pour Macgregor *et al.* (2007) les erreurs attribuables à ces approches sont plus dues aux puces elles mêmes qu'à la construction des pools. Le choix des ADN à inclure dans les pools est d'autant plus important que dans le cas d'une maladie multifactorielle comme le CCR sporadique, il peut modifier totalement les résultats et mettre en avant un gène, ou un marqueur de prédisposition qui ne serait pas ressorti si les pools avaient été conçus différemment.

C'est dans cette optique que, dans un premier temps, nous avons réalisé des pools d'ADN à partir d'individus sélectionnés sur l'âge d'apparition du cancer pour les patients, et sur l'âge de recrutement pour les témoins. Dans la catégorie « patients », nous avons constitué une première catégorie de pools avec les patients ayant déclaré leur cancer avant l'âge de 60 ans (moyenne d'âge= 54 ans [40-60]) et une deuxième catégorie de pools avec des patients ayant développé beaucoup plus tardivement cette pathologie (moyenne d'âge = 76 ans [69-99]). En fixant la limite d'âge à 60 ans nous espérons pouvoir séparer les vrais cas sporadiques de cancer colorectal des cas dont la précocité était telle qu'elle pouvait faire penser à des formes familiales inconnues. Notre hypothèse était basée sur l'obtention d'une forme de CCR à composante à la fois génétique et

environnementale tandis que la composante génétique devait être beaucoup plus importante dans les formes précoces de CCR sporadique (<60 ans). Ces deux groupes d'individus ont été opposés à des pools d'ADN constitués des individus témoins les plus âgés de notre cohorte. Ce choix de témoins reposait sur l'idée que plus un individu témoin est âgé, plus le risque qu'il développe un CCR sporadique est faible.

Dans un deuxième temps, nous avons constitué des pools d'ADN basés sur l'origine géographique des individus. Nous avons ainsi choisi de constituer des pools patients et témoins avec des individus issus depuis au moins deux générations de la région Vendée. Le choix de cette région s'est fait selon deux critères : le premier est le fort taux d'individus issus de cette région dans notre recrutement, et le second critère renvoie à l'histoire de cette région dont la population originelle présente toutes les caractéristiques d'une population à effet fondateur. Le fond génétique des patients et des témoins issus de cette région est théoriquement commun ou presque, ce qui signifie que les marqueurs spécifiques de la pathologie pourraient ressortir plus facilement.

Sur le plan pratique et technique la constitution des pools implique l'élaboration d'un protocole très précis et minutieux, depuis le choix du nombre d'individus constituant chacun des pools, jusqu'au mélange et au dépôt des ADN poolés. Selon les publications, le nombre d'individus constituant un pool varie : en effet, pour Barratt *et al.* (2002) ils peuvent être constitués de 50 individus et pour Moskvina *et al.* (2005) ils peuvent contenir jusqu'à 716 individus. Toutefois, la plupart des auteurs s'accordent sur le fait qu'un pool doit être constitué d'environ 100 individus pour avoir une significativité maximale des résultats. Dans notre étude, nous n'avons pas pu atteindre cette limite pour certains de nos pools compte tenu du nombre d'individus dont nous disposions, du nombre de réplicats que nous voulions réaliser, et du nombre de puces dont nous disposions. Afin que chaque ADN des individus constituant le pool joue un rôle équivalent dans celui-ci, il faut que tous les ADN soient équidossés et présents en qualité équivalente. Ainsi chaque ADN d'un pool subit en premier lieu une amplification afin de déterminer s'il peut être amplifié correctement par PCR. Dans notre étude, tous les ADN dont la qualité était douteuse (mauvaise amplification lors de la PCR en temps réel test) ont été éliminés de la constitution des pools. Enfin, afin de s'affranchir d'un manque de reproductibilité intra-pool, nous avons répliqué l'hybridation des pools comme le conseillent Macgregor *et al.* (2006) ou encore Le Hellard *et al.* (2002). Cette réplification des pools nous a permis de faire des moyennes de fréquences alléliques par pool.

IV.3.2. Analyse informatique

Au moment de la réalisation de ces travaux, l'analyse bioinformatique de ces données n'était pas évidente, car aucun logiciel en accès libre n'avait été mis au point pour gérer une quantité aussi

importante de données et adapté à nos besoins. C'est pourquoi, après avoir hybridé les puces et fait ressortir les loci les plus significatifs, nous avons restreint nos investigations aux loci contenant des gènes ayant déjà été décrits dans la littérature comme associés au cancer. Cette limite volontaire nous a permis de ne pas trop nous éparpiller dans l'analyse de gènes sans intérêt et de nous focaliser sur de bons candidats, ce qui rapproche notre technique de sélection d'une approche « gènes candidats ». En procédant de la sorte, nous avons délibérément occulté, pour l'instant du moins, la possibilité d'identifier des gènes dont l'implication dans la carcinogenèse est encore méconnue. De plus, cette association entre le gène et le cancer n'est que « virtuelle », car elle porte seulement sur la présence du nom du gène et du mot « cancer » dans le titre et/ou le résumé des articles publiés. En effet, certaines publications peuvent employer ces deux mots dans un même paragraphe, même s'ils n'ont pas de lien. Cette méthode, même si elle est discutable, permet une discrimination assez stricte des gènes de prédisposition retenus.

De la même manière, pour un même gène dans nos analyses, l'association d'un gène au CCR a parfois été indiquée par les OR de plusieurs SNP. Aussi, afin de ne pas multiplier les analyses ultérieures, nous avons fait le choix de n'étudier par la suite qu'un seul SNP par gène. Ce SNP, le plus représentatif de l'association du gène au CCR, devait à notre sens soit être intragénique, soit être en déséquilibre de liaison avec un SNP intragénique. Cette sélection nous a donc fait éliminer un bon nombre de SNP intergéniques, qui pouvaient intervenir dans des systèmes de régulation à distance ou dans des mécanismes encore inconnus à ce jour. La meilleure illustration de ce type de SNP intergénique est sûrement le SNP rs10505477 et les autres SNP en déséquilibre de liaison avec lui, tous situés dans le locus 8q24.21. Cette région pauvre en gènes a été montrée récemment comme étant associée à la prédisposition à de nombreux cancers comme le côlon (ZANKE *et al.* 2007), la prostate et le sein (SCHUMACHER *et al.* 2007), or, à ce jour, aucun mécanisme n'a été identifié qui puisse expliquer cette prédisposition. Notre choix de n'étudier que des SNP intragéniques, ou liés à des SNP intragéniques, a été opéré précisément pour contourner ce type d'écueil symbolisé par le SNP rs10505477.

IV.3.3. Validation de la stratégie

Afin de valider notre approche, différentes stratégies ont été employées. La première a consisté à comparer nos résultats obtenus par SNP-MaP à ceux des études GWA classiques sur le CCR publiés par d'autres équipes pour six polymorphismes présents sur la puce Affymetrix 250K Nsp (Tableau 32). Sur les six polymorphismes étudiés, deux (rs10505477 et rs7841264) ont été trouvés significativement associés au CCR à la fois par le génotypage individuel des autres études et par notre allélotypage SNP-MaP. Pour trois autres SNP (rs4148157, rs7816535, rs6470531), ni le génotypage

individuel des autres équipes, ni le résultat de SNP-MaP de notre laboratoire, n'ont montré d'association significative au CCR. En revanche, notre approche a montré une association significative du sixième SNP au CCR (rs11785277), alors qu'aucune autre étude ne montrait une telle association. Cette dernière comparaison signifie que notre approche pourrait induire de faux résultats positifs. Afin de déterminer la validité du résultat relatif au SNP rs11785277 et connaître son statut, faux-positif ou phénomène populationnel dû aux formes de cancers étudiées, il faudrait génotyper individuellement ce SNP sur toute notre cohorte. Nous pourrions ainsi en déduire l'association au cancer colorectal de ce SNP pour notre cohorte et la comparer à celle obtenue dans la littérature. Si le résultat obtenu se rapproche de celui de la littérature, cela indiquerait que l'association mise en avant par l'approche SNP-MaP est un résultat faux-positif. Par contre, si l'association reste identique à celle trouvée initialement, le phénomène populationnel serait démontré. Pour étayer cette seconde hypothèse, on peut noter que les résultats publiés par Schafamyer *et al.* (2009) correspondent au test d'association pour des patients atteints de cancers colorectaux sans distinction du type (PAF, HNPCC, sporadique,...), alors que notre population n'est constituée que de formes sporadiques de CCR.

Dans le même souci de valider notre méthode, nous avons comparé des résultats de génotypage individuel déjà disponibles au laboratoire aux résultats de la méthode SNP-MaP. Parmi ces SNP, des génotypes individuels de tous nos individus de notre étude pour le SNP rs10505477 localisé en 8q24.21 avaient été montrés associés au cancer colorectal dans une étude à laquelle notre équipe a collaboré (ZANKE *et al.* 2007). Dans cette publication, le SNP rs10505477 était trouvé associé au CCR avec un OR global, de 1,17 ([1,12-1,23], p-value=3,16.10⁻¹¹) en tenant compte de toutes les populations de répliation étudiées (soit 7 populations). Ce SNP était inclus à la puce Affymetrix 250K Nsp que nous avons hybridée pour notre méthode SNP-MaP, il était donc facile pour nous de vérifier cette association. Par l'approche SNP-MaP, nous trouvons un OR de 1,32 ([1,19-1.45] p-value=2,8.10⁻⁸) soit une valeur proche des résultats de la publication de Zanke *et al.* (2007). Par génotypage individuel sur l'ensemble des individus de notre population, ce même SNP rs10505477 est associé avec un OR=1,13 ([1,00-1,27], p-value=0,04). Certes les résultats de SNP-MaP et du génotypage individuel de ce SNP sur notre population ne sont pas identiques, mais la tendance reste la même, à savoir un effet prédisposant significatif du SNP au CCR sporadique. Une explication probable à cette différence peut tenir dans le fait que les pools ne constituent qu'une fraction de notre population totale et donc surestiment les OR obtenus. Il est à noter que le SNP rs10505477 ne fait pas partie de la liste des SNP candidats que nous avons retenus au terme de nos analyses par la méthode SNP-MaP sur notre population d'étude. Ceci tient à la méthode que nous avons appliquée. En effet, le SNP rs10505477 étant intergénique et n'ayant pas été rapporté dans la littérature comme

un polymorphisme associé au CCR au moment de nos analyses, il ne pouvait en tout état de cause être sélectionné comme SNP candidat. De plus, comme nous l'avons vu, ce SNP, lorsqu'il est génotypé individuellement sur notre population, n'est pas fortement associé au CCR (OR=1,13, [1,00-1,27], p-value=0,04). Cela peut expliquer pourquoi ce SNP n'est pas ressorti lors notre d'analyse dérivée de la méthode SNP-MaP qui ne permet de détecter que des associations plus franches.

Deux autres SNP (rs1801282, rs1800566), dont nous disposons des données de génotypage individuel du fait de nos travaux antérieurs sur les polymorphismes de prédisposition au CCR sporadique (Articles 2), ont également été soumis à cette validation (Tableau 33). Pour ces deux SNP, comme pour le rs10505477 du locus 8q24.21, les OR obtenus par l'approche SNP-MaP (OR=0,89 et OR=0,98 respectivement) et par génotypage individuel (OR=0,95 et OR=1,07 respectivement) sont proches et non significatifs pour les deux méthodes employées. Ces résultats tendent à confirmer la validité de l'approche SNP-MaP qui, sur ces exemples du moins, ne fait pas ressortir de fausses associations.

Ces différentes étapes nous ont permis de valider l'utilisation de la méthode SNP-MaP. En effet, pour la majorité des SNP nous avons obtenu une concordance des significativités et des non significativités entre la méthode par allélotypage en pool et génotypage individuel. Pour les SNP trouvés significativement associés, nous avons pu constater que l'approche SNP-MaP surestimait leur effet. Pour les SNP trouvés associés par l'approche SNP-MaP, il faut tout de même garder une certaine réserve, car dans ces étapes de validation nous avons identifié un résultat qui pourrait être un faux-positif. Afin de pallier à ce problème, tous les résultats positifs par cette approche devront subir un génotypage individuel. Une limite de notre étape de validation réside dans le nombre limité de SNP pour lesquels nous avons pu comparer directement les associations déterminées à partir des génotypages individuels et de l'allélotypage par SNP-MaP. Ce travail de validation, bien que nécessaire, serait beaucoup trop onéreux, au regard du bénéfice escompté.

IV.3.4. Résultats

Après avoir sélectionné les sept SNP qui nous semblaient prédisposer le plus au CCR sporadique en appliquant notre méthode de sélection, nous avons réalisé un génotypage individuel de chacun d'entre eux. Comme les étapes de validation nous laissaient l'augurer, le génotypage individuel a permis de montrer des faux positifs dans notre sélection de polymorphismes de prédisposition (5 sur 7). Ces résultats confirment le fait que le génotypage individuel des SNP trouvés associés par la méthode SNP-MaP est indispensable. Pour les deux SNP dont l'association au CCR sporadique a été validée par génotypage individuel, nous avons constaté une surestimation de leur effet par SNP-MaP : pour le rs2306907 du gène *CDH13*, nous avons calculé des OR=1,54 par SNP-MaP

et 1,24 par génotypage individuel ; pour le rs3760454 du gène *UTP6*, nous avons calculé des OR=0,43 par SNP-MaP et 0,83 par génotypage individuel. Ces différences entre les OR obtenus par les deux techniques peuvent être expliquées par la constitution des pools qui ne représentent qu'une fraction de nos cohortes (73% des patients et 43% des témoins) et qui ont été choisis afin de faire ressortir un marqueur de prédisposition dépendant d'un phénotype particulier du CCR (endophénotype), ou caractéristique d'une localisation géographique.

Une autre explication possible pour expliquer ces variations d'OR entre les deux méthodes de génotypage individuel et par pool est basée sur la correction des données de génotypage des pools d'ADN par un facteur de correction appelé k . Ce facteur k a été mis au point pour passer outre l'hybridation différentielle des allèles qui est similaire à l'amplification différentielle observée en séquençage ou génotypage classique. En effet, pour une dose équivalente d'allèle de deux allèles différents qui ne sont pas marqués par la même molécule fluorescente, la fluorescence des signaux de ces allèles n'est pas équivalente, du fait des différences de chimies et de réponses du matériel à la fluorescence. Pour pallier à ce problème, il faut au préalable connaître les données de fluorescence pour un individu hétérozygote, afin d'en déduire le facteur k par la méthode mathématique décrite par Simpson *et al.* (2005). Cette correction par le facteur k a été réalisée par de nombreux auteurs (BUTCHER *et al.* 2005; KIROV *et al.* 2006; LE HELLARD *et al.* 2002) dans leurs études. Dans la nôtre, nous n'avons pas réalisé cette correction, car nous ne disposons pas des données de fluorescence d'un individu hétérozygote pour chaque SNP de la puce. De plus, Macgregor *et al.* (2006) et Meaburn *et al.* (2005) conseillent de n'utiliser ce facteur que lorsqu'il faut comparer les fréquences alléliques obtenues par pool d'ADN à celles obtenues par le génotypage individuel. L'impossibilité de calculer le facteur k est peut-être une autre raison pour laquelle nos résultats diffèrent selon l'approche individuelle ou par pool. Mathématiquement parlant, il ne devrait pas y avoir de différences entre ces résultats (résultats corrigés par le facteur k et non corrigés), car cette correction s'applique aux fréquences alléliques. Or, toutes les données entrant dans le calcul de l'OR d'un SNP donné sont corrigées par le même facteur k . Ce calcul devrait donc produire les mêmes résultats avec n'importe quelle donnée de génotypage.

IV.3.5. Réplication des associations au CCR pour le SNP

rs2306907 du gène *CDH13* et rs3760454 du gène *UTP6*

Afin de valider les deux associations mises en évidence par SNP-MaP sur notre population d'étude, nous avons collaboré avec l'équipe canadienne de B. Zanke dans le but de répliquer le test d'association des deux SNP (rs2306907 et rs3760454) sur leur propre population. Les résultats obtenus sur la population ARTIC de Zanke *et al.* ne montrent aucune association significative pour les

deux SNP (rs2306907, OR=1,02 [0,88-1,17], p-value=0,81 et rs3760454, OR=0,98 [0,87-1,10], p-value=0,78). Cette absence de réplication pourrait venir du fait que cette population est vraiment différente de la nôtre, même si elle est constituée de Caucasiens. En effet, notre population a subi un processus de recrutement strict en ce qui concerne l'origine ethnique des participants et le type de cancer des patients. En d'autres termes, notre cohorte est constituée uniquement d'individus caucasiens français atteints de cancers colorectaux sporadiques, alors que la population de réplication n'est pas limitée à un seul type de cancer, ce qui peut influencer grandement sur les résultats. A titre d'exemple, la population ARTIC de Zanke *et al.* réunit à la fois des CCR sporadiques et des CCR plutôt familiaux dont les gènes en cause sont encore inconnus. La non réplication observée sur les deux SNP rs2306907 de *CDH13* et rs3760454 de *UTP6* peut venir en partie de cette différence de recrutement. Afin de vérifier cette hypothèse, il faudrait connaître la proportion de cancers sporadiques de la population de réplication et calculer les OR pour ces individus. Il est bien sûr nécessaire de recommencer cette réplication sur d'autres populations, à savoir aussi bien des populations regroupant différentes formes de cancers colorectaux que des populations essentiellement constituées de formes sporadiques de cancer colorectal. C'est seulement en multipliant les réplifications que les associations seront véritablement éprouvées et que nous pourrions répondre à la seconde hypothèse.

IV.3.6. Corrélation entre le génotype et l'expression de l'ARNm des gènes d'intérêt, *CDH13* et *UTP6*

En parallèle de la réplication des SNP, nous avons mené une analyse de l'expression des ARNm des gènes *UTP6* et *CDH13* dans les tissus sains et tumoraux de patients atteints de CCR. Nous avons ainsi montré que l'expression du gène *UTP6* ne varie quasiment pas entre les tissus sains et tumoraux, alors que l'expression du gène *CDH13* est fortement augmentée dans les tissus tumoraux. De plus, ces taux d'expression ont pu être corrélés aux génotypes des SNP rs2306907 et rs3760454, qui, sur notre population, ont été montrés en association avec la pathologie. Cette analyse entre l'ARNm du gène *UTP6* et son SNP n'a pas été convaincante, alors que pour le gène *CDH13*, une corrélation semble se dessiner. Ce résultat est encourageant mais est à pondérer par le faible nombre d'individus analysés (36 patients).

La corrélation observée pour le gène *CDH13*, qui montre une perte d'expression dans les tissus tumoraux lorsque l'allèle variant est présent, est intéressante en ce sens où le SNP rs2306907 est localisé dans le site accepteur d'épissage de l'exon 6 du gène *CDH13* situé à 4 bases en 5'. Il serait possible que la différence d'expression du gène soit due à la présence de l'allèle variant du SNP, qui, par sa position modifierait le processus d'épissage. Cette option a été vérifiée par RT-PCR en

amplifiant la région codante autour de l'exon 6 d'une part et la totalité du gène d'autre part. Nous n'avons pu mettre en évidence cependant aucun transcrit alternatif.

Ces différents résultats prouvent que le SNP rs2306907 du gène *CDH13* n'est pas la cause de sa variation d'expression entre différents tissus tumoraux, mais qu'il est simplement un marqueur probablement en déséquilibre de liaison avec le SNP causal.

Il est bien entendu que ces résultats ne donnent qu'un aperçu de la corrélation entre l'expression de ces gènes et le génotype des variants identifiés, car, par manque de temps, nous n'avons pu étudier plus de couples de tissus tumoraux/sains. Ces résultats ne représentent donc qu'une tendance, qu'il faudrait confirmer par une investigation plus poussée.

IV.3.7. Analyse par le logiciel GenePool

Bien après que nous ayons réalisé nos analyses à partir de notre propre protocole SNP-MaP, nous avons eu connaissance du logiciel GenePool, spécifiquement dédié à l'analyse des pools d'ADN (PEARSON *et al.* 2007).

Le logiciel GenePool permet une analyse plus robuste et moins fastidieuse que celle que nous avons développée, du fait de l'automatisation de toutes les étapes allant du fichier généré par le fabricant des puces (fichier CEL) à la sélection des SNP candidats. De manière assez simple, ce logiciel permet d'identifier des loci de susceptibilité au CCR indiqués par une succession de SNP peu distants et tous associés au CCR. Un même gène peut ainsi être désigné par la succession de plusieurs SNP aussi bien intergéniques qu'intragéniques. La probabilité que la région où le gène induise une réelle susceptibilité au CCR augmente naturellement proportionnellement au nombre de SNP trouvés associés dans le locus. L'analyse par GenePool peut très bien être appliquée en complément de la méthode que nous avons développée, car elle n'est pas tout à fait basée sur la même méthode de discrimination des SNP et ne repose pas sur la détermination d'allélotypes.

IV.4. Les cadhérines et le rôle de la cadhérine-13 dans la prédisposition au cancer colorectal

Nos résultats par l'approche SNP-MaP ont mis en évidence l'association de la cadhérine-13 (*CDH13*) dans la prédisposition au cancer colorectal. Cette association a été confirmée par l'analyse effectuée avec le logiciel GenePool qui a également indiqué le possible rôle prédisposant d'autres cadhérines (*CDH2*, *CDH19*, *CDH18*, *CDH9*) ou de gènes liés aux cadhérines soit en appartenant à la même famille (*DSC3*, *PCDH10*, *PCDH15*,...), soit en appartenant à la même voie (*CTNNA2*). C'est donc tout naturellement que nous nous sommes intéressés à la fonction des cadhérines dans la cellule et plus particulièrement dans la cellule cancéreuse.

La relation la plus connue entre les cadhérines et le cancer est illustrée par l'implication de la cadhérine-1 (ou E-cadhérine) dans la carcinogenèse gastrique. Environ 1% des cancers gastriques sont des formes héréditaires diffuses dues à des mutations germinales dans le gène *CDH1*. L'E-cadhérine est connue pour établir et maintenir la polarisation des cellules épithéliales grâce à des complexes d'adhésion intercellulaire (PEDRAZZANI *et al.* 2007). En effet, l'E-cadhérine, comme la plupart des membres de la superfamille des cadhérines comportant plus de 350 membres répartis dans environ 30 espèces différentes (HULPIAU and VAN ROY 2009), joue un rôle dans l'adhésion intercellulaire. La fonction d'adhésion de l'E-cadhérine et son rôle dans le cancer sont dus à son organisation protéique en trois domaines : extracellulaire, inter-membranaire et intracellulaire. Le domaine extracellulaire est formé par cinq « domaines cadhérines » répétés en tandem contenant chacun environ 110 acides aminés et intervenant dans les interactions intercellulaires dépendantes du calcium (Ca^{2+}). Le domaine cytoplasmique, quant à lui, interagit avec les filaments d'actine du cytosquelette par les caténines (α , β , γ et p120^{ctn}) qui vont jouer un rôle dans la régulation de la voie de signalisation intracellulaire. En effet, le complexe cadhérine-caténine va jouer un rôle dans la promotion tumorale au travers de la voie de signalisation Wnt (Figure 52) (Chapitre I : II.3.1.1. La voie WNT/ β -caténine) (JEANES *et al.* 2008; PEDRAZZANI *et al.* 2007).

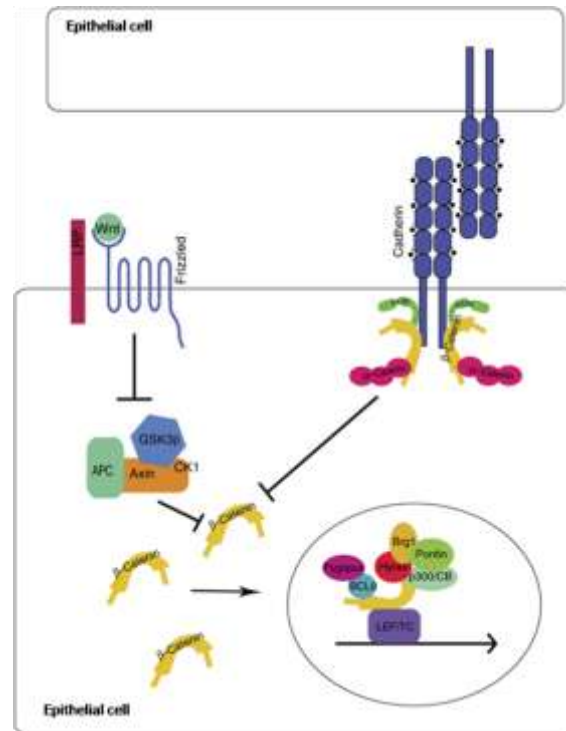


Figure 52 : Signalisation et régulation de la β -caténine impliquant l'E-cadhérine. La β -caténine cytosolique libre est régulée par la combinaison au complexe de destruction (APC, Axin, GSK3 β et CK1) et par la liaison au domaine cytoplasmique des cadhérines. Les deux mécanismes limitent l'accumulation de β -caténine libre et tendent à inhiber la signalisation par la voie Wnt (JEANES *et al.* 2008).

Il a été proposé que la cadhérine-13 (T- ou H-cadhérine) serait, quant à elle, un suppresseur de tumeur. En effet, la perte d'expression du gène *CDH13* par méthylation de son promoteur a été montrée comme pouvant promouvoir la tumorigenèse et l'invasion dans de nombreux cancers, comme le cancer colorectal (QIAN *et al.* 2007; TOYOOKA *et al.* 2002). La H-cadhérine est une exception dans la superfamille des cadhérines, puisqu'elle ne possède pas de domaine intracellulaire capable de s'ancrer aux filaments d'actine et elle est liée à la membrane plasmique via un glycosylphosphatidylinositol (GPI) (RANSCHT and DOURS-ZIMMERMANN 1991). L'analyse structurale du gène *CDH13* montre qu'il code pour une cadhérine « classique » comparable à CDH1, à ceci près qu'elle est dénuée d'une partie intracellulaire. Récemment, différentes études GWA ont montré l'implication du gène *CDH13* dans différentes pathologies complexes : Thomas *et al.* (2008) suggèrent que ce gène serait un candidat potentiel dans la susceptibilité au cancer de la prostate, tandis que d'autres équipes ont mis en avant son rôle probable dans différentes pathologies comme le mélanome (KUPHAL *et al.* 2009) ou dans les déficits de l'attention et l'hyperactivité (FRANKE *et al.* 2009).

Il a été montré que dans les cancers, les cellules subissent une transition morphologique les faisant passer de cellules épithéliales à mésenchymateuses. Cette transition, due à une commutation (ou *switch*) de cadhérines dans les cellules, entraîne une perte de polarité des cellules et le gain du pouvoir invasif (HAZAN *et al.* 2004; WHEELLOCK *et al.* 2008). Ce terme de *switch* est employé lorsque

l'expression de la E-cadhérine diminue, ou ne change pas, alors que l'expression d'une autre cadhérine augmente. Par exemple, Riou *et al.* (2006) ont constaté dans le carcinome hépatocellulaire une surexpression de la H-cadhérine (CDH13) au détriment de la E-cadhérine (CDH1). Ce phénomène pourrait être une des explications permettant de comprendre l'augmentation de l'expression de *CDH13* observée dans notre analyse alors que d'autres études, au contraire, ont montré la diminution de son expression (LEE 1996). Cette diminution d'expression est sans doute la conséquence de la méthylation de son promoteur (JIN *et al.* 2008). De plus, ce *switch* entre cadhérines peut également expliquer la prédisposition liée au gène *CDH13* au CCR qui semble ressortir de notre étude. En effet, les cancers colorectaux surviennent dans les cellules épithéliales où ces cellules vont subir une transition épithélio-mésenchymateuse probablement due à un *switch* entre différentes cadhérines. Cette commutation entre cadhérines va libérer de la β -caténine, qui va activer la voie de signalisation Wnt (Chapitre I : II.3.1.1. La voie WNT/ β -caténine). En outre, comme *CDH13* ne possède pas de domaine intracellulaire, le taux de β -caténines libres sera encore plus important que si le *switch* survenait entre deux cadhérines possédant un domaine intracellulaire. La piste de la commutation entre cadhérines apparaît par conséquent très prometteuse. Ce mécanisme, très hypothétique, nécessite néanmoins une validation biologique. En tout état de cause, il ne permettrait, au mieux, que d'expliquer une toute petite partie du très complexe phénomène de prédisposition, en ce sens où il devrait concerner un phénotype précis de CCR.

En outre, puisque nous avons retrouvé, à la fois par l'approche SNP-MaP et par l'analyse GenePool, plusieurs cadhérines (*CDH13*, *CDH2*, *CDH19*) associées au cancer colorectal, il serait intéressant d'effectuer une analyse la plus exhaustive possible des différents membres de la famille. Connaissant désormais maintenant la fonction de certaines cadhérines dans différents cancers, il semblerait légitime d'étudier les trois cadhérines potentiellement associées, voire également d'autres cadhérines comme la E-cadhérine, car elle pourrait être impliquée dans le *switch* des cadhérines. Dans le même ordre d'idée, d'autres membres de la superfamille des cadhérines comme les protocadhérines, les desmocollines ou les desmoglénines apparaissent comme de très bons candidats dans la perspective d'une étude exhaustive qui semble justifiée au regard de nos analyses. En effet, outre *CDH13*, on retrouve quatre autres cadhérines (*CDH18*, *CDH19*, *CDH2*, *CDH9*), mais aussi des protocadhérines (*PCDH10*, *PCDH15*, *PCDH17*, *PCDH20*, *PCDH9*) et une desmocolline (*DSC3*) qui ressortent lors de l'analyse croisée de la totalité des gènes significativement associés au CCR (4560 gènes) par notre méthode et celle reposant sur GenePool.

V. Conclusion et perspectives

L'identification des gènes de prédisposition dans les maladies complexes comme le cancer est un défi pour la génétique. Sur le plan technologique tout d'abord, l'investigation de nombreux gènes a été possible par l'apparition des puces à ADN et des techniques de génotypage haut débit. Sur le plan scientifique, la difficulté principale est l'intégration aux analyses de la contribution des facteurs environnementaux, en particulier de l'alimentation.

L'intérêt de l'identification de gènes de faible prédisposition peut permettre la mise en place de génotypage « préventif » dans les populations générales comme on peut le voir proposé actuellement sur internet par certaines sociétés, qui proposent une surveillance adaptée en cas d'identification de marqueurs génétiques de prédisposition. Cet intérêt peut paraître cependant un peu prématuré du fait du manque de validation d'un certain nombre de marqueurs testés. Il est plus intéressant, compte tenu de l'impact de certains facteurs environnementaux, de proposer des mesures diététiques ou d'hygiène de vie, que l'on connaît déjà comme bénéfiques pour la population générale.

On trouvera ailleurs l'intérêt des études comme celle que nous avons présentée ici. L'approche pangénomique doit permettre surtout une meilleure compréhension du mécanisme génétique de la carcinogenèse colorectale. Le chemin est parfois compliqué, notamment avec l'identification d'un marqueur génétique comme le polymorphisme trouvé en 8q24.21 (SCHUMACHER *et al.* 2007; ZANKE *et al.* 2007) dans les cancers colorectaux et de la prostate, dont on ne connaît pas le mécanisme d'action. Toutefois, dans d'autres pathologies, des gènes de prédisposition ont pu être identifiés avec succès, comme par exemple le gène *RELN* dans l'otosclérose (SCHRAUWEN *et al.* 2009) ou les gènes *KLF7* et *CREB1* dans le retard mental chez les enfants (BUTCHER *et al.* 2005).

Dans notre étude, l'approche originale choisie consistant à hybrider des pools d'ADN sur des puces pangénomiques est une solution économiquement satisfaisante à ce défi. Nous avons pu montrer que le gène *CDH13* prédisposerait au cancer colorectal, mais que le SNP rs2306907 ne serait qu'un marqueur de la pathologie. Compte tenu de ces résultats il faudrait réaliser d'autres répliques sur des populations caucasiennes, afin non seulement de valider ce SNP comme un marqueur du cancer colorectal, mais également afin de détecter le variant responsable de la prédisposition. Il faudrait dans un premier temps densifier le nombre de SNP analysés dans le gène et sa région afin de détecter le SNP causal. Une autre alternative pour étudier la prédisposition au CCR sporadique de *CDH13* serait de réaliser de l'immunohistochimie sur sa protéine pour connaître son expression dans les tissus cancéreux et sains, afin de valider son potentiel prédisposant au niveau cellulaire. De plus il serait également intéressant de poursuivre l'investigation sur les autres membres

de la famille des cadhérines, car leur rôle connu dans différents cancers, ajoutés au nombre de membres retrouvés dans nos analyses, en font de très bons candidats de prédisposition au CCR.

Le but ultime de notre étude serait de pouvoir réaliser un génotypage individuel par des puces pangénomiques de l'intégralité de notre population. Ainsi pourrions-nous réaliser l'étude complète des voies métaboliques et corréler ces résultats avec les données d'habitudes de vie dont nous disposons.

ANNEXES

QUESTIONNAIRE - HABITUDES DE VIE à compléter par le patient

NOM (3 premières lettres) : **PRENOM** (3 premières lettres) : **Date de naissance** :

Lieu de naissance des grands parents (ville ou région ou département) :

A défaut, lieu de naissance des parents :

Attention : Une seule réponse par question donc une seule case cochée par question

Etes vous :

Non-fumeur Fumeur actuel Ancien fumeur date d'arrêt :

Pratiquez vous une activité physique (marche, vélo, sports...) :

Moins de 1 heure / semaine 1 à 3 heures / semaine Plus de 3 heures / semaine

A propos de votre consommation d'alcool, en buvez vous :

Jamais Occasionnellement 2 à 3 fois / semaine Tous les jours

A propos de votre consommation de viande rouge (bœuf, agneau...), en mangez vous :

1 fois / semaine ou moins 2 à 4 fois / semaine Plus de 5 fois / semaine

A propos de votre consommation de volaille et/ou viande blanche (poulet, pintade, dinde, lapin, veau...), en mangez vous :

1 fois / semaine ou moins 2 à 4 fois / semaine Plus de 5 fois / semaine

A propos de votre consommation de charcuterie, en mangez vous :

1 fois / semaine ou moins 2 à 4 fois / semaine Plus de 5 fois / semaine

A propos de votre consommation de poisson, en mangez vous :

1 fois / semaine ou moins 2 à 4 fois / semaine Plus de 5 fois / semaine

Votre mode de cuisson privilégié (pour viande, volaille et poisson) est :

À la poêle ou au four avec matière grasse A la poêle ou au four sans matière grasse
Grillade Vapeur

A propos de votre consommation de fruits, en mangez vous :

Moins de 1/ jour 1/ jour 2/ jour ou plus

A propos de votre consommation de légumes, en mangez vous :

Moins de 1 fois / jour 1 fois / jour 2 fois / jour ou plus

A propos de votre consommation de pâtisserie, viennoiserie, glace, en mangez vous :

Moins de 1 fois / semaine 1 fois / semaine 2 fois / semaine ou plus

A propos de votre consommation en produits laitiers (lait, fromage, yaourt, crème dessert...), en mangez vous :

1 fois / jour ou moins 2 fois / jour 3 fois / jour ou plus

Consommez vous préférentiellement :

Du pain blanc Du pain complet

SVP, assurez vous d'avoir répondu à toutes les questions et merci pour votre participation.

FEUILLE DE RECUEIL DES DONNEES à compléter par le médecin

Constitution d'une banque d'ADN et étude des polymorphismes des gènes de susceptibilité aux cancers colorectaux sporadiques

DONNEES DEMOGRAPHIQUES

NOM : **Prénom :** **Date de naissance :**
 (3 premières lettres) (3 premières lettres) **Sexe :** M F

CARACTERISTIQUES AU DIAGNOSTIC

Année du diagnostic : **Poids (kg) :** **Taille (m) :**
Diabète : oui non **Dyslipidémie :** oui non
ATCD significatifs :

CARACTERISTIQUES TUMORALES

A - Données macroscopiques

* Localisation tumorale :

Rectum Côlon sigmoïde Côlon gauche
 Côlon transverse Côlon droit
*** Aspect :** Bourgeonnant Infiltrant Plan
*** Perforation :** Absente Présente
*** Polypes synchrones :** Absents Présents
*** Métastases synchrones :** Absentes Présentes
Localisations :

B - Données histologiques

*** Type :** Adénocarcinome Bien Moyennement Peu différencié

COMPOSANTE COLLOÏDE MUQUEUSE : %

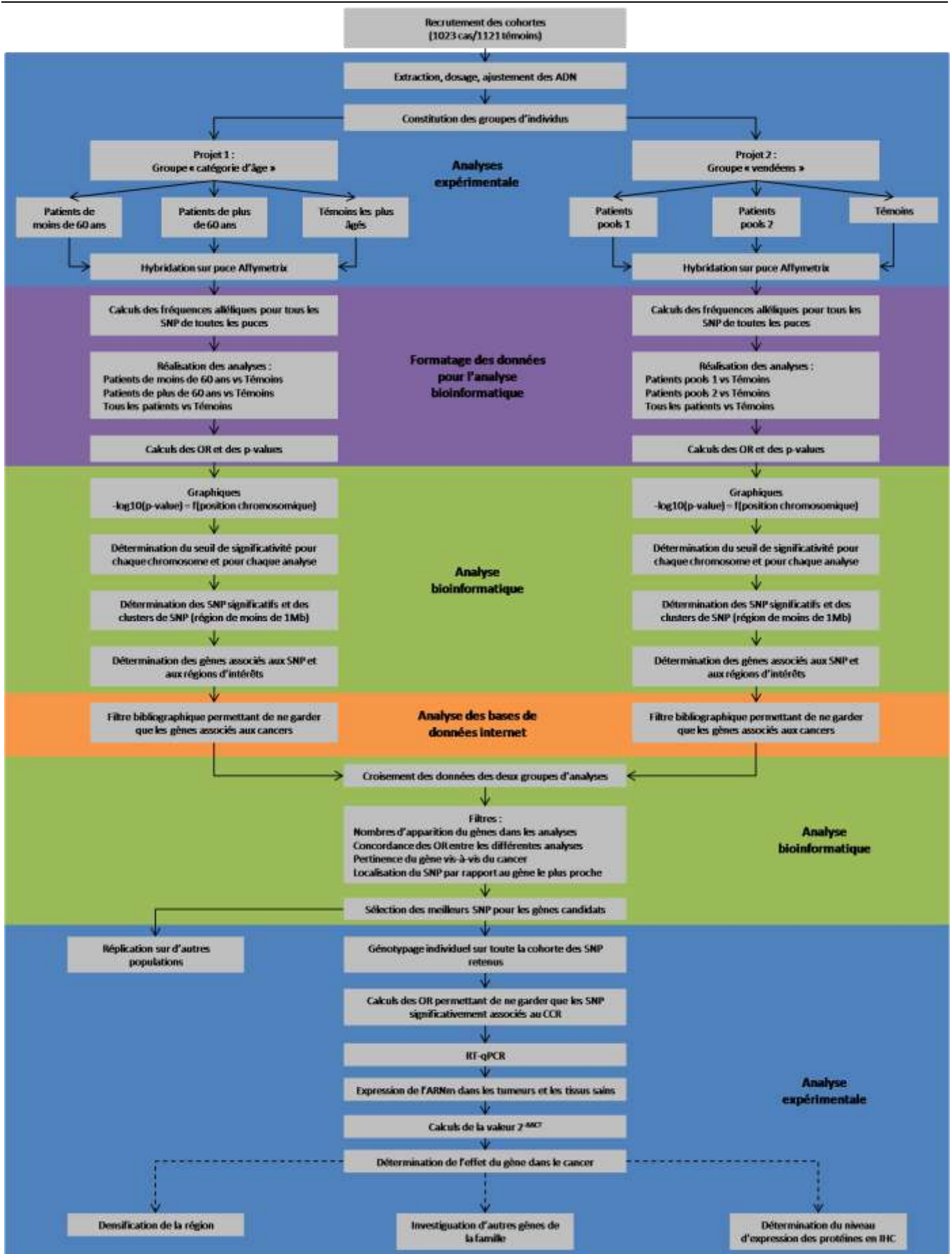
* Envahissement en profondeur :

Intramuqueux (Tis) Limité à la sous-muqueuse (T1)
 Limité à la musculieuse (T2) Envahissant la sous-séreuse (T3)
 Envahissement de la séreuse ou des organes de voisinage (T4)

* Extension ganglionnaire :

Nombre de ganglions prélevés : Nombre de ganglions envahis :
*** Emboles vasculaires :** Absents Présents
*** Engainements péri-nerveux :** Absents Présents
Stade pTNM :

Nom de l'investigateur : **Centre investigateur :**



Annexe 3 : Organigramme représentant les différentes opérations réalisées dans l'approche SNP-Map.

RÉFÉRENCES

- 2003 The International HapMap Project. *Nature* **426**: 789-796.
- AALTONEN, L. A., P. PELTOMAKI, F. S. LEACH, P. SISTONEN, L. PYLKANEN *et al.*, 1993 Clues to the pathogenesis of familial colorectal cancer. *Science* **260**: 812-816.
- AARONSON, S. A., 1991 Growth factors and cancer. *Science* **254**: 1146-1153.
- ABDEL-RAHMAN, S. Z., A. S. SOLIMAN, M. L. BONDY, S. OMAR, S. A. EL-BADAWY *et al.*, 2000 Inheritance of the 194Trp and the 399Gln variant alleles of the DNA repair gene XRCC1 are associated with increased risk of early-onset colorectal carcinoma in Egypt. *Cancer Lett* **159**: 79-86.
- ABDEL-RAHMAN, W. M., I. B. GEORGIADES, L. J. CURTIS, M. J. ARENDS and A. H. WYLLIE, 1999 Role of BAX mutations in mismatch repair-deficient colorectal carcinogenesis. *Oncogene* **18**: 2139-2142.
- AGGARWAL, B. B., and S. SHISHODIA, 2006 Molecular targets of dietary agents for prevention and therapy of cancer. *Biochem Pharmacol* **71**: 1397-1421.
- AGUILERA, A., and B. GOMEZ-GONZALEZ, 2008 Genome instability: a mechanistic view of its causes and consequences. *Nat Rev Genet* **9**: 204-217.
- AL-TASSAN, N., N. H. CHMIEL, J. MAYNARD, N. FLEMING, A. L. LIVINGSTON *et al.*, 2002 Inherited variants of MYH associated with somatic G:C-->T:A mutations in colorectal tumors. *Nat Genet* **30**: 227-232.
- ALEXANDER, J., T. WATANABE, T. T. WU, A. RASHID, S. LI *et al.*, 2001 Histopathological identification of colon cancer with microsatellite instability. *Am J Pathol* **158**: 527-535.
- ALVAREZ-LEON, E. E., B. ROMAN-VINAS and L. SERRA-MAJEM, 2006 Dairy products and health: a review of the epidemiological evidence. *Br J Nutr* **96 Suppl 1**: S94-99.
- AMEMORI, S., A. OOTANI, S. AOKI, T. FUJISE, R. SHIMODA *et al.*, 2007 Adipocytes and preadipocytes promote the proliferation of colon cancer cells in vitro. *Am J Physiol Gastrointest Liver Physiol* **292**: G923-929.
- AMES, B. N., 2006 Low micronutrient intake may accelerate the degenerative diseases of aging through allocation of scarce micronutrients by triage. *Proc Natl Acad Sci U S A* **103**: 17589-17594.
- ANDREYEV, H. J., A. R. NORMAN, D. CUNNINGHAM, J. OATES, B. R. DIX *et al.*, 2001 Kirsten ras mutations in patients with colorectal cancer: the 'RASCAL II' study. *Br J Cancer* **85**: 692-696.
- ANTI, M., G. MARRA, F. ARMELAO, G. M. BARTOLI, R. FICARELLI *et al.*, 1992 Effect of omega-3 fatty acids on rectal mucosal cell proliferation in subjects at risk for colon cancer. *Gastroenterology* **103**: 883-891.
- ARAKI, A., T. HOSOI, H. ORIMO and H. ITO, 2005 Association of plasma homocysteine with serum interleukin-6 and C-peptide levels in patients with type 2 diabetes. *Metabolism* **54**: 809-814.
- ARAND, M., R. MUHLBAUER, J. HENGSTLER, E. JAGER, J. FUCHS *et al.*, 1996 A multiplex polymerase chain reaction protocol for the simultaneous analysis of the glutathione S-transferase GSTM1 and GSTT1 polymorphisms. *Anal Biochem* **236**: 184-186.
- ARETZ, S., S. UHLHAAS, H. GOERGENS, K. SIBERG, M. VOGEL *et al.*, 2006 MUTYH-associated polyposis: 70 of 71 patients with biallelic mutations present with an attenuated or atypical phenotype. *Int J Cancer* **119**: 807-814.
- ARIGA, H., Y. IMAMURA and S. M. IGUCHI-ARIGA, 1989 DNA replication origin and transcriptional enhancer in c-myc gene share the c-myc protein binding sequences. *Embo J* **8**: 4273-4279.
- ARMITAGE, P., and R. DOLL, 1954 The age distribution of cancer and a multi-stage theory of carcinogenesis. *Br J Cancer* **8**: 1-12.
- ATTARDI, L. D., and R. A. DEPINHO, 2004 Conquering the complexity of p53. *Nat Genet* **36**: 7-8.
- AUWERX, J., 2002 Nuclear receptors. I. PPAR gamma in the gastrointestinal tract: gain or pain? *Am J Physiol Gastrointest Liver Physiol* **282**: G581-585.
- BADAWI, A. F., S. J. STERN, N. P. LANG and F. F. KADLUBAR, 1996 Cytochrome P-450 and acetyltransferase expression as biomarkers of carcinogen-DNA adduct levels and human cancer susceptibility. *Prog Clin Biol Res* **395**: 109-140.
- BAI, H., S. JONES, X. GUAN, T. M. WILSON, J. R. SAMPSON *et al.*, 2005 Functional characterization of two human MutY homolog (hMYH) missense mutations (R227W and V232F) that lie within the putative hMSH6 binding domain and are associated with hMYH polyposis. *Nucleic Acids Res* **33**: 597-604.
- BAKER, S. J., E. R. FEARON, J. M. NIGRO, S. R. HAMILTON, A. C. PREISINGER *et al.*, 1989 Chromosome 17 deletions and p53 gene mutations in colorectal carcinomas. *Science* **244**: 217-221.
- BALAGUER, F., S. CASTELLVI-BEL, A. CASTELLS, M. ANDREU, J. MUNOZ *et al.*, 2007 Identification of MYH mutation carriers in colorectal cancer: a multicenter, case-control, population-based study. *Clin Gastroenterol Hepatol* **5**: 379-387.
- BALDWIN, A. S., JR., 1996 The NF-kappa B and I kappa B proteins: new discoveries and insights. *Annu Rev Immunol* **14**: 649-683.
- BARBACID, M., 1987 ras genes. *Annu Rev Biochem* **56**: 779-827.
- BARDELLI, A., S. SAHA, J. A. SAGER, K. E. ROMANS, B. XIN *et al.*, 2003 PRL-3 expression in metastatic cancers. *Clin Cancer Res* **9**: 5607-5615.
- BARDOU, M., S. MONTEMBAULT, V. GIRAUD, A. BALIAN, E. BOROTTO *et al.*, 2002 Excessive alcohol consumption favours high risk polyp or colorectal cancer occurrence among patients with adenomas: a case control study. *Gut* **50**: 38-42.
- BARRATT, B. J., F. PAYNE, H. E. RANCE, S. NUTLAND, J. A. TODD *et al.*, 2002 Identification of the sources of error in allele frequency estimations from pooled DNA indicates an optimal experimental design. *Ann Hum Genet* **66**: 393-405.
- BARTSCH, H., U. NAIR, A. RISCH, M. ROJAS, H. WIKMAN *et al.*, 2000 Genetic polymorphism of CYP genes, alone or in combination, as a risk modifier of tobacco-related cancers. *Cancer Epidemiol Biomarkers Prev* **9**: 3-28.
- BAUMANN, J., F. VON BRUCHHAUSEN and G. WURM, 1980 Flavonoids and related compounds as inhibition of arachidonic acid peroxidation. *Prostaglandins* **20**: 627-639.

- BAYLIN, S. B., and J. G. HERMAN, 2000 DNA hypermethylation in tumorigenesis: epigenetics joins genetics. *Trends Genet* **16**: 168-174.
- BEAUNE, P., 1999 les cytochromes P450 humains : applications en toxicologie. *Med Ther* **4**: 18-38.
- BEAUNE, P., and M. A. LORIOT, 2000 Bases moléculaire de la susceptibilité aux xénobiotiques : aspects métaboliques. *Médecine/Sciences* 2000 **16**: 1051-1056.
- BELL, D. A., E. A. STEPHENS, T. CASTRANIO, D. M. UMBACH, M. WATSON *et al.*, 1995 Polyadenylation polymorphism in the acetyltransferase 1 gene (NAT1) increases risk of colorectal cancer. *Cancer Res* **55**: 3537-3542.
- BELOT, A., P. GROSCLAUDE, N. BOSSARD, E. JOUGLA, E. BENHAMOU *et al.*, 2008 Cancer incidence and mortality in France over the period 1980-2005. *Rev Epidemiol Sante Publique* **56**: 159-175.
- BENITO, E., E. CABEZA, V. MORENO, A. OBRADOR and F. X. BOSCH, 1993 Diet and colorectal adenomas: a case-control study in Majorca. *Int J Cancer* **55**: 213-219.
- BERESFORD, S. A., K. C. JOHNSON, C. RITENBAUGH, N. L. LASSER, L. G. SNETSELAAR *et al.*, 2006 Low-fat dietary pattern and risk of colorectal cancer: the Women's Health Initiative Randomized Controlled Dietary Modification Trial. *Jama* **295**: 643-654.
- BERNDT, S. I., E. A. PLATZ, M. D. FALLIN, L. W. THUITA, S. C. HOFFMAN *et al.*, 2006 Genetic variation in the nucleotide excision repair pathway and colorectal cancer risk. *Cancer Epidemiol Biomarkers Prev* **15**: 2263-2269.
- BERNDT, S. I., E. A. PLATZ, M. D. FALLIN, L. W. THUITA, S. C. HOFFMAN *et al.*, 2007 Mismatch repair polymorphisms and the risk of colorectal cancer. *Int J Cancer* **120**: 1548-1554.
- BEROUD, C., and T. SOUSSI, 1996 APC gene: database of germline and somatic mutations in human tumors and cell lines. *Nucleic Acids Res* **24**: 121-124.
- BEROUKHIM, R., M. LIN, Y. PARK, K. HAO, X. ZHAO *et al.*, 2006 Inferring loss-of-heterozygosity from unpaired tumors using high-density oligonucleotide SNP arrays. *PLoS Comput Biol* **2**: e41.
- BERTARIO, L., A. RUSSO, P. SALA, L. VARESCO, M. GIAROLA *et al.*, 2003 Multiple approach to the exploration of genotype-phenotype correlations in familial adenomatous polyposis. *J Clin Oncol* **21**: 1698-1707.
- BIGLER, J., C. M. ULRICH, T. KAWASHIMA, J. WHITTON and J. D. POTTER, 2005 DNA repair polymorphisms and risk of colorectal adenomatous or hyperplastic polyps. *Cancer Epidemiol Biomarkers Prev* **14**: 2501-2508.
- BIGNON, Y. J., 2004 *Genetic predisposition to cancer: biological basis of cancer predisposition*, London.
- BINGHAM, S. A., N. E. DAY, R. LUBEN, P. FERRARI, N. SLIMANI *et al.*, 2003 Dietary fibre in food and protection against colorectal cancer in the European Prospective Investigation into Cancer and Nutrition (EPIC): an observational study. *Lancet* **361**: 1496-1501.
- BINGHAM, S. A., B. PIGNATELLI, J. R. POLLOCK, A. ELLUL, C. MALAVEILLE *et al.*, 1996 Does increased endogenous formation of N-nitroso compounds in the human colon explain the association between red meat and colon cancer? *Carcinogenesis* **17**: 515-523.
- BISHOP, D. T., 2004 *Genetic predisposition to cancer: an introduction*, London.
- BJELLAND, S., and E. SEEBERG, 2003 Mutagenicity, toxicity and repair of DNA base damage induced by oxidation. *Mutat Res* **531**: 37-80.
- BJORK, J., J. NILSSON, R. HULTCRANTZ and C. JOHANSSON, 1993 Growth-regulatory effects of sensory neuropeptides, epidermal growth factor, insulin, and somatostatin on the non-transformed intestinal epithelial cell line IEC-6 and the colon cancer cell line HT 29. *Scand J Gastroenterol* **28**: 879-884.
- BLASCO, M. A., 2005 Telomeres and human disease: ageing, cancer and beyond. *Nat Rev Genet* **6**: 611-622.
- BLOBE, G. C., W. P. SCHIEMANN and H. F. LODISH, 2000 Role of transforming growth factor beta in human disease. *N Engl J Med* **342**: 1350-1358.
- BLOUNT, B. C., M. M. MACK, C. M. WEHR, J. T. MACGREGOR, R. A. HIATT *et al.*, 1997 Folate deficiency causes uracil misincorporation into human DNA and chromosome breakage: implications for cancer and neuronal damage. *Proc Natl Acad Sci U S A* **94**: 3290-3295.
- BOLAND, C. R., 2006 Decoding hereditary colorectal cancer. *N Engl J Med* **354**: 2815-2817.
- BOLAND, C. R., J. SATO, H. D. APPELMAN, R. S. BRESALIER and A. P. FEINBERG, 1995 Microallelotyping defines the sequence and tempo of allelic losses at tumour suppressor gene loci during colorectal cancer progression. *Nat Med* **1**: 902-909.
- BOS, J. L., 1988 The ras gene family and human carcinogenesis. *Mutat Res* **195**: 255-271.
- BOS, J. L., E. R. FEARON, S. R. HAMILTON, M. VERLAAN-DE VRIES, J. H. VAN BOOM *et al.*, 1987 Prevalence of ras gene mutations in human colorectal cancers. *Nature* **327**: 293-297.
- BOSTICK, R. M., 1997 Human studies of calcium supplementation and colorectal epithelial cell proliferation. *Cancer Epidemiol Biomarkers Prev* **6**: 971-980.
- BOURNE, H. R., 1987 Discovery of a new oncogene in pituitary tumours? *Nature* **330**: 517-518.
- BOURNE, H. R., D. A. SANDERS and F. McCORMICK, 1990 The GTPase superfamily: a conserved switch for diverse cell functions. *Nature* **348**: 125-132.
- BOWERS, K., D. ALBANES, P. LIMBURG, P. PIETINEN, P. R. TAYLOR *et al.*, 2006 A prospective study of anthropometric and clinical measurements associated with insulin resistance syndrome and colorectal cancer in male smokers. *Am J Epidemiol* **164**: 652-664.
- BOYLE, P., and J. S. LANGMAN, 2000 ABC of colorectal cancer: Epidemiology. *Bmj* **321**: 805-808.
- BRENNAN, P., 2002 Gene-environment interaction and aetiology of cancer: what does it mean and how can we measure it? *Carcinogenesis* **23**: 381-387.

- BREUER-KATSCHINSKI, B., K. NEMES, A. MARR, B. RUMP, B. LEIENDECKER *et al.*, 2000 Alcohol and cigarette smoking and the risk of colorectal adenomas. *Dig Dis Sci* **45**: 487-493.
- BREUER-KATSCHINSKI, B., K. NEMES, A. MARR, B. RUMP, B. LEIENDECKER *et al.*, 2001 Colorectal adenomas and diet: a case-control study. Colorectal Adenoma Study Group. *Dig Dis Sci* **46**: 86-95.
- BRODERICK, P., L. CARVAJAL-CARMONA, A. M. PITTMAN, E. WEBB, K. HOWARTH *et al.*, 2007 A genome-wide association study shows that common alleles of SMAD7 influence colorectal cancer risk. *Nat Genet* **39**: 1315-1317.
- BROOKS, L. D., 2003 SNPs: why do we care? *Methods Mol Biol* **212**: 1-14.
- BRYAN, T. M., and T. R. CECH, 1999 Telomerase and the maintenance of chromosome ends. *Curr Opin Cell Biol* **11**: 318-324.
- BULOW, S., T. BERK and K. NEALE, 2006 The history of familial adenomatous polyposis. *Fam Cancer* **5**: 213-220.
- BURGERING, B. M., R. H. MEDEMA, J. A. MAASSEN, M. L. VAN DE WETERING, A. J. VAN DER EB *et al.*, 1991 Insulin stimulation of gene expression mediated by p21ras activation. *Embo J* **10**: 1103-1109.
- BURKITT, D. P., 1971 Possible relationships between bowel cancer and dietary habits. *Proc R Soc Med* **64**: 964-965.
- BUSKENS, C. J., B. P. VAN REES, A. SIVULA, J. B. REITSMA, C. HAGLUND *et al.*, 2002 Prognostic significance of elevated cyclooxygenase 2 expression in patients with adenocarcinoma of the esophagus. *Gastroenterology* **122**: 1800-1807.
- BUSSEY, H., 1975 *Familial polyposis coli*, Baltimore.
- BUSSTRA, M. C., C. L. SIEZEN, M. J. GRUBBEN, H. J. VAN KRANEN, F. M. NAGENGAST *et al.*, 2003 Tissue levels of fish fatty acids and risk of colorectal adenomas: a case-control study (Netherlands). *Cancer Causes Control* **14**: 269-276.
- BUTCHER, L. M., E. MEABURN, J. KNIGHT, P. C. SHAM, L. C. SCHALKWYK *et al.*, 2005 SNPs, microarrays and pooled DNA: identification of four loci associated with mild mental impairment in a sample of 6000 children. *Hum Mol Genet* **14**: 1315-1325.
- BUTLER, W. J., P. RYAN and I. C. ROBERTS-THOMSON, 2001 Metabolic genotypes and risk for colorectal cancer. *J Gastroenterol Hepatol* **16**: 631-635.
- CAMPA, D., B. PARDINI, A. NACCARATI, L. VODICKOVA, J. NOVOTNY *et al.*, 2008 A gene-wide investigation on polymorphisms in the ABCG2/BRCP transporter and susceptibility to colorectal cancer. *Mutat Res* **645**: 56-60.
- CAPEAU, J., 2003 Voies de signalisation de l'insuline : mécanismes affectés dans l'insulino-résistance. *Médecine/Sciences* **19**: 834-839.
- CASPARI, R., W. FRIEDL, M. MANDL, G. MOSLEIN, M. KADMON *et al.*, 1994 Familial adenomatous polyposis: mutation at codon 1309 and early onset of colon cancer. *Lancet* **343**: 629-632.
- CHAI, F., A. EVDOKIOU, G. P. YOUNG and P. D. ZALEWSKI, 2000 Involvement of p21(Waf1/Cip1) and its cleavage by DEVD-caspase during apoptosis of colorectal cancer cells induced by butyrate. *Carcinogenesis* **21**: 7-14.
- CHAN, A. O., R. R. BROADDUS, P. S. HOULIHAN, J. P. ISSA, S. R. HAMILTON *et al.*, 2002 CpG island methylation in aberrant crypt foci of the colorectum. *Am J Pathol* **160**: 1823-1830.
- CHARBONNIER, F., G. RAUX, Q. WANG, N. DROUOT, F. CORDIER *et al.*, 2000 Detection of exon deletions and duplications of the mismatch repair genes in hereditary nonpolyposis colorectal cancer families using multiplex polymerase chain reaction of short fluorescent fragments. *Cancer Res* **60**: 2760-2763.
- CHEADLE, J. P., and J. R. SAMPSON, 2007 MUTYH-associated polyposis--from defect in base excision repair to clinical genetic testing. *DNA Repair (Amst)* **6**: 274-279.
- CHEN, J., E. GIOVANNUCCI, S. E. HANKINSON, J. MA, W. C. WILLETT *et al.*, 1998a A prospective study of methylenetetrahydrofolate reductase and methionine synthase gene polymorphisms, and risk of colorectal adenoma. *Carcinogenesis* **19**: 2129-2132.
- CHEN, J., E. GIOVANNUCCI, K. KELSEY, E. B. RIMM, M. J. STAMPFER *et al.*, 1996 A methylenetetrahydrofolate reductase polymorphism and the risk of colorectal cancer. *Cancer Res* **56**: 4862-4864.
- CHEN, J., C. KYTE, M. VALCIN, W. CHAN, J. G. WETMUR *et al.*, 2004 Polymorphisms in the one-carbon metabolic pathway, plasma folate levels and colorectal cancer in a prospective study. *Int J Cancer* **110**: 617-620.
- CHEN, J., M. J. STAMPFER, H. L. HOUGH, M. GARCIA-CLOSAS, W. C. WILLETT *et al.*, 1998b A prospective study of N-acetyltransferase genotype, red meat intake, and risk of colorectal cancer. *Cancer Res* **58**: 3307-3311.
- CHEN, W. Y., X. ZENG, M. G. CARTER, C. N. MORRELL, R. W. CHIU YEN *et al.*, 2003 Heterozygous disruption of Hic1 predisposes mice to a gender-dependent spectrum of malignant tumors. *Nat Genet* **33**: 197-202.
- CHENG, J., K. OGAWA, K. KURIKI, Y. YOKOYAMA, T. KAMIYA *et al.*, 2003 Increased intake of n-3 polyunsaturated fatty acids elevates the level of apoptosis in the normal sigmoid colon of patients polypectomized for adenomas/tumors. *Cancer Lett* **193**: 17-24.
- CHITTENDEN, T. W., E. A. HOWE, A. C. CULHANE, R. SULTANA, J. M. TAYLOR *et al.*, 2008 Functional classification analysis of somatically mutated genes in human breast and colorectal cancers. *Genomics* **91**: 508-511.
- CHO, E., S. A. SMITH-WARNER, D. SPIEGELMAN, W. L. BEESON, P. A. VAN DEN BRANDT *et al.*, 2004 Dairy foods, calcium, and colorectal cancer: a pooled analysis of 10 cohort studies. *J Natl Cancer Inst* **96**: 1015-1022.
- CHOI, S. W., K. J. LEE, Y. A. BAE, K. O. MIN, M. S. KWON *et al.*, 2002 Genetic classification of colorectal cancer based on chromosomal loss and microsatellite instability predicts survival. *Clin Cancer Res* **8**: 2311-2322.
- CHOI, S. W., and J. B. MASON, 2002 Folate status: effects on pathways of colorectal carcinogenesis. *J Nutr* **132**: 2413S-2418S.
- CHOW, E., L. LIPTON, E. LYNCH, R. D'SOUZA, C. ARAGONA *et al.*, 2006 Hyperplastic polyposis syndrome: phenotypic presentations and the role of MBD4 and MYH. *Gastroenterology* **131**: 30-39.
- CHUNG, D. C., and A. K. RUSTGI, 2003 The hereditary nonpolyposis colorectal cancer syndrome: genetics and clinical implications. *Ann Intern Med* **138**: 560-570.
- CLARK, W. H., JR., D. E. ELDER, D. T. GUERRY, L. E. BRAITMAN, B. J. TROCK *et al.*, 1989 Model predicting survival in stage I melanoma based on tumor progression. *J Natl Cancer Inst* **81**: 1893-1904.

- CLEMENTE, C. G., M. C. MIHM, JR., R. BUFALINO, S. ZURRIDA, P. COLLINI *et al.*, 1996 Prognostic value of tumor infiltrating lymphocytes in the vertical growth phase of primary cutaneous melanoma. *Cancer* **77**: 1303-1310.
- COMA, 1998 Nutritional aspects of the development of cancer. Report of the Working Group on Diet and Cancer of the Committee on Medical Aspects of Food and Nutrition Policy. *Rep Health Soc Subj (Lond)* **48**: i-xiv, 1-274.
- COMINGS, D. E., 1973 A general theory of carcinogenesis. *Proc Natl Acad Sci U S A* **70**: 3324-3328.
- COOGAN, P. F., L. ROSENBERG, C. LOUIK, A. G. ZAUBER, P. D. STOLLEY *et al.*, 2000 NSAIDs and risk of colorectal cancer according to presence or absence of family history of the disease. *Cancer Causes Control* **11**: 249-255.
- COOPER, J. A., D. F. BOWEN-POPE, E. RAINES, R. ROSS and T. HUNTER, 1982 Similar effects of platelet-derived growth factor and epidermal growth factor on the phosphorylation of tyrosine in cellular proteins. *Cell* **31**: 263-273.
- CORDON-CARDO, C., and C. PRIVES, 1999 At the crossroads of inflammation and tumorigenesis. *J Exp Med* **190**: 1367-1370.
- COTTON, S. C., L. SHARP, J. LITTLE and N. BROCKTON, 2000 Glutathione S-transferase polymorphisms and colorectal cancer: a HuGE review. *Am J Epidemiol* **151**: 7-32.
- COURTNEY, E. D., S. MATTHEWS, C. FINLAYSON, D. DI PIERRO, A. BELLUZZI *et al.*, 2007 Eicosapentaenoic acid (EPA) reduces crypt cell proliferation and increases apoptosis in normal colonic mucosa in subjects with a history of colorectal adenomas. *Int J Colorectal Dis* **22**: 765-776.
- COUSSENS, L. M., and Z. WERB, 2001 Inflammatory cells and cancer: think different! *J Exp Med* **193**: F23-26.
- COUSSENS, L. M., and Z. WERB, 2002 Inflammation and cancer. *Nature* **420**: 860-867.
- COX, D. G., C. PONTES, E. GUINO, M. NAVARRO, A. OSORIO *et al.*, 2004 Polymorphisms in prostaglandin synthase 2/cyclooxygenase 2 (PTGS2/COX2) and risk of colorectal cancer. *Br J Cancer* **91**: 339-343.
- CRANDALL, C. J., 1999 Estrogen replacement therapy and colon cancer: a clinical review. *J Womens Health Gend Based Med* **8**: 1155-1166.
- CRIPPS, W., 1882 Two cases of disseminated polyps of the rectum. *Trans Pathol Soc London* **33**.
- CROSS, A. J., J. R. POLLOCK and S. A. BINGHAM, 2003 Haem, not protein or inorganic iron, is responsible for endogenous intestinal N-nitrosation arising from red meat. *Cancer Res* **63**: 2358-2360.
- CROSS, A. J., and R. SINHA, 2004 Meat-related mutagens/carcinogens in the etiology of colorectal cancer. *Environ Mol Mutagen* **44**: 44-55.
- CUMMINGS, J. H., 1981 Dietary fibre and large bowel cancer. *Proc Nutr Soc* **40**: 7-14.
- CUNNINGHAM, J. M., E. R. CHRISTENSEN, D. J. TESTER, C. Y. KIM, P. C. ROCHE *et al.*, 1998 Hypermethylation of the hMLH1 promoter in colon cancer with microsatellite instability. *Cancer Res* **58**: 3455-3460.
- CURTIN, K., W. Y. LIN, R. GEORGE, M. KATORY, J. SHORTO *et al.*, 2009 Meta Association of Colorectal Cancer Confirms Risk Alleles at 8q24 and 18q21. *Cancer Epidemiol Biomarkers Prev*.
- DALLE, S., T. MARTIN-DENAVIT and L. THOMAS, 2006 [Genotypic hypervariability of melanoma: a therapeutic challenge]. *Med Sci (Paris)* **22**: 178-182.
- DALLOSSO, A. R., S. DOLWANI, N. JONES, S. JONES, J. COLLEY *et al.*, 2008 Inherited predisposition to colorectal adenomas caused by multiple rare alleles of MUTYH but not OGG1, NUDT1, NTH1 or NEIL 1, 2 or 3. *Gut* **57**: 1252-1255.
- DANNHARDT, G., and W. KIEFER, 2001 Cyclooxygenase inhibitors--current status and future prospects. *Eur J Med Chem* **36**: 109-126.
- DAVID, S. S., V. L. O'SHEA and S. KUNDU, 2007 Base-excision repair of oxidative DNA damage. *Nature* **447**: 941-950.
- DAVIDSON, N. O., 2007 Genetic testing in colorectal cancer: who, when, how and why. *Keio J Med* **56**: 14-20.
- DAVIES, H., G. R. BIGNELL, C. COX, P. STEPHENS, S. EDKINS *et al.*, 2002 Mutations of the BRAF gene in human cancer. *Nature* **417**: 949-954.
- DAVIS, O. S., R. PLOMIN and L. C. SCHALKWYK, 2009 The SNPMAp package for R: a framework for genome-wide association using DNA pooling on microarrays. *Bioinformatics* **25**: 281-283.
- DE JONG, M. M., I. M. NOLTE, G. J. TE MEERMAN, W. T. VAN DER GRAAF, E. G. DE VRIES *et al.*, 2002 Low-penetrance genes and their involvement in colorectal cancer susceptibility. *Cancer Epidemiol Biomarkers Prev* **11**: 1332-1352.
- DE LA CHAPELLE, A., 2004 Genetic predisposition to colorectal cancer. *Nat Rev Cancer* **4**: 769-780.
- DE LA CHAPELLE, A., and P. PELTOMAKI, 1995 Genetics of hereditary colon cancer. *Annu Rev Genet* **29**: 329-348.
- DEAKIN, M., J. ELDER, C. HENDRICKSE, D. PECKHAM, D. BALDWIN *et al.*, 1996 Glutathione S-transferase GSTT1 genotypes and susceptibility to cancer: studies of interactions with GSTM1 in lung, oral, gastric and colorectal cancers. *Carcinogenesis* **17**: 881-884.
- DEDOUSSIS, G. V., D. B. PANAGIOTAKOS, C. PITSAVOS, C. CHRYSOHOOU, J. SKOUMAS *et al.*, 2005 An association between the methylenetetrahydrofolate reductase (MTHFR) C677T mutation and inflammation markers related to cardiovascular disease. *Int J Cardiol* **100**: 409-414.
- DELATTRE, O., S. OLSCHWANG, D. J. LAW, T. MELOT, Y. REMVIKOS *et al.*, 1989 Multiple genetic alterations in distal and proximal colorectal cancer. *Lancet* **2**: 353-356.
- DENG, G., I. BELL, S. CRAWLEY, J. GUM, J. P. TERDIMAN *et al.*, 2004 BRAF mutation is frequently present in sporadic colorectal cancer with methylated hMLH1, but not in hereditary nonpolyposis colorectal cancer. *Clin Cancer Res* **10**: 191-195.
- DENG, G., A. CHEN, E. PONG and Y. S. KIM, 2001 Methylation in hMLH1 promoter interferes with its binding to transcription factor CBF and inhibits gene expression. *Oncogene* **20**: 7120-7127.
- DERYNYCK, R., and X. H. FENG, 1997 TGF-beta receptor signaling. *Biochim Biophys Acta* **1333**: F105-150.

- DIAZ, G. D., C. PARASKEVA, M. G. THOMAS, L. BINDERUP and A. HAGUE, 2000 Apoptosis is induced by the active metabolite of vitamin D3 and its analogue EB1089 in colorectal adenoma and carcinoma cells: possible implications for prevention and therapy. *Cancer Res* **60**: 2304-2312.
- DIHLMANN, S., A. SIERMANN and M. VON KNEBEL DOEBERITZ, 2001 The nonsteroidal anti-inflammatory drugs aspirin and indomethacin attenuate beta-catenin/TCF-4 signaling. *Oncogene* **20**: 645-653.
- DIMBERG, J., A. HUGANDER, A. SIRSIJO and P. SODERKVIST, 2001 Enhanced expression of cyclooxygenase-2 and nuclear beta-catenin are related to mutations in the APC gene in human colorectal cancer. *Anticancer Res* **21**: 911-915.
- DIONIGI, G., V. BIANCHI, F. ROVERA, L. BONI, M. ANNONI *et al.*, 2007 Genetic alteration in hereditary colorectal cancer. *Surg Oncol* **16 Suppl 1**: S11-15.
- DOCHERTY, S. J., L. M. BUTCHER, L. C. SCHALKWYK and R. PLOMIN, 2007 Applicability of DNA pools on 500 K SNP microarrays for cost-effective initial screens in genomewide association studies. *BMC Genomics* **8**: 214.
- DOXSEY, S., 1998 The centrosome--a tiny organelle with big potential. *Nat Genet* **20**: 104-106.
- DRANOFF, G., 2004 Cytokines in cancer pathogenesis and cancer therapy. *Nat Rev Cancer* **4**: 11-22.
- DUBOIS, R. N., 2003 Leukotriene A4 signaling, inflammation, and cancer. *J Natl Cancer Inst* **95**: 1028-1029.
- DUGGAN, D., S. L. ZHENG, M. KNOWLTON, D. BENITEZ, L. DIMITROV *et al.*, 2007 Two genome-wide association studies of aggressive prostate cancer implicate putative prostate tumor suppressor gene DAB2IP. *J Natl Cancer Inst* **99**: 1836-1844.
- EASTON, D. F., 2004 *Genetic predisposition to cancer: from families to chromosomes: genetic linkage and association studies for finding cancer-predisposition genes*, London.
- EASTWOOD, M. A., and E. R. MORRIS, 1992 Physical properties of dietary fiber that influence physiological function: a model for polymers along the gastrointestinal tract. *Am J Clin Nutr* **55**: 436-442.
- EBERHART, C. E., R. J. COFFEY, A. RADHIKA, F. M. GIARDIELLO, S. FERRENBACH *et al.*, 1994 Up-regulation of cyclooxygenase 2 gene expression in human colorectal adenomas and adenocarcinomas. *Gastroenterology* **107**: 1183-1188.
- EDELMANN, L., and W. EDELMANN, 2004 Loss of DNA mismatch repair function and cancer predisposition in the mouse: animal models for human hereditary nonpolyposis colorectal cancer. *Am J Med Genet* **129C**: 91-99.
- EDEN, A., F. GAUDET, A. WAGHMARE and R. JAENISCH, 2003 Chromosomal instability and tumors promoted by DNA hypomethylation. *Science* **300**: 455.
- EL-BAYOUMY, K., Y. H. CHAE, P. UPADHYAYA, C. MESCHTER, L. A. COHEN *et al.*, 1992 Inhibition of 7,12-dimethylbenz(a)anthracene-induced tumors and DNA adduct formation in the mammary glands of female Sprague-Dawley rats by the synthetic organoselenium compound, 1,4-phenylenebis(methylene)selenocyanate. *Cancer Res* **52**: 2402-2407.
- ENOMOTO, M., M. KONISHI, T. IWAMA, J. UTSUNOMIYA, K. I. SUGIHARA *et al.*, 2000 The relationship between frequencies of extracolonic manifestations and the position of APC germline mutation in patients with familial adenomatous polyposis. *Jpn J Clin Oncol* **30**: 82-88.
- EPPERT, K., S. W. SCHERER, H. OZCELIK, R. PIRONE, P. HOODLESS *et al.*, 1996 MADR2 maps to 18q21 and encodes a TGFbeta-regulated MAD-related protein that is functionally mutated in colorectal carcinoma. *Cell* **86**: 543-552.
- ESHLEMAN, J. R., E. Z. LANG, G. K. BOWERFIND, R. PARSONS, B. VOGELSTEIN *et al.*, 1995 Increased mutation rate at the hprt locus accompanies microsatellite instability in colon cancer. *Oncogene* **10**: 33-37.
- ESTELLER, M., M. F. FRAGA, M. GUO, J. GARCIA-FONCILLAS, I. HEDENFALK *et al.*, 2001 DNA methylation patterns in hereditary human cancers mimic sporadic tumorigenesis. *Hum Mol Genet* **10**: 3001-3007.
- EVA, A., K. C. ROBBINS, P. R. ANDERSEN, A. SRINIVASAN, S. R. TRONICK *et al.*, 1982 Cellular genes analogous to retroviral onc genes are transcribed in human tumour cells. *Nature* **295**: 116-119.
- FANTUZZI, G., 2005 Adipose tissue, adipokines, and inflammation. *J Allergy Clin Immunol* **115**: 911-919; quiz 920.
- FARRINGTON, S. M., A. TENESA, R. BARNETSON, A. WILTSHIRE, J. PRENDERGAST *et al.*, 2005 Germline susceptibility to colorectal cancer due to base-excision repair gene defects. *Am J Hum Genet* **77**: 112-119.
- FARROW, D. C., T. L. VAUGHAN, P. D. HANSTEN, J. L. STANFORD, H. A. RISCH *et al.*, 1998 Use of aspirin and other nonsteroidal anti-inflammatory drugs and risk of esophageal and gastric cancer. *Cancer Epidemiol Biomarkers Prev* **7**: 97-102.
- FEARON, E., 1998 *The genetic basis of human cancer: Tumor suppressor genes*, New York.
- FEARON, E. R., 1997 Human cancer syndromes: clues to the origin and nature of cancer. *Science* **278**: 1043-1050.
- FEARON, E. R., and B. VOGELSTEIN, 1990 A genetic model for colorectal tumorigenesis. *Cell* **61**: 759-767.
- FELSHER, D. W., and J. M. BISHOP, 1999 Transient excess of MYC activity can elicit genomic instability and tumorigenesis. *Proc Natl Acad Sci U S A* **96**: 3940-3944.
- FIALA, E. S., M. E. STARETZ, G. A. PANDYA, K. EL-BAYOUMY and S. R. HAMILTON, 1998 Inhibition of DNA cytosine methyltransferase by chemopreventive selenium compounds, determined by an improved assay for DNA cytosine methyltransferase and DNA cytosine methylation. *Carcinogenesis* **19**: 597-604.
- FICARI, F., A. CAMA, R. VALANZANO, M. C. CURIA, R. PALMIROTTA *et al.*, 2000 APC gene mutations and colorectal adenomatosis in familial adenomatous polyposis. *Br J Cancer* **82**: 348-353.
- FISHEL, R., M. K. LESCOE, M. R. RAO, N. G. COPELAND, N. A. JENKINS *et al.*, 1994 The human mutator gene homolog MSH2 and its association with hereditary nonpolyposis colon cancer. *Cell* **77**: 167.
- FLEISCHMANN, C., J. PETO, J. CHEADLE, B. SHAH, J. SAMPSON *et al.*, 2004 Comprehensive analysis of the contribution of germline MYH variation to early-onset colorectal cancer. *Int J Cancer* **109**: 554-558.
- FLIER, J. S., P. USHER and A. C. MOSES, 1986 Monoclonal antibody to the type I insulin-like growth factor (IGF-I) receptor blocks IGF-I receptor-mediated DNA synthesis: clarification of the mitogenic mechanisms of IGF-I and insulin in human skin fibroblasts. *Proc Natl Acad Sci U S A* **83**: 664-668.

- FODDE, R., R. SMITS and H. CLEVERS, 2001 APC, signal transduction and genetic instability in colorectal cancer. *Nat Rev Cancer* **1**: 55-67.
- FORRESTER, K., C. ALMOGUERA, K. HAN, W. E. GRIZZLE and M. PERUCHO, 1987 Detection of high incidence of K-ras oncogenes during human colon tumorigenesis. *Nature* **327**: 298-303.
- FRANCESCHI, S., 1999 Nutrients and food groups and large bowel cancer in Europe. *Eur J Cancer Prev* **8 Suppl 1**: S49-52.
- FRANCESCHI, S., and C. LA VECCHIA, 1998 Colorectal cancer and hormone replacement therapy: an unexpected finding. *Eur J Cancer Prev* **7**: 427-438.
- FRANKE, B., B. M. NEALE and S. V. FARAONE, 2009 Genome-wide association studies in ADHD. *Hum Genet.*
- FRIEDBERG, E. C., 2003 DNA damage and repair. *Nature* **421**: 436-440.
- FRIEDENREICH, C., T. NORAT, K. STEINDORF, M. C. BOUTRON-RUAULT, T. PISCHON *et al.*, 2006 Physical activity and risk of colon and rectal cancers: the European prospective investigation into cancer and nutrition. *Cancer Epidemiol Biomarkers Prev* **15**: 2398-2407.
- FRIEDL, W., R. CASPARI, M. SENGTELLER, S. UHLHAAS, C. LAMBERTI *et al.*, 2001 Can APC mutation analysis contribute to therapeutic decisions in familial adenomatous polyposis? Experience from 680 FAP families. *Gut* **48**: 515-521.
- FROSST, P., H. J. BLOM, R. MILOS, P. GOYETTE, C. A. SHEPPARD *et al.*, 1995 A candidate genetic risk factor for vascular disease: a common mutation in methylenetetrahydrofolate reductase. *Nat Genet* **10**: 111-113.
- FUCHS, C. S., E. L. GIOVANNUCCI, G. A. COLDITZ, D. J. HUNTER, M. J. STAMPFER *et al.*, 1999 Dietary fiber and the risk of colorectal cancer and adenoma in women. *N Engl J Med* **340**: 169-176.
- FUKASAWA, K., 2007 Oncogenes and tumour suppressors take on centrosomes. *Nat Rev Cancer* **7**: 911-924.
- FUSARO, R. M., S. J. LEMON and H. T. LYNCH, 1996 Muir-Torre syndrome and defective DNA mismatch repair genes. *J Am Acad Dermatol* **35**: 493-494.
- GAFA, R., I. MAESTRI, M. MATTEUZZI, A. SANTINI, S. FERRETTI *et al.*, 2000 Sporadic colorectal adenocarcinomas with high-frequency microsatellite instability. *Cancer* **89**: 2025-2037.
- GANN, P. H., J. E. MANSON, R. J. GLYNN, J. E. BURING and C. H. HENNEKENS, 1993 Low-dose aspirin and incidence of colorectal tumors in a randomized trial. *J Natl Cancer Inst* **85**: 1220-1224.
- GARCIA RODRIGUEZ, L. A., and C. HUERTA-ALVAREZ, 2000 Reduced incidence of colorectal adenoma among long-term users of nonsteroidal antiinflammatory drugs: a pooled analysis of published studies and a new population-based study. *Epidemiology* **11**: 376-381.
- GARDNER, E. J., 1951 A genetic and clinical study of intestinal polyposis, a predisposing factor for carcinoma of the colon and rectum. *Am J Hum Genet* **3**: 167-176.
- GARLAND, C. F., and F. C. GARLAND, 1980 Do sunlight and vitamin D reduce the likelihood of colon cancer? *Int J Epidemiol* **9**: 227-231.
- GATALICA, Z., and E. TORLAKOVIC, 2008 Pathology of the hereditary colorectal carcinoma. *Fam Cancer* **7**: 15-26.
- GAWRONSKA-SZKLARZ, B., J. LUBINSKI, J. KLADNY, G. KURZAWSKI, D. BIELICKI *et al.*, 1999 Polymorphism of GSTM1 gene in patients with colorectal cancer and colonic polyps. *Exp Toxicol Pathol* **51**: 321-325.
- GEBERT, J. F., C. DUPON, M. KADMON, M. HAHN, C. HERFARTH *et al.*, 1999 Combined molecular and clinical approaches for the identification of families with familial adenomatous polyposis coli. *Ann Surg* **229**: 350-361.
- GEBHARDT, C., J. NEMETH, P. ANGEL and J. HESS, 2006 S100A8 and S100A9 in inflammation and cancer. *Biochem Pharmacol* **72**: 1622-1631.
- GERHARDSSON DE VERDIER, M., U. HAGMAN, G. STEINECK, A. RIEGER and S. E. NORELL, 1990 Diet, body mass and colorectal cancer: a case-referent study in Stockholm. *Int J Cancer* **46**: 832-838.
- GERSON, S. L., 2004 MGMT: its role in cancer aetiology and cancer therapeutics. *Nat Rev Cancer* **4**: 296-307.
- GHOUSSAINI, M., H. SONG, T. KOESSLER, A. A. AL OLAMA, Z. KOTE-JARAI *et al.*, 2008 Multiple loci with different cancer specificities within the 8q24 gene desert. *J Natl Cancer Inst* **100**: 962-966.
- GIARDIELLO, F. M., G. J. OFFERHAUS and R. N. DUBOIS, 1995 The role of nonsteroidal anti-inflammatory drugs in colorectal cancer prevention. *Eur J Cancer* **31A**: 1071-1076.
- GIARDIELLO, F. M., G. M. PETERSEN, S. PIANTADOSI, S. B. GRUBER, E. I. TRABOULSI *et al.*, 1997 APC gene mutations and extraintestinal phenotype of familial adenomatous polyposis. *Gut* **40**: 521-525.
- GILLET, L. C., and O. D. SCHARER, 2006 Molecular mechanisms of mammalian global genome nucleotide excision repair. *Chem Rev* **106**: 253-276.
- GIOVANNUCCI, E., 2001a Insulin, insulin-like growth factors and colon cancer: a review of the evidence. *J Nutr* **131**: 3109S-3120S.
- GIOVANNUCCI, E., 2001b An updated review of the epidemiological evidence that cigarette smoking increases risk of colorectal cancer. *Cancer Epidemiol Biomarkers Prev* **10**: 725-731.
- GIOVANNUCCI, E., A. ASCHERIO, E. B. RIMM, G. A. COLDITZ, M. J. STAMPFER *et al.*, 1995a Physical activity, obesity, and risk for colon cancer and adenoma in men. *Ann Intern Med* **122**: 327-334.
- GIOVANNUCCI, E., E. B. RIMM, A. ASCHERIO, M. J. STAMPFER, G. A. COLDITZ *et al.*, 1995b Alcohol, low-methionine-low-folate diets, and risk of colon cancer in men. *J Natl Cancer Inst* **87**: 265-273.
- GIOVANNUCCI, E., E. B. RIMM, M. J. STAMPFER, G. A. COLDITZ, A. ASCHERIO *et al.*, 1994 Aspirin use and the risk for colorectal cancer and adenoma in male health professionals. *Ann Intern Med* **121**: 241-246.
- GIOVANNUCCI, E., M. J. STAMPFER, G. A. COLDITZ, D. J. HUNTER, C. FUCHS *et al.*, 1998 Multivitamin use, folate, and colon cancer in women in the Nurses' Health Study. *Ann Intern Med* **129**: 517-524.

- GLICK, A. B., W. C. WEINBERG, I. H. WU, W. QUAN and S. H. YUSPA, 1996 Transforming growth factor beta 1 suppresses genomic instability independent of a G1 arrest, p53, and Rb. *Cancer Res* **56**: 3645-3650.
- GOALSTONE, M. L., K. WALL, J. W. LEITNER, T. KUROWSKI, N. RUDERMAN *et al.*, 1999 Increased amounts of farnesylated p21Ras in tissues of hyperinsulinaemic animals. *Diabetologia* **42**: 310-316.
- GOLIAS, C. H., A. CHARALABOPOULOS and K. CHARALABOPOULOS, 2004 Cell proliferation and cell cycle control: a mini review. *Int J Clin Pract* **58**: 1134-1141.
- GOODE, E. L., J. D. POTTER, W. R. BAMLET, D. N. RIDER and J. BIGLER, 2006 Inherited variation in carcinogen-metabolizing enzymes and risk of colorectal polyps. *Carcinogenesis*.
- GOODMAN, J. E., E. D. BOWMAN, S. J. CHANOCK, A. J. ALBERG and C. C. HARRIS, 2004 Arachidonate lipoxygenase (ALOX) and cyclooxygenase (COX) polymorphisms and colon cancer risk. *Carcinogenesis* **25**: 2467-2472.
- GOODMAN, J. E., L. E. MECHANIC, B. T. LUKE, S. AMBS, S. CHANOCK *et al.*, 2006 Exploring SNP-SNP interactions and colon cancer risk using polymorphism interaction analysis. *Int J Cancer* **118**: 1790-1797.
- GOSS, K. H., and J. GRODEN, 2000 Biology of the adenomatous polyposis coli tumor suppressor. *J Clin Oncol* **18**: 1967-1979.
- GRADY, W. M., and J. M. CARETHERS, 2008 Genomic and epigenetic instability in colorectal cancer pathogenesis. *Gastroenterology* **135**: 1079-1099.
- GRADY, W. M., A. RAJPUT, L. MYEROFF, D. F. LIU, K. KWON *et al.*, 1998 Mutation of the type II transforming growth factor-beta receptor is coincident with the transformation of human colon adenomas to malignant carcinomas. *Cancer Res* **58**: 3101-3104.
- GREENBERG, E. R., J. A. BARON, D. H. FREEMAN, JR., J. S. MANDEL and R. HAILE, 1993 Reduced risk of large-bowel adenomas among aspirin users. The Polyp Prevention Study Group. *J Natl Cancer Inst* **85**: 912-916.
- GRODEN, J., A. THLIVERIS, W. SAMOWITZ, M. CARLSON, L. GELBERT *et al.*, 1991 Identification and characterization of the familial adenomatous polyposis coli gene. *Cell* **66**: 589-600.
- GRUBER, S. B., V. MORENO, L. S. ROZEK, H. S. RENNERT, F. LEJBKOWICZ *et al.*, 2007 Genetic Variation in 8q24 Associated with Risk of Colorectal Cancer. *Cancer Biol Ther* **6**.
- GUNTER, M. J., and M. F. LEITZMANN, 2006 Obesity and colorectal cancer: epidemiology, mechanisms and candidate genes. *J Nutr Biochem* **17**: 145-156.
- GUNTER, M. J., R. STOLZENBERG-SOLOMON, A. J. CROSS, M. F. LEITZMANN, S. WEINSTEIN *et al.*, 2006 A prospective study of serum C-reactive protein and colorectal cancer risk in men. *Cancer Res* **66**: 2483-2487.
- GUO, Y. S., S. NARAYAN, C. YALLAMPALLI and P. SINGH, 1992 Characterization of insulinlike growth factor I receptors in human colon cancer. *Gastroenterology* **102**: 1101-1108.
- GUTIERREZ, L., A. I. MAGEE, C. J. MARSHALL and J. F. HANCOCK, 1989 Post-translational processing of p21ras is two-step and involves carboxyl-methylation and carboxy-terminal proteolysis. *Embo J* **8**: 1093-1098.
- HALL, M. N., H. CAMPOS, H. LI, H. D. SESSO, M. J. STAMPFER *et al.*, 2007 Blood levels of long-chain polyunsaturated fatty acids, aspirin, and the risk of colorectal cancer. *Cancer Epidemiol Biomarkers Prev* **16**: 314-321.
- HALUSHKA, M. K., L. P. WALKER and P. V. HALUSHKA, 2003 Genetic variation in cyclooxygenase 1: effects on response to aspirin. *Clin Pharmacol Ther* **73**: 122-130.
- HAMPE, J., A. FRANKE, P. ROSENSTIEL, A. TILL, M. TEUBER *et al.*, 2007 A genome-wide association scan of nonsynonymous SNPs identifies a susceptibility variant for Crohn disease in ATG16L1. *Nat Genet* **39**: 207-211.
- HANSEN, R., M. SAEBO, C. F. SKJELBRED, B. A. NEXO, P. C. HAGEN *et al.*, 2005 GPX Pro198Leu and OGG1 Ser326Cys polymorphisms and risk of development of colorectal adenomas and colorectal cancer. *Cancer Lett* **229**: 85-91.
- HARRIS, M. J., M. COGGAN, L. LANGTON, S. R. WILSON and P. G. BOARD, 1998 Polymorphism of the Pi class glutathione S-transferase in normal populations and cancer patients. *Pharmacogenetics* **8**: 27-31.
- HATTERSLEY, A. T., and M. I. MCCARTHY, 2005 What makes a good genetic association study? *Lancet* **366**: 1315-1323.
- HAZAN, R. B., R. QIAO, R. KEREN, I. BADANO and K. SUYAMA, 2004 Cadherin switch in tumor progression. *Ann N Y Acad Sci* **1014**: 155-163.
- HEAVEY, P. M., D. MCKENNA and I. R. ROWLAND, 2004 Colorectal Cancer and the Relationship Between Genes and the Environment. *Nutr Cancer* **48**: 124-141.
- HELBOCK, H. J., K. B. BECKMAN, M. K. SHIGENAGA, P. B. WALTER, A. A. WOODALL *et al.*, 1998 DNA oxidation matters: the HPLC-electrochemical detection assay of 8-oxo-deoxyguanosine and 8-oxo-guanine. *Proc Natl Acad Sci U S A* **95**: 288-293.
- HELDIN, C. H., K. MIYAZONO and P. TEN DIJKE, 1997 TGF-beta signalling from cell membrane to nucleus through SMAD proteins. *Nature* **390**: 465-471.
- HERMAN, J. G., A. MERLO, L. MAO, R. G. LAPIDUS, J. P. ISSA *et al.*, 1995 Inactivation of the CDKN2/p16/MTS1 gene is frequently associated with aberrant DNA methylation in all common human cancers. *Cancer Res* **55**: 4525-4530.
- HERMAN, J. G., A. UMAR, K. POLYAK, J. R. GRAFF, N. AHUJA *et al.*, 1998 Incidence and functional consequences of hMLH1 promoter hypermethylation in colorectal carcinoma. *Proc Natl Acad Sci U S A* **95**: 6870-6875.
- HIRSCHHORN, J. N., and M. J. DALY, 2005 Genome-wide association studies for common diseases and complex traits. *Nat Rev Genet* **6**: 95-108.
- HODA, M. R., S. J. KEELY, L. S. BERTELSEN, W. G. JUNGER, D. DHARMASENA *et al.*, 2007 Leptin acts as a mitogenic and antiapoptotic factor for colonic cancer cells. *Br J Surg* **94**: 346-354.
- HOEIJMAKERS, J. H., 2001 Genome maintenance mechanisms for preventing cancer. *Nature* **411**: 366-374.
- HOFSETH, L. J., S. P. HUSSAIN, G. N. WOGAN and C. C. HARRIS, 2003a Nitric oxide in cancer and chemoprevention. *Free Radic Biol Med* **34**: 955-968.

- HOFSETH, L. J., S. SAITO, S. P. HUSSAIN, M. G. ESPEY, K. M. MIRANDA *et al.*, 2003b Nitric oxide-induced cellular stress and p53 activation in chronic inflammation. *Proc Natl Acad Sci U S A* **100**: 143-148.
- HONG, Y. C., K. H. LEE, W. C. KIM, S. K. CHOI, Z. H. WOO *et al.*, 2005 Polymorphisms of XRCC1 gene, alcohol consumption and colorectal cancer. *Int J Cancer* **116**: 428-432.
- HOTAMISLIGIL, G. S., 2006 Inflammation and metabolic disorders. *Nature* **444**: 860-867.
- HOULSTON, R. S., A. COLLINS, J. SLACK and N. E. MORTON, 1992 Dominant genes for colorectal cancer are not rare. *Ann Hum Genet* **56**: 99-103.
- HUBNER, R. A., and R. S. HOULSTON, 2009 Folate and colorectal cancer prevention. *Br J Cancer* **100**: 233-239.
- HUDSON, J. D., M. A. SHOAIIBI, R. MAESTRO, A. CARNERO, G. J. HANNON *et al.*, 1999 A proinflammatory cytokine inhibits p53 tumor suppressor activity. *J Exp Med* **190**: 1375-1382.
- HUGHES, R., A. J. CROSS, J. R. POLLOCK and S. BINGHAM, 2001 Dose-dependent effect of dietary meat on endogenous colonic N-nitrosation. *Carcinogenesis* **22**: 199-202.
- HULPIAU, P., and F. VAN ROY, 2009 Molecular evolution of the cadherin superfamily. *Int J Biochem Cell Biol* **41**: 349-369.
- HUNTER, D. J., P. KRAFT, K. B. JACOBS, D. G. COX, M. YEAGER *et al.*, 2007 A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. *Nat Genet* **39**: 870-874.
- HUNTER, T., and J. A. COOPER, 1985 Protein-tyrosine kinases. *Annu Rev Biochem* **54**: 897-930.
- HURLSTONE, D. P., M. KARAJEH, S. S. CROSS, M. E. McALINDON, S. BROWN *et al.*, 2005 The role of high-magnification-chromoscopic colonoscopy in hereditary nonpolyposis colorectal cancer screening: a prospective "back-to-back" endoscopic study. *Am J Gastroenterol* **100**: 2167-2173.
- HUSSAIN, S. P., L. J. HOFSETH and C. C. HARRIS, 2003 Radical causes of cancer. *Nat Rev Cancer* **3**: 276-285.
- HUSSAIN, S. P., G. E. TRIVERS, L. J. HOFSETH, P. HE, I. SHAIKH *et al.*, 2004 Nitric oxide, a mediator of inflammation, suppresses tumorigenesis. *Cancer Res* **64**: 6849-6853.
- IACOPETTA, B., 2003 TP53 mutation in colorectal cancer. *Hum Mutat* **21**: 271-276.
- ILETT, K. F., B. M. DAVID, P. DETCHON, W. M. CASTLEDEN and R. KWA, 1987 Acetylation phenotype in colorectal carcinoma. *Cancer Res* **47**: 1466-1469.
- ILYAS, M., J. STRAUB, I. P. TOMLINSON and W. F. BODMER, 1999 Genetic pathways in colorectal and other cancers. *Eur J Cancer* **35**: 1986-2002.
- INIGUEZ, M. A., A. RODRIGUEZ, O. V. VOLPERT, M. FRESNO and J. M. REDONDO, 2003 Cyclooxygenase-2: a therapeutic target in angiogenesis. *Trends Mol Med* **9**: 73-78.
- INNOCENTI, F., C. GRIMSLEY, S. DAS, J. RAMIREZ, C. CHENG *et al.*, 2002 Haplotype structure of the UDP-glucuronosyltransferase 1A1 promoter in different ethnic groups. *Pharmacogenetics* **12**: 725-733.
- ITO, Y., K. SUZUKI, K. TAMAKOSHI, K. WAKAI, M. KOJIMA *et al.*, 2005 Colorectal cancer and serum C-reactive protein levels: a case-control study nested in the JACC Study. *J Epidemiol* **15 Suppl 2**: S185-189.
- ITZKOWITZ, S. H., and X. YIO, 2004 Inflammation and cancer IV. Colorectal cancer in inflammatory bowel disease: the role of inflammation. *Am J Physiol Gastrointest Liver Physiol* **287**: G7-17.
- JAIN, R. K., 2002 Tumor angiogenesis and accessibility: role of vascular endothelial growth factor. *Semin Oncol* **29**: 3-9.
- JALLEPALLI, P. V., and C. LENGAUER, 2001 Chromosome segregation and cancer: cutting through the mystery. *Nat Rev Cancer* **1**: 109-117.
- JANSSEN, I., P. T. KATZMARZYK and R. ROSS, 2004 Waist circumference and not body mass index explains obesity-related health risk. *Am J Clin Nutr* **79**: 379-384.
- JASS, J. R., 2004 Role of the pathologist in the diagnosis of hereditary non-polyposis colorectal cancer. *Dis Markers* **20**: 215-224.
- JASS, J. R., 2006 Hereditary Non-Polyposis Colorectal Cancer: the rise and fall of a confusing term. *World J Gastroenterol* **12**: 4943-4950.
- JASS, J. R., 2008 Colorectal polyposes: from phenotype to diagnosis. *Pathol Res Pract* **204**: 431-447.
- JASS, J. R., and S. M. STEWART, 1992 Evolution of hereditary non-polyposis colorectal cancer. *Gut* **33**: 783-786.
- JEANES, A., C. J. GOTTARDI and A. S. YAP, 2008 Cadherins and cancer: how does cadherin dysfunction promote tumor progression? *Oncogene* **27**: 6920-6929.
- JHUN, B. H., J. L. MEINKOTH, J. W. LEITNER, B. DRAZNIN and J. M. OLEFSKY, 1994 Insulin and insulin-like growth factor-I signal transduction requires p21ras. *J Biol Chem* **269**: 5699-5704.
- JIANG, J., V. GAJALAKSHMI, J. WANG, K. KURIKI, S. SUZUKI *et al.*, 2005 Influence of the C161T but not Pro12Ala polymorphism in the peroxisome proliferator-activated receptor-gamma on colorectal cancer in an Indian population. *Cancer Sci* **96**: 507-512.
- JIN, M. J., K. CHEN, L. SONG, C. H. FAN, Q. CHEN *et al.*, 2005 The association of the DNA repair gene XRCC3 Thr241Met polymorphism with susceptibility to colorectal cancer in a Chinese population. *Cancer Genet Cytogenet* **163**: 38-43.
- JIN, Z., Y. CHENG, A. OLARU, T. KAN, J. YANG *et al.*, 2008 Promoter hypermethylation of CDH13 is a common, early event in human esophageal adenocarcinogenesis and correlates with clinical risk factors. *Int J Cancer* **123**: 2331-2336.
- JOHANNSDOTTIR, J. T., J. G. JONASSON, J. T. BERGTHORSSON, L. T. AMUNDADOTTIR, J. MAGNUSSON *et al.*, 2000 The effect of mismatch repair deficiency on tumorigenesis; microsatellite instability affecting genes containing short repeated sequences. *Int J Oncol* **16**: 133-139.
- JOHN, B. J., S. IRUKULLA, A. M. ABULAFI, D. KUMAR and M. A. MENDALL, 2006 Systematic review: adipose tissue, obesity and gastrointestinal diseases. *Aliment Pharmacol Ther* **23**: 1511-1523.

- JOHNSON, I. T., 2001 Mechanisms and anticarcinogenic effects of diet-related apoptosis in the intestinal mucosa. *Nutr Res Rev* **14**: 229-256.
- JOHNSON, I. T., 2007 Phytochemicals and cancer. *Proc Nutr Soc* **66**: 207-215.
- JOHNSON, I. T., and E. K. LUND, 2007 Review article: nutrition, obesity and colorectal cancer. *Aliment Pharmacol Ther* **26**: 161-181.
- JOHNSON, I. T., G. WILLIAMSON and S. R. MUSK, 1994 Anticarcinogenic factors in plant foods: a new class of nutrients? *Nutr Res Rev* **7**: 175-204.
- JONES, J. I., and D. R. CLEMMONS, 1995 Insulin-like growth factors and their binding proteins: biological actions. *Endocr Rev* **16**: 3-34.
- JONES, S., P. EMMERSON, J. MAYNARD, J. M. BEST, S. JORDAN *et al.*, 2002 Biallelic germline mutations in MYH predispose to multiple colorectal adenoma and somatic G:C-->T:A mutations. *Hum Mol Genet* **11**: 2961-2967.
- JUNG, B., R. T. DOCTOLERO, A. TAJIMA, A. K. NGUYEN, T. KEKU *et al.*, 2004 Loss of activin receptor type 2 protein expression in microsatellite unstable colon cancers. *Gastroenterology* **126**: 654-659.
- JUNG, Y., J. S. ISAACS, S. LEE, J. TREPEL, Z. G. LIU *et al.*, 2003a Hypoxia-inducible factor induction by tumour necrosis factor in normoxic cells requires receptor-interacting protein-dependent nuclear factor kappa B activation. *Biochem J* **370**: 1011-1017.
- JUNG, Y. J., J. S. ISAACS, S. LEE, J. TREPEL and L. NECKERS, 2003b IL-1beta-mediated up-regulation of HIF-1alpha via an NFkappaB/COX-2 pathway identifies HIF-1 as a critical link between inflammation and oncogenesis. *Faseb J* **17**: 2115-2117.
- JUNIEN, C., 2001 [Colon cancer and nutritional genetics: modifier genes]. *Ann Med Interne (Paris)* **152**: 337-351.
- KADIYSKA, T. K., D. V. KONSTANTINOVA, V. R. ATANASOV, I. M. KREMENSKY and V. I. MITEV, 2007 Frequency and application of the hot spot BRAF gene mutation (p.V600E) in the diagnostic strategy for Hereditary Nonpolyposis Colorectal Cancer. *Cancer Detect Prev* **31**: 254-256.
- KAIRUPAN, C. F., C. J. MELDRUM, R. CROOKS, E. A. MILWARD, A. D. SPIGELMAN *et al.*, 2005 Mutation analysis of the MYH gene in an Australian series of colorectal polyposis patients with or without germline APC mutations. *Int J Cancer* **116**: 73-77.
- KAMPMAN, E., M. L. SLATTERY, J. BIGLER, M. LEPPERT, W. SAMOWITZ *et al.*, 1999 Meat consumption, genetic susceptibility, and colon cancer risk: a United States multicenter case-control study. *Cancer Epidemiol Biomarkers Prev* **8**: 15-24.
- KANE, M. F., M. LODA, G. M. GAIDA, J. LIPMAN, R. MISHRA *et al.*, 1997 Methylation of the hMLH1 promoter correlates with lack of expression of hMLH1 in sporadic colon tumors and mismatch repair-defective human tumor cell lines. *Cancer Res* **57**: 808-811.
- KARIN, M., and Y. BEN-NERIAH, 2000 Phosphorylation meets ubiquitination: the control of NF-[kappa]B activity. *Annu Rev Immunol* **18**: 621-663.
- KARIN, M., Y. CAO, F. R. GRETEN and Z. W. LI, 2002 NF-kappaB in cancer: from innocent bystander to major culprit. *Nat Rev Cancer* **2**: 301-310.
- KEPPLER, D., 1999 Export pumps for glutathione S-conjugates. *Free Radic Biol Med* **27**: 985-991.
- KHANNA, K. K., and S. P. JACKSON, 2001 DNA double-strand breaks: signaling, repair and the cancer connection. *Nat Genet* **27**: 247-254.
- KHONG, H. T., and N. P. RESTIFO, 2002 Natural selection of tumor variants in the generation of "tumor escape" phenotypes. *Nat Immunol* **3**: 999-1005.
- KIM, J. C., S. A. ROH, K. H. KOO, I. H. KA, H. C. KIM *et al.*, 2004 Genotyping possible polymorphic variants of human mismatch repair genes in healthy Korean individuals and sporadic colorectal cancer patients. *Fam Cancer* **3**: 129-137.
- KIM, Y. H., K. J. WOO, J. H. LIM, S. KIM, T. J. LEE *et al.*, 2005 8-Hydroxyquinoline inhibits iNOS expression and nitric oxide production by down-regulating LPS-induced activity of NF-kappaB and C/EBPbeta in Raw 264.7 cells. *Biochem Biophys Res Commun* **329**: 591-597.
- KINZLER, K. W., and B. VOGELSTEIN, 1996 Lessons from hereditary colorectal cancer. *Cell* **87**: 159-170.
- KINZLER, K. W., and B. VOGELSTEIN, 1997 Cancer-susceptibility genes. Gatekeepers and caretakers. *Nature* **386**: 761, 763.
- KINZLER, K. W., and B. VOGELSTEIN, 1998 Landscaping the cancer terrain. *Science* **280**: 1036-1037.
- KIROV, G., I. NIKOLOV, L. GEORGIEVA, V. MOSKVINA, M. J. OWEN *et al.*, 2006 Pooled DNA genotyping on Affymetrix SNP genotyping arrays. *BMC Genomics* **7**: 27.
- KIRSCH, D. G., and M. B. KASTAN, 1998 Tumor-suppressor p53: implications for tumor development and prognosis. *J Clin Oncol* **16**: 3158-3168.
- KISS, I., J. SANDOR, G. PAJKOS, B. BOGNER, G. HEGEDUS *et al.*, 2000 Colorectal cancer risk in relation to genetic polymorphism of cytochrome P450 1A1, 2E1, and glutathione-S-transferase M1 enzymes. *Anticancer Res* **20**: 519-522.
- KISSEBAH, A. H., 1996 Intra-abdominal fat: is it a major factor in developing diabetes and coronary artery disease? *Diabetes Res Clin Pract* **30 Suppl**: 25-30.
- KNUDSEN, A. L., M. L. BISGAARD and S. BULOW, 2003 Attenuated familial adenomatous polyposis (AFAP). A review of the literature. *Fam Cancer* **2**: 43-55.
- KNUDSON, A. G., 1996 Hereditary cancer: two hits revisited. *J Cancer Res Clin Oncol* **122**: 135-140.
- KNUDSON, A. G., JR., 1971 Mutation and cancer: statistical study of retinoblastoma. *Proc Natl Acad Sci U S A* **68**: 820-823.
- KOENUMA, M., T. YAMORI and T. TSURUO, 1989 Insulin and insulin-like growth factor 1 stimulate proliferation of metastatic variants of colon carcinoma 26. *Jpn J Cancer Res* **80**: 51-58.
- KONDO, Y., and J. P. ISSA, 2004 Epigenetic changes in colorectal cancer. *Cancer Metastasis Rev* **23**: 29-39.

- KONG, T., H. K. ELTZSCHIG, J. KARHAUSEN, S. P. COLGAN and C. S. SHELLEY, 2004 Leukocyte adhesion during hypoxia is mediated by HIF-1-dependent induction of beta2 integrin gene expression. *Proc Natl Acad Sci U S A* **101**: 10440-10445.
- KORINEK, V., N. BARKER, P. J. MORIN, D. VAN WICHEN, R. DE WEGER *et al.*, 1997 Constitutive transcriptional activation by a beta-catenin-Tcf complex in APC-/- colon carcinoma. *Science* **275**: 1784-1787.
- KRUPA, R., and J. BLASIAK, 2004 An association of polymorphism of DNA repair genes XRCC1 and XRCC3 with colorectal cancer. *J Exp Clin Cancer Res* **23**: 285-294.
- KÜRY, S., B. BUECHER, S. ROBIOU-DU-PONT, C. SCOUL, H. COLMAN *et al.*, 2008 Low-penetrance alleles predisposing to sporadic colorectal cancers: a French case-controlled genetic association study. *BMC Cancer* **8**: 326.
- KÜRY, S., B. BUECHER, S. ROBIOU-DU-PONT, C. SCOUL, H. COLMAN *et al.*, 2007a The thorough screening of the MUTYH gene in a large French cohort of sporadic colorectal cancers. *Genet Test* **11**: 373-379.
- KÜRY, S., B. BUECHER, S. ROBIOU-DU-PONT, C. SCOUL, V. SEBILLE *et al.*, 2007b Combinations of cytochrome P450 gene polymorphisms enhancing the risk for sporadic colorectal cancer related to red meat consumption. *Cancer Epidemiol Biomarkers Prev* **16**: 1460-1467.
- KUNE, G. A., and L. VITETTA, 1992 Alcohol consumption and the etiology of colorectal cancer: a review of the scientific evidence from 1957 to 1991. *Nutr Cancer* **18**: 97-111.
- KUPHAL, S., A. C. MARTYN, J. PEDLEY, L. M. CROWTHER, V. F. BONAZZI *et al.*, 2009 H-Cadherin expression reduces invasion of malignant melanoma. *Pigment Cell Melanoma Res*.
- LADERO, J. M., J. F. GONZALEZ, J. BENITEZ, E. VARGAS, M. J. FERNANDEZ *et al.*, 1991 Acetylator polymorphism in human colorectal carcinoma. *Cancer Res* **51**: 2098-2100.
- LAKEN, S. J., G. M. PETERSEN, S. B. GRUBER, C. ODDDOUX, H. OSTRER *et al.*, 1997 Familial colorectal cancer in Ashkenazim due to a hypermutable tract in APC. *Nat Genet* **17**: 79-83.
- LAMLUM, H., M. ILYAS, A. ROWAN, S. CLARK, V. JOHNSON *et al.*, 1999 The type of somatic mutation at APC in familial adenomatous polyposis is determined by the site of the germline mutation: a new facet to Knudson's 'two-hit' hypothesis. *Nat Med* **5**: 1071-1075.
- LAMLUM, H., A. PAPADOPOULOU, M. ILYAS, A. ROWAN, C. Gillet *et al.*, 2000 APC mutations are sufficient for the growth of early colorectal adenomas. *Proc Natl Acad Sci U S A* **97**: 2225-2228.
- LAMMI, L., S. ARTE, M. SOMER, H. JARVINEN, P. LAHERMO *et al.*, 2004 Mutations in AXIN2 cause familial tooth agenesis and predispose to colorectal cancer. *Am J Hum Genet* **74**: 1043-1050.
- LANDI, S., F. GEMIGNANI, F. BOTTARI, L. GIOIA-PATRICOLA, E. GUINO *et al.*, 2006 Polymorphisms within inflammatory genes and colorectal cancer. *J Negat Results Biomed* **5**: 15.
- LANDI, S., F. GEMIGNANI, V. MORENO, L. GIOIA-PATRICOLA, A. CHABRIER *et al.*, 2005 A comprehensive analysis of phase I and phase II metabolism gene polymorphisms and risk of colorectal cancer. *Pharmacogenet Genomics* **15**: 535-546.
- LANDI, S., V. MORENO, L. GIOIA-PATRICOLA, E. GUINO, M. NAVARRO *et al.*, 2003 Association of common polymorphisms in inflammatory genes interleukin (IL)6, IL8, tumor necrosis factor alpha, NFKB1, and peroxisome proliferator-activated receptor gamma with colorectal cancer. *Cancer Res* **63**: 3560-3566.
- LANG, N. P., M. A. BUTLER, J. MASSENGILL, M. LAWSON, R. C. STOTTS *et al.*, 1994 Rapid metabolic phenotypes for acetyltransferase and cytochrome P4501A2 and putative exposure to food-borne heterocyclic amines increase the risk for colorectal cancer or polyps. *Cancer Epidemiol Biomarkers Prev* **3**: 675-682.
- LANG, N. P., D. Z. CHU, C. F. HUNTER, D. C. KENDALL, T. J. FLAMMANG *et al.*, 1986 Role of aromatic amine acetyltransferase in human colorectal cancer. *Arch Surg* **121**: 1259-1261.
- LARSSON, S. C., and A. WOLK, 2006 Meat consumption and risk of colorectal cancer: a meta-analysis of prospective studies. *Int J Cancer* **119**: 2657-2664.
- LATHAM, P., E. K. LUND, J. C. BROWN and I. T. JOHNSON, 2001 Effects of cellular redox balance on induction of apoptosis by eicosapentaenoic acid in HT29 colorectal adenocarcinoma cells and rat colon in vivo. *Gut* **49**: 97-105.
- LAURENT-PUIG, P., C. BEROUJ and T. SOUSSI, 1998 APC gene: database of germline and somatic mutations in human tumors and cell lines. *Nucleic Acids Res* **26**: 269-270.
- LE HELLARD, S., S. J. BALLEREAU, P. M. VISSCHER, H. S. TORRANCE, J. PINSON *et al.*, 2002 SNP genotyping on pooled DNAs: comparison of genotyping technologies and a semi automated method for data storage and analysis. *Nucleic Acids Res* **30**: e74.
- LE MARCHAND, L., T. DONLON, J. H. HANKIN, L. N. KOLONEL, L. R. WILKENS *et al.*, 2002a B-vitamin intake, metabolic genes, and colorectal cancer risk (United States). *Cancer Causes Control* **13**: 239-248.
- LE MARCHAND, L., T. DONLON, A. SEIFRIED, R. KAAKS, S. RINALDI *et al.*, 2002b Association of a common polymorphism in the human GH1 gene with colorectal neoplasia. *J Natl Cancer Inst* **94**: 454-460.
- LE MARCHAND, L., L. R. WILKENS, L. N. KOLONEL, J. H. HANKIN and L. C. LYU, 1997 Associations of sedentary lifestyle, obesity, smoking, alcohol use, and diabetes with the risk of colorectal cancer. *Cancer Res* **57**: 4787-4794.
- LEE, I. M., and R. S. PAFFENBARGER, JR., 1992 Quetelet's index and risk of colon cancer in college alumni. *J Natl Cancer Inst* **84**: 1326-1331.
- LEE, S. W., 1996 H-cadherin, a novel cadherin with growth inhibitory functions and diminished expression in human breast cancer. *Nat Med* **2**: 776-782.
- LEEK, R. D., and A. L. HARRIS, 2002 Tumor-associated macrophages in breast cancer. *J Mammary Gland Biol Neoplasia* **7**: 177-189.
- LEES, E., 1995 Cyclin dependent kinase regulation. *Curr Opin Cell Biol* **7**: 773-780.

- LEITNER, J. W., T. KLINE, K. CAREL, M. GOALSTONE and B. DRAZNIN, 1997 Hyperinsulinemia potentiates activation of p21Ras by growth factors. *Endocrinology* **138**: 2211-2214.
- LENGAUER, C., K. W. KINZLER and B. VOGELSTEIN, 1998 Genetic instabilities in human cancers. *Nature* **396**: 643-649.
- LEVI, F., C. PASCHE, F. LUCCHINI and C. LA VECCHIA, 2000 Selected micronutrients and colorectal cancer. a case-control study from the canton of Vaud, Switzerland. *Eur J Cancer* **36**: 2115-2119.
- LI, Q., and I. M. VERMA, 2002 NF-kappaB regulation in the immune system. *Nat Rev Immunol* **2**: 725-734.
- LICHTENSTEIN, P., N. V. HOLM, P. K. VERKASALO, A. ILIADOU, J. KAPRIO *et al.*, 2000 Environmental and heritable factors in the causation of cancer--analyses of cohorts of twins from Sweden, Denmark, and Finland. *N Engl J Med* **343**: 78-85.
- LIN, E. Y., V. GOUON-EVANS, A. V. NGUYEN and J. W. POLLARD, 2002 The macrophage growth factor CSF-1 in mammary gland development and tumor progression. *J Mammary Gland Biol Neoplasia* **7**: 147-162.
- LIN, E. Y., and J. W. POLLARD, 2004 Role of infiltrated leucocytes in tumour growth and spread. *Br J Cancer* **90**: 2053-2058.
- LIN, H. J., N. M. PROBST-HENSCH, S. A. INGLES, C. Y. HAN, B. K. LIN *et al.*, 1995 Glutathione transferase (GSTM1) null genotype, smoking, and prevalence of colorectal adenomas. *Cancer Res* **55**: 1224-1226.
- LIN, H. J., N. M. PROBST-HENSCH, A. D. LOUIE, I. H. KAU, J. S. WITTE *et al.*, 1998a Glutathione transferase null genotype, broccoli, and lower prevalence of colorectal adenomas. *Cancer Epidemiol Biomarkers Prev* **7**: 647-652.
- LIN, K. M., M. SHASHIDHARAN, C. A. TERNENT, A. G. THORSON, G. J. BLATCHFORD *et al.*, 1998b Colorectal and extracolonic cancer variations in MLH1/MSH2 hereditary nonpolyposis colorectal cancer kindreds and the general population. *Dis Colon Rectum* **41**: 428-433.
- LINDBLOM, A., P. TANNERGARD, B. WERELIUS and M. NORDENSKJOLD, 1993 Genetic mapping of a second locus predisposing to hereditary non-polyposis colon cancer. *Nat Genet* **5**: 279-282.
- LINDOR, N. M., K. RABE, G. M. PETERSEN, R. HAILE, G. CASEY *et al.*, 2005 Lower cancer incidence in Amsterdam-I criteria families without mismatch repair deficiency: familial colorectal cancer type X. *Jama* **293**: 1979-1985.
- LINGLE, W. L., S. L. BARRETT, V. C. NEGRON, A. B. D'ASSORO, K. BOENEMAN *et al.*, 2002 Centrosome amplification drives chromosomal instability in breast tumor development. *Proc Natl Acad Sci U S A* **99**: 1978-1983.
- LIPKIN, M., B. REDDY, H. NEWMARK and S. A. LAMPRECHT, 1999 Dietary factors in human colorectal cancer. *Annu Rev Nutr* **19**: 545-586.
- LIPTON, L., S. E. HALFORD, V. JOHNSON, M. R. NOVELLI, A. JONES *et al.*, 2003 Carcinogenesis in MYH-associated polyposis follows a distinct genetic pathway. *Cancer Res* **63**: 7595-7599.
- LIPTON, L., and I. TOMLINSON, 2004 The multiple colorectal adenoma phenotype and MYH, a base excision repair gene. *Clin Gastroenterol Hepatol* **2**: 633-638.
- LIU, W. M., X. DI, G. YANG, H. MATSUZAKI, J. HUANG *et al.*, 2003 Algorithms for large-scale genotyping microarrays. *Bioinformatics* **19**: 2397-2403.
- LIVAK, K. J., and T. D. SCHMITTGEN, 2001 Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* **25**: 402-408.
- LODISH, H., 2003 *Molecular cell biology*. W. H. Freeman and Co, New York.
- LOGAN, R. F., J. LITTLE, P. G. HAWTIN and J. D. HARDCASTLE, 1993 Effect of aspirin and non-steroidal anti-inflammatory drugs on colorectal adenomas: case-control study of subjects participating in the Nottingham faecal occult blood screening programme. *Bmj* **307**: 285-289.
- LOVMAR, L., A. AHLFORD, M. JONSSON and A. C. SYVANEN, 2005 Silhouette scores for assessment of SNP genotype clusters. *BMC Genomics* **6**: 35.
- LU, H., W. OUYANG and C. HUANG, 2006 Inflammation, a key event in cancer development. *Mol Cancer Res* **4**: 221-233.
- LU, J., C. JIANG, M. KAECK, H. GANTHER, S. VADHANAVIKIT *et al.*, 1995 Dissociation of the genotoxic and growth inhibitory effects of selenium. *Biochem Pharmacol* **50**: 213-219.
- LUND, E. K., S. G. WHARF, S. J. FAIRWEATHER-TAIT and I. T. JOHNSON, 1999 Oral ferrous sulfate supplements increase the free radical-generating capacity of feces from healthy volunteers. *Am J Clin Nutr* **69**: 250-255.
- LYNCH, H. T., C. R. BOLAND, G. GONG, T. G. SHAW, P. M. LYNCH *et al.*, 2006 Phenotypic and genotypic heterogeneity in the Lynch syndrome: diagnostic, surveillance and management implications. *Eur J Hum Genet* **14**: 390-402.
- LYNCH, H. T., and A. DE LA CHAPELLE, 1999 Genetic susceptibility to non-polyposis colorectal cancer. *J Med Genet* **36**: 801-818.
- LYNCH, H. T., and A. DE LA CHAPELLE, 2003 Hereditary colorectal cancer. *N Engl J Med* **348**: 919-932.
- LYNCH, H. T., and A. J. KRUSH, 1971 Cancer family "G" revisited: 1895-1970. *Cancer* **27**: 1505-1511.
- LYNCH, H. T., J. F. LYNCH, P. M. LYNCH and T. ATTARD, 2008 Hereditary colorectal cancer syndromes: molecular genetics, genetic counseling, diagnosis and management. *Fam Cancer* **7**: 27-39.
- LYNCH, H. T., B. D. RILEY, S. M. WEISSMAN, S. M. CORONEL, Y. KINARSKY *et al.*, 2004a Hereditary nonpolyposis colorectal carcinoma (HNPCC) and HNPCC-like families: Problems in diagnosis, surveillance, and management. *Cancer* **100**: 53-64.
- LYNCH, H. T., M. W. SHAW, C. W. MAGNUSON, A. L. LARSEN and A. J. KRUSH, 1966 Hereditary factors in cancer. Study of two large midwestern kindreds. *Arch Intern Med* **117**: 206-212.
- LYNCH, H. T., T. G. SHAW and J. F. LYNCH, 2004b Inherited predisposition to cancer: a historical overview. *Am J Med Genet* **129C**: 5-22.
- MA, J., M. J. STAMPFER, B. CHRISTENSEN, E. GIOVANNUCCI, D. J. HUNTER *et al.*, 1999 A polymorphism of the methionine synthase gene: association with plasma folate, vitamin B12, homocyst(e)ine, and colorectal cancer risk. *Cancer Epidemiol Biomarkers Prev* **8**: 825-829.

- MACARTHUR, M., G. L. HOLD and E. M. EL-OMAR, 2004 Inflammation and Cancer II. Role of chronic inflammation and cytokine gene polymorphisms in the pathogenesis of gastrointestinal malignancy. *Am J Physiol Gastrointest Liver Physiol* **286**: G515-520.
- MACAULAY, V. M., 1992 Insulin-like growth factors and cancer. *Br J Cancer* **65**: 311-320.
- MACDONALD, R. S., W. H. THORNTON, JR. and T. L. BEAN, 1993 Insulin and IGE-1 receptors in a human intestinal adenocarcinoma cell line (CACO-2): regulation of Na⁺ glucose transport across the brush border. *J Recept Res* **13**: 1093-1113.
- MACGREGOR, S., 2007 Most pooling variation in array-based DNA pooling is attributable to array error rather than pool construction error. *Eur J Hum Genet* **15**: 501-504.
- MACGREGOR, S., P. M. VISSCHER and G. MONTGOMERY, 2006 Analysis of pooled DNA samples on high density arrays without prior knowledge of differential hybridization rates. *Nucleic Acids Res* **34**: e55.
- MACMILLAN, J. C., J. W. HUDSON, S. BULL, J. W. DENNIS and C. J. SWALLOW, 2001 Comparative expression of the mitotic regulators SAK and PLK in colorectal cancer. *Ann Surg Oncol* **8**: 729-740.
- MAEDA, H., and T. AKAIKE, 1998 Nitric oxide and oxygen radicals in infection, inflammation, and cancer. *Biochemistry (Mosc)* **63**: 854-865.
- MALKHOSYAN, S., N. RAMPINO, H. YAMAMOTO and M. PERUCHO, 1996 Frameshift mutator mutations. *Nature* **382**: 499-500.
- MARCHAND, L. L., 1999 Combined influence of genetic and dietary factors on colorectal cancer incidence in Japanese Americans. *J Natl Cancer Inst Monogr*: 101-105.
- MARCHINI, J., B. HOWIE, S. MYERS, G. McVEAN and P. DONNELLY, 2007 A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet* **39**: 906-913.
- MARGISON, G. P., and M. F. SANTIBANEZ-KOREF, 2002 O6-alkylguanine-DNA alkyltransferase: role in carcinogenesis and chemotherapy. *Bioessays* **24**: 255-266.
- MARKOWITZ, S., J. WANG, L. MYEROFF, R. PARSONS, L. SUN *et al.*, 1995 Inactivation of the type II TGF-beta receptor in colon cancer cells with microsatellite instability. *Science* **268**: 1336-1338.
- MARTINEZ-LOPEZ, E., A. ABAD, A. FONT, M. MONZO, I. OJANGUREN *et al.*, 1998 Allelic loss on chromosome 18q as a prognostic marker in stage II colorectal cancer. *Gastroenterology* **114**: 1180-1187.
- MARUGAME, T., E. TSUJI, H. INOUE, S. SHINOMIYA, C. KIYOHARA *et al.*, 2000 Methylenetetrahydrofolate reductase polymorphism and risk of colorectal adenomas. *Cancer Lett* **151**: 181-186.
- MASSAGUE, J., 1998 TGF-beta signal transduction. *Annu Rev Biochem* **67**: 753-791.
- MCCULLOUGH, M. L., A. S. ROBERTSON, C. RODRIGUEZ, E. J. JACOBS, A. CHAO *et al.*, 2003 Calcium, vitamin D, dairy products, and risk of colorectal cancer in the Cancer Prevention Study II Nutrition Cohort (United States). *Cancer Causes Control* **14**: 1-12.
- MEABURN, E., L. M. BUTCHER, L. LIU, C. FERNANDES, V. HANSEN *et al.*, 2005 Genotyping DNA pools on microarrays: tackling the QTL problem of large samples and large numbers of SNPs. *BMC Genomics* **6**: 52.
- MECKLIN, J. P., 2008 The implications of genetics in colorectal cancer. *Ann Oncol* **19 Suppl 5**: v87-90.
- MECKLIN, J. P., H. J. JARVINEN, A. HAKKILUOTO, H. HALLIKAS, K. M. HILTUNEN *et al.*, 1995 Frequency of hereditary nonpolyposis colorectal cancer. A prospective multicenter study in Finland. *Dis Colon Rectum* **38**: 588-593.
- MEZELIO, D., 1721 De excrescentals verrucosa critosis in intestinis crassis dysenteriam passi abservatis. *Ast Med Berolinensium* **4**: 68-71.
- MICHALIK, L., B. DESVERGNE and W. WAHLI, 2004 Peroxisome-proliferator-activated receptors and cancers: complex stories. *Nat Rev Cancer* **4**: 61-70.
- MIYAKI, M., M. KONISHI, K. TANAKA, R. KIKUCHI-YANOSHITA, M. MURAOKA *et al.*, 1997 Germline mutation of MSH6 as the cause of hereditary nonpolyposis colorectal cancer. *Nat Genet* **17**: 271-272.
- MOBARHAN, S., 1999 Calcium and the colon: recent findings. *Nutr Rev* **57**: 124-126.
- MOFFATT, M. F., M. KABESCH, L. LIANG, A. L. DIXON, D. STRACHAN *et al.*, 2007 Genetic variants regulating ORMDL3 expression contribute to the risk of childhood asthma. *Nature* **448**: 470-473.
- MOINOVA, H. R., W. D. CHEN, L. SHEN, D. SMIRAGLIA, J. OLECHNOWICZ *et al.*, 2002 HLTf gene silencing in human colon cancer. *Proc Natl Acad Sci U S A* **99**: 4562-4567.
- MOLLER, D. E., and J. S. FLIER, 1991 Insulin resistance--mechanisms, syndromes, and implications. *N Engl J Med* **325**: 938-948.
- MONONEN, N., E. H. SEPPALA, P. DUGGAL, V. AUTIO, T. IKONEN *et al.*, 2006 Profiling genetic variation along the androgen biosynthesis and metabolism pathways implicates several single nucleotide polymorphisms and their combinations as prostate cancer risk factors. *Cancer Res* **66**: 743-747.
- MONTEIRO, A. N., 2003 BRCA1: the enigma of tissue-specific tumor development. *Trends Genet* **19**: 312-315.
- MOORE, J., and P. COWLED, 1999 Hereditary non-polyposis colorectal cancer syndrome. *Aust N Z J Surg* **69**: 6-13.
- MORENO, V., F. GEMIGNANI, S. LANDI, L. GIOIA-PATRICOLA, A. CHABRIER *et al.*, 2006 Polymorphisms in genes of nucleotide and base excision repair: risk and prognosis of colorectal cancer. *Clin Cancer Res* **12**: 2101-2108.
- MORIN, P. J., A. B. SPARKS, V. KORINEK, N. BARKER, H. CLEVERS *et al.*, 1997 Activation of beta-catenin-Tcf signaling in colon cancer by mutations in beta-catenin or APC. *Science* **275**: 1787-1790.
- MORT, R., L. MO, C. McEWAN and D. W. MELTON, 2003 Lack of involvement of nucleotide excision repair gene polymorphisms in colorectal cancer. *Br J Cancer* **89**: 333-337.
- MOSKIVINA, V., N. NORTON, N. WILLIAMS, P. HOLMANS, M. OWEN *et al.*, 2005 Streamlined analysis of pooled genotype data in SNP-based association studies. *Genet Epidemiol* **28**: 273-282.
- MUNKHOLM, P., 2003 Review article: the incidence and prevalence of colorectal cancer in inflammatory bowel disease. *Aliment Pharmacol Ther* **18 Suppl 2**: 1-5.

- MURATA, M., M. TAGAWA, S. WATANABE, H. KIMURA, T. TAKESHITA *et al.*, 1999 Genotype difference of aldehyde dehydrogenase 2 gene in alcohol drinkers influences the incidence of Japanese colorectal cancer patients. *Jpn J Cancer Res* **90**: 711-719.
- NACCARATI, A., B. PARDINI, K. HEMMINKI and P. VODICKA, 2007 Sporadic colorectal cancer and individual susceptibility: a review of the association studies investigating the role of DNA repair genetic polymorphisms. *Mutat Res* **635**: 118-145.
- NANDA, K., L. A. BASTIAN, V. HASSELBLAD and D. L. SIMEL, 1999 Hormone replacement therapy and the risk of colorectal cancer: a meta-analysis. *Obstet Gynecol* **93**: 880-888.
- NASTRUZZI, C., R. CORTESI, E. ESPOSITO, E. MENEGATTI, O. LEONI *et al.*, 2000 In vitro antiproliferative activity of isothiocyanates and nitriles generated by myrosinase-mediated hydrolysis of glucosinolates from seeds of cruciferous vegetables. *J Agric Food Chem* **48**: 3572-3575.
- NATHAN, C., 2002 Points of control in inflammation. *Nature* **420**: 846-852.
- NELSON, D. R., D. C. ZELDIN, S. M. HOFFMAN, L. J. MALTAIS, H. M. WAIN *et al.*, 2004 Comparison of cytochrome P450 (CYP) genes from the mouse and human genomes, including nomenclature recommendations for genes, pseudogenes and alternative-splice variants. *Pharmacogenetics* **14**: 1-18.
- NELSON, R. L., 2001 Iron and colorectal cancer risk: human studies. *Nutr Rev* **59**: 140-148.
- NG, P. C., and S. HENIKOFF, 2006 Predicting the effects of amino acid substitutions on protein function. *Annu Rev Genomics Hum Genet* **7**: 61-80.
- NICOLAIDES, N. C., S. J. LITTMAN, P. MODRICH, K. W. KINZLER and B. VOGELSTEIN, 1998 A naturally occurring hPMS2 mutation can confer a dominant negative mutator phenotype. *Mol Cell Biol* **18**: 1635-1641.
- NICOLAIDES, N. C., N. PAPADOPOULOS, B. LIU, Y. F. WEI, K. C. CARTER *et al.*, 1994 Mutations of two PMS homologues in hereditary nonpolyposis colon cancer. *Nature* **371**: 75-80.
- NIEUWENHUIS, M. H., and H. F. VASEN, 2007 Correlations between mutation site in APC and phenotype of familial adenomatous polyposis (FAP): a review of the literature. *Crit Rev Oncol Hematol* **61**: 153-161.
- NKONDJOCK, A., B. SHATENSTEIN, P. MAISONNEUVE and P. GHADIRIAN, 2003 Assessment of risk associated with specific fatty acids and colorectal cancer among French-Canadians in Montreal: a case-control study. *Int J Epidemiol* **32**: 200-209.
- NORAT, T., S. BINGHAM, P. FERRARI, N. SLIMANI, M. JENAB *et al.*, 2005 Meat, fish, and colorectal cancer risk: the European Prospective Investigation into cancer and nutrition. *J Natl Cancer Inst* **97**: 906-916.
- NORAT, T., A. LUKANOVA, P. FERRARI and E. RIBOLI, 2002 Meat consumption and colorectal cancer risk: dose-response meta-analysis of epidemiological studies. *Int J Cancer* **98**: 241-256.
- NORBURY, C. J., and I. D. HICKSON, 2001 Cellular responses to DNA damage. *Annu Rev Pharmacol Toxicol* **41**: 367-401.
- NOTREDAME, C., D. G. HIGGINS and J. HERINGA, 2000 T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J Mol Biol* **302**: 205-217.
- NYBERG, K. A., R. J. MICHELSON, C. W. PUTNAM and T. A. WEINERT, 2002 Toward maintaining the genome: DNA damage and replication checkpoints. *Annu Rev Genet* **36**: 617-656.
- O'DRISCOLL, M., and P. A. JEGGO, 2006 The role of double-strand break repair - insights from human genetics. *Nat Rev Genet* **7**: 45-54.
- OGUNWOB, O. O., and I. L. BEALES, 2007 The anti-apoptotic and growth stimulatory actions of leptin in human colon cancer cells involves activation of JNK mitogen activated protein kinase, JAK2 and PI3 kinase/Akt. *Int J Colorectal Dis* **22**: 401-409.
- OKADA, F., 2002 Inflammation and free radicals in tumor development and progression. *Redox Rep* **7**: 357-368.
- OLSCHWANG, S., H. BLANCHE, C. DE MONCUIT and G. THOMAS, 2007 Similar colorectal cancer risk in patients with monoallelic and biallelic mutations in the MYH gene identified in a population with adenomatous polyposis. *Genet Test* **11**: 315-320.
- ORTEGA, S., M. MALUMBRES and M. BARBACID, 2002 Cyclin D-dependent kinases, INK4 inhibitors and cancer. *Biochim Biophys Acta* **1602**: 73-87.
- OTANI, T., M. IWASAKI, S. SASAZUKI, M. INOUE and S. TSUGANE, 2006 Plasma C-reactive protein and risk of colorectal cancer in a nested case-control study: Japan Public Health Center-based prospective study. *Cancer Epidemiol Biomarkers Prev* **15**: 690-695.
- Ou, J., R. C. NIESSEN, J. VONK, H. WESTERS, R. M. HOFSTRA *et al.*, 2008 A database to support the interpretation of human mismatch repair gene variants. *Hum Mutat* **29**: 1337-1341.
- PALMQVIST, R., G. HALLMANS, S. RINALDI, C. BIESSY, R. STENLING *et al.*, 2002 Plasma insulin-like growth factor 1, insulin-like growth factor binding protein 3, and risk of colorectal cancer: a prospective study in northern Sweden. *Gut* **50**: 642-646.
- PARK, C. H., J. Y. CHANG, E. R. HAHM, S. PARK, H. K. KIM *et al.*, 2005a Quercetin, a potent inhibitor against beta-catenin/Tcf signaling in SW480 colon cancer cells. *Biochem Biophys Res Commun* **328**: 227-234.
- PARK, K. S., J. W. MOK and J. C. KIM, 1999 The 677C > T mutation in 5,10-methylenetetrahydrofolate reductase and colorectal cancer risk. *Genet Test* **3**: 233-236.
- PARK, M., 1998 *The genetic basis of human cancer: Oncogenes*, New York.
- PARK, S. Y., S. P. MURPHY, L. R. WILKENS, A. M. NOMURA, B. E. HENDERSON *et al.*, 2007 Calcium and Vitamin D Intake and Risk of Colorectal Cancer: The Multiethnic Cohort Study. *Am J Epidemiol*.
- PARK, Y., D. J. HUNTER, D. SPIEGELMAN, L. BERGKVIST, F. BERRINO *et al.*, 2005b Dietary fiber intake and risk of colorectal cancer: a pooled analysis of prospective cohort studies. *Jama* **294**: 2849-2857.
- PARSONS, R., G. M. LI, M. J. LONGLEY, W. H. FANG, N. PAPADOPOULOS *et al.*, 1993 Hypermutability and mismatch repair deficiency in RER+ tumor cells. *Cell* **75**: 1227-1236.

- PEARSON, J. V., M. J. HUENTELMAN, R. F. HALPERIN, W. D. TEMBE, S. MELQUIST *et al.*, 2007 Identification of the genetic basis for complex disorders by use of pooling-based genomewide single-nucleotide-polymorphism association studies. *Am J Hum Genet* **80**: 126-139.
- PEDRAZZANI, C., G. CORSO, D. MARRELLI and F. ROVIELLO, 2007 E-cadherin and hereditary diffuse gastric cancer. *Surgery* **142**: 645-657.
- PELTOMAKI, P., 1997 DNA mismatch repair gene mutations in human cancer. *Environ Health Perspect* **105 Suppl 4**: 775-780.
- PELTOMAKI, P., L. A. AALTONEN, P. SISTONEN, L. PYLKKANEN, J. P. MECKLIN *et al.*, 1993 Genetic mapping of a locus predisposing to human colorectal cancer. *Science* **260**: 810-812.
- PELTOMAKI, P., and A. DE LA CHAPPELLE, 1997 Mutations predisposing to hereditary nonpolyposis colorectal cancer. *Adv Cancer Res* **71**: 93-119.
- PENNA, C. P., E. TIRET, A. H. KARTHEUSER, S. OLSCHWANG and R. PARC, 1992 [Value of screening of familial adenomatous polyposis for the prevention of colorectal cancer]. *Gastroenterol Clin Biol* **16**: 210-214.
- PEREZ-LOSADA, J., J. H. MAO and A. BALMAIN, 2005 Control of genomic instability and epithelial tumor development by the p53-Fbxw7/Cdc4 pathway. *Cancer Res* **65**: 6488-6492.
- PERKINS, N. D., 2000 The Rel/NF-kappa B family: friend and foe. *Trends Biochem Sci* **25**: 434-440.
- PERTEA, M., X. LIN and S. L. SALZBERG, 2001 GeneSplicer: a new computational method for splice site prediction. *Nucleic Acids Res* **29**: 1185-1190.
- PETERS, U., R. SINHA, N. CHATTERJEE, A. F. SUBAR, R. G. ZIEGLER *et al.*, 2003 Dietary fibre and colorectal adenoma in a colorectal cancer early detection programme. *Lancet* **361**: 1491-1495.
- PETRENKO, O., and U. M. MOLL, 2005 Macrophage migration inhibitory factor MIF interferes with the Rb-E2F pathway. *Mol Cell* **17**: 225-236.
- PETRI, N., C. TANNERGREN, B. HOLST, F. A. MELLON, Y. BAO *et al.*, 2003 Absorption/metabolism of sulforaphane and quercetin, and regulation of phase II enzymes, in human jejunum in vivo. *Drug Metab Dispos* **31**: 805-813.
- PHILIP, M., D. A. ROWLEY and H. SCHREIBER, 2004 Inflammation as a tumor promoter in cancer induction. *Semin Cancer Biol* **14**: 433-439.
- PHILLIPS, R. K., M. H. WALLACE, P. M. LYNCH, E. HAWK, G. B. GORDON *et al.*, 2002 A randomised, double blind, placebo controlled study of celecoxib, a selective cyclooxygenase 2 inhibitor, on duodenal polyposis in familial adenomatous polyposis. *Gut* **50**: 857-860.
- PIETINEN, P., N. MALILA, M. VIRTANEN, T. J. HARTMAN, J. A. TANGREA *et al.*, 1999 Diet and risk of colorectal cancer in a cohort of Finnish men. *Cancer Causes Control* **10**: 387-396.
- PIHAN, G. A., A. PUROHIT, J. WALLACE, R. MALHOTRA, L. LIOTTA *et al.*, 2001 Centrosome defects can account for cellular and genetic changes that characterize prostate cancer progression. *Cancer Res* **61**: 2212-2219.
- PISCHON, T., P. H. LAHMANN, H. BOEING, C. FRIEDENREICH, T. NORAT *et al.*, 2006 Body size and risk of colon and rectal cancer in the European Prospective Investigation Into Cancer and Nutrition (EPIC). *J Natl Cancer Inst* **98**: 920-931.
- PLEDGER, W. J., C. D. STILES, H. N. ANTONIADES and C. D. SCHER, 1977 Induction of DNA synthesis in BALB/c 3T3 cells by serum components: reevaluation of the commitment process. *Proc Natl Acad Sci U S A* **74**: 4481-4485.
- POIROT, O., E. O'TOOLE and C. NOTREDAME, 2003 Tcoffee@igs: A web server for computing, evaluating and combining multiple sequence alignments. *Nucleic Acids Res* **31**: 3503-3506.
- POLAKIS, P., 1999 The oncogenic activation of beta-catenin. *Curr Opin Genet Dev* **9**: 15-21.
- POLLAK, M. N., J. F. PERDUE, R. G. MARGOLESE, K. BAER and M. RICHARD, 1987 Presence of somatomedin receptors on primary human breast and colon carcinomas. *Cancer Lett* **38**: 223-230.
- POLLARD, J. W., 2004 Tumour-educated macrophages promote tumour progression and metastasis. *Nat Rev Cancer* **4**: 71-78.
- POULLIS, A., R. FOSTER, A. SHETTY, M. K. FAGERHOL and M. A. MENDALL, 2004 Bowel inflammation as measured by fecal calprotectin: a link between lifestyle factors and colorectal cancer risk. *Cancer Epidemiol Biomarkers Prev* **13**: 279-284.
- POYNTER, J. N., J. C. FIGUEIREDO, D. V. CONTI, K. KENNEDY, S. GALLINGER *et al.*, 2007 Variants on 9p24 and 8q24 are associated with risk of colorectal cancer: results from the Colon Cancer Family Registry. *Cancer Res* **67**: 11128-11132.
- PRESCOTT, S. M., and F. A. FITZPATRICK, 2000 Cyclooxygenase-2 and carcinogenesis. *Biochim Biophys Acta* **1470**: M69-78.
- PURDIE, C. A., J. O'GRADY, J. PIRIS, A. H. WYLLIE and C. C. BIRD, 1991 p53 expression in colorectal tumors. *Am J Pathol* **138**: 807-813.
- QIAN, Z. R., T. SANO, K. YOSHIMOTO, S. L. ASA, S. YAMADA *et al.*, 2007 Tumor-specific downregulation and methylation of the CDH13 (H-cadherin) and CDH1 (E-cadherin) genes correlate with aggressiveness of human pituitary adenomas. *Mod Pathol* **20**: 1269-1277.
- RAJAGOPALAN, H., A. BARDELLI, C. LENGAUER, K. W. KINZLER, B. VOGELSTEIN *et al.*, 2002 Tumorigenesis: RAF/RAS oncogenes and mismatch-repair status. *Nature* **418**: 934.
- RAJAGOPALAN, H., P. V. JALLEPALLI, C. RAGO, V. E. VELCULESCU, K. W. KINZLER *et al.*, 2004 Inactivation of hCDC4 can cause chromosomal instability. *Nature* **428**: 77-81.
- RAJAH, R., B. VALENTINIS and P. COHEN, 1997 Insulin-like growth factor (IGF)-binding protein-3 induces apoptosis and mediates the effects of transforming growth factor-beta1 on programmed cell death through a p53- and IGF-independent mechanism. *J Biol Chem* **272**: 12181-12188.
- RAMPINO, N., H. YAMAMOTO, Y. IONOV, Y. LI, H. SAWAI *et al.*, 1997 Somatic frameshift mutations in the BAX gene in colon cancers of the microsatellite mutator phenotype. *Science* **275**: 967-969.

- RANSCHT, B., and M. T. DOURS-ZIMMERMANN, 1991 T-cadherin, a novel cadherin cell adhesion molecule in the nervous system lacks the conserved cytoplasmic region. *Neuron* **7**: 391-402.
- RAUNIO, H., K. HUSGAFVEL-PURSAINEN, S. ANTTILA, E. HIETANEN, A. HIRVONEN *et al.*, 1995 Diagnosis of polymorphisms in carcinogen-activating and inactivating enzymes and cancer susceptibility--a review. *Gene* **159**: 113-121.
- RECHLER, M. M., 1997 Growth inhibition by insulin-like growth factor (IGF) binding protein-3--what's IGF got to do with it? *Endocrinology* **138**: 2645-2647.
- REMYKOS, Y., N. VOGT, M. MULERIS, R. J. SALMON, B. MALFOY *et al.*, 1995 DNA-repeat instability is associated with colorectal cancers presenting minimal chromosome rearrangements. *Genes Chromosomes Cancer* **12**: 272-276.
- REYA, T., and H. CLEVERS, 2005 Wnt signalling in stem cells and cancer. *Nature* **434**: 843-850.
- RIOU, P., R. SAFFROY, C. CHENAILLER, B. FRANC, C. GENTILE *et al.*, 2006 Expression of T-cadherin in tumor cells influences invasive potential of human hepatocellular carcinoma. *Faseb J* **20**: 2291-2301.
- RISCH, N., and K. MERIKANGAS, 1996 The future of genetic studies of complex human diseases. *Science* **273**: 1516-1517.
- RISCH, N. J., 2000 Searching for genetic determinants in the new millennium. *Nature* **405**: 847-856.
- ROBERTS-THOMSON, I. C., P. RYAN, K. K. KHOO, W. J. HART, A. J. McMICHAEL *et al.*, 1996 Diet, acetylator phenotype, and risk of colorectal neoplasia. *Lancet* **347**: 1372-1374.
- ROEDIGER, W. E., 1990 The starved colon--diminished mucosal nutrition, diminished absorption, and colitis. *Dis Colon Rectum* **33**: 858-862.
- RONDINONE, C. M., 2006 Adipocyte-derived hormones, cytokines, and mediators. *Endocrine* **29**: 81-90.
- ROSS, R., J. GLOMSET, B. KARIYA and L. HARKER, 1974 A platelet-dependent serum factor that stimulates the proliferation of arterial smooth muscle cells in vitro. *Proc Natl Acad Sci U S A* **71**: 1207-1210.
- ROSS, R., L. LEGER, D. MORRIS, J. DE GUISE and R. GUARDO, 1992 Quantification of adipose tissue by MRI: relationship with anthropometric variables. *J Appl Physiol* **72**: 787-795.
- ROWLAND, I. R., H. WISEMAN, T. A. SANDERS, H. ADLERCREUTZ and E. A. BOWEY, 2000 Interindividual variation in metabolism of soy isoflavones and lignans: influence of habitual diet on equol production by the gut microflora. *Nutr Cancer* **36**: 27-32.
- RUBINFELD, B., I. ALBERT, E. PORFIRI, C. FIOLE, S. MUNEMITSU *et al.*, 1996 Binding of GSK3beta to the APC-beta-catenin complex and regulation of complex assembly. *Science* **272**: 1023-1026.
- RUSTGI, A. K., 2007 The genetics of hereditary colon cancer. *Genes Dev* **21**: 2525-2538.
- SACHSE, C., G. SMITH, M. J. WILKIE, J. H. BARRETT, R. WAXMAN *et al.*, 2002 A pharmacogenetic study to investigate the role of dietary carcinogens in the etiology of colorectal cancer. *Carcinogenesis* **23**: 1839-1849.
- SAHA, S., A. BARDELLI, P. BUCKHAULTS, V. E. VELCULESCU, C. RAGO *et al.*, 2001 A phosphatase associated with metastasis of colorectal cancer. *Science* **294**: 1343-1346.
- SAJI, H., M. KOIKE, T. YAMORI, S. SAJI, M. SEIKI *et al.*, 2001 Significant correlation of monocyte chemoattractant protein-1 expression with neovascularization and progression of breast carcinoma. *Cancer* **92**: 1085-1091.
- SALAHSHOR, S., U. KRESSNER, L. PAHLMAN, B. GLIMELIUS, G. LINDMARK *et al.*, 1999 Colorectal cancer with and without microsatellite instability involves different genes. *Genes Chromosomes Cancer* **26**: 247-252.
- SALOVAARA, R., A. LOUKOLA, P. KRISTO, H. KAARIAINEN, H. AHTOLA *et al.*, 2000 Population-based molecular detection of hereditary nonpolyposis colorectal cancer. *J Clin Oncol* **18**: 2193-2200.
- SAMOWITZ, W. S., H. ALBERTSEN, J. HERRICK, T. R. LEVIN, C. SWEENEY *et al.*, 2005 Evaluation of a large, population-based sample supports a CpG island methylator phenotype in colon cancer. *Gastroenterology* **129**: 837-845.
- SAMOWITZ, W. S., K. CURTIN, H. H. LIN, M. A. ROBERTSON, D. SCHAFFER *et al.*, 2001 The colon cancer burden of genetically defined hereditary nonpolyposis colon cancer. *Gastroenterology* **121**: 830-838.
- SAMPSON, J. R., S. DOLWANI, S. JONES, D. ECCLES, A. ELLIS *et al.*, 2003 Autosomal recessive colorectal adenomatous polyposis due to inherited mutations of MYH. *Lancet* **362**: 39-41.
- SAMPSON, J. R., S. JONES, S. DOLWANI and J. P. CHEADLE, 2005 MutYH (MYH) and colorectal cancer. *Biochem Soc Trans* **33**: 679-683.
- SANDHU, M. S., I. R. WHITE and K. MCPHERSON, 2001 Systematic review of the prospective cohort studies on meat consumption and colorectal cancer risk: a meta-analytical approach. *Cancer Epidemiol Biomarkers Prev* **10**: 439-446.
- SANJOAQUIN, M. A., P. N. APPLEBY, M. THOROGOOD, J. I. MANN and T. J. KEY, 2004 Nutrition, lifestyle and colorectal cancer incidence: a prospective investigation of 10998 vegetarians and non-vegetarians in the United Kingdom. *Br J Cancer* **90**: 118-121.
- SATO, N., K. MIZUMOTO, M. NAKAMURA, N. MAEHARA, Y. A. MINAMISHIMA *et al.*, 2001 Correlation between centrosome abnormalities and chromosomal instability in human pancreatic cancer cells. *Cancer Genet Cytogenet* **126**: 13-19.
- SAUNDERS, W. S., M. SHUSTER, X. HUANG, B. GHARAIBEH, A. H. ENYENIHI *et al.*, 2000 Chromosomal instability and cytoskeletal defects in oral cancer cells. *Proc Natl Acad Sci U S A* **97**: 303-308.
- SCALBERT, A., I. T. JOHNSON and M. SALTMARSH, 2005 Polyphenols: antioxidants and beyond. *Am J Clin Nutr* **81**: 215S-217S.
- SCHAFMAYER, C., S. BUCH, H. VOLZKE, W. VON SCHONFELS, J. H. EGBERTS *et al.*, 2009 Investigation of the colorectal cancer susceptibility region on chromosome 8q24.21 in a large German case-control sample. *Int J Cancer* **124**: 75-80.
- SCHOEN, R. E., C. M. TANGEN, L. H. KULLER, G. L. BURKE, M. CUSHMAN *et al.*, 1999 Increased blood glucose and insulin, body size, and incident colorectal cancer. *J Natl Cancer Inst* **91**: 1147-1154.
- SCHRAUWEN, I., M. EALY, M. J. HUENTELMAN, M. THYS, N. HOMER *et al.*, 2009 A genome-wide analysis identifies genetic variants in the RELN gene associated with otosclerosis. *Am J Hum Genet* **84**: 328-338.

- SCHUMACHER, F. R., H. S. FEIGELSON, D. G. COX, C. A. HAIMAN, D. ALBANES *et al.*, 2007 A common 8q24 variant in prostate and breast cancer from a large nested case-control study. *Cancer Res* **67**: 2951-2956.
- SCHWARTE-WALDHOFF, I., S. KLEIN, S. BLASS-KAMPMANN, A. HINTELMANN, C. EILERT *et al.*, 1999 DPC4/SMAD4 mediated tumor suppression of colon carcinoma cells is associated with reduced urokinase expression. *Oncogene* **18**: 3152-3158.
- SELKE, B., I. DURAND, J. P. MARISSAL, D. CHEVALIER and T. LEBRUN, 2003 Cost of colorectal cancer in France in 1999. *Gastroenterol Clin Biol* **27**: 22-27.
- SEMENZA, G. L., 1999 Regulation of mammalian O2 homeostasis by hypoxia-inducible factor 1. *Annu Rev Cell Dev Biol* **15**: 551-578.
- SEN, S., H. ZHOU and R. A. WHITE, 1997 A putative serine/threonine kinase encoding gene BTAK on chromosome 20q13 is amplified and overexpressed in human breast cancer cell lines. *Oncogene* **14**: 2195-2200.
- SERIL, D. N., J. LIAO, G. Y. YANG and C. S. YANG, 2003 Oxidative stress and ulcerative colitis-associated carcinogenesis: studies in humans and animal models. *Carcinogenesis* **24**: 353-362.
- SESINK, A. L., D. S. TERMONT, J. H. KLEIBEUKER and R. VAN DER MEER, 1999 Red meat and colon cancer: the cytotoxic and hyperproliferative effects of dietary heme. *Cancer Res* **59**: 5704-5709.
- SHACTER, E., and S. A. WEITZMAN, 2002 Chronic inflammation and cancer. *Oncology (Williston Park)* **16**: 217-226, 229; discussion 230-212.
- SHADAN, F. F., C. CUNNINGHAM and C. R. BOLAND, 2002 JC virus: a biomarker for colorectal cancer? *Med Hypotheses* **59**: 667-669.
- SHAM, P., J. S. BADER, I. CRAIG, M. O'DONOVAN and M. OWEN, 2002 DNA Pooling: a tool for large-scale association studies. *Nat Rev Genet* **3**: 862-871.
- SHANE, B., 1995 Folate chemistry and metabolism, pp. 1-22, edited by I. H. A. D. FOLATE. Marcel Dekker, New-York.
- SHANNON, B., S. GNANASAMPANTHAN, J. BEILBY and B. IACOPETTA, 2002 A polymorphism in the methylenetetrahydrofolate reductase gene predisposes to colorectal cancers with microsatellite instability. *Gut* **50**: 520-524.
- SHARP, L., and J. LITTLE, 2004 Polymorphisms in genes involved in folate metabolism and colorectal neoplasia: a HuGE review. *Am J Epidemiol* **159**: 423-443.
- SHARPE, C. R., J. SIEMIATYCKI and B. RACHET, 2002 Effects of alcohol consumption on the risk of colorectal cancer among men by anatomical subsite (Canada). *Cancer Causes Control* **13**: 483-491.
- SHIBATA, D., M. A. PEINADO, Y. IONOV, S. MALKHOSYAN and M. PERUCHO, 1994 Genomic instability in repeated sequences is an early somatic event in colorectal tumorigenesis that persists after transformation. *Nat Genet* **6**: 273-281.
- SHIBATA, D., M. A. REALE, P. LAVIN, M. SILVERMAN, E. R. FEARON *et al.*, 1996 The DCC protein and prognosis in colorectal cancer. *N Engl J Med* **335**: 1727-1732.
- SHIBUTA, K., T. NAKASHIMA, M. ABE, M. MASHIMO, M. MORI *et al.*, 1994 Molecular genotyping for N-acetylation polymorphism in Japanese patients with colorectal cancer. *Cancer* **74**: 3108-3112.
- SHIFMAN, S., M. JOHANNESON, M. BRONSTEIN, S. X. CHEN, D. A. COLLIER *et al.*, 2008 Genome-wide association identifies a common variant in the reelin gene that increases the risk of schizophrenia only in women. *PLoS Genet* **4**: e28.
- SHIH, I. M., W. ZHOU, S. N. GOODMAN, C. LENGAUER, K. W. KINZLER *et al.*, 2001 Evidence that genetic instability occurs at an early stage of colorectal tumorigenesis. *Cancer Res* **61**: 818-822.
- SHUREIQI, I., and S. M. LIPPMAN, 2001 Lipoxygenase modulation to reverse carcinogenesis. *Cancer Res* **61**: 6307-6312.
- SIEBER, O. M., K. M. HOWARTH, C. THIRLWELL, A. ROWAN, N. MANDIR *et al.*, 2004 Myh deficiency enhances intestinal tumorigenesis in multiple intestinal neoplasia (ApcMin/+) mice. *Cancer Res* **64**: 8876-8881.
- SIEBER, O. M., L. LIPTON, M. CRABTREE, K. HEINIMANN, P. FIDALGO *et al.*, 2003 Multiple colorectal adenomas, classic adenomatous polyposis, and germ-line mutations in MYH. *N Engl J Med* **348**: 791-799.
- SIEBER, O. M., I. P. TOMLINSON and H. LAMLUM, 2000 The adenomatous polyposis coli (APC) tumour suppressor--genetics, function and disease. *Mol Med Today* **6**: 462-469.
- SIEZEN, C. L., H. B. BUENO-DE-MESQUITA, P. H. PEETERS, N. R. KRAM, M. VAN DOESELAAR *et al.*, 2006 Polymorphisms in the genes involved in the arachidonic acid-pathway, fish consumption and the risk of colorectal cancer. *Int J Cancer* **119**: 297-303.
- SIEZEN, C. L., A. I. VAN LEEUWEN, N. R. KRAM, M. E. LUKEN, H. J. VAN KRANEN *et al.*, 2005 Colorectal adenoma risk is modified by the interplay between polymorphisms in arachidonic acid pathway genes and fish consumption. *Carcinogenesis* **26**: 449-457.
- SILVI, S., C. J. RUMNEY, A. CRESCI and I. R. ROWLAND, 1999 Resistant starch modifies gut microflora and microbial metabolism in human flora-associated rats inoculated with faeces from Italian and UK donors. *J Appl Microbiol* **86**: 521-530.
- SIMOPOULOS, A. P., 2006 Evolutionary aspects of diet, the omega-6/omega-3 ratio and genetic variation: nutritional implications for chronic diseases. *Biomed Pharmacother* **60**: 502-507.
- SIMPSON, C. L., J. KNIGHT, L. M. BUTCHER, V. K. HANSEN, E. MEABURN *et al.*, 2005 A central resource for accurate allele frequency estimation from pooled DNA genotyped on DNA microarrays. *Nucleic Acids Res* **33**: e25.
- SKJELBRED, C. F., M. SAEBO, H. WALLIN, B. A. NEXO, P. C. HAGEN *et al.*, 2006 Polymorphisms of the XRCC1, XRCC3 and XPD genes and risk of colorectal adenoma and carcinoma, in a Norwegian cohort: a case control study. *BMC Cancer* **6**: 67.
- SLABY, O., M. SVOBODA, P. FABIAN, T. SMERDOVA, D. KNOFLICKOVA *et al.*, 2007 Altered expression of miR-21, miR-31, miR-143 and miR-145 is related to clinicopathologic features of colorectal cancer. *Oncology* **72**: 397-402.
- SLATTERY, M. L., K. ANDERSON, K. CURTIN, K. N. MA, D. SCHAFFER *et al.*, 2001 Dietary intake and microsatellite instability in colon tumors. *Int J Cancer* **93**: 601-607.

- SLATTERY, M. L., K. CURTIN, K. ANDERSON, K. N. MA, L. BALLARD *et al.*, 2000a Associations between cigarette smoking, lifestyle factors, and microsatellite instability in colon tumors. *J Natl Cancer Inst* **92**: 1831-1836.
- SLATTERY, M. L., S. EDWARDS, K. CURTIN, K. MA, R. EDWARDS *et al.*, 2003 Physical activity and colorectal cancer. *Am J Epidemiol* **158**: 214-224.
- SLATTERY, M. L., J. D. POTTER, A. COATES, K. N. MA, T. D. BERRY *et al.*, 1997 Plant foods and colon cancer: an assessment of specific foods and their related nutrients (United States). *Cancer Causes Control* **8**: 575-590.
- SLATTERY, M. L., J. D. POTTER, K. N. MA, B. J. CAAN, M. LEPPERT *et al.*, 2000b Western diet, family history of colorectal cancer, NAT2, GSTM-1 and risk of colon cancer. *Cancer Causes Control* **11**: 1-8.
- SLATTERY, M. L., J. D. POTTER, W. SAMOWITZ, D. SCHAFFER and M. LEPPERT, 1999 Methylenetetrahydrofolate reductase, diet, and risk of colon cancer. *Cancer Epidemiol Biomarkers Prev* **8**: 513-518.
- SLATTERY, M. L., W. SAMOWITZ, K. CURTIN, K. N. MA, M. HOFFMAN *et al.*, 2004a Associations among IRS1, IRS2, IGF1, and IGF1BP3 genetic polymorphisms and colorectal cancer. *Cancer Epidemiol Biomarkers Prev* **13**: 1206-1214.
- SLATTERY, M. L., W. SAMOWITZ, M. HOFFMAN, K. N. MA, T. R. LEVIN *et al.*, 2004b Aspirin, NSAIDs, and colorectal cancer: possible involvement in an insulin-related pathway. *Cancer Epidemiol Biomarkers Prev* **13**: 538-545.
- SLUPSKA, M. M., C. BAIKALOV, W. M. LUTHER, J. H. CHIANG, Y. F. WEI *et al.*, 1996 Cloning and sequencing a human homolog (hMYH) of the *Escherichia coli* mutY gene whose function is required for the repair of oxidative DNA damage. *J Bacteriol* **178**: 3885-3892.
- SMITH, A. J., H. S. STERN, M. PENNER, K. HAY, A. MITRI *et al.*, 1994 Somatic APC and K-ras codon 12 mutations in aberrant crypt foci from human colons. *Cancer Res* **54**: 5527-5530.
- SMITH, W. J., L. E. UNDERWOOD and D. R. CLEMMONS, 1995 Effects of caloric or protein restriction on insulin-like growth factor-I (IGF-I) and IGF-binding proteins in children and adults. *J Clin Endocrinol Metab* **80**: 443-449.
- SMYTH, M. J., E. CRETNEY, M. H. KERSHAW and Y. HAYAKAWA, 2004 Cytokines in cancer immunity and immunotherapy. *Immunol Rev* **202**: 275-293.
- SOUSI, T., 2000 [Record no. 39: p53 (TP53)]. *Bull Cancer* **87**: 691-692.
- SOUZA, R. F., R. APPEL, J. YIN, S. WANG, K. N. SMOLINSKI *et al.*, 1996 Microsatellite instability in the insulin-like growth factor II receptor gene in gastrointestinal tumours. *Nat Genet* **14**: 255-257.
- SOUZA, R. F., J. YIN, K. N. SMOLINSKI, T. T. ZOU, S. WANG *et al.*, 1997 Frequent mutation of the E2F-4 cell cycle gene in primary human gastrointestinal tumors. *Cancer Res* **57**: 2350-2353.
- SPARKS, A. B., P. J. MORIN, B. VOGELSTEIN and K. W. KINZLER, 1998 Mutational analysis of the APC/beta-catenin/Tcf pathway in colorectal cancer. *Cancer Res* **58**: 1130-1134.
- SPIELMAN, R. S., R. E. MCGINNIS and W. J. EWENS, 1993 Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am J Hum Genet* **52**: 506-516.
- SPIRIO, L., L. NELSON, K. WARD, R. BURT, R. WHITE *et al.*, 1993 A CA-repeat polymorphism close to the adenomatous polyposis coli (APC) gene offers improved diagnostic testing for familial APC. *Am J Hum Genet* **52**: 286-296.
- STEELE, V. E., E. T. HAWK, J. L. VINER and R. A. LUBET, 2003 Mechanisms and applications of non-steroidal anti-inflammatory drugs in the chemoprevention of cancer. *Mutat Res* **523-524**: 137-144.
- STIEGER, B., and P. J. MEIER, 1998 Bile acid and xenobiotic transporters in liver. *Curr Opin Cell Biol* **10**: 462-467.
- STOPPA-LYONNET, D., and G. LENOIR, 2005 [Cancer genetic predisposition: current events and perspectives 2005]. *Med Sci (Paris)* **21**: 962-968.
- STOREY, J. D., and R. TIBSHIRANI, 2003 Statistical significance for genomewide studies. *Proc Natl Acad Sci U S A* **100**: 9440-9445.
- STOSS, O., P. STOILOV, A. M. HARTMANN, O. NAYLER and S. STAMM, 1999 The in vivo minigene approach to analyze tissue-specific splicing. *Brain Res Brain Res Protoc* **4**: 383-394.
- SU, L. J., and L. ARAB, 2001 Nutritional status of folate and colon cancer risk: evidence from NHANES I epidemiologic follow-up study. *Ann Epidemiol* **11**: 65-72.
- SU, L. K., B. VOGELSTEIN and K. W. KINZLER, 1993 Association of the APC tumor suppressor protein with catenins. *Science* **262**: 1734-1737.
- SUGIMURA, T., K. WAKABAYASHI, H. OHGAKI, S. TAKAYAMA, M. NAGAO *et al.*, 1990 Heterocyclic amines produced in cooked food: unavoidable xenobiotics. *Princess Takamatsu Symp* **21**: 279-288.
- SUN, Y., X. M. TANG, E. HALF, M. T. KUO and F. A. SINICROPE, 2002 Cyclooxygenase-2 overexpression reduces apoptotic susceptibility by inhibiting the cytochrome c-dependent apoptotic pathway in human colon cancer cells. *Cancer Res* **62**: 6323-6328.
- SWEENEY, C., L. LI, R. SHANMUGAM, P. BHAT-NAKSHATRI, V. JAYAPRAKASAN *et al.*, 2004 Nuclear factor-kappaB is constitutively activated in prostate cancer in vitro and is overexpressed in prostatic intraepithelial neoplasia and adenocarcinoma of the prostate. *Clin Cancer Res* **10**: 5501-5507.
- SWEENEY, M. R., J. MCPARTLIN and J. SCOTT, 2007 Folic acid fortification and public health: report on threshold doses above which unmetabolised folic acid appear in serum. *BMC Public Health* **7**: 41.
- TABOR, H. K., N. J. RISCH and R. M. MYERS, 2002 Candidate-gene approaches for studying complex genetic traits: practical considerations. *Nat Rev Genet* **3**: 391-397.
- TAK, P. P., and G. S. FIRESTEIN, 2001 NF-kappaB: a key role in inflammatory diseases. *J Clin Invest* **107**: 7-11.
- TAKAGI, Y., H. KOHMURA, M. FUTAMURA, H. KIDA, H. TANEMURA *et al.*, 1996 Somatic alterations of the DPC4 gene in human colorectal cancers in vivo. *Gastroenterology* **111**: 1369-1372.
- TAKAKU, K., M. OSHIMA, H. MIYOSHI, M. MATSUI, M. F. SELDIN *et al.*, 1998 Intestinal tumorigenesis in compound mutant mice of both *Dpc4* (*Smad4*) and *Apc* genes. *Cell* **92**: 645-656.

- TATSUKA, M., H. KATAYAMA, T. OTA, T. TANAKA, S. ODASHIMA *et al.*, 1998 Multinuclearity and increased ploidy caused by overexpression of the aurora- and lpl1-like midbody-associated protein mitotic kinase in human cancer cells. *Cancer Res* **58**: 4811-4816.
- THIAGALINGAM, S., C. LENGAUER, F. S. LEACH, M. SCHUTTE, S. A. HAHN *et al.*, 1996 Evaluation of candidate tumour suppressor genes on chromosome 18 in colorectal cancers. *Nat Genet* **13**: 343-346.
- THIBODEAU, S. N., G. BREN and D. SCHAID, 1993 Microsatellite instability in cancer of the proximal colon. *Science* **260**: 816-819.
- THIBODEAU, S. N., A. J. FRENCH, J. M. CUNNINGHAM, D. TESTER, L. J. BURGART *et al.*, 1998 Microsatellite instability in colorectal cancer: different mutator phenotypes and the principal involvement of hMLH1. *Cancer Res* **58**: 1713-1718.
- THIBODEAU, S. N., A. J. FRENCH, P. C. ROCHE, J. M. CUNNINGHAM, D. J. TESTER *et al.*, 1996 Altered expression of hMSH2 and hMLH1 in tumors with microsatellite instability and genetic alterations in mismatch repair genes. *Cancer Res* **56**: 4836-4840.
- THOMAS, G., K. B. JACOBS, M. YEAGER, P. KRAFT, S. WACHOLDER *et al.*, 2008 Multiple loci identified in a genome-wide association study of prostate cancer. *Nat Genet* **40**: 310-315.
- THUNE, I., and E. LUND, 1996 Physical activity and risk of colorectal cancer in men and women. *Br J Cancer* **73**: 1134-1140.
- TOMATIS, L., and H. BARTSCH, 1990 The contribution of experimental studies to risk assessment of carcinogenic agents in humans. *Exp Pathol* **40**: 251-266.
- TOMLINSON, I. P., N. E. BECK, K. NEALE and W. F. BODMER, 1996 Variants at the secretory phospholipase A2 (PLA2G2A) locus: analysis of associations with familial adenomatous polyposis and sporadic colorectal tumours. *Ann Hum Genet* **60**: 369-376.
- TOMLINSON, I. P., E. WEBB, L. CARVAJAL-CARMONA, P. BRODERICK, K. HOWARTH *et al.*, 2008 A genome-wide association study identifies colorectal cancer susceptibility loci on chromosomes 10p14 and 8q23.3. *Nat Genet* **40**: 623-630.
- TOMONAGA, T., K. MATSUSHITA, M. ISHIBASHI, M. NEZU, H. SHIMADA *et al.*, 2005 Centromere protein H is up-regulated in primary human colorectal cancer and its overexpression induces aneuploidy. *Cancer Res* **65**: 4683-4689.
- TOMONAGA, T., K. MATSUSHITA, S. YAMAGUCHI, T. OOHASHI, H. SHIMADA *et al.*, 2003 Overexpression and mistargeting of centromere protein-A in human primary colorectal cancer. *Cancer Res* **63**: 3511-3516.
- TOYOOKA, S., K. O. TOYOOKA, K. HARADA, K. MIYAJIMA, P. MAKARLA *et al.*, 2002 Aberrant methylation of the CDH13 (H-cadherin) promoter region in colorectal cancers and adenomas. *Cancer Res* **62**: 3382-3386.
- TOYOTA, M., N. AHUJA, M. OHE-TOYOTA, J. G. HERMAN, S. B. BAYLIN *et al.*, 1999a CpG island methylator phenotype in colorectal cancer. *Proc Natl Acad Sci U S A* **96**: 8681-8686.
- TOYOTA, M., C. HO, N. AHUJA, K. W. JAIR, Q. LI *et al.*, 1999b Identification of differentially methylated sequences in colorectal cancer by methylated CpG island amplification. *Cancer Res* **59**: 2307-2312.
- TRAN, T. T., D. NAIGAMWALLA, A. I. OPRESCU, L. LAM, G. MCKEOWN-EYSSEN *et al.*, 2006 Hyperinsulinemia, but not other factors associated with insulin resistance, acutely enhances colorectal epithelial proliferation in vivo. *Endocrinology* **147**: 1830-1837.
- TRANAH, G. J., J. BUGNI, E. GIOVANNUCCI, J. MA, C. FUCHS *et al.*, 2006 O6-methylguanine-DNA methyltransferase Leu84Phe and Ile143Val polymorphisms and risk of colorectal cancer in the Nurses' Health Study and Physicians' Health Study (United States). *Cancer Causes Control* **17**: 721-731.
- TRANAH, G. J., E. GIOVANNUCCI, J. MA, C. FUCHS, S. E. HANKINSON *et al.*, 2004 XRCC2 and XRCC3 polymorphisms are not associated with risk of colorectal adenoma. *Cancer Epidemiol Biomarkers Prev* **13**: 1090-1091.
- TRAYHURN, P., and J. H. BEATTIE, 2001 Physiological role of adipose tissue: white adipose tissue as an endocrine and secretory organ. *Proc Nutr Soc* **60**: 329-339.
- TROWELL, H., and D. BURKITT, 1986 Physiological role of dietary fiber: a ten-year review. *Bol Asoc Med P R* **78**: 541-544.
- TRUNINGER, K., M. MENIGATTI, J. LUZ, A. RUSSELL, R. HAIDER *et al.*, 2005 Immunohistochemical analysis reveals high frequency of PMS2 defects in colorectal cancer. *Gastroenterology* **128**: 1160-1171.
- TSENG, M., S. C. MURRAY, L. L. KUPPER and R. S. SANDLER, 1996 Micronutrients and the risk of colorectal adenomas. *Am J Epidemiol* **144**: 1005-1014.
- TSUJII, M., S. KAWANO and R. N. DuBOIS, 1997 Cyclooxygenase-2 expression in human colon cancer cells increases metastatic potential. *Proc Natl Acad Sci U S A* **94**: 3336-3340.
- TUYNMAN, J. B., M. P. PEPPELENBOSCH and D. J. RICHEL, 2004 COX-2 inhibition as a tool to treat and prevent colorectal cancer. *Crit Rev Oncol Hematol* **52**: 81-101.
- ULRICH, C. M., J. BIGLER, R. BOSTICK, L. FOSDICK and J. D. POTTER, 2002 Thymidylate synthase promoter polymorphism, interaction with folate intake, and risk of colorectal adenomas. *Cancer Res* **62**: 3361-3364.
- ULRICH, C. M., J. BIGLER, R. SPARKS, J. WHITTON, J. G. SIBERT *et al.*, 2004 Polymorphisms in PTGS1 (=COX-1) and risk of colorectal polyps. *Cancer Epidemiol Biomarkers Prev* **13**: 889-893.
- ULRICH, C. M., E. KAMPMAN, J. BIGLER, S. M. SCHWARTZ, C. CHEN *et al.*, 1999 Colorectal adenomas and the C677T MTHFR polymorphism: evidence for gene-environment interaction? *Cancer Epidemiol Biomarkers Prev* **8**: 659-668.
- UNDERWOOD, L. E., J. P. THISSEN, S. LEMOZY, J. M. KETELSLEGGERS and D. R. CLEMMONS, 1994 Hormonal and nutritional regulation of IGF-I and its binding proteins. *Horm Res* **42**: 145-151.
- VAN DER LOGT, E. M., S. M. BERGEVOET, H. M. ROELOFS, Z. VAN HOOUDONK, R. H. TE MORSCHÉ *et al.*, 2004 Genetic polymorphisms in UDP-glucuronosyltransferases and glutathione S-transferases and colorectal cancer risk. *Carcinogenesis* **25**: 2407-2415.
- VASEN, H. F., 2000 Clinical diagnosis and management of hereditary colorectal cancer syndromes. *J Clin Oncol* **18**: 81S-92S.

- VASEN, H. F., J. P. MECKLIN, P. M. KHAN and H. T. LYNCH, 1991 The International Collaborative Group on Hereditary Non-Polyposis Colorectal Cancer (ICG-HNPCC). *Dis Colon Rectum* **34**: 424-425.
- VASEN, H. F., P. WATSON, J. P. MECKLIN and H. T. LYNCH, 1999 New clinical criteria for hereditary nonpolyposis colorectal cancer (HNPCC, Lynch syndrome) proposed by the International Collaborative group on HNPCC. *Gastroenterology* **116**: 1453-1456.
- VEIGL, M. L., L. KASTURI, J. OLECHNOWICZ, A. H. MA, J. D. LUTTERBAUGH *et al.*, 1998 Biallelic inactivation of hMLH1 by epigenetic gene silencing, a novel mechanism causing human MSI cancers. *Proc Natl Acad Sci U S A* **95**: 8698-8702.
- VERHOEVEN, D. T., R. A. GOLDBOHM, G. VAN POPPEL, H. VERHAGEN and P. A. VAN DEN BRANDT, 1996 Epidemiological studies on brassica vegetables and cancer risk. *Cancer Epidemiol Biomarkers Prev* **5**: 733-748.
- VIATOUR, P., M. P. MERVILLE, V. BOURS and A. CHARIOT, 2005 Phosphorylation of NF-kappaB and I kappa B proteins: implications in cancer and inflammation. *Trends Biochem Sci* **30**: 43-52.
- VOGELSTEIN, B., E. R. FEARON, S. R. HAMILTON, S. E. KERN, A. C. PREISINGER *et al.*, 1988 Genetic alterations during colorectal-tumor development. *N Engl J Med* **319**: 525-532.
- VOGELSTEIN, B., E. R. FEARON, S. E. KERN, S. R. HAMILTON, A. C. PREISINGER *et al.*, 1989 Allelotype of colorectal carcinomas. *Science* **244**: 207-211.
- VOORRIPS, L. E., R. A. GOLDBOHM, G. VAN POPPEL, F. STURMANS, R. J. HERMUS *et al.*, 2000 Vegetable and fruit consumption and risks of colon and rectal cancer in a prospective cohort study: The Netherlands Cohort Study on Diet and Cancer. *Am J Epidemiol* **152**: 1081-1092.
- WALON, C., A. KARTHEUSER, G. MICHILS, M. SMAERS, N. LANNON *et al.*, 1997 Novel germline mutations in the APC gene and their phenotypic spectrum in familial adenomatous polyposis kindreds. *Hum Genet* **100**: 601-605.
- WANG, G. L., B. H. JIANG, E. A. RUE and G. L. SEMENZA, 1995a Hypoxia-inducible factor 1 is a basic-helix-loop-helix-PAS heterodimer regulated by cellular O₂ tension. *Proc Natl Acad Sci U S A* **92**: 5510-5514.
- WANG, J., L. SUN, L. MYEROFF, X. WANG, L. E. GENTRY *et al.*, 1995b Demonstration that mutation of the type II transforming growth factor beta receptor inactivates its tumor suppressor activity in replication error-positive colon carcinoma cells. *J Biol Chem* **270**: 22044-22049.
- WANG, X., H. W. CHEUNG, A. C. CHUN, D. Y. JIN and Y. C. WONG, 2008 Mitotic checkpoint defects in human cancers and their implications to chemotherapy. *Front Biosci* **13**: 2103-2114.
- WANNAMETHEE, S. G., A. G. SHAPER and M. WALKER, 2001 Physical activity and risk of cancer in middle-aged men. *Br J Cancer* **85**: 1311-1316.
- WARD, M. H., A. J. CROSS, H. DIVAN, M. KULLDORFF, S. NOWELL-KADLUBAR *et al.*, 2007 Processed meat intake, CYP2A6 activity and risk of colorectal adenoma. *Carcinogenesis* **28**: 1210-1216.
- WARGOVICH, M. J., V. W. ENG, H. L. NEWMARK and W. R. BRUCE, 1983 Calcium ameliorates the toxic effect of deoxycholic acid on colonic epithelium. *Carcinogenesis* **4**: 1205-1207.
- WARREN, R. S., H. YUAN, M. R. MATLI, N. FERRARA and D. B. DONNER, 1996 Induction of vascular endothelial growth factor by insulin-like growth factor 1 in colorectal carcinoma. *J Biol Chem* **271**: 29483-29488.
- WARTHIN, A., 1913 Hereditary with reference to carcinoma as shown by the study of the cases examined in the pathological laboratory of the University of Michigan, 1895-1913. *Arch Intern Med* **12**: 546-555.
- WASYLYK, C., J. SCHNEIKERT and B. WASYLYK, 1990 Oncogene v-jun modulates DNA replication. *Oncogene* **5**: 1055-1058.
- WATKINS, L. F., L. R. LEWIS and A. E. LEVINE, 1990 Characterization of the synergistic effect of insulin and transferrin and the regulation of their receptors on a human colon carcinoma cell line. *Int J Cancer* **45**: 372-375.
- WCRF, 1997 *Food, Nutrition and prevention of cancer: A Global Perspective*. WCRF/AICR, Washington, DC.
- WEBB, E. L., M. F. RUDD and R. S. HOULSTON, 2006 Colorectal cancer risk in monoallelic carriers of MYH variants. *Am J Hum Genet* **79**: 768-771; author reply 771-762.
- WEI, E. K., J. MA, M. N. POLLAK, N. RIFAI, C. S. FUCHS *et al.*, 2005 A prospective study of C-peptide, insulin-like growth factor-I, insulin-like growth factor binding protein-1, and the risk of colorectal cancer in women. *Cancer Epidemiol Biomarkers Prev* **14**: 850-855.
- WEINSTEIN, B., A. M. CAROTHERS, R. SANTELLA and F. PERERA, 1995 *The molecular basis of cancer: Molecular mechanisms of mutagenesis and multistage carcinogenesis*, Philadelphia.
- WHANGER, P. D., 2004 Selenium and its relationship to cancer: an update. *Br J Nutr* **91**: 11-28.
- WHEELOCK, M. J., Y. SHINTANI, M. MAEDA, Y. FUKUMOTO and K. R. JOHNSON, 2008 Cadherin switching. *J Cell Sci* **121**: 727-735.
- WICKING, C., L. A. SIMMS, T. EVANS, M. WALSH, K. CHAWENGSAKSOPHAK *et al.*, 1998 CDX2, a human homologue of Drosophila caudal, is mutated in both alleles in a replication error positive colorectal cancer. *Oncogene* **17**: 657-659.
- WICKRAMASINGHE, S. N., and S. FIDA, 1994 Bone marrow cells from vitamin B12- and folate-deficient patients misincorporate uracil into DNA. *Blood* **83**: 1656-1661.
- WIJNEN, J., H. VAN DER KLIFT, H. VASEN, P. M. KHAN, F. MENKO *et al.*, 1998 MSH2 genomic deletions are a frequent cause of HNPCC. *Nat Genet* **20**: 326-328.
- WILLETT, W. C., 2001 Diet and cancer: one view at the start of the millennium. *Cancer Epidemiol Biomarkers Prev* **10**: 3-8.
- WILLIAMS, C. S., M. MANN and R. N. DUBOIS, 1999 The role of cyclooxygenases in inflammation, cancer, and development. *Oncogene* **18**: 7908-7916.
- WOOD, L. D., D. W. PARSONS, S. JONES, J. LIN, T. SJOBLOM *et al.*, 2007 The genomic landscapes of human breast and colorectal cancers. *Science* **318**: 1108-1113.
- WOOD, R. D., M. MITCHELL, J. SGOUROS and T. LINDAHL, 2001 Human DNA repair genes. *Science* **291**: 1284-1289.

- WU, A. H., D. SHIBATA, M. C. YU, M. Y. LAI and R. K. ROSS, 2001a Dietary heterocyclic amines and microsatellite instability in colon adenocarcinomas. *Carcinogenesis* **22**: 1681-1684.
- WU, K., E. GIOVANNUCCI, C. BYRNE, E. A. PLATZ, C. FUCHS *et al.*, 2006 Meat mutagens and risk of distal colon adenoma in a cohort of U.S. men. *Cancer Epidemiol Biomarkers Prev* **15**: 1120-1125.
- WU, Y., M. J. BERENDS, R. H. SIMONS, R. G. MENSINK, E. VERLIND *et al.*, 2001b A role for MLH3 in hereditary nonpolyposis colorectal cancer. *Nat Genet* **29**: 137-138.
- WURZELMANN, J. I., A. SILVER, D. M. SCHREINEMACHERS, R. S. SANDLER and R. B. EVERSON, 1996 Iron intake and the risk of colorectal cancer. *Cancer Epidemiol Biomarkers Prev* **5**: 503-507.
- XU, J., J. LOWEY, F. WIKLUND, J. SUN, F. LINDMARK *et al.*, 2005 The interaction of four genes in the inflammation pathway significantly predicts prostate cancer risk. *Cancer Epidemiol Biomarkers Prev* **14**: 2563-2568.
- XUE, L., M. LIPKIN, H. NEWMARK and J. WANG, 1999 Influence of dietary calcium and vitamin D on diet-induced epithelial cell hyperproliferation in mice. *J Natl Cancer Inst* **91**: 176-181.
- YAGI, O. K., Y. AKIYAMA, T. NOMIZU, T. IWAMA, M. ENDO *et al.*, 1998 Proapoptotic gene BAX is frequently mutated in hereditary nonpolyposis colorectal cancers but not in adenomas. *Gastroenterology* **114**: 268-274.
- YAMADA, Y., N. HAMAJIMA, T. KATO, H. IWATA, Y. YAMAMURA *et al.*, 2003 Association of a polymorphism of the phospholipase D2 gene with the prevalence of colorectal cancer. *J Mol Med* **81**: 126-131.
- YAMAGUCHI, T., T. IJIMA, T. MORI, K. TAKAHASHI, H. MATSUMOTO *et al.*, 2006 Accumulation profile of frameshift mutations during development and progression of colorectal cancer from patients with hereditary nonpolyposis colorectal cancer. *Dis Colon Rectum* **49**: 399-406.
- YAMAMOTO, H., H. HANAFUSA, M. OUCHIDA, M. YANO, H. SUZUKI *et al.*, 2005 Single nucleotide polymorphisms in the EXO1 gene and risk of colorectal cancer in a Japanese population. *Carcinogenesis* **26**: 411-416.
- YAMAMOTO, H., H. SAWAI and M. PERUCHO, 1997 Frameshift somatic mutations in gastrointestinal cancer of the microsatellite mutator phenotype. *Cancer Res* **57**: 4420-4426.
- YANG, C. R., S. L. HSIEH, F. M. HO and W. W. LIN, 2005 Decoy receptor 3 increases monocyte adhesion to endothelial cells via NF-kappa B-dependent up-regulation of intercellular adhesion molecule-1, VCAM-1, and IL-8 expression. *J Immunol* **174**: 1647-1656.
- YE, W., A. ROMELSIJ, K. AUGUSTSSON, H. O. ADAMI and O. NYREN, 2003 No excess risk of colorectal cancer among alcoholics followed for up to 25 years. *Br J Cancer* **88**: 1044-1046.
- YEH, C. C., L. L. HSIEH, R. TANG, C. R. CHANG-CHIEH and F. C. SUNG, 2005 MS-920: DNA repair gene polymorphisms, diet and colorectal cancer risk in Taiwan. *Cancer Lett* **224**: 279-288.
- YEO, G., and C. B. BURGE, 2004 Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J Comput Biol* **11**: 377-394.
- YIN, M. J., Y. YAMAMOTO and R. B. GAYNOR, 1998 The anti-inflammatory agents aspirin and salicylate inhibit the activity of I(kappa)B kinase-beta. *Nature* **396**: 77-80.
- YUEN, K. W., C. D. WARREN, O. CHEN, T. KWOK, P. HIETER *et al.*, 2007 Systematic genome instability screens in yeast and their potential relevance to cancer. *Proc Natl Acad Sci U S A* **104**: 3925-3930.
- ZANKE, B. W., C. M. GREENWOOD, J. RANGREJ, R. KUSTRA, A. TENESA *et al.*, 2007 Genome-wide association scan identifies a colorectal cancer susceptibility locus on chromosome 8q24. *Nat Genet* **39**: 989-994.
- ZHANG, H., A. AHMADI, G. ARBMAN, J. ZDOLSEK, J. CARSTENSEN *et al.*, 1999 Glutathione S-transferase T1 and M1 genotypes in normal mucosa, transitional mucosa and colorectal adenocarcinoma. *Int J Cancer* **84**: 135-138.
- ZHANG, M. Q., 1998 Statistical features of human exons and their flanking regions. *Hum Mol Genet* **7**: 919-932.
- ZHONG, S., A. H. WYLLIE, D. BARNES, C. R. WOLF and N. K. SPURR, 1993 Relationship between the GSTM1 genetic polymorphism and susceptibility to bladder, breast and colon cancer. *Carcinogenesis* **14**: 1821-1824.
- ZHOU, J., T. SCHMID and B. BRUNE, 2003 Tumor necrosis factor-alpha causes accumulation of a ubiquitinated form of hypoxia inducible factor-1alpha through a nuclear factor-kappaB-dependent pathway. *Mol Biol Cell* **14**: 2216-2225.
- ZHOU, M., S. SHELDON, N. AKEL and A. A. KILLEEN, 1999 Chromosomal aneuploidy in leukemic blast crisis: a potential source of error in interpretation of bone marrow engraftment analysis by VNTR amplification. *Mol Diagn* **4**: 153-157.
- ZHOU, X. L., T. DJUREINOVIC, B. WERELIUS, G. LINDMARK, X. F. SUN *et al.*, 2005 Germline mutations in the MYH gene in Swedish familial and sporadic colorectal cancer. *Genet Test* **9**: 147-151.
- ZHU, L., 2005 Tumour suppressor retinoblastoma protein Rb: a transcriptional regulator. *Eur J Cancer* **41**: 2415-2427.
- ZHU, M., and S. ZHAO, 2007 Candidate gene identification approach: progress and challenges. *Int J Biol Sci* **3**: 420-427.

Etude d'association cas-témoins pour l'identification de gènes de prédisposition aux cancers colorectaux sporadiques

Les cancers colorectaux (CCR) constituent la troisième pathologie cancéreuse la plus fréquente en France. On distingue les formes familiales (25%) et les formes sporadiques (75%). Les premières sont dues à des mutations délétères dans des gènes de prédisposition majeure connus, alors que les secondes sont induites par l'effet combiné de facteurs de risque génétiques et environnementaux dont la nature exacte reste encore inconnue. Afin de déterminer la composante génétique des CCR sporadiques, nous avons développé une étude d'association cas-témoins portant sur des variants génétiques à faible pénétrance. Par une approche « gènes candidats », nous avons montré l'effet prédisposant aux CCR de cinq SNP (*Single Nucleotide Polymorphism*) localisés dans les gènes *PTGS1*, *IL8*, *MTHFR*, *PLA2G2A*, *PPARG*, presque tous directement impliqués dans des phénomènes inflammatoires procarcinogènes. En nous focalisant sur des voies métaboliques, nous avons également mis en avant une interaction gène-environnement entre des SNP de cytochromes P450 et une forte consommation de viande rouge. Par une approche pangénomique réalisée sur des pools d'ADN, nous avons finalement montré l'association possible au risque de CCR, dans notre population, de deux autres SNP, situés dans *CDH13* et *UTP6*, l'effet prédisposant probable du SNP de *CDH13* étant renforcé par une corrélation entre génotype et taux d'expression ARNm. L'appartenance de *CDH13* à la superfamille des cadhérines, dont certains membres ont un rôle évident dans le cancer, nous pousse à poursuivre nos investigations en analysant l'implication aux CCR d'autres SNP des gènes des cadhérines.

Mots clés : Etude d'association – Cancer colorectal – Polymorphisme – Gène candidat – Etude pangénomique – Pool d'ADN

Case-Control association study for the identification of genes predisposing to sporadic colorectal cancers

Colorectal cancers (CRC) constitute the third most frequent cancer in France. Two distinct forms of CRC can be distinguished: the family forms (25%), due to deleterious mutations in known high-risk genes, and sporadic forms (75%), induced by the combined effect of both genetic and environmental risk factors not identified yet. In order to determine the genetic component of sporadic CRC, we conducted a case-control association study using low-penetrance genetic variants. By a "candidate gene approach", we demonstrated the CRC-predisposing effect of 5 SNPs (Single Nucleotide Polymorphism) located genes *PTGS1*, *IL8*, *MTHFR*, *PLA2G2A*, *PPARG*, almost all involved in procarcinogenetic inflammatory events. By focusing on metabolic pathway, we highlighted a gene-environment interaction between cytochromes P450 SNPs and great red meat consumption. By a genome-wide association study based on DNA pools, we eventually showed, in our population, the involvement of two others SNPs located in *CDH13* and *UTP6*, the predisposing effect of the *CDH13* SNP being strengthened by a genotype-mRNA expression rate correlation. The fact that *CDH13* belongs to the cadherin superfamily, whose members have a clear role in cancer, leads us to continue our investigations by analyzing the involvement of other SNPs of cadherin genes in CRC.

Key words : Association study – Colorectal cancer – Polymorphism – Candidate gene – Genome-wide study – DNA pool

Discipline : Génétique Moléculaire

EA Biométadys, Faculté de Médecine de NANTES, 1 rue Gaston Veil, 44035 NANTES cedex