

UNIVERSITE DE NANTES
FACULTE DE MEDECINE

**Identification d'un gène humain localisé en 17q12,
codant pour un membre de la famille F-box (Fbxo47).
Étude des caractéristiques de la protéine.**

THESE DE DOCTORAT

École Doctorale CHIMIE-BIOLOGIE
Discipline : Médecine, Biologie, Santé
Spécialité : Génétique moléculaire

Présentée et soutenue publiquement par

Barbara SIMON-KAYSER

Le 29 mai 2006, devant le jury ci-dessous

Président François VALETTE, Directeur de Recherche INSERM, Nantes

Rapporteurs Pascal REYNIER, Professeur, Université d'Angers
Bertrand DAIGNAN-FORNIER, Directeur de Recherche CNRS,
Bordeaux

Examineurs François VALETTE, Directeur de Recherche INSERM,
Nantes

Directeur de thèse Stéphane BEZIEAU , MCU-PH, Nantes.

ABBREVIATIONS

Les abréviations citées ici, sont uniquement celles qui ne sont pas transcrites dans le texte.

ADN, ADNc: acide désoxyribonucléique, ADN complémentaire

APS : persulfate d'ammonium

ARN, ARNm, ARNt : acide ribonucléique, acide ribonucléique messager, acide ribonucléique de transfert.

ATP, ADP, AMP : adénosine triphosphate, adénosine diphosphate, adénosine monophosphate.

BET : bromure d'éthydiu

BSA : sérum-albumine bovine

DNase : désoxyribonucléase

dNTP : désoxyribonucléoside-5'-triphosphate

EDTA : Acide éthylène diamine tétraacétique

GFP : "green fluorescent protein"

Hsp : "Heat-shock protein"

IPTG : isopentényl- β -D-thiogalactopyranoside

LB : Luria broth LB

mt : mitochondrial

NADH, NADPH :

ORF : Open Reading Frame : phase de lecture ouverte

pb : paire de bases

PBS : tampon phosphate salin

PCR : réaction de polymérisation en chaîne

PM : poids moléculaire

RNase : ribonucléase

RT-PCR : transcription inverse puis PCR

SDS : dodécylsulfate de sodium

SSC : standard sodium citrate

SVF : sérum de veau foetal

TCA : acide trichloroacétique

Tm : température d'annealing

Tris : N-tris(hydroxyméthyl)aminométhane

TE : Tris-EDTA

UV : ultraviolet

UNITES

°C : degré Celsius
cm : centimètre
g : gramme
g : unité de gravitation
kb : kilobase
kDa : kilodalton
kV : kilovolt
M : molaire
mA : milliampère
ml : millilitre
mm : millimètre
mA : milliampère
Ci : curie

mM : millimolaire
ng : nanogramme
nm : nanomètre
pmol : picomole
rpm : tours par minute
U : unité
V : volt
µg : microgramme
µl : microlitre
µm : micromètre
µM : micromolaire
cfu : unité de colonies formées

NOMENCLATURE

Pour une question d'harmonisation, la nomenclature suivante a été adoptée : GENE – Protéine

δ[X-Y] Orf138: dénomme la protéine Orf138 où les résidus X à Y ont été supprimés

Δ[X-Y] ORF138 dénomme le gène ORF138 où les nucléotides X à Y ont été supprimés

CxA : dénomme la substitution du résidu cystéine en un résidu alanine à la position x

SOMMAIRE

LISTE DES TABLEAUX.....	7
LISTE DES FIGURES	8
AVANT-PROPOS.....	10
CHAPITRE 1 INTRODUCTION.....	12
A. CONTEXTE DE LA RECHERCHE D'UN GENE CANDIDAT	13
I. LA REGION 17Q12, CANCER COLORECTAL ET CANCER PAPILLAIRE RENAL	13
I.1. <i>Étude des cancers colorectaux familiaux au CHU de Nantes</i>	13
I.2. <i>17q12 et cancers papillaires rénaux</i>	15
I.3. <i>17q11-q23 : une région d'intérêt dans le cancer</i>	17
II. LES ACTEURS DE LA CANCEROGENESE : ONCOGENES, GENES SUPPESSEURS DE TUMEUR ET AUTRES.	19
II.1. <i>Oncogènes et Gain de Fonction</i>	19
II.2. <i>Gènes suppresseurs de tumeurs et perte de fonction</i>	20
II.3. <i>Les autres acteurs de la cancérogenèse</i>	21
B. LE SYSTEME UBIQUITINE.....	22
I. MODIFICATION UBIQUITINE ET PROCESSUS CELLULAIRES	22
I.1. <i>L'ubiquitine</i>	22
I.2. <i>L'ubiquitination</i>	23
I.3. <i>Conséquence sur le substrat ubiquitiné</i>	24
II. LES LIGASES E3S : DEUX GRANDES CLASSES	26
II.1. <i>Définition</i>	27
II.2. <i>Les ligases de type HECT</i>	29
II.3. <i>Les ligases de type RING</i>	29
C. LES PROTEINES F-BOX.....	34
I. DEFINITION ET CLASSIFICATION	34
I.1. <i>Historique : l'hypothèse F-box</i>	34
I.2. <i>Caractéristiques du motif F-box</i>	35
I.3. <i>Le domaine d'interaction spécifique du substrat</i>	36
II. LOCALISATION CELLULAIRE	40
III. REGULATION	42
III.1. <i>Régulation transcriptionnelle</i>	42
III.2. <i>Régulation post-traductionnelle</i>	42
IV. QUELQUES FONCTIONS DES PROTEINES F-BOX	43
IV.1. <i>F-box et cycle cellulaire</i>	43
IV.2. <i>Régulation de la biogenèse mitochondriale</i>	47
IV.3. <i>Mort cellulaire programmée</i>	48
IV.4. <i>Exception à la voie ubiquitine-protéasome</i>	50
V. PROTEINES F-BOX ET PATHOLOGIES HUMAINES	53
V.1. <i>Cancer : dérégulation du cycle cellulaire</i>	53
V.2. <i>Subversion virale</i>	57
V.3. <i>Autres pathologies</i>	57
D. OBJECTIFS DE LA THESE	58
I. IMPORTANCE DE LA RECHERCHE SUR LA VOIE UBIQUITINE	58
II. BUT DU TRAVAIL DE THESE.....	59
CHAPITRE 2 MATERIELS ET METHODES	60
A. ÉTUDE DE LA REGION GENOMIQUE DE FBXO47.....	61

I. ÉTUDE DE PERTE D'HETEROZYGOTIE LOH	61
B. CLONAGE DU GENE FBXO47	64
I. CLONAGE IN SILICO	64
I.1. Recherche d'une structure exon / intron et phase ouverte de lecture.....	64
I.2. Analyse des régions non traduites	65
II. CLONAGE IN VITRO.....	65
II.1. PCR exon-connection.....	65
II.2. Race-PCR.....	66
C. ETUDE DE L'EXPRESSION DU GENE FBXO47	69
I. NORTHERN BLOT	69
II.1. Membranes utilisées.....	69
I.2. Choix et marquage de la Sonde.....	69
I.3. Conditions d'hybridation.....	70
I.4. Visualisation.....	70
II. RT-PCR PANEL.....	70
II.1. Panels utilisés	70
II.2. Conditions d'amplification (PCR)	71
II.3. Visualisation.....	72
III. EXPRESSION DIFFERENTIELLE : CANCER PROFILING ARRAY (CPA).....	72
III.1. Principe	72
III.2. Choix et Marquage de la sonde	73
III.3. Hybridation.....	73
III.4. Visualisation et Quantification	73
D. ÉTUDE DE LA PROTEINE FBXO47	74
I. ANALYSE BIOINFORMATIQUE DES MOTIFS PROTEIQUES ET CARACTERISTIQUES PHYSICO-CHIMIQUES	74
I.1. Caractéristiques physico-chimiques.....	74
I.2. Caractéristiques protéiques.....	74
II. LOCALISATION CELLULAIRE	76
II.1. Clonage des protéines d'intérêt dans des vecteurs d'expression	76
II.2. Culture cellulaire et transfection transitoire.....	77
II.3. Marquage du noyau et des mitochondries	77
II.4. Lecture des lames et Analyse	77
III. IDENTIFICATION D'INTERACTIONS PROTEIQUES.....	78
III.1. Co-précipitation / Interaction INVITRO.....	78
III.2. Double Hybride / Interaction IN VIVO.....	80
CHAPITRE 3 RESULTATS.....	88
A. IDENTIFICATION D'UN NOUVEAU GENE LOCALISE EN 17Q12.....	89
I. ÉTUDE DE REARRANGEMENTS GENOMIQUES DANS LA REGION 17Q12.....	89
II. ANALYSE BIOINFORMATIQUE DE LA REGION ENCADRANT LE MARQUEUR D17S250.....	91
II.1. Définition de la zone d'étude	91
II.2. Annotation des séquences répétées	91
II.3. Prédiction d'exons et de gènes pour la région d'intérêt	93
II.4. Synthèse des résultats : définition d'une ORF potentielle.....	98
III. VALIDATION BIOLOGIQUE DES TRANSCRITS POTENTIELS.....	99
III.1. PCR exon-connection : structure de la phase ouverte de lecture	99
III.2. Structure des extrémités 5' et 3' non codantes	100
B. CARACTERISATION FONCTIONNELLE DU TRANSCRIT.....	105
I. COMPARAISON AUX BANQUES D'EST	105
II. EXPRESSION DE L'ARNM	105
II.1. Northern Blot	105
II.2. RT-PCR - banque de tissus	106
II.3. Expression différentielle.....	107

III.	ÉTUDE BIOINFORMATIQUE DU PRODUIT CODE PAR LE TRANSCRIT	112
III.1.	<i>Caractérisation bioinformatique de propriétés physico-chimiques</i>	112
III.2.	<i>Identification d'un domaine F-Box</i>	113
III.3.	<i>Nomenclature</i>	116
III.4.	<i>Identification de signaux/motifs protéiques</i>	117
III.5.	<i>Recherche d'un peptide-signal et/ou d'un signal d'adressage</i>	118
III.6.	<i>Structure secondaire de Fbxo47</i>	121
C.	RECHERCHE DE LA FONCTION BIOLOGIQUE DE LA PROTEINE FBXO47	124
I.	ETUDE DE LA FONCTIONNALITE DU DOMAINE F-BOX.....	124
I.1.	<i>Interaction in vitro</i>	124
I.2.	<i>Interaction in vivo</i>	126
II.	LOCALISATION MITOCHONDRIALE DE LA PROTEINE FBXO47	127
III.	RECHERCHE DE PARTENAIRES PROTEIQUES DE FBXO47	129
III.1.	<i>Amplification et titrage de la banque d'ADNc de rein</i>	129
III.2.	<i>Détermination du nombre de co-transformations</i>	130
III.3.	<i>Criblage primaire</i>	130
III.4.	<i>Criblage secondaire</i>	130
III.5.	<i>Validations des interactions positives</i>	131
III.6.	<i>Analyse et identification d'interacteurs protéique de Fbxo47</i>	132
	CHAPITRE 4 DISCUSSION	136
A.	CLONAGE DU GENE FBXO47	137
I.	DE LA PRESENCE D'UN GENE SUPPESSEUR DE TUMEUR EN 17Q12	137
I.1.	<i>Perte d'hétérozygotie en 17q12 et cancer papillaire rénal</i>	137
I.2.	<i>Expression différentielle tissu normal/tissu tumoral</i>	138
I.3.	<i>FBXO47 est-il un gène suppresseur de tumeur?</i>	139
II.	PROFIL D'EXPRESSION DU GENE FBXO47	140
II.1.	<i>Pertinence des techniques d'étude d'expression de gènes</i>	141
II.2.	<i>Synthèse : Profil d'expression du gène FBXO47</i>	142
B.	CARACTERISTIQUES DE LA PROTEINE FBXO47	144
I.	FBXO47, UN NOUVEAU MEMBRE DE LA FAMILLE DES F-BOX	144
I.1.	<i>Fonctionnalité du domaine F-Box</i>	145
I.2.	<i>Le second domaine d'interaction protéique</i>	146
I.3.	<i>L'appartenance à une ligase E3 ?</i>	147
II.	FBXO47 : UNE PROTEINE MITOCHONDRIALE	147
II.1.	<i>Le signal d'adressage mitochondrial de type pré-séquence</i>	147
II.2.	<i>Localisation cellulaire de Fbxo47</i>	148
II.3.	<i>Le système F-box et la mitochondrie</i>	151
C.	QUEL ROLE POUR FBXO47 ?	154
I.	TECHNIQUE DU DOUBLE-HYBRIDE BACTERIEN.....	154
I.1.	<i>Double-Hybride Bactérie versus Levure</i>	154
I.2.	<i>Choix de la banque testée</i>	155
I.2.	<i>Limites de la technique</i>	155
II.	GFM1, PROTEINE CIBLE DE FBXO47?	159
II.1.	<i>Présentation du facteur d'élongation mitochondrial G1</i>	159
II.2.	<i>Conséquences sur la fonction de Fbxo47</i>	161
	CHAPITRE 5 CONCLUSION ET PERSPECTIVES	164
	REFERENCES BIBLIOGRAPHIQUES	168
	ANNEXES	199

Liste des tableaux

Tableau 1. Présentation des gènes localisés en 17q11-q23.....	18
Tableau 2. Fréquences de perte d'hétérozygotie des huit microsatellites étudiés sur les 15 cas de cancers papillaires rénaux.....	90
Tableau 3. Répartition globale des séquences répétées dans la séquence analysée de 100 kb.....	92
Tableau 4. Prédiction des exons sur la séquence de 100 kb (brin « forward »).....	94
Tableau 5. Prédiction des exons sur la séquence de 100 kb (brin « reverse »).....	95
Tableau 6. Comparaison entre la séquence génomique contenant les exons candidats sur le chromosome 17 humain et le chromosome 11 de la Souris.....	97
Tableau 7. Résultat des trois itérations effectuées à l'aide du programme PSI-BLAST.....	115
Tableau 8. Compilation des résultats pour les modifications post-traductionnelles de Fbxo47, obtenus par les programmes Prosite, NetPhos 2.0 et NetNGlyc.....	118
Tableau 9. Résultats de comparaison des séquences des interacteurs positifs validés aux banques de données nucléotidiques via le programme BlastN.....	133
Tableau 10. Résultats de comparaison des séquences des interacteurs positifs validés aux banques de données protéiques via le programme TblastX.....	134
Tableau 11. Comparaison des résultats obtenus pour l'étude de l'expression de FBXO47 dans des tissus normaux adultes par les trois techniques employées : Northern Blot, RT-PCR, et CPA II (hybridation sur ADNc).....	143

Liste des figures

Figure 1. Illustration de la théorie de Knudson ou « Two-Hit hypothesis »	20
Figure 2. Représentation de la structure tertiaire de l'ubiquitine.	22
Figure 3. Réactions enzymatiques du système ubiquitine.....	23
Figure 4. Les différents types d'ubiquitination et leurs conséquences	24
Figure 5. Représentation de la dégradation protéosomale ubiquitine dépendante.....	26
Figure 6. Différents mécanismes de transfert de l'ubiquitine sur son substrat.	28
Figure 7. Structure tertiaire du domaine HECT.	28
Figure 8. Le domaine RING des E3 ligases.....	30
Figure 9. La super famille des ubiquitine ligases de type RING.	32
Figure 10. Similarité de structure des domaines de liaisons à la sous-unité E3,.....	33
Figure 11. Caractéristiques du motif F-box.	37
Figure 12. Structures des domaines de protéines F-box de mammifères ¹⁰⁵	38
Figure 13. Répartition de la localisation sub-cellulaire des trois classes de protéines F-box selon l'analyse de leur séquence protéique via le programme IPSORTII.....	41
Figure 14. F-Box et Cycle cellulaire d'après Cardozo et al. ¹³⁶	46
Figure 15. Apoptose et ubiquitination d'après Lee et al ¹⁵¹	50
Figure 16. Rôles de la protéine F-box Emi1 dans la régulation du cycle cellulaire.....	52
Figure 17. Principe méthodologique de la RACE-PCR.....	68
Figure 18. Système Double-Hybride du <i>BacterioMatch System II</i>	82
Figure 20. Validation des interacteurs positifs dans la technique de double hybride.	87
Figure 21. Résultats de l'analyse LOH des 15 cas de pRCC à l'aide de huit marqueurs microsatellites répartis sur le chromosome 17.	90
Figure 22. Représentation de la région génomique d'intérêt pour notre recherche de gène candidat.	92
Figure 23. Définition <i>IN SILICO</i> de la structure du gène candidat (31kb, 11 exons).....	98
Figure 24. PCR exon-connection.	100
Figure 25. Localisation des oligonucléotides utilisés pour la RACE-PCR.....	101
Figure 26. Résultats obtenus par RACE-PCR.....	102
Figure 27. Représentation de la structure génomique du gène FBXO47.....	104
Figure 28. Expression tissulaire du gène d'intérêt.	107

Figure 29. Profil d'expression différentielle sur des ADNC de tissus sains et tumoraux (N et T) et sur des lignées cellulaires.	110
Figure 30. Semi-Quantification des résultats obtenus au PhosphorImager sur Cancer Profiling Array II avec la sonde gène-spécifique.	111
Figure 31. Séquence consensus du motif F-Box,	116
Figure 32. Alignement multiple de séquences F-Box pour quelques membres représentatifs de la superfamille des protéines F-box.	116
Figure 33. Résultat de la soumission de notre séquence protéique au programme PSORT II.	120
Figure 34. Prédiction de structures secondaires pour la protéine Fbxo47.	122
Figure 35. ADNc et séquence protéique déduite du gène humain FBXO47.	123
Figure 36. Interaction Fbxo47-Skp1 <i>in vitro</i> (SDS-Page et autoradiographie des protéines radiomarquées).	125
Figure 37. Test d'Interaction Skp1-Fbxo47 (A) et Skp1- Δ [1-155]Fbxo47 (B) par le système double-hybride bactérien.	125
Figure 38. Localisation sub-cellulaire de la protéine Fbxo47 après transfection de cellules COS.	128
Figure 39. Structure exons / introns des différents ARNm variants de NR4A1 et alignement avec l'ADNc contenu dans les clones C201 et C286.	133
Figure 40. Visualisation de l'interaction Gfm1 et formes avec et sans F-box de la protéine Fbxo47.	135
Figure 41. Distribution taxonomique des protéines F-Box selon la base de données InterPro.	145
Figure 42. L'importation des protéines à la mitochondrie.	149

Avant-propos

Le LEPA, Laboratoire d'Etude du Polymorphisme de l'ADN (UPRES EA3823) a pour principal thème de recherche, l'étude de la prédisposition génétique aux cancers colorectaux sporadiques. Le LEPA dispose d'une banque d'ADN, issus de prélèvements sanguins de deux grands groupes : les « cas » bien caractérisés de cancers colorectaux non familiaux et les « témoins » indemnes de toute pathologie de ce type. Au cours de mon DEA, j'ai participé à ce projet d'identification des prédispositions génétiques au cancer colorectal dans la population générale. J'ai développé et évalué les techniques de génotypage les mieux adaptées à l'étude génétique de 2000 prélèvements et de 50 marqueurs génétiques.

D'autre part, le laboratoire de recherche appartient au Service de génétique médicale du CHU de Nantes. La section oncogénétique moléculaire de cette unité de diagnostic des maladies génétiques recherche notamment les instabilités microsatellites au niveau tumoral dans le syndrome de Lynch. C'est au cours d'une de ces analyses, qu'une délétion sur le chromosome 17 a été mise en évidence sur du matériel tumoral. Après une étude bibliographique de l'implication de cette région chromosomique dans les cancers, nous avons décidé de rechercher et caractériser un éventuel gène suppresseur de tumeur.

Parallèlement, les services d'urologie et d'anatomopathologie nous ont demandé d'étudier cette même région chromosomique dans les cas de cancers papillaires rénaux, renforçant ainsi notre suspicion de l'existence d'un gène impliqué dans la prédisposition au cancer en 17q12.

Le résultat de cette recherche nous a permis d'identifier un gène codant pour une protéine F-box, famille dont certains membres, connus et caractérisés, sont soit des oncogènes, soit des gènes suppresseurs de tumeur.

La première partie de ce mémoire est consacrée à une présentation générale de l'importance de la région 17q12 dans certains cancers puis à la présentation des notions d'oncogènes et de gènes suppresseurs de tumeur. J'aborderais ensuite le système ubiquitine et plus particulièrement les protéines F-box et leurs différents rôles au sein de la cellule.

Je détaillerais dans un deuxième chapitre les différentes techniques utilisées au cours de mon travail de thèse. Le chapitre suivant sera consacré aux résultats de ces investigations d'une part sur le gène cloné, et d'autre part à propos de la protéine codée. L'ensemble des données obtenues fera ensuite l'objet d'une discussion.

Enfin, je tenterai de rassembler l'ensemble des résultats afin d'en dégager des perspectives et des voies d'investigations futures visant à caractériser le rôle du gène et de la protéine dans la régulation cellulaire.

CHAPITRE 1

INTRODUCTION

A. Contexte de la recherche d'un gène candidat

I. La région 17q12, cancer colorectal et cancer papillaire rénal

I.1. Étude des cancers colorectaux familiaux au CHU de Nantes

I.1.1 Présentation des cancers colorectaux familiaux : syndrome de Lynch

On estime qu'environ 5% à 8% des cancers colorectaux (CCR) sont d'origine génétique, c'est-à-dire qu'ils sont liés à la présence d'un gène de prédisposition muté chez l'individu atteint. Cette mutation est héréditaire (elle est dite germinale ou constitutionnelle par opposition aux mutations dites « somatiques » qui ne sont présentes que dans les cellules tumorales). Dans ce contexte, le cancer colorectal apparaît alors comme une maladie héréditaire de transmission autosomique dominante.

Les cancers colorectaux familiaux se regroupent en deux grands syndromes : la polypose adénomateuse familiale (PAF) et le syndrome HNPCC (*Hereditary Non Polyposis Colorectal Cancer*) ou syndrome de Lynch. L'identification des gènes responsables de ces syndromes a permis d'établir des stratégies de diagnostic voire de dépistage des formes génétiques de CCR. Elle a aussi permis de mieux comprendre les voies de carcinogénèse impliquées dans les formes communes non génétiques de CCR.

La polypose adénomateuse familiale (1% des CCR) représente une entité bien définie et facilement reconnaissable compte tenu de son phénotype spécifique. On retrouve une mutation du gène APC dans environ 70% des PAF typiques. Son identification permet de réaliser un dépistage familial simple et fiable. De plus, la mise en évidence de corrélations génotype-phénotype a permis de mieux ajuster la prise en charge clinique des patients.

Les syndromes HNPCC représentent la forme la plus fréquente des CCR héréditaires (5 à 10%). Plusieurs gènes responsables, appartenant à une même famille (gènes MMR *MisMatch Repair*), ont été identifiés. Les syndromes HNPCC posent le problème de leur reconnaissance au sein de l'ensemble des CCR sporadiques. En effet, leur phénotype n'est pas spécifique même s'il existe des caractéristiques cliniques évocatrices comme le caractère familial, l'âge précoce, le caractère pluri-focal, la localisation colique droite, l'association à d'autres tumeurs (notamment de l'endomètre). Or, il est actuellement bien établi que le

dépistage du syndrome de Lynch permet d'en réduire drastiquement la morbidité et la mortalité ¹.

Le laboratoire de génétique moléculaire du CHU de Nantes réalise en routine le test de dépistage du syndrome de Lynch.

I.1.2. Principe de l'analyse moléculaire HNPCC

Dans le syndrome HNPCC, la tumeur se caractérise dans 80% des cas par un phénotype moléculaire MSI+ (*MicroSatellite loci Instability*) encore appelée RER+ (*Replication Error*) correspondant à l'existence d'allèles surnuméraires résultant de l'inactivation de l'un des gènes MMR impliqués dans le processus de réparation des mésappariements de l'ADN (hMSH2 ou hMLH1 le plus souvent, hPMS1, MPS2 et hMSH6 beaucoup plus rarement).

Les microsatellites, petites séquences d'ADN non codantes réparties aléatoirement dans le génome, sont les témoins privilégiés d'un dysfonctionnement cellulaire au cours de la réplication, étape intervenant avant la division cellulaire et assurant le dédoublement de l'ADN. En effet, au cours de la réplication de l'ADN, se glissent fréquemment des erreurs d'appariement des bases induisant des erreurs de séquence au sein de l'ADN répliqué. Lors du processus normal dans la cellule, ces erreurs sont automatiquement corrigées par un ensemble de protéines codées par les gènes du système MMR. Lorsque ce système est défaillant, on retrouve très facilement au niveau des microsatellites répliqués, des erreurs se traduisant le plus souvent par un nombre de répétition des nucléotides non conservé (par rapport à la séquence d'origine). Ces variations de séquence, au niveau des microsatellites, identifiables par des techniques de biologie moléculaire (type PCR), se nomment instabilité microsatellitaire ou IMS.

Comme recommandé par la conférence de consensus de Bethesda ², le laboratoire du CHU de Nantes utilise cinq marqueurs (3 microsatellites et 2 répétitions de mononucléotides) situés dans les gènes de susceptibilité à différents cancers pour déterminer le phénotype RER de la tumeur. Ces 5 marqueurs sont les suivants :

- Bat 25 : répétitions de A situé dans un intron du proto-oncogène c-kit
- Bat 26 : répétitions de A situé dans un intron du gène hMSH2
- D5S346 : répétitions de CA situé en 5q21 dans le gène APC
- D17S250 : répétitions de CA localisé en 17q12 sans gène connu à proximité.
- D2S123 : répétitions de CA situé en 2p16 dans le gène hMSH2.

En janvier 2002, au CHU, lors de l'analyse d'une tumeur colorectale dans le cadre d'une suspicion de syndrome HNPCC, le profil microsatellite d'un patient présentait une perte d'hétérozygotie au niveau du marqueur D17S250. Si les délétions du marqueur D5S346, proche d'APC, sont relativement fréquentes dans les cancers colorectaux sporadiques, une perte d'allèle au niveau du marqueur D17S250 était observée pour la première fois dans l'expérience du laboratoire.

Ainsi, nous nous sommes intéressés à caractériser un éventuel gène suppresseur de tumeur dans cette région. Parallèlement, le service d'urologie s'intéressait également aux pertes d'hétérozygotie en 17q rapportées dans certains cas de cancers rénaux dans la littérature.

1.2. 17q12 et cancers papillaires rénaux

1.2.1. Les cancers rénaux

Les carcinomes rénaux cellulaires (RCC) forment un groupe hétérogène de tumeurs dans lequel on peut distinguer 4 sous-types présentant des différences histologiques et pathologiques :

- RCC à cellules claires représentant 70-80% de la totalité des cancers rénaux
- RCC de type papillaire (10-15%) nommés pRCC
- RCC chromophobes (< 5%)
- Carcinome du tube de Bellini (<1%), les plus rares

De plus, il existe une sous-catégorisation dans les RCC papillaires, proposée par Delahunt et Eble ³, ceux de type 1 possédant de petites cellules avec un large cytoplasme et un petit noyau, et ceux de type 2 présentant de grandes cellules au cytoplasme éosinophile avec un large noyau sphérique et un nucléole proéminent.

1.2.2. Etude de la littérature concernant les anomalies chromosomiques dans les pRCC.

Comme la plupart des cancers, le cancer du rein existe sous une forme familiale et sous une forme sporadique non héritée. Chaque type de RCC « hérité » est associé à un type histologique décrit ci-dessus. Par exemple, le RCC à cellules claires héréditaire est associé au

syndrome Von Hippel Lindau avec des délétions partielles ou des mutations dans le gène VHL. Néanmoins, même si dans quelques cas les gènes impliqués dans les formes familiales sont altérés, les gènes responsables de la carcinogenèse dans les cas sporadiques sont encore mal connus. Les RCC papillaires (pRCC) sont caractérisés par des trisomies des chromosomes 7 et 17 ainsi que par la perte du chromosome Y, et quelques gains additionnels⁴⁻⁵. En fait, les données récentes de la littérature sont contradictoires autant sur la caractérisation des pRCC que sur leur classement moléculaire, en sous-type 1 et 2.

En effet, classiquement, le chromosome 17 est tripliqué dans les pRCC, or plusieurs études portant sur des marqueurs microsatellites localisés en 17 p ou 17q montrent soit des pertes d'hétérozygotie (LOH *Loss of Heterozygosity*) soit des AI (*Allelic Imbalance*). L'AI comprend à la fois la perte d'hétérozygotie et le gain sélectif d'allèle.

La première étude datant de 1998, porte sur l'incidence de MSI ou de LOH sur 22 cas de pRCC avec 3 marqueurs sur le chromosome 17. Les résultats montrent que 8 patients présentent soit une MSI soit une LOH en 17q, indiquant la présence putative d'un gène suppresseur de tumeur à ce locus⁶.

La seconde étude menée par Salama et ses collègues, repose sur l'analyse de 7 patients par FISH (*Fluorescent in Situ Hybridation*). Ils n'ont pas mis en évidence de trisomie ni pour le chromosome 7 ni pour le 17, et ils constatent même une monosomie du 7 et du 17 dans 67% et 100% des cas étudiés, respectivement⁷.

Enfin, une série de tumeurs de type 1 (21 cas) et de type 2 (14 cas) de pRCC a été analysée pour la présence ou non d'une AI sur 11 chromosomes d'intérêt dont le bras long du chromosome 17. Les résultats indiquent que les tumeurs de type 1 présentaient une AI en 17q, bien plus fréquemment que les tumeurs de type 2 (78,6% versus 12,5%, $p = 0,006$)⁸.

Au vu de ces données plus ou moins contradictoires, nous avons décidé, en collaboration avec le service d'urologie, de tester un certain nombre de cas de pRCC afin d'évaluer leur statut en 17q.

I.3. 17q11-q23 : une région d'intérêt dans le cancer

La région d'intérêt porte sur la bande q11-q23 du chromosome 17. Le tableau 1 présente l'ensemble des gènes connus dans cette région. Les anomalies chromosomiques au niveau de ces loci sont communes à de nombreux cancers. La détection d'une perte ou d'un gain de matériel génomique peut conduire, respectivement à l'identification d'un gène suppresseur de tumeur ou d'un oncogène.

Le gène HER2 (ou ERBB2/cERB2/HER2/NEU), localisé dans la bande 17q12, est l'oncogène le plus fréquemment amplifié (15 à 30%) dans le cancer du sein ⁹ et il est également communément amplifié dans d'autres types de cancers ¹⁰⁻¹². Dans le cancer du sein, la région autour de HER2 contient d'autres gènes présentant un nombre de copie altéré ¹³. Une étude récente a permis d'identifier 5 gènes à proximité immédiate de HER2 : PSMB3, RPL19, PPARBP, PPP1R1B, et GRB7 ¹⁴. Un autre gène, TOP2A, adjacent à l'oncogène HER2, a été trouvé amplifié ou délété avec une fréquence identique dans le cancer du sein ¹³. Des données similaires existent pour les cancers gastriques, avec une co-amplification de plusieurs oncogènes potentiels incluant HER2, PPP1R1B, GRB7, et SARD3 ¹⁵.

Il existe également quelques études sur la région 17q12-q23, qui mettent en évidence la probable existence de gènes suppresseurs de tumeurs, impliqués dans des cas sporadiques de cancers. La perte d'hétérozygotie en 17q12-q21 a été bien documentée dans les tumeurs épithéliales ovariennes ¹⁶⁻¹⁷, dans le carcinome laryngé ¹⁸⁻²¹, dans le cancer de la prostate ²², et dans les tumeurs endocrines du pancréas ²³. Dans deux familles présentant des tumeurs de Wilms, un locus nommé FWT1 a été localisé par analyse de liaison à l'intérieur de la région 17q12-q21 ²⁴⁻²⁶. Malgré ces nombreuses observations, très peu de gènes suppresseurs de tumeur ont été caractérisés dans cette région chromosomique. L'équipe de Li et Chen, en 2004, a décrit un gène suppresseur de tumeur potentiel, nommé LCRG1 (*Laryngeal Carcinoma Related Gene 1*), localisé en 17q21 ¹⁹.

Nom du gène	Localisation	Identité/Fonction	Oncogène/ Gène suppresseur de tumeur
USP6 ou TRE2	17p13 ou 17q11	inconnue	Oncogène ²⁷
LASP1	17q11-q21	Protéine LIM SH3	Fusion avec MLL leucémie myéloïde ²⁸
TAF2N	17q11.1-q11.2	Liaison ADN ?	Oncogène ²⁹
<i>NF1</i>	<i>17q11.2</i>	<i>Protéine GTPase</i>	<i>Neurofibromatose Gène suppresseur de tumeur</i> ³⁰
THRA ou c-ERBA	17q11.2	Récepteur hormone thyroïdienne	Oncogène ³¹
RAR α	17q12	Récepteur nucléaire acide rétinoïque/ réponse hormonale	Oncogène Leucémie promyélocytaire/fusion PML ³²
<i>WT4 ou FWT1</i>	<i>17q12-q21</i>	<i>inconnue</i>	<i>Tumeur de Wilms</i> ²⁵
CCR7	17q12-q21.2	Récepteur couplé à la protéine G voie chémokine	Oncogène ³³
<i>BRCA1</i>	<i>17q21</i>	<i>Ubiquitine ligase</i>	<i>Gène suppresseur de tumeur</i> ³⁴
<i>PHB</i>	<i>17q21</i>	<i>Protéine chaperonne/activateur transcriptionnel</i>	<i>Gène suppresseur de tumeur</i> ³⁵
MLLT6	17q21	Régulateur transcriptionnel ?	Oncogène Fusion ALL-1 Leucémie ³⁶
<i>LCRG1</i>	<i>17q21</i>	<i>inconnue</i>	<i>Gène suppresseur de tumeur ?</i> ¹⁹
ETV4	17q21	Enhancer gènes famille métallo- protéases	Oncogène ³⁷
TOP2A	17q21-q22	Topoisomérase	Oncogène ³⁸
ERBB2	17q21.1	Récepteur tyrosine kinase (signalisation IL6)	Oncogène ³⁹
M17S2	17q21.1	inconnue	? régulation BRCA1
CA9	17q21.2	Anhydrase carbonique	Oncogène ⁴⁰
<i>NME1</i>	<i>17q21.3</i>	<i>Kinase nucléoside diphosphate</i>	<i>Gène suppresseur de tumeur</i> ⁴¹
<i>NME2</i>	<i>17q21.3</i>	<i>Kinase nucléoside diphosphate</i>	<i>Gène suppresseur de tumeur</i> ⁴¹
COL1A1	17q21.31-q22.05	Collagène	Ostéogenèse imparfaite ⁴²
<i>ABCC3</i>	<i>17q22</i>	<i>Transporteur S-gluthatione</i>	<i>Gène suppresseur de tumeur</i> ⁴³
BCL5	17q22	inconnue	Réarrangement MYC ⁴⁴
CD79B	17q23	Immunoglobuline	Oncogène ⁴⁵

Tableau 1. Présentation des gènes localisés en 17q11-q23.

Les gènes sont présentés suivant leur localisation sur le chromosome 17, du centromère au télomère. Les gènes suppresseurs de tumeur sont indiqués en italique.

II. Les acteurs de la cancérogenèse : Oncogènes, Gènes Suppresseurs de tumeur et autres.

II.1. Oncogènes et Gain de Fonction

Du point de vue de la nomenclature, on désigne par c-onc, un oncogène cellulaire présent dans le génome de la cellule cancéreuse, et par v-onc un oncogène viral. Dans le génome humain, les proto-oncogènes participent normalement à la prolifération et à la croissance cellulaire. Par simplification, on désignera ces proto-oncogènes sous le terme d'oncogène, et lorsqu'ils sont transformés et entrent dans le processus de cancérisation, sous le terme d'oncogènes activés.

II.1.1. Classification et fonction des oncogènes

Les oncogènes interviennent à différents niveaux dans la transmission du signal (quelque soit sa nature) qui déclenche ou maintient la prolifération cellulaire. Ces différentes étapes permettent de définir 5 grandes classes d'oncogènes :

- les facteurs de croissance,
- les récepteurs aux facteurs de croissance,
- les oncogènes liant le GTP,
- les transducteurs de signaux intracellulaires, les GEF (*Guanine Exchanging Factor*),
- les facteurs de transcription nucléaires, les composants du réseau cyclines/kinase dépendantes des cyclines/inhibiteurs de kinases.

II.1.2. Activation des oncogènes

La fonction des oncogènes est fortement liée à la voie de signalisation moléculaire des facteurs de croissance. Une perturbation de cette cascade métabolique peut s'exercer à tout niveau par des oncogènes activés, conduisant en règle générale à une amplification du signal de prolifération cellulaire transmis jusqu'au noyau⁴⁶⁻⁴⁷.

Les mutations activatrices des oncogènes sont quasiment toujours somatiques (à l'inverse des gènes suppresseurs de tumeur). Cette activation peut s'effectuer par mutation ponctuelle au niveau du proto-oncogène (gènes de la famille ras, gène RET), par délétion

(EGFR), remaniement de structure (fusion bcr/abl1), amplification (HER-2/neu), ou insertion activatrice d'une séquence promotrice d'origine virale à proximité d'oncogènes (c-myc et c-myb). En plus de ces caractéristiques, les oncogènes activés possèdent un phénotype dominant au niveau cellulaire, se traduisant par le fait qu'un seul allèle muté est suffisant pour obtenir un effet oncogénique par gain de fonction.

II.2. Gènes suppresseurs de tumeurs et perte de fonction

II.2.1. Définition et théorie de Knudson

Contrairement aux oncogènes activés pour lesquels un allèle muté suffit à produire une activité transformante, il faut que les deux allèles d'un gène suppresseur de tumeur soient perdus ou inactivés pour que le processus oncogénique se développe. Il s'agit de la théorie dite des « deux évènements » (« *two-hits* ») de Knudson⁴⁸. Cette hypothèse évoque la nécessité de la survenue de deux évènements mutationnels pour que le cancer se développe chez un sujet sans antécédent familial, alors que pour un sujet ayant des antécédents familiaux, une seule mutation peut suffire (Figure 1). Il a ensuite été proposé que ces deux évènements pourraient concerner les deux allèles d'un même gène.

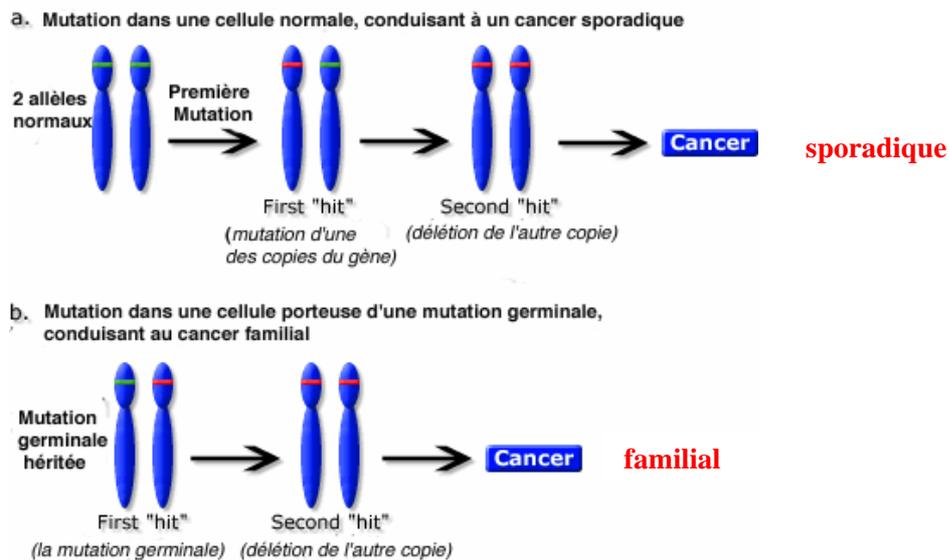


Figure 1. Illustration de la théorie de Knudson ou « Two-Hit hypothesis »

- mutation somatique et cancer sporadique
- mutation germinale et cancer familial

II.2.2. Classification et fonction des gènes suppresseurs de tumeur.

Il est très difficile de classer les gènes suppresseurs de tumeur. En effet, à la différence des oncogènes, le plus souvent ces gènes ne possèdent pas d'homologie avec des gènes de fonction connue, ou même entre eux, permettant de les regrouper par familles fonctionnelles.

Ces gènes peuvent avoir une fonction dans la transcription, le contrôle du cycle cellulaire, la réparation de l'ADN et l'apoptose. Pour cette raison, la compréhension du mécanisme d'action est passée par l'intégration du produit de ces gènes dans les voies métaboliques de la surface de la cellule au noyau.

Il s'agit souvent de gènes indispensables à la vie car les souris nullizygotés pour un gène suppresseur de tumeur ne sont pas viables en général (APC, NF2, WF1, RB1 et NF1). Un certain nombre de gènes suppresseurs de tumeur agissent comme régulateurs au niveau des points de contrôle du cycle cellulaire (RB1, p53).

II.3. Les autres acteurs de la cancérogenèse

II.3.1. Les gènes de l'apoptose

Les facteurs pro-apoptotiques (BAX ou BID) et anti-apoptotiques (BCL-2) assurent l'homéostasie cellulaire. Si BAX est en excès, la cellule exécute la commande programmée de mort cellulaire par l'intermédiaire de la voie des caspases et d'un dérèglement des mitochondries. Au contraire, un excès de BCL-2 inhibe ce programme et les cellules survivent.

II.3.2. Les gènes de détoxification

Les gènes de susceptibilité au cancer, impliqués dans les interactions entre le génome d'un individu et son environnement suscitent un intérêt grandissant pour l'identification des carcinogènes en cause et pour l'identification des individus à risque élevé.

Ces gènes codent principalement pour les enzymes de détoxification de la cellule et comprennent les cytochromes P450 (enzymes de phase I) et les Glutathion-S-Transférase ou GST (enzymes de phase II).

Pour pouvoir entraîner des lésions de l'ADN et initier le processus de transformation maligne, les carcinogènes subissent souvent une étape d'activation métabolique par des enzymes de phase I. Les enzymes de phase II vont conjuguer les carcinogènes obtenus (formes

électrophiles) à des ligands endogènes comme le glutathion réduit (GSH). Une fois liés, ces électrophiles vont devenir plus solubles dans l'eau et vont être éliminés par la voie des acides mercapturiques ce qui permet de contrecarrer leur toxicité. La formation de ce conjugué constituerait en outre une sorte de « signal moléculaire » rendant possible l'exportation hors de la cellule via des transporteurs membranaires ⁴⁹.

On comprend que l'absence ou le défaut d'activité enzymatique de l'un de ces gènes, dû au polymorphisme génétique, pourrait favoriser l'apparition de mutations somatiques d'autant plus que l'organisme dispose simultanément d'enzymes de phase I très actives capables de générer des intermédiaires électrophiles du métabolisme.

B. Le système ubiquitine

I. Modification ubiquitine et processus cellulaires

I.1. L'ubiquitine

L'ubiquitine (Figure 2) est le prototype d'une famille de protéines qui montrent de remarquables similarités de structure, mais des séquences variables.

L'ubiquitine est l'une des protéines les mieux conservée chez les Eucaryotes (les Procaryotes n'ont pas de molécule fonctionnellement analogue à l'ubiquitine). C'est un polypeptide de 76 acides aminés qui est présent en grande quantité dans toutes les cellules. Cette petite protéine existe sous forme libre, mais elle est destinée à être conjuguée, de façon post-traductionnelle, à une large variété de substrats.

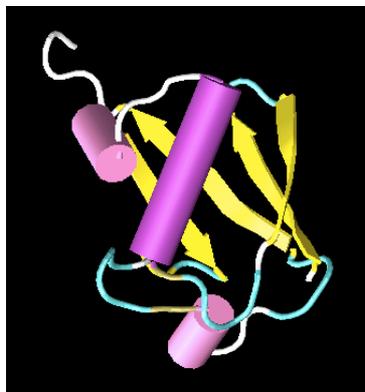


Figure 2. Représentation de la structure tertiaire de l'ubiquitine.

Les hélices α sont représentées par les cylindres violets et, les feuillets β par les flèches jaunes.

I.2. L'ubiquitination

L'ubiquitination des substrats est conduite par une cascade enzymatique hautement conservée⁵⁰: $E1 \rightarrow E2 \rightarrow E3$ (Figure 3).

- 1) Tout d'abord l'activation d'un résidu glycine en C-terminal de l'ubiquitine de façon ATP-dépendante par une enzyme E1 « Activating Enzyme » permet la formation d'un pont thioester avec un résidu cystéine d'E1.
- 2) L'ubiquitine est alors transférée sur un résidu cystéine d'une E2 « Conjugating Enzyme », également via un pont thioester.
- 3) Finalement, l'ubiquitine est transférée sur son substrat grâce à une E3 « ubiquitine ligase », qui permet la formation d'un lien isopeptide entre le résidu glycine de l'ubiquitine et le groupe amino- ϵ d'un résidu lysine du substrat.

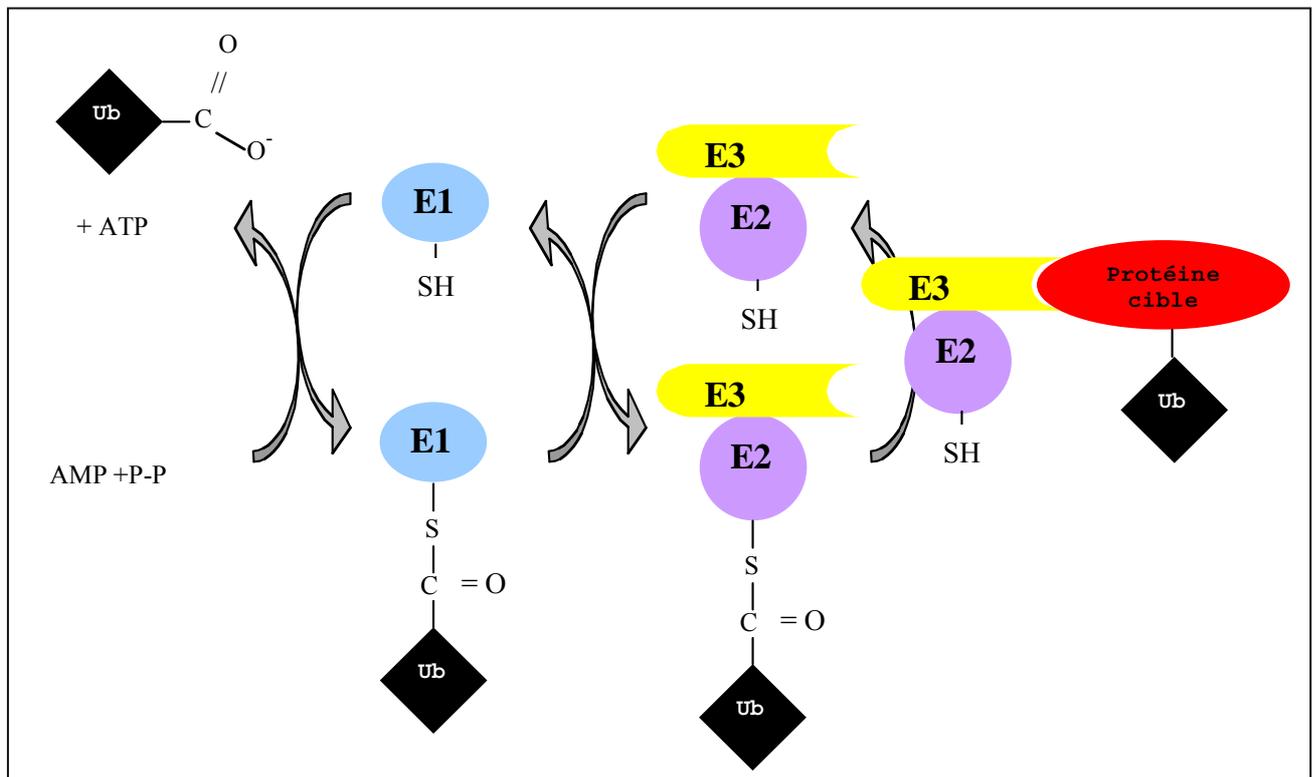


Figure 3. Réactions enzymatiques du système ubiquitine.

E1 : Enzyme activatrice, E2 : Enzyme conjuguatrice, E3 : ubiquitine ligase, Ub : ubiquitine

Bien que certaines E2s puissent ubiquitiner le substrat directement, dans la plupart des cas, c'est l'ubiquitine ligase E3 qui assure la reconnaissance et la conjugaison du substrat. La cascade d'ubiquitination est une hiérarchie de spécificité enzymatique. Chez la levure, une unique enzyme E1 interagit avec une paire d'E2s qui en retour peut se lier avec un grand

nombre d'E3s ⁵¹. Cette stratégie a probablement évolué pour optimiser le nombre de substrats à ubiquitiner tout en minimisant le nombre de composants du cœur du système. Ainsi, une stratégie impliquant un petit nombre de E1s et E2s couplés à une très grande famille d'E3s capitalise l'efficacité du système d'ubiquitination.

I.3. Conséquence sur le substrat ubiquitiné

Il existe trois types d'ubiquitination : mono-ubiquitination, multi-ubiquitination et poly-ubiquitination (Figure 4).

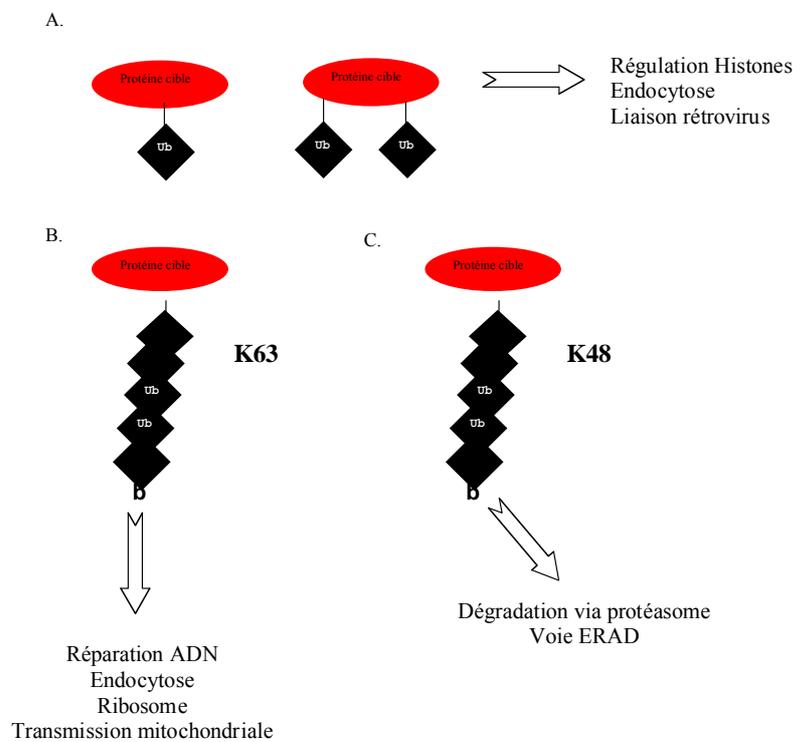


Figure 4. Les différents types d'ubiquitination et leurs conséquences

- (A) : Mono-ubiquitination sur une ou plusieurs lysines
- (B) : Poly-ubiquitination sur le résidu K63
- (C) : Poly-ubiquitination sur le résidu K48

I.3.1. Poly-ubiquitination K48

Dégradation par le protéasome 26S

La poly-ubiquitination via le résidu lysine 48 (K48), avec au moins 4 unités ubiquitine, cible la protéine modifiée pour sa dégradation par le protéasome 26S.

Ceci constitue le système de dégradation, essentiel et sélectif, des protéines cytosoliques et nucléaires chez les Eucaryotes. Cette machinerie dédiée à la dégradation des protéines intracellulaires est hautement conservée au travers de l'évolution. De multiples répétitions d'ubiquitination aboutissent à l'attachement d'une chaîne polyubiquitine où chaque résidu glycine C-terminal d'ubiquitine est attaché à un résidu lysine spécifique de la molécule d'ubiquitine précédente. Quelques réactions nécessitent une E4, de façon à stimuler l'ubiquitination ⁵².

Une fois ajoutées à leur substrat protéique, les chaînes polyubiquitine agissent comme un signal de dégradation de la protéine pour le protéasome 26S qui reconnaît spécifiquement les substrats portant un minimum de quatre molécules d'ubiquitine liées K48 ⁵¹ (Figure 5). L'action du protéasome 26S génère plusieurs types de produits : des peptides libres, de petits peptides encore liés à l'ubiquitine via leur résidu lysine et des chaînes poly-ubiquitine. Les deux derniers produits sont convertis en molécules d'ubiquitine libres et réutilisables par l'action d'hydrolases ubiquitine C-terminale ou d'isopeptidases. Quelques isopeptidases pourraient également dé-assembler certains conjugués protéine-ubiquitine. Ces enzymes de dé-ubiquitination ont été identifiées, certaines ont la capacité de réduire ou de prévenir la dégradation d'une protéine spécifique ⁵³⁻⁵⁴. Les petits peptides générés par ce processus peuvent alors être dégradés ultérieurement en acides aminés libres par des peptidases cytosoliques.

Les exceptions à cette voie sont les protéines non-ubiquitinées reconnues par le protéasome ⁵⁵.

Voie ERAD (ER-Associated Degradation)

En plus de son rôle dans la dégradation des protéines nucléaires et cytosoliques, la poly-ubiquitination sur le résidu K48 est impliquée dans la voie de dégradation associée au réticulum endoplasmique (voie ERAD). Cela constitue un mécanisme de contrôle qualité ciblant les protéines N-glycosylées mal conformées et permettant leur transfert rétrograde du réticulum endoplasmique au cytosol suivi par leur dégradation protéosomale ⁵⁶.

I.3.2. Poly-ubiquitination K63

La chaîne ubiquitine peut être constituée sur le résidu lysine 63 des molécules d'ubiquitine. Dans ce cas, la voie suivie par le substrat n'est pas celle du protéasome. Ce type de liaison a été largement étudié chez la levure où plusieurs études montrent que des souches K63Ub (c'est-à-dire bloquées pour l'assemblage de la chaîne ubiquitine via K63) présentaient des défauts dans la réparation de l'ADN post-réplivative ⁵⁷, dans la réponse au stress ⁵⁸, dans le partage de l'ADN mitochondrial ⁵⁹, dans l'endocytose de certaines protéines de la membrane plasmique ⁶⁰ et dans les fonctions ribosomales ⁶¹.

I.3.3. Mono-ubiquitination

Plusieurs protéines mono-ubiquitinées ont été identifiées, *in vitro* et *in vivo*, et il apparaît maintenant, que ce phénomène agit comme un régulateur de la localisation et de l'activité de protéines cellulaires diverses. La mono-ubiquitination est impliquée dans au moins trois fonctions cellulaires différentes : la régulation des histones, l'endocytose et la liaison des rétrovirus à la membrane plasmique ⁶².

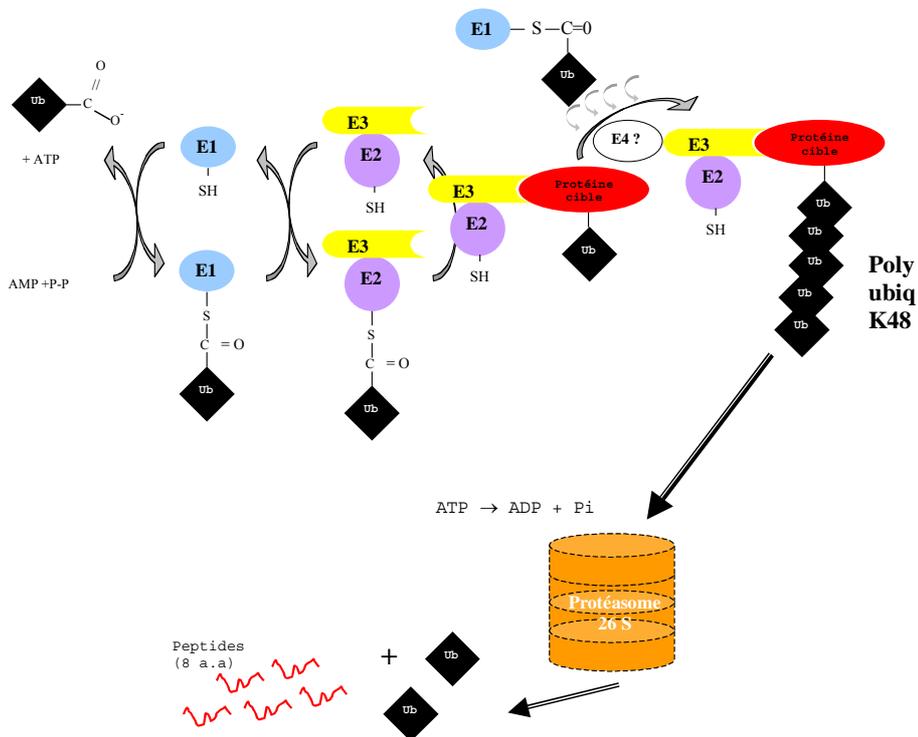


Figure 5. Représentation de la dégradation protéasomale ubiquitine dépendante.
 E1 : Enzyme activatrice, E2 : Enzyme conjugatrice, E3 : ubiquitine ligase, Ub : ubiquitine

II. Les ligases E3s : deux grandes classes

II.1. Définition

La définition des E3s dans la littérature reste confuse pour deux raisons principales :

- Tout d'abord, il existe une large variété de mécanismes par lesquels différents types de E3s réalisent la ligation protéine-ubiquitine. Dans certains cas, la protéine-cible est liée directement à une E3, tandis que dans d'autres, le substrat à ubiquitiner est lié à la ligase via une molécule adaptatrice. Le mécanisme de transfert de l'ubiquitine activée peut également être différent selon les types d'E3s. Ainsi, certaines E3s acceptent directement l'ubiquitine activée de la part d'une E2 et s'y lient sous forme d'un intermédiaire thiolester avant de la transférer sur la protéine-cible, tandis que d'autres E3s vont simplement faciliter, de façon conformationnelle, le transfert direct de l'ubiquitine activée d'E2 vers la protéine-cible (Figure 6).
- Il existe une absence d'homologie de séquence entre les différentes classes d'E3s. De plus, certaines E3s sont associées dans de grands complexes multi sous-unités, dans lesquels on ne sait pas toujours quel composant est responsable de l'activité ubiquitine ligase.

Nous retiendrons la définition proposée par Hersko et Ciechanover selon laquelle une E3 est une enzyme qui lie, directement ou indirectement, une protéine substrat spécifique et permet le transfert de l'ubiquitine, directement ou indirectement, à partir d'un intermédiaire thiolester jusqu'à une liaison amide avec les protéines ou les chaînes polyubiquitine⁵¹.

Actuellement, les E3s se répartissent en deux grandes classes majoritaires : les E3s à domaine HECT (*Homologous to E6AP Carboxy Terminus*) et les E3s à domaine RING (*Really Interesting New Gene*).

Notons qu'il existe une troisième classe, les E3s à domaine U-box (*UFD2 Homology*), plus petite et encore mal définie quant à son rôle de E3 ou de E4⁶³.

Les domaines HECT, RING et U-box présentent une propriété biochimique commune de liaison à l'enzyme E2.

L'activité ubiquitine ligase est également associée avec plusieurs autres protéines à domaine RING comme c-Cbl, Mdm2, Hrd1/der3, Ubr1, Brca1 et Bard1. Cependant, toutes les protéines contenant un domaine RING ne sont pas des enzymes E3s⁶⁴.

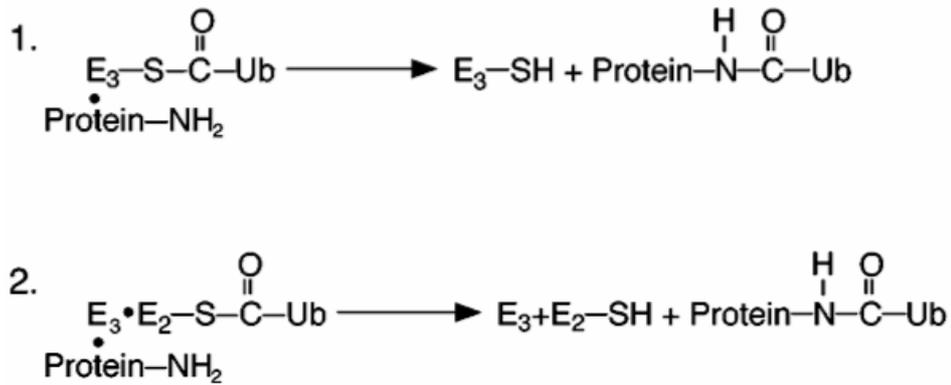


Figure 6. Différents mécanismes de transfert de l'ubiquitine sur son substrat.

1. Transfert direct (HECT): L'intermédiaire thiolester E3-ubiquitine est le donneur pour la formation de la liaison amide avec la protéine substrat
2. Transfert indirect (RING): L'E3-protéine substrat se lie à l'intermédiaire thiolester E2-ubiquitine E2 et l'ubiquitine est transférée directement de l'E2 sur la protéine cible.

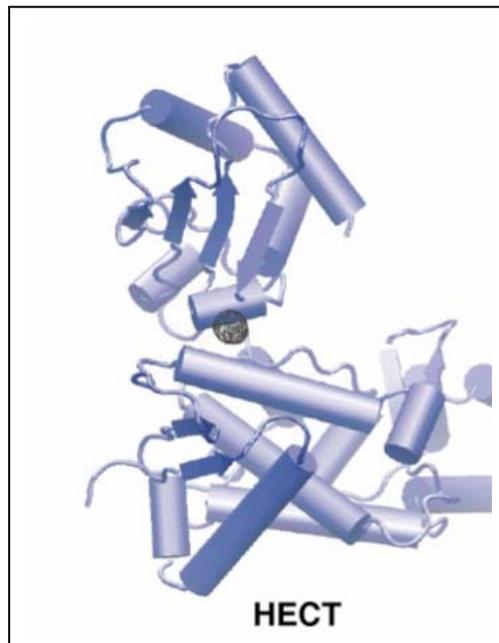


Figure 7. Structure tertiaire du domaine HECT.

II.2. Les ligases de type HECT

Les E3s portant un domaine C-terminal HECT de 350 acides aminés (*homologous to E6-AP- C-terminus*) constituent la classe HECT des enzymes E3s (Figure 7), qui transfère directement l'ubiquitine à son substrat via un résidu C-terminal cystéine hautement conservé⁶⁵. E6-AP, le membre le plus connu de ce groupe, est détourné par la protéine E6 du papillomavirus pour cibler de façon aberrante la protéine suppresseur de tumeur p53 vers la dégradation. De nombreuses protéines à domaine HECT ont été identifiées, comme RSP5, UFD4 chez *S.cerevisiae*, Pub1 chez *S.pombe* et NEDD4 chez l'homme. Les E3s à domaine HECT contiennent plusieurs domaines d'interaction protéique WW (ainsi nommés pour les deux résidus tryptophane conservés) qui reconnaissent des séquences riches en proline contenant un motif PY (XPPXY)⁶⁶. Par exemple, chez les mammifères, l'E3 HECT Nedd4 lie la séquence riche en proline du canal sodique épithélial humain (EnaC) via ses domaines WW, et le cible alors pour sa dégradation⁶⁷.

II.3. Les ligases de type RING

II.3.1. L'archétype : le complexe SCF

L'identification du domaine RING dans les E3 de type SCF (*Skp1-Cdc53-F-box protein*) et autres ligases non-HECT a suggéré que ce domaine intervient également pour ubiquitiner des protéines cibles. En réalité, il ne porte pas directement l'activité catalytique de transfert de l'ubiquitine, mais permet une conformation telle que l'E2 puisse transférer directement l'ubiquitine sur le substrat. Le domaine RING est un motif en doigt de zinc défini par un ensemble de huit résidus cystéine et histidine suivant le motif consensus : CX2CX (9-39)CX(1-3)HX(2-3)C/HX2CX(4-48)CX2C(4, 6, 12) (Figure 8). Ce domaine existe dans des centaines de protéines eucaryotes différentes⁶⁴.

On peut distinguer deux sous-types de motif RING :

- 7 cystéines et 1 histidine en position 4 = RING-HC
- 6 cystéines et 2 histidines en position 4 et 5 = RING-H2

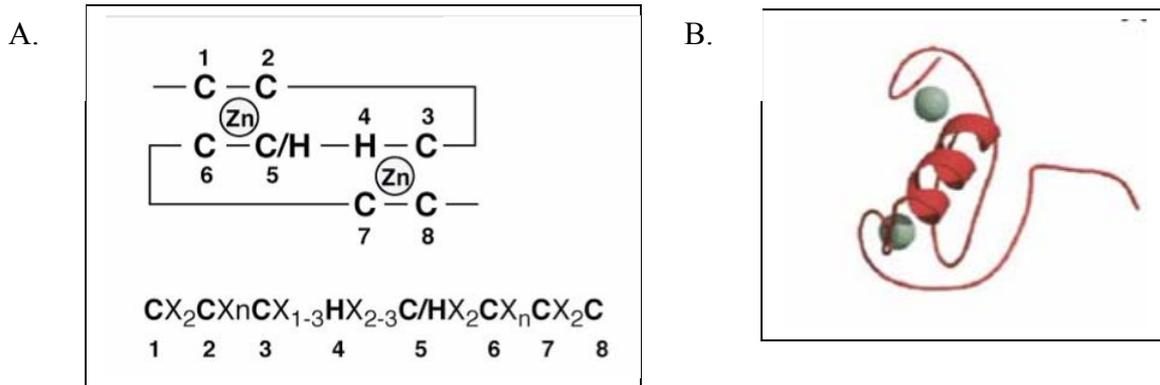


Figure 8. Le domaine RING des E3 ligases.

- A. Séquence consensus RING (X désigne n'importe quel résidu et n un nombre de répétition).
 B. Structure tertiaire. Les sphères grises représentent les atomes de zinc.

Le complexe SCF, sous-famille des E3s, a été historiquement découvert et étudié chez la levure *S.cerevisiae*. Le complexe SCF est composé d'au moins 4 unités : Cul1/Cdc53, Hrt1/Rbx1/Roc1, Skp1 et d'une protéine F-Box⁶⁸⁻⁷⁰. Toutes ces sous-unités connues sont conservées à travers l'évolution. La protéine RING Hrt1 et la culline Cul1 forment le cœur catalytique de l'ubiquitine ligase, qui se lie et active l'E2 Cdc34 ou Ubch5⁷¹⁻⁷². Hrt1 est une petite protéine RING de 121 acides aminés qui se lie directement à la culline et à Cdc34, vraisemblablement par le domaine CH. Skp1 se lie au cœur Cdc53/Hrt1 d'une part et d'autre part recrute parmi un large spectre, une protéine adaptatrice F-Box, qui assure la spécificité de substrat (Figure 9).

Le complexe SCF servirait comme un prototype pour d'autres ubiquitine ligases modulaires basées sur une culline⁷³⁻⁷⁴.

II.3.2. Les complexes « SCF-like »

De multiples cullines sont présentes dans toutes les espèces allant de 4 chez la levure jusqu'à au moins 9 pour les mammifères⁷⁵⁻⁷⁶. Étant donné que la famille de protéines Rbx lie tous les membres connus de la famille culline (excepté Cul4b), des équipes ont émis l'hypothèse que chacun puisse former des ubiquitine ligases⁷⁷⁻⁷⁸.

Il existe au moins 5 autres cullines chez l'homme, qui ne se lient pas à Skp1 (Figures 9 et 10).

EC₂S = EloC-EloB-Cul2-Rbx1-protéine SOCSbox

La culline Cul2 humaine se lie à la protéine ElonginC pour former un complexe du type Skp1-Cdc53/Cul1. Ce complexe se lie également à la protéine ElonginB et à une famille d'adaptateurs spécifiques de substrat contenant un motif SOCS-box (*Suppressor Of Cytokine Signaling*). La protéine SOCS-Box, produit du gène Von Hippel-Lindau (VHL), par l'intermédiaire des protéines Elongine B et Elongine C forme un complexe (Cul2-VBC) présentant une activité ubiquitine ligase ⁷⁹⁻⁸⁰.

Le gène VHL sous-régule les ARNms induits sous hypoxie en contrôlant la protéolyse des facteurs de transcription hypoxie régulés, HIF1 α et HIF2 α ⁸¹⁻⁸².

BC₃B = BTB-Cul3-Rbx1-BTB

La culline Cul3 a été identifiée par sa capacité à former une ligase E3 via sa liaison avec Rbx1 et une protéine contenant un domaine BTB. Les protéines BTB combine dans un seul polypeptide, à la fois un domaine adaptateur de type Skp1 et un domaine de reconnaissance du substrat ⁸³. Un tel complexe a été suspecté d'intervenir dans l'ubiquitination de la cycline E et dans le contrôle de la phase S des cellules de mammifères ⁸⁴⁻⁸⁵.

Autres complexes cullines

Trois autres cullines ont été partiellement caractérisées : Cul4, Cul5 et Cul7.

La protéine Cul4 humaine apparaît participer à de nombreux complexes ubiquitine ligases. Par exemple, elle régule la réparation de l'ADN par excision de nucléotide en ciblant pour la dégradation de Dbd2, une protéine de liaison à l'ADN endommagé par les UV ⁸⁶⁻⁸⁷.

La culline Cul5 semble se lier aux protéines SOCS-box (en plus de Cul2), notamment avec la protéine VHL ⁸⁸. Il existe au moins une protéine SOCS-box à domaine LRR, Muf1, qui ne se lie qu'avec la culline 5 ⁸⁹.

La culline la plus récemment identifiée, Cul7 forme une E3 « SCF-like » avec les protéines Roc1, Skp1 et les protéines F-Box Fbxw6/Fbxo29 ⁷⁷. Bien que les cibles de ces complexes ne soient pas connues, une première étude montre que des souris Cul7^{-/-} meurent à la naissance avec de sévères déficiences cardiaques ⁹⁰.

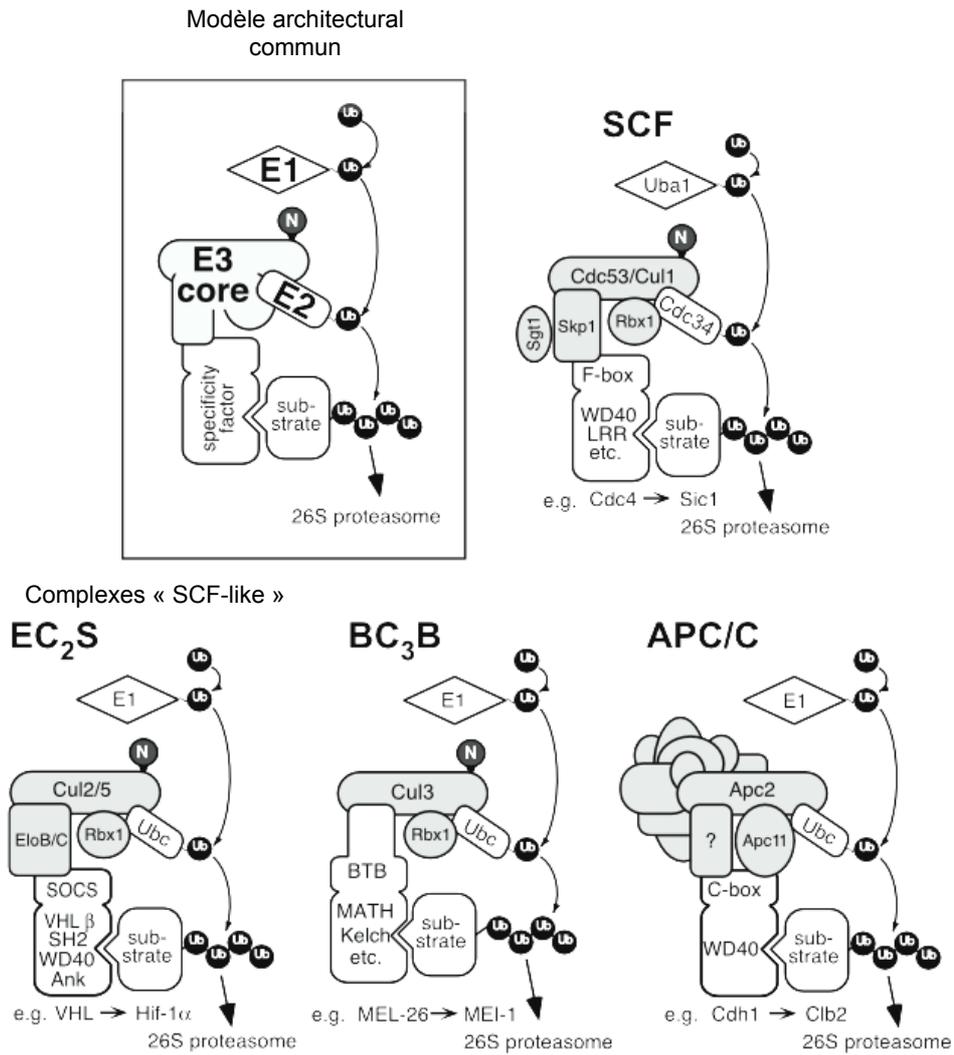


Figure 9. La super famille des ubiquitine ligases de type RING.

L'architecture générique culline-RING H2 est présentée en haut à gauche, le complexe SCF en haut à droite, et en bas sont présentés trois complexes « SCF-like ».

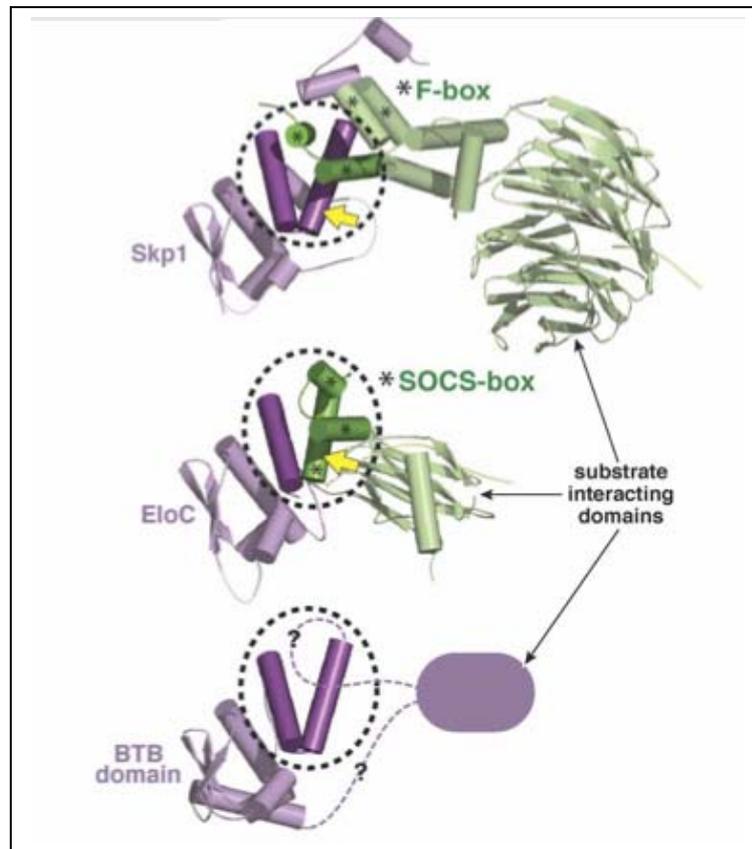


Figure 10. Similarité de structure des domaines de liaisons à la sous-unité E3.
(cercles pointillés) des différentes molécules adaptatrices spécifiques du substrat (F-box, SOCS-box,...) dans les différents complexes de type SCF.

APC/C (Anaphase Promoting Complex/Cyclosome)(Figure 9)

Une autre protéine E3 importante de type RING est le complexe de promotion de l'anaphase ou APC/C qui est nécessaire à la transition métaphase/anaphase et à la sortie de mitose ⁹¹. APC/C est une ubiquitine ligase multimérique qui consiste en au moins 12 protéines, incluant un membre de la famille culline APC2 et une protéine RING-H2 APC11. Comme SCF, APC/C cible différentes classes de substrats via une association avec une des au moins trois sous-unités variables de protéines adaptatrices contenant un motif WD40 nommées CDC20/Fzy, CDH1/HCT1/Fzr et AMA1 ⁹².

Les variants de sous-unité du coeur SCF servent ainsi à élaborer des fonctions « SCF-like » par des interactions combinatoires avec différentes classes de protéines adaptatrices.

Complexes non-cullines

Une nouvelle classe d'ubiquitine ligase (E3) a été récemment découverte. Elle est basée sur l'association de Skp1, d'une protéine F-box (Ebi), d'une protéine à domaine RING non culline (Siah), et d'une protéine de type Sgt1 (SIP) ⁹³⁻⁹⁴. Dans les cellules humaines, Ebi reconnaît et ubiquitine la β caténine en parallèle de la voie phosphorylation dépendante SCF ^{β T_{rcp}}. Puisque l'expression de Siah est induite à la suite d'un stress génotoxique de façon p53 dépendante, ce complexe pourrait jouer un rôle dans l'inhibition de la prolifération cellulaire dans ces conditions ⁹³⁻⁹⁴.

C. Les Protéines F-box

I. Définition et Classification

I.1. Historique : l'hypothèse F-box

Les premières découvertes proviennent de l'étude de mutants de la division cellulaire (cdc) chez la levure *S.cerevisiae* : cdc34, cdc4 et cdc53. Ces mutants arrêtent leur cycle cellulaire avec un ADN non répliqué et de nombreux bourgeons à la température non permissive car ils sont défailants pour la dégradation de l'inhibiteur de kinase cycline-

dépendant Cdc28 de type b nommé Sic1, lequel empêche l'entrée en phase S ⁹⁵. Sic1 est dégradé lorsqu'il est phosphorylé par la cycline Cdc28 en fin de phase G1 ⁹⁶.

D'autres analyses chez la levure révèlent que Cdc53, Cdc4 et deux autres protéines, Skp1 et Rbx1, forment un complexe ubiquitine ligase E3 qui agit de concert avec l'E2 Cdc34 pour réguler la transition de phase G1/S ⁶⁸⁻⁹⁷.

La première identification de Skp1 a été faite à l'intérieur d'un complexe comprenant la cycline A-Cdk2 humaine et une protéine nommée Skp2 ⁹⁸. Skp1 est alors isolée indépendamment de son interaction génétique et physique avec yCdc4, yCtf13 et hCyclineF ⁹⁹.

Un alignement de séquence protéique entre les protéines Skp2, Cdc4 et Cycline F fait apparaître un motif conservé de 40 résidus servant de site de liaison avec Skp1 : le domaine F-Box ⁹⁹. Cette observation conduit à « **l'hypothèse F-Box** », qui postule que Skp1 permet de lier des protéines contenant ce domaine au cœur du complexe d'ubiquitination et qu'en retour, les protéines F-Box recrutent les substrats pour la protéolyse ubiquitine médiée via leur domaine d'interaction protéine-protéine spécifique, comme WD40 (répétitions de résidus tryptophane) pour Cdc4 ou LRR (*Leucine rich repeat*) pour Skp2 ⁹⁹.

Le premier lien direct entre Cdc53 et la dégradation protéique avait été mis en évidence par la découverte que Cdc53 interagit avec une fraction de la cycline G1, Cln2 dont l'instabilité et l'ubiquitination dépendaient de Cdc53 *in vivo* ¹⁰⁰. L'interaction physique et génétique entre Cdc4, Cdc34 et Cdc53 suggérait alors que ces composants pouvaient former un complexe E2-E3 ⁹⁸⁻¹⁰⁰. Au même moment, une étude chez *C.Elegans* montrait qu'un orthologue de Cdc53, Cul1 était un régulateur négatif de la division cellulaire, suggérant que Cdc53 jouait un rôle conservé dans la dégradation protéique et le contrôle du cycle cellulaire ¹⁰¹. Les études de différentes équipes ont démontré des interactions directes entre Skp1, cdc53, cdc34 et des protéines F-Box variées. Elles ont également reconstitué une activité ubiquitine ligase spécifique contre la protéine Sic1 phosphorylée ⁶⁹⁻⁷⁰.

Ces complexes spécifiques sont alors désignés SCF = Skp1-Culline-F-Box protéine, avec la protéine F-Box spécifique indiquée en exposant SCF^{Cdc4}, SCF^{Grr1} ...

I.2. Caractéristiques du motif F-box

Le motif F-box est le domaine de liaison des protéines F-box à la protéine Skp1. Il est composé d'environ 50 acides aminés. La séquence consensus (Figure 11A) proposée par

Kipreos et Pagano a été obtenue en alignant 230 séquences pour créer un profil F-box dans la base de données Pfam (<http://pfam.wuslt.edu/cgi-bin/getdesc?name=f-box>)⁷⁴. Au regard ce profil, on s'aperçoit qu'il existe très peu de positions invariantes. Les positions les plus conservées sont les positions 8 (92% des 234 séquences alignées montrent une leucine ou une méthionine), 9 (92% proline), 16 (86% isoleucine ou valine), 20 (81% leucine ou méthionine), et la position 32 (92% sérine ou cystéine).

La structure secondaire du motif F-box contient trois hélices α principales H1, H2 et H3 et une hélice α H0. (Figure 11B).

Le motif F-box se lie à Skp1 à travers une interface bipartite. Une des faces de la F-box interagit avec les hélices H5 et H6 du cœur BTB/POZ et avec l'hélice H7 de Skp1, tandis que la face opposée interagit avec l'hélice H8. La figure 11C montre la modélisation de l'interaction Skp1-protéine F-box pour deux membres différents de la famille F-box : Skp2 et β Trcp1. Le mode de liaison reste très similaire bien que ces deux protéines n'aient pas le même domaine d'interaction pour le substrat¹⁰².

1.3. Le domaine d'interaction spécifique du substrat

1.3.1. Nomenclature et Classification

Le nom des protéines F-box a été donné par l'équipe de Bai sur la base de la présence du motif d'interaction protéique présent dans la cycline F⁹⁹.

Les protéines F-box sont définies comme des adaptateurs entre protéines. Elles contiennent deux motifs d'interaction protéique : le domaine F-Box, plus fréquemment situé en N-terminal et un autre motif présent le plus souvent en C-terminal dont les deux plus fréquents chez l'homme sont des répétitions riches en leucine (LRRs) et les répétitions de tryptophane (WD40).

La nomenclature pour les protéines F-box humaines a été proposée par le HUGO (*HUMAN Genome Organization*) suivant le profil proposé par les équipes de Cenciarelli¹⁰³ et de Winston¹⁰⁴ :

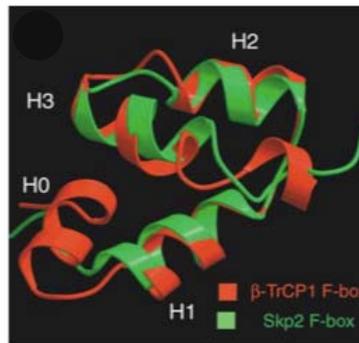
- FBXL pour une protéine contenant une F-Box et un domaine LRR.
- FBXW pour une protéine contenant une F-Box et un domaine WD40.
- FBXO pour une protéine contenant une F-Box et un autre domaine différent des deux précédents, ou non identifié.

A.

```

k P F P L L R L P e E I L r K I L e k L D P i D L L r L R K V S K K W R s L V D s l n i w f k f I e
s s s I s d m l K l i k e v f k h M p f k E R f n F s l t C R R F K r i i k k k f k i r k L l
r f n i d v n i r r s l i i k f l n l q l r d l f k d
a s y e i v s t r
q
    
```

B.



C.

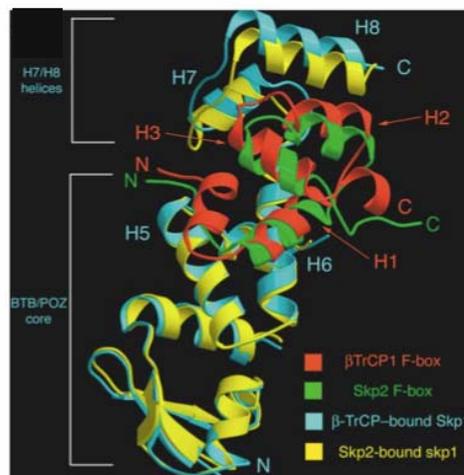


Figure 11. Caractéristiques du motif F-box.

(A) Séquence consensus du motif F-Box.

Celle-ci a été déterminée à partir de l'alignement de 234 séquences utilisées pour créer le profil F-box dans la base Pfam (<http://pfam.wuols.edu/cgi-bin/getdesc?name=F-box>.) Les résidus en majuscules, gras et soulignés sont trouvés dans plus de 40% des séquences F-box, en gras seulement dans 20 à 40%, en minuscules et gras dans 15 à 19%, et les minuscules uniquement dans 10 à 14% des F-Box ⁷⁴.

(B) Structure tri-dimensionnelle du motif F-box ¹⁰²

(C) Modélisation de l'interaction Skp1/Skp2 et Skp1/beta-Trcp via le motif F-box ¹⁰².

En 2004, Jin *et al* ont réalisé une analyse systématique des protéines F-box présentes dans la banque de données GenBank (pour les mammifères) ¹⁰⁵. Cette analyse a permis d'identifier 68 gènes humains et 74 gènes de souris codant pour des motifs F-box reconnaissables (Figure 12).

La phylogénie des seules séquences contenant le domaine F-box, qui est le meilleur marqueur d'une signature évolutive de la famille, montre deux groupes majeurs de protéines F-box, révélant ainsi une divergence d'évolution. Les différents domaines d'interaction protéiques (LRR, WD40 ou autres) sont répartis dans les deux groupes indiquant un domaine similaire, mais éliminant l'hypothèse que tous les membres de la sous-famille FBXW, par exemple, aient un ancêtre commun.

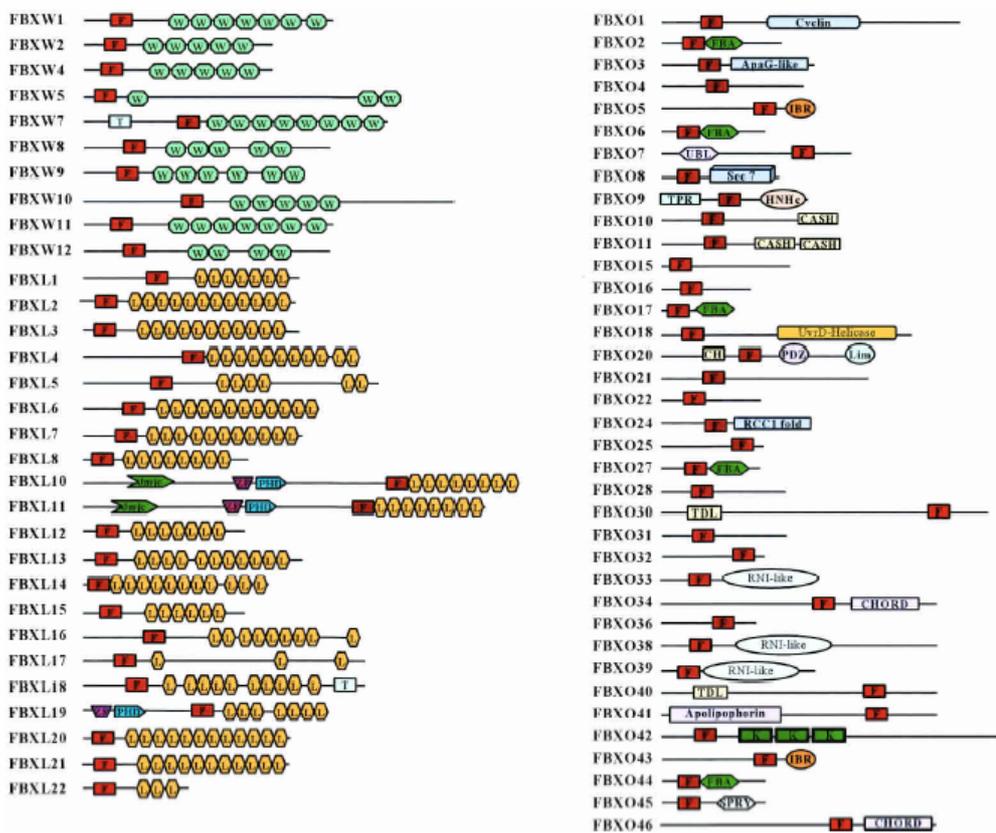


Figure 12. Structures des domaines de protéines F-box de mammifères ¹⁰⁵.

Les domaines identifiés par les algorithmes des programmes Pfam ou SMART incluent le motif F-box (F)(rouge), les répétitions WD40 (WD)(vert clair), les répétitions riches en leucine (L)(jaune), les domaines transmembranaires (T), le domaine associé F-box (FBA), le domaine « between-ring » (IBR), le domaine de protéines de liaison au carbohydrates et des hydrolases de sucre (CASH), les répétitions kelch (K), le domaine d'homologie calponine (CH), le domaine trouvé dans la famille des métalloenzymes cuppine (Jmjc), le domaine présent dans PSD-95, Dlg et ZO-1 (PDZ), le domaine de liaison au zinc trouvé dans Lin-11, Isl-1 et Mec-3 (Lim), la famille des nucléases HNH (HNHc), le nouveau domaine eucaryote de liaison au zinc (CHORD), et les répétitions de tétratrico peptides (TPR).

Les domaines suivants ont été trouvés avec la base de données SCOP (*Structural Classification of Proteins*) : ApaG-like, Apolipophorin, Ubl, TDL, RNI-like et RCC1 ¹⁰⁵.

1.3.2. Signaux protéiques déclenchant l'ubiquitination

La reconnaissance du substrat par une ligase E3 repose sur son niveau de modification. Ce dernier peut être dirigé par phosphorylation ou glycosylation dans le cas d'une reconnaissance par une protéine F-Box.

Phosphorylation

La phosphorylation est le type de modification de substrat le plus courant pour la reconnaissance par une ligase E3, et plus particulièrement par les complexes de type SCF.

Par exemple, tous les substrats connus de la protéine F-box β Trcp contiennent une similarité forte avec le motif canonique DGSXXS, identifié à l'origine dans I κ B α , à l'intérieur duquel 2 résidus sérine doivent être phosphorylés pour une reconnaissance efficace¹⁰⁶. Ce sont les kinases IKK (*I κ B Kinase*) pour I κ B α et Gsk3 pour la β caténine qui régulent leur dégradation phosphorylation-dépendante via SCF ^{β Trcp}.

De la même façon, la protéine F-box Skp2 reconnaît la protéine p27 et la cible pour sa dégradation, uniquement lorsqu'elle est phosphorylée sur son résidu thréonine 187¹⁰⁷.

Le complexe SCF^{Fbw7} se lie et ubiquitine efficacement les formes phosphorylées (Thr380) de la cycline E. Néanmoins, bien que l'interaction Fbw7-cycline E dépende de cette phosphorylation, la phosphorylation sur un autre site, Thr 62, influence directement ou indirectement la reconnaissance de la cycline E. En fait, le résidu Thr380 est phosphorylé par Gsk3 en combinaison avec Cdk2, celle-ci phosphorylant par ailleurs le résidu Ser384, qui facilite ainsi la phosphorylation de Thr380 par Gsk3¹⁰⁸.

De plus, la multiplicité des signaux ciblant les protéines vers une dégradation via l'ubiquitine, est soulignée par le fait que la phosphorylation de certaines protéines (sur des sites particuliers) les protège contre leur dégradation. Ainsi, la phosphorylation du résidu sérine 3 du proto-oncogène c-mos supprime son ubiquitination et sa dégradation¹⁰⁹.

À travers les différents profils de phosphorylation programmant différentes protéines pour leur dégradation, il apparaît que l'événement régulateur initial de la protéolyse ubiquitine médiée est porté par une protéine kinase, tandis que le rôle de l'ubiquitine ligase E3 est de reconnaître les formes phosphorylées des protéines cibles.

Glycosylation

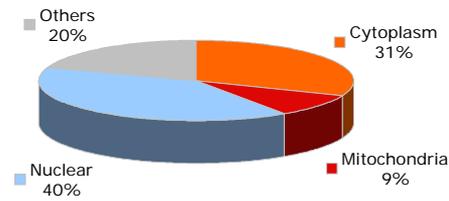
Au contraire des autres systèmes classiques de reconnaissance de substrat, la protéine F-box Fbxo2 (ou Fbs2) reconnaît des substrats N-glycosylés¹¹⁰. Elle se lie, et ubiquitine spécifiquement des protéines portant dans leur partie N-terminale, des oligosaccharides riches en mannose. Les substrats connus du complexe SCF^{Fbx2} sont des protéines de la voie ERAD (pré-intégrine β 1, Cfr et la sous-unité α du TCR), suggérant ainsi un rôle pour Fbxo2 dans le contrôle de la qualité des protéines à travers l'élimination des glycoprotéines plasmatiques aberrantes (mal assemblées ou mal conformées)¹¹¹.

II. Localisation cellulaire

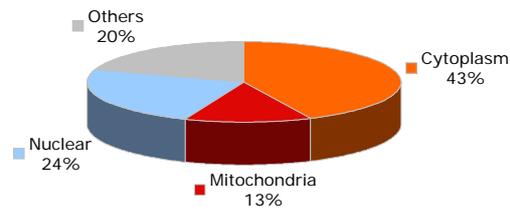
Il existe un nombre limité d'études sur la localisation sub-cellulaire des protéines F-box, et dans presque tous les cas, ces analyses ont été effectuées par surexpression de protéines marquées¹⁰³⁻¹⁰⁴. La distribution identique entre les constructions protéiques avec ou sans motif F-box, démontre que la présence du domaine et l'interaction avec Skp1 ne déterminent pas leur localisation cellulaire.

Nous avons réalisé une analyse systématique de l'ensemble des protéines F-box humaines présentes dans la base de donnée NCBI, pour la prédiction de localisation cellulaire via le programme IPSORTII. Les résultats sont présentés dans la figure 13. Ainsi, parmi les trois sous-familles de protéines F-box, la plupart sont nucléaires ou cytoplasmiques, mais il existe une part non négligeable située dans d'autres compartiments cellulaires, comme la mitochondrie (9 à 13 % selon la famille).

Famille
FBXLs (34)



Famille
FBXWs
(13)



Famille
FBXOs (41)

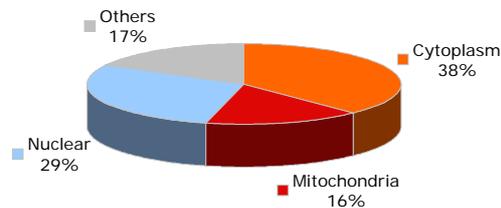


Figure 13. Répartition de la localisation sub-cellulaire des trois classes de protéines F-box selon l'analyse de leur séquence protéique via le programme IPSORTII.

L'analyse a porté sur 88 protéines F-box humaines au total dont 41 Fbxos, 34 Fbxls et 13 Fbxws.

Dans la catégorie « others » sont regroupés les compartiments cellulaires suivants : golgi, cytosquelette, peroxyosome, réticulum endoplasmique, membrane plasmique, voies de sécrétion, extracellulaire.

III. Régulation

Comme les autres protéines, les F-box sont régulées par différents mécanismes et à plusieurs niveaux. Dans cette partie, nous ne citerons que l'exemple de la double régulation de la protéine F-box Skp2, car il en existe trop pour pouvoir les décrire dans leur intégralité.

Bien que la quantité des composants Skp1, Cul1 et Rbx1 du système SCF^{Skp2} ne varie pratiquement pas durant le cycle cellulaire, le taux de Skp2 varie significativement. Skp2 est d'abord détectable pendant la transition G1-S, s'accumule pendant les phases S et G2, puis disparaît lors de la phase M. Ce sont à la fois des régulations transcriptionnelles et post-traductionnelles qui produisent ce profil cycle-dépendant.

III.1. Régulation transcriptionnelle

Des stimuli mitogènes extracellulaires induisent la sortie de quiescence (Go) des cellules et l'entrée en phase G1. L'activation de la voie Ras-Raf-MAPK augmente l'activité des promoteurs activateurs GADB dépendants¹¹². L'équipe d'Imaki, par une analyse du promoteur de SKP2, a identifié une séquence de 105 pb essentielle pour l'activité transcriptionnelle de SKP2. Cette zone contient un motif CACTTCCG, similaire au site de fixation de la protéine GA-binding (GABP). De plus, la surexpression de GADP augmente l'activité du promoteur SKP2, tandis que la suppression par ARN interférence de GABP provoque une diminution de l'activité du promoteur SKP2. Ainsi, la liaison cycle-dépendante de GABP sur le promoteur SKP2 régule l'expression de SKP2 durant le cycle cellulaire et plus particulièrement de la phase G1 à la phase S¹¹³.

III.2. Régulation post-traductionnelle

Le taux de protéine Skp2 est faible pendant la phase G1 du cycle cellulaire, période pendant laquelle le complexe APC/C est actif. De plus, Skp2 contient un motif canonique D-box ou *Destruction-box* (R-X-X-L-X-X-X-X-N/D/E), domaine toujours présent dans les cibles d'APC/C¹¹⁴. Le motif D-box est également connu dans d'autres substrats d'APC/C, comme la cycline B par exemple, comme nécessaire pour leur poly-ubiquitination par APC/C

mais pas pour leur liaison à APC¹¹⁵. Récemment, deux équipes (Wei *et al* et Bashir *et al*) ont montré que la disparition de Skp2 en phase G1 était due à sa poly-ubiquitination par APC/C (et son co-activateur Cdh1) de façon D-box spécifique¹¹⁶⁻¹¹⁷.

Ainsi, la régulation de Skp2 est multiple mais coordonnée de façon à ce que la protéine soit présente et active uniquement dans les phases du cycle cellulaire où elle joue un rôle prépondérant.

IV. Quelques fonctions des protéines F-box

Presque toutes les protéines F-box connues entrent dans la composition d'une ubiquitine ligase de type SCF. Elles sont responsables de la reconnaissance du ou des substrat(s) de manière spécifique et phosphorylation dépendante. Les protéines cibles des protéines F-box sont le plus souvent ubiquitinées avec une chaîne polyubiquitine type K48 c'est-à-dire destinées à être dégradées par le protéasome 26S. Les fonctions décrites ci-dessous, de manière non exhaustive, appartiennent à cette voie protéolytique spécifique.

IV.1. F-box et cycle cellulaire

La progression du cycle cellulaire eucaryote est contrôlée par une série de kinases cycline-dépendantes (CDKs). L'activité de ces enzymes est régulée par plusieurs mécanismes incluant l'association avec des sous-unités régulatrices (cyclines), la phosphorylation et déphosphorylation, et l'interaction avec des inhibiteurs de CDK (CKIs)¹¹⁸. Les taux de cyclines, CKIs et autres régulateurs du cycle cellulaire oscillent pendant le cycle comme le résultat d'une protéolyse périodique¹¹⁹. La voie protéolytique ubiquitine-protéasome conduit la dégradation de telles protéines régulatrices à courte durée de vie et ainsi contrôle leur concentration intracellulaire¹²⁰⁻¹²¹. La rapidité et la spécificité de substrat pour la dégradation protéique générée par cette voie est cohérente avec un tel rôle. Bien qu'il existe des similarités de structure entre les cyclines, ces protéines sont dégradées par la voie ubiquitine-protéasome à différents stades du cycle cellulaire, suggérant que la machinerie d'ubiquitination pour chaque cycline est distincte et hautement spécifique. De plus, les effecteurs protéolytiques sont eux-même réprimés par protéolyse dans un ensemble complexe de boucles rétroactives de régulation.

Les quatre complexes SCF^{F-box} connus dans la régulation du cycle cellulaire chez les mammifères sont SCF^{Skp2}, SCF^{Fbw7}, SCF^{βTrcp} et SCF^{Tome-1}. Il y a également un complexe « SCF-like » : APC/C, une protéine F-box Emi-1 n'appartenant pas à un complexe SCF (exception).

Phase G1 :

Les protéines p21 et p27 sont des inhibiteurs clés de Cdk1 et Cdk2. SCF^{Skp2} est responsable de la dégradation ubiquitine-dépendante de p27¹⁰⁷⁻¹²² et p21⁷⁶⁻¹²³ lors du point de contrôle G1/S. D'une part, ce complexe permet la transition en phase S en dégradant ces inhibiteurs de CDKs, et d'autre part, il permet de maintenir des activités Cdk1 et Cdk2 en maintenant un environnement cellulaire sans p21 et p27 dans la suite du cycle (Figure 14 a et b).

Une autre ligase E3, APC/C, agit lors de cette phase. APC/C^{Cdh1} maintient l'atténuation de Cdk1 et Cdk2 lors de la phase G1 en ciblant simultanément Skp2¹¹⁶⁻¹¹⁷ et son cofacteur Cks1 pour leur dégradation, et continue la dégradation des cyclines A et B initiée pendant la mitose. La cycline A se lie à Cdk1 et Cdk2, tandis que la cycline B active uniquement Cdk1. La première action d'APC/C^{Cdh1} préserve ainsi l'inhibition de Cdk1 et Cdk2 par p27 et p21, et la seconde maintient des taux faibles de cyclines A et B pendant la phase G1. De plus, APC/C^{Cdh1} dégrade la phosphatase Cdc25a (activatrice de Cdk1) pendant cette même période.

Phases G1-S, S et G2.

Un changement dans la régulation de Cdk1 et Cdk2 impliquant deux autres protéines F-box survient pendant la transition G1-S et les phases S et G2 (Figure 14 b). D'abord, la protéine F-box Emi1 s'accumule du fait de sa transcription induite par le facteur de transcription E2F (lequel induit également la transcription des gènes codant les cyclines A, B et E)¹²⁴. Cette accumulation inhibe, de façon non protéolytique, le complexe APC/C^{Cdh1}¹²⁵. Ceci provoque simultanément la stabilisation de Skp2, et des cyclines A et B (puisqu'elles ne sont plus dégradées par APC/C^{Cdh1}). Comment le taux de Cdk1 est-il plus faible alors que ses sous-unités activatrices cycline sont à la fois transcrites et non dégradées ?

Pour que l'activité Cdk1 soit maximale, la phosphatase Cdc25a doit être active pour contrecarrer la phosphorylation inactivatrice de Cdk1 via la kinase Wee1.

Le complexe SCF^{βTrcp} dégrade Cdc25a durant les phases S et G2¹²⁶⁻¹²⁷. Ceci permet d'expliquer le maintien de l'activité Cdk2 et la baisse de l'activité Cdk1. En effet, d'une part, les activités Cdk2-cyclineE et Cdk2-cyclineA sont maintenues par l'action de Skp2 sur p27 et

p21 et de Emi1 sur APC/CCdh1, et d'autre part, une faible activité de Cdk1 est maintenue via SCF^{βTrcp} en dépit du fait que les cyclines A et B ne soit plus déstabilisées par APC/C^{Cdh1} et que p21 et p27 soient dégradées par Skp2.

Transition G2-M.

La baisse du taux de Cdk2 est maintenant effectuée par la dégradation via SCF^{Fbxw7} de son cofacteur cycline E¹²⁸⁻¹³⁰, bien qu'une activité Cdk2 résiduelle subsiste par la présence de cycline A. L'activité Cdk1 est, quant à elle, maintenue par l'inhibition continue d'APC/C^{Cdh1}, celle APC/C^{Cdc20}, par Emi1, mais aussi par la re-stabilisation de Cdc25a qui n'est plus dégradée par SCF^{βTrcp} (Figure 14 c). Ce complexe continue pourtant à jouer un rôle pendant cette phase à travers la dégradation de l'inhibiteur de Cdk1, Wee1¹³¹. Citons également la protéine F-box Tome1, antagoniste de Wee1, qui apparaît promouvoir l'activité Cdk1¹³².

Mitose.

En fin de mitose, les taux de Cdk1 et de Cdk2 sont faibles. Le blocage d'APC/C est levé par la dégradation d'Emi1 par SCF^{Fbxw7}¹³³⁻¹³⁴. Le complexe APC/C^{Cdc20} est maintenant actif, et en plus de promouvoir l'induction de l'anaphase par la destruction de la sécurine, il induit activement la dégradation des cyclines A et B, réduisant ainsi l'activité Cdk1 (Figure 14 d). Cdk1 active APC/C en phosphorylant plusieurs de ses sous-unités, contribuant ainsi à sa propre régulation¹³⁵.

En résumé, Skp2 est un activateur de Cdk1 et Cdk2, Fbxw7 est un inhibiteur de Cdk2, βTrcp contribue à l'élimination de Cdk1 en phase S et G2, à son activation à la transition G2-M et enfin, à son élimination de nouveau à la sortie de mitose.

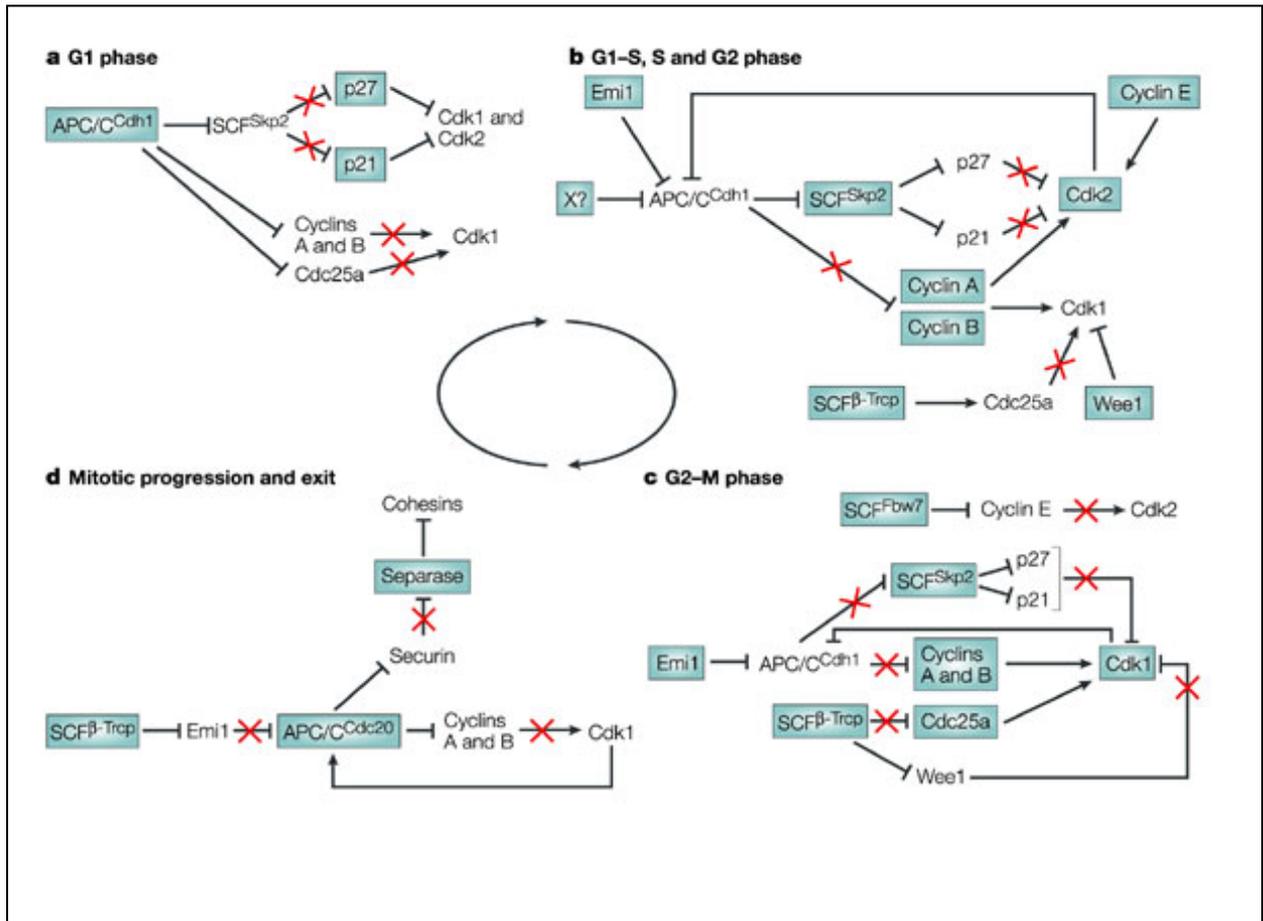


Figure 14. F-Box et Cycle cellulaire d'après Cardozo et al. ¹³⁶

Les kinases dépendantes des cyclines (CDK) sont les régulateurs clés du cycle cellulaire. Pendant la phase G1, Cdk1 et Cdk2 sont nécessaires afin de prévenir une synthèse prématurée d'ADN (a). Au point de contrôle G1-S, où débutent la réplication de l'ADN et la duplication des centrosomes, l'activité Cdk2 augmente fortement (b). Ceci conduit à un pic de la synthèse d'ADN et de la duplication des centrosomes en G2. Après ce pic, l'activité Cdk2 baisse en G2 tardive, et celle de Cdk1 augmente, permettant l'établissement d'un environnement cellulaire mitotique par la phosphorylation de ses substrats (c). L'anaphase est marquée par une baisse brutale de l'activité Cdk1, et à la fois Cdk1 et Cdk2 sont maintenus inactifs jusqu'à la transition G1-S.

X ? représente un facteur inconnu. Les rectangles bleus représentent les formes actives des protéines.

IV.2. Régulation de la biogenèse mitochondriale

Les mitochondries sont des compartiments cellulaires très dynamiques. Dans la plupart des types cellulaires eucaryotes, elles se déplacent continuellement le long du réseau formé par le cytosquelette, se divisent et fusionnent fréquemment. Leur forme varie selon le type cellulaire, le statut physiologique et les conditions nutritionnelles¹³⁷⁻¹³⁸.

Des études récentes ont montré que la dynamique mitochondriale est cruciale pour une grande variété de fonctions cellulaires. Par exemple, la fusion mitochondriale est nécessaire pour la synthèse du sperme chez la drosophile¹³⁹ et pour le maintien de l'intégrité de l'ADN mitochondrial chez la levure¹⁴⁰, mais elle joue également un rôle important dans la formation des réseaux mitochondriaux intracellulaires qui permettent la dissipation de l'énergie à travers la cellule¹⁴¹. Quant à la fission mitochondriale, elle constitue une étape importante dans l'apoptose, et est nécessaire au développement embryonnaire chez le nématode¹⁴².

La machinerie de ces deux mécanismes a été bien étudiée chez la levure. Elle repose sur 6 protéines formant le cœur du dispositif¹⁴³. La fusion mitochondriale est gouvernée par une GTPase de la membrane mitochondriale externe, nommée Fzo1¹⁴⁴, qui interagit avec la protéine adaptatrice Ugo1 (membrane externe)¹⁴⁵ et s'associe à la protéine Mgm1 située dans la membrane interne pour coordonner la fusion des quatre membranes de deux mitochondries juxtaposées¹⁴⁶. La fission mitochondriale est conduite par la GTPase Dnm1¹⁴⁷, la protéine adaptatrice Mdv1 et la protéine de la membrane externe Fis1.

Ces mécanismes de dynamique mitochondriale sont beaucoup moins connus chez les mammifères. Néanmoins, il existe des orthologues de ces protéines chez les mammifères : Drp1 (*DNMI*), hFis1 (*FIS1*), Mfn1 et Mfn2 (*FZO1*), Opa1 (*MGMI*). Comme chez la levure, Opa1 nécessite Mfn1 pour promouvoir la fusion mitochondriale¹⁴⁸.

En 2003, l'équipe de Fritz a établi le premier lien entre dégradation protéolytique ubiquitine-dépendante et mitochondrie. Ils ont identifié, lors d'une analyse systématique du génome de levure, une protéine F-box, appelée Mdm30. Celle-ci est capable de s'associer à Skp1 et Cdc53 pour former une ligase E3 de type SCF. Ce polypeptide de 598 acides aminés est localisé à la fois dans le cytosol et dans la mitochondrie. Dans ces mêmes travaux, l'étude de mutants et de doubles mutants révèle que Mdm30 est nécessaire au maintien de la compétence de fusion de la mitochondrie mais qu'elle ne fait pas directement partie de la machinerie dédiée à cet événement. En fait, Mdm30 semble contrôler le taux de protéine Fzo1 puisque celui-ci se corrèle au niveau d'expression de MDM30. Ainsi, deux rôles possibles

(non exclusifs l'un l'autre) pour la protéine F-box Mdm30 dans la fusion mitochondriale ont été proposés :

- Premièrement, Mdm30 peut réguler le taux de Fzo1 nécessaire à la balance des événements fusion-fission, équilibre critique pour le maintien de la morphologie tubulaire des mitochondries.
- Deuxièmement, les agrégats mitochondriaux formés lorsque Fzo1 est surexprimée, provoquent une accumulation de complexes de fusion non productifs. Mdm30 pourrait être nécessaire à leur élimination et ainsi parer l'agrégation mitochondriale ¹⁴⁹.

Neutzer *et al* en 2005, apportent une subtilité supplémentaire dans le rôle de régulation de Fzo1 par Mdm30. Il semblerait que Mdm30 régule la dégradation de Fzo1 uniquement lors du fonctionnement fusion-fission de la mitochondrie en conditions normales, une autre voie semble impliquée lorsque une forte sous-régulation de Fzo1 est nécessaire, par exemple lors de la fragmentation mitochondriale induite par l'activation de la voie de signalisation par phéromone ¹⁵⁰.

IV.3. Mort cellulaire programmée

L'élimination des cellules via l'apoptose ou mort cellulaire programmée, est un processus, évolutivement très conservé, central de la biologie cellulaire, du développement embryonnaire à l'homéostasie du système immunitaire. Une dérégulation dans ce programme n'interfère pas seulement avec la physiologie normale mais conduit souvent à des pathologies. Les relations entre apoptose et ubiquitination sont diverses, nombreuses et complexes (compilées par Lee et Peter en 2003)¹⁵¹. La figure 15 ¹⁵¹ présente un schéma général récapitulatif des cibles connues de l'ubiquitination dans les voies apoptotiques.

Les trois types d'ubiquitination sont impliqués : mono-ubiquitination, poly-ubiquitination K63 et poly-ubiquitination K48.

La poly-ubiquitination K48, impliquant une dégradation protéosomale, concerne les protéines IκBα, IAPs (*Inhibitor of Apoptosis*), c-FLIP, Bid, Bcl2, Bax et p53. La poly-ubiquitination K63 de Traf-6 permet l'activation des kinases IκB (IKK). La mono-ubiquitination régule les caspases 3 et 7, l'histone 2A et la localisation subcellulaire de DEDD.

Étant donné les diverses cibles de l'ubiquitination, il n'est pas surprenant de voir qu'il y a également plusieurs types de ligases E3s impliquées : de type HECT comme Nedd4 et de type RING, comme SCF^{βTrep}, SCF^{Morgue}, Mdm2 ou IAP.

Nous ne décrivons ici que les protéines F-box jouant un rôle dans le processus apoptotique.

La protéine F-box β Trcp, à travers le complexe SCF ^{β Trcp}, participe à l'activation du facteur de transcription NF- κ B en ciblant son inhibiteur I κ B pour une ubiquitination et sa dégradation par le protéasome. La protéine NF- κ B est connue comme un régulateur majeur des activités anti-apoptotiques. Classiquement, NF- κ B est décrit comme une molécule hétérodimérique constituée de membres de la famille Rel (p65, RelB, NF- κ B1, NF- κ B2 et c-Rel). Tous les hétérodimères NF- κ B sont séquestrés dans le cytosol par la liaison à leur inhibiteur (I κ B), qui masque les signaux de localisation nucléaire de NF- κ B. Sous l'action de signaux variés anti-apoptotiques ou pro-survie, I κ B est phosphorylé par le complexe IKK. Cette modification post-traductionnelle entraîne la reconnaissance d'I κ B par le complexe SCF ^{β Trcp} et sa dégradation protéosomale. NF κ B, libre, est alors transloqué du cytoplasme vers le noyau où il peut activer des gènes pro-inflammatoires et des gènes de survie¹⁵²⁻¹⁵³. En plus de l'activation NF κ B, le complexe SCF ^{β Trcp} permet la maturation des protéines précurseurs p105/p100 en p50/p52, par ubiquitination de leur domaine C-terminal. La dégradation de ceux-ci rend leurs domaines transactivateurs N-terminaux libres.

Un second complexe SCF, contenant la protéine F-box Morgue, a été décrit comme intervenant dans l'apoptose chez la Drosophile. La protéine Morgue cible la protéine DIAP-1 (analogue d'IAP-1) pour sa dégradation ubiquitine dépendante¹⁵⁴. DIAP-1 est une ligase E3 qui ubiquitine Dronc (homologue de la caspase 9), facteur pro-apoptotique¹⁵⁵.

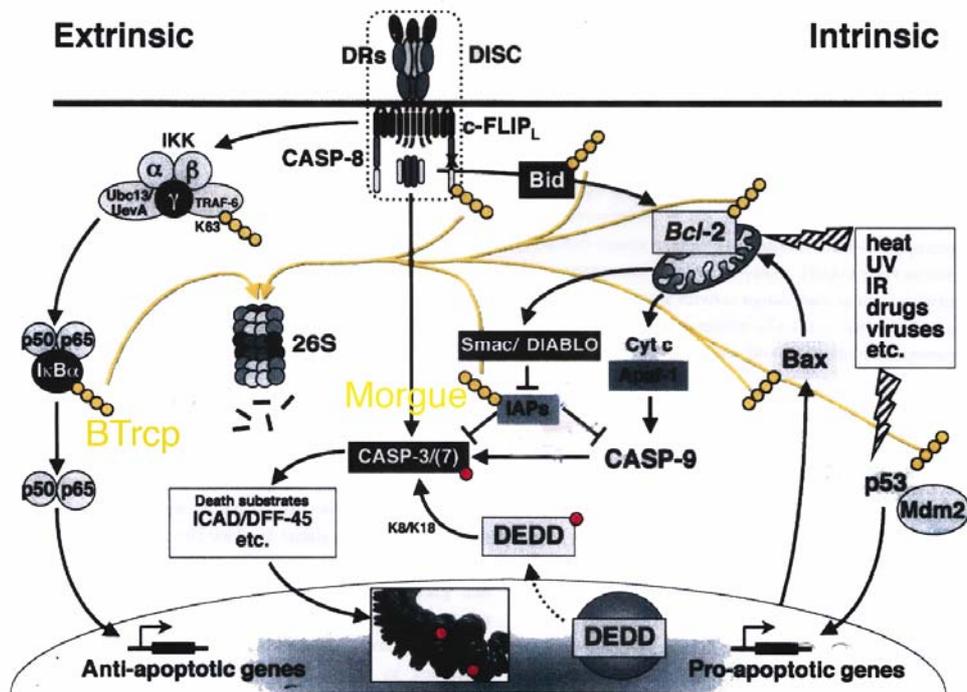


Figure 15. Apoptose et ubiquitination d'après Lee et al ¹⁵¹.

Les cibles de la poly-ubiquitination dans les voies intrinsèques et extrinsèques apoptotiques sont désignées par quatre billes jaunes, celles de la mono-ubiquitination par une bille rouge. Bien qu'usuellement les protéines poly-ubiquitinées soient destinées à la protéolyse via le protéasome 26S (flèches jaunes), l'activation de la kinase d'IκB (IKK) est dépendante d'une poly-ubiquitination de type K63 sur le facteur 6 TRAF-6. Les complexes SCF impliqués dans l'apoptose sont écrits en jaunes (**βTrcp** et **Morgue**) DRs (*death receptors*), DISC (*death-inducing signaling complex*), CASP (*caspase*), c-FLIP (*cellular Fas-associated death domain-like interleukin-1α converting enzyme inhibitory protein*), IAP (*inhibitor of apoptosis*), DEDD (*death effector domain-containing DNA-binding protein*), p50/p65 (*nuclear factor κB NFκB heterodimer*), DFF (*DNAse fragmentation factor*), Ubc (*ubiquitin conjugating enzyme*), Apaf-1 (*apoptosis protease activating factor-1*), cyt c (*cytochrome c*), Smac (*second mitochondrial activator of apoptosis*), UV (*ultraviolet*), IR (*infrared*).

IV.4. Exception à la voie ubiquitine-protéasome

Certaines protéines F-box ont été décrites comme fonctionnant dans des systèmes biochimiques autre que la protéolyse ubiquitine-dépendante via le protéasome 26S. Nous en présentons quelques-unes ci-dessous.

Tout d'abord, chez la levure, la protéine Cft13 contient un domaine F-box divergent qui n'est pas retenu par les algorithmes de recherche type Prosite ou Pfam, mais qui a été montré comme nécessaire et suffisant pour la liaison à Skp1 ¹⁵⁶. Cft13 est un composant du complexe

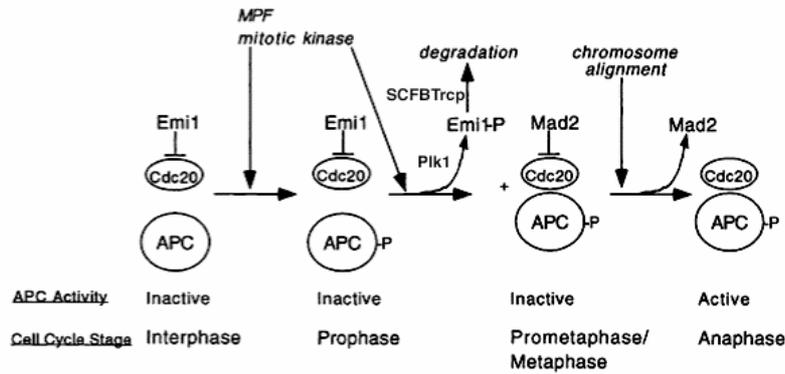
kinétochore CBF3, lequel relie les chromosomes mitotiques condensés aux microtubules. La liaison à Skp1 facilite la phosphorylation de Cft13 par une kinase inconnue, ceci permettant à Cft13 de s'assembler à l'intérieur du complexe CBF3¹⁵⁷. La perte totale de CFT13 est létale, tandis qu'à la température permissive, les mutants sensibles à la température *cft13*, présentent un phénotype avec des chromosomes mal ségrégués¹⁵⁸.

Le deuxième exemple dans lequel une protéine F-Box fonctionne dans un contexte biochimique différent de la protéolyse a été décrit chez *C.elegans* avec la protéine FOG-2. Cette protéine F-box, forme un complexe avec la protéine de liaison à l'ARN GLD-1 par l'intermédiaire de son domaine FTH/DUF38 et /ou sa séquence C-terminale¹⁵⁹. FOG-2 est requis pour la spermatogenèse chez les hermaphrodites *C.elegans*. Dans les lignées germinales, le complexe FOG-2/GRD1 se fixe sur la région 3' non traduite de l'ARNm tra-2 et inhibe sa translation, permettant ainsi la spermatogenèse¹⁵⁹⁻¹⁶⁰. Actuellement la fonction du motif F-box de FOG-2 n'est pas connue.

Le dernier exemple proposé concerne la protéine F-box humaine Emi1. Cette protéine a été découverte lors d'un criblage de banque d'ADNc d'œufs de Xénope par double-hybride¹⁶¹. Emi1 est une protéine de 392 acides aminés contenant un domaine F-box, une région de liaison au zinc (ZBR) et cinq sites possibles de phosphorylation par des Cdk¹²⁵. De façon étonnante, Emi-1 se lie à Skp1 mais pas aux autres composants connus du système SCF, tel Rbx1, ainsi son appartenance à un complexe E3 ligase, reste incertain à ce jour. L'équipe de Reimann a montré que Emi1 régule négativement l'activité du complexe APC/C en se liant à sa sous-unité activatrice Cdc20 via son domaine ZBR¹²⁵ (Figure 16A). Emi1 régule donc la progression de la phase précoce de la mitose en empêchant une activation prématurée du complexe APC/C. La sortie de mitose est permise par la destruction d'Emi1 qui est phosphorylée par la kinase Polo-like (Plk1) et ainsi, reconnue et ubiquitinée par SCF^{βTrcp}. La dégradation ubiquitine-protéasome d'Emi1 permet l'activation d'APC/C et la sortie de mitose¹⁶²⁻¹⁶³.

En 2002, dans une autre étude sur le rôle d'Emi1, Reimann *et al* ont montré que cette protéine F-box faisait partie du CFS (*cytostatic factor system*) des œufs de vertébrés. En effet, Emi-1 est nécessaire et suffisant (*in vitro*) pour bloquer les œufs non fécondés en métaphase de méiose II (Figure 16B)¹⁶⁴. Le rôle de l'interaction Emi1/Skp1 reste à découvrir, puisque les fonctions connues d'Emi-1 sont liées à ses autres domaines d'interaction protéique.

A.



B.

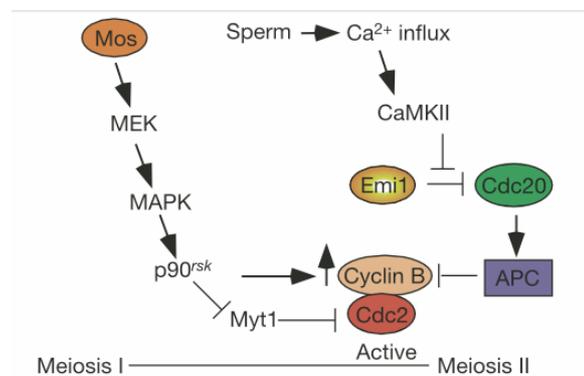


Figure 16. Rôles de la protéine F-box Emi1 dans la régulation du cycle cellulaire.

- A. Modèle complété de la régulation du complexe APC/C par Emi1 proposé par Reimann ¹²⁵.
- B. Modèle de l'action des voies Emi-1 et Mos/MAPK dans le blocage en métaphase de méiose II proposé par Reimann ¹⁶⁴. Mos active l'activité cyclineB/Cdc2 en régulant négativement la kinase Myt1 et en augmentant la synthèse de cycline B durant la transition méiose I- méiose II. Emi-1 inhibe la destruction de la cycline B via APC en méiose II, stabilisant ainsi l'activité cycline B/Cdc2. La fécondation provoque un afflux de calcium activant la CaMKII qui phosphoryle Emi-1 et lève l'inhibition d'APC/C.

V. Protéines F-Box et pathologies humaines

Puisque la voie de dégradation ubiquitine-protéasome régule la plupart des mécanismes cellulaires, il paraît évident que sa dérégulation, directe ou indirecte, puisse provoquer des défaillances conduisant à diverses pathologies chez l'homme.

V.1. Cancer : dérégulation du cycle cellulaire

Puisque les régulateurs de la croissance et du cycle cellulaire sont, pour la plupart, sous le contrôle du système SCF, il n'est pas surprenant que les voies SCF aient été fréquemment liées aux cancers humains. Nous présentons ici, les trois exemples les plus étudiés de protéines F-box dans les voies de cancérogenèse.

V.1.1. Skp2 et Cancer

Skp2 est une protéine F-box à domaine LRR qui a des effets pléiotropiques dans le contrôle du cycle cellulaire. En effet, cette protéine participe à la dégradation de substrats variés, comme E2F1, p27, p130, hOrc1 ou encore Myc.

E2F1 est un des nombreux membres de la famille E2F qui dirige l'expression des gènes de la transition G1/S¹⁶⁵. Il apparaît que la stabilité de E2F1 est déterminée à la fois par la disponibilité de Skp2, laquelle est elle-même cycle régulée avec un pic en phase S, et par la liaison de E2F1 à Rb, laquelle empêche sa reconnaissance par SCF^{Skp2}¹⁶⁶⁻¹⁶⁷.

La protéine p27 est un inhibiteur des CDK dont la dégradation est critique pour l'entrée en phase S. Des travaux de 1999 montrent que la reconnaissance de p27 par SCF^{Skp2} nécessite la phosphorylation CDK dépendante de p27 sur le résidu thréonine 187¹⁶⁸. La dégradation de p27 requiert également sa liaison avec la protéine Cks1 qui est un composant essentiel des complexes CDK. Comme Skp2, Cks1 est cycle régulée avec un pic en phase G1, tel que l'élimination de p27 ne soit permise jusqu'à la transition G1/S¹⁶⁹. Ainsi, Skp2, Cks1 et p27 phosphorylée forment un complexe trimérique dans lequel Cks1 se lie étroitement à Skp2 augmentant ainsi l'affinité de Skp2 pour l'épitote pThr187 de p27.

La protéine p21^{Cip1} est dégradée par SCF^{Skp2} de façon similaire, Cks1 et phosphorylation dépendantes¹²³.

Des découvertes récentes indiquent que Skp2 joue un rôle dans l'activation transcriptionnelle via la protéine Myc. Myc est un facteur de transcription proto-oncogène surexprimé ou sujet à des mutations stabilisatrices dans au moins 30% des cancers humains ¹⁷⁰. Skp2 se lie à la boîte Myc II (MBII) et au domaine bHLH de myc, apparemment de façon constitutive.

De plus, Cull1 et de nombreuses sous-unités du protéasome s'associent sur la chromatine sur des promoteurs Myc-dépendants ¹⁷¹. Il apparaît que l'activation et la dégradation couplée de Myc sont cruciales pour l'activité biologique de Skp2 puisque la surexpression de Skp2 dans des fibroblastes Myc -/- conduit l'induction de la phase S en dépit de la dégradation de p27 ¹⁷¹.

Puisque l'ubiquitination SCF^{Skp2} dépendante de Myc n'a pas été montrée *in vitro*, il est possible que la reconnaissance de Myc nécessite un cofacteur inconnu, de la même façon que la reconnaissance p27 requiert Cks1. Remarquons enfin, qu'il est probable qu'il existe des voies parallèles contribuant à l'activation de Myc puisque des souris Skp2-/- sont viables et ne développent que peu de tumeurs, alors que Myc est fortement oncogénique lorsqu'elle est entièrement dérégulée ¹²².

La protéine Skp2 est souvent surexprimée dans de nombreux cancers et notamment les cancers primaires du sein, en parallèle avec une réduction des taux d'expression de protéine p27. De façon concordante avec son rôle de proto-oncogène, la surexpression de Skp2 coopère avec des mutations du gène Ras pour la transformation de fibroblastes de souris et la promotion de tumeurs dans des souris transgéniques ¹⁷²⁻¹⁷³.

Enfin, Skp2 est impliquée dans l'ubiquitination et la co-activation transcriptionnelle de l'oncoprotéine c-myc. Dans ce cas, le rôle oncogène de Skp2 serait indirect et dépendrait plus probablement de son lien avec la voie Myc.

V.1.2. Fbxw7 et Cancer

La protéine Fbxw7 est l'homologue humain de Cdc4 (ou Ago), ses substrats principaux sont les protéines cycline E, Notch et Myc.

SCF^{Fbxw7} se lie et ubiquitine efficacement les formes phosphorylées de cycline E. Bien que l'interaction cycline E/Fbxw7 dépende de la phosphorylation de la thréonine 380, la phosphorylation sur au moins un autre site, la thréonine 62, influence directement ou indirectement la reconnaissance de la cycline E. Ces phosphorylations sont réalisées par la kinase Gsk3.

L'instabilité de Myc nécessite la phosphorylation des résidus thréonine 58 et sérine 62 situés dans la boîte Myc I (MBI), point chaud de mutation oncogénique¹⁷⁰. SCF^{Fbxw7} cible Myc pour sa dégradation de façon phospho - MBI dépendante¹²⁹⁻¹⁷⁴.

Le récepteur transmembranaire LIN-12/Notch, est clivé par des protéases lors de la liaison d'un ligand. Le fragment intracellulaire du récepteur (Notch^{ICD}) est alors transloqué vers le noyau où il va activer ses gènes cibles¹⁷⁵. Plusieurs équipes dont celle de Wu, a montré que l'activité de Notch^{ICD} est atténuée par Fbxw7, de façon phosphorylation dépendante¹⁷⁶⁻¹⁷⁸. En réalité, Fbxw7 est seulement une des au moins 4 ligases E3 contrôlant l'activité de Notch¹⁷⁹.

La composition de la cohorte des substrats principaux de Fbxw7 montre que celle-ci réside à un nœud du contrôle de la croissance et du cycle cellulaire.

Depuis quelques années, des corrélations similaires entre le statut mutationnel de Fbxw7 et le cancer sont mises en évidence. Les études initiales effectuées par les équipes de Stromaier et Moberg ont détecté des mutations de Fbxw7 associées à des taux importants de la protéine cycline E dans des cancers du sein et de l'ovaire¹²⁹⁻¹³⁰. Significativement, une forte activité de la cycline E constitue un indicateur pronostique pauvre dans la clinique du cancer du sein¹⁸⁰.

De la même façon, des analyses récentes ont mis en évidence une relation étroite entre mutations dans le gène FBXW7 et cancer du côlon puisque 22 tumeurs primaires sur 190 analysées portent des mutations somatiques ainsi que 4 polypes bénins sur 58 étudiés¹⁸¹. Dans ce travail, l'équipe montre également que l'inactivation de FBXW7 peut induire une instabilité chromosomique. En effet, la délétion homozygote de FBXW7 dans des lignées cellulaires euploïdes provoque la formation d'un micronucleus et une progression rapide vers l'aneuploïdie due au moins en partie à l'accumulation de la cycline E.

De façon remarquable, de nombreuses mutations dans FBXW7 associées au cancer sont localisées dans des résidus de la région liant le substrat. Comme le système SCF^{Fbxw7} fonctionne probablement comme un dimère, ce type de mutation pourrait agir, en partie, comme dominant négatif, étant donné qu'une réduction de 40% des taux de Fbxw7 réalisée par la technique « siRNA » (*silencing RNA*) reproduit le phénotype d'instabilité chromosomique¹⁸¹.

V.1.3. β Trcp et Cancer

La protéine F-Box + WD40, β -Trcp a constitué la première voie SCF décryptée chez les mammifères. Des études initiales ont révélé que β -Trcp catalyse l'ubiquitination phosphorylo-dépendante de $\text{I}\kappa\text{B}\alpha$, inhibiteur de $\text{NF}\kappa\text{B}$, ciblé par les kinases IKKs et de la β caténine, facteur de transcription proto-oncogène, ciblé par le complexe GSK3/APC axine¹⁰⁴⁻¹⁰⁶⁻¹⁸². Finalement, le répertoire des substrats de β -Trcp a été récemment étendu à deux autres régulateurs importants du cycle cellulaire ; la phosphatase Cdc25A qui active le complexe CDK de phase S pour permettre la réplication de l'ADN⁹¹⁻¹²⁶⁻¹²⁷ et la protéine F-Box Emi1, qui inhibe le complexe APC/C par liaison à sa sous-unité Cdc20¹³⁴.

La principale fonction cellulaire de β Trcp consistant à activer la voie de survie cellulaire $\text{NF-}\kappa\text{B}$ (par la dégradation de $\text{I}\kappa\text{B}$), sa dérégulation pourrait potentiellement contribuer à la tumorigenèse. En effet, la littérature récente montre une surexpression de la protéine β Trcp dans un certain nombre de lignées cellulaires cancéreuses et dans des tumeurs primaires¹⁸³. Par exemple, la surexpression de β Trcp est détectée dans la plupart des lignées cellulaires issues de tissu humain de sein et de prostate comparée aux lignées cellulaires non transformées issues de tissu d'origine similaire. De plus, des taux élevés de β Trcp ont également été observés dans un certain nombre d'échantillons de tissus tumoraux provenant de patients porteurs de cancers primaires de la prostate, du sein et de l'estomac¹⁸⁴. La protéine β Trcp est surexprimée dans les tumeurs de la peau chimiquement induites chez la souris, et cette surexpression coïncide avec la dégradation accélérée de $\text{I}\kappa\text{B}$ *in vivo*, l'activation constitutive de l'activation de la signalisation $\text{NF-}\kappa\text{B}$ dans ces mêmes tumeurs¹⁸⁵. Ainsi, cette surexpression et l'activation de la fonction β Trcp peut jouer un rôle significatif dans la tumorigenèse via l'activation de la voie anti-apoptotique dépendante de $\text{NF-}\kappa\text{B}$.

La protéine β Trcp régule également la stabilité de la β caténine. Bien que, hypothétiquement (par des mutations dans β Trcp), cette voie pourrait causer la stabilisation de la β caténine dans des tumeurs, elle est souvent inactivée par d'autres phénomènes incluant des mutations directes sur les sites de phosphorylation de la β caténine et l'inactivation de la kinase Gsk3- β , la β caténine devenant ainsi insensible à la dégradation via β Trcp. Les mutations dans le gène β TRCP sont rares et seulement un petit nombre a été trouvé dans des cancers colorectaux sans mutation pour la β caténine et APC¹⁰⁴.

V.2. Subversion virale

Le système ubiquitine protéasome joue un rôle important dans l'immunité cellulaire. Dans les cellules infectées par des virus, les antigènes viraux sont ciblés pour leur dégradation par la voie ubiquitine-protéasome. Ainsi, comme la plupart des processus cellulaires, la dégradation par la voie de l'ubiquitine, et notamment les systèmes SCFs, est détournée par un grand nombre de virus pour échapper à la surveillance immunitaire.

De nombreux exemples ont été décrits, nous présentons ici celui du virus VIH.

La dégradation protéosomale de la protéine membranaire CD4 survient à travers la formation d'un complexe entre CD4, la protéine virale Vpu et la protéine F-Box β Trcp,¹⁸⁶. Vpu est une phosphoprotéine membranaire de 81 acides aminés qui interagit avec le récepteur CD4. Dans les cellules infectées par le VIH, l'expression du précurseur de la glycoprotéine d'enveloppe virale gp160 provoque la formation d'un complexe stable CD4-gp160 retenu dans le réticulum endoplasmique. La protéine Vpu exprimée dans le même compartiment est alors associée avec le CD4 cytoplasmique et recrute β Trcp à la membrane du réticulum endoplasmique. β Trcp recrute ensuite le complexe SCF via Skp1, conduisant ainsi à l'ubiquitination et à la dégradation de CD4.

Vpu semble également inhiber de façon compétitive, la dégradation SCF ^{β Trcp} dépendante d' $\text{I}\kappa\text{B}\alpha$, de la β caténine et d'autres substrats¹⁸⁷⁻¹⁸⁸.

V.3. Autres pathologies

V.3.1. Désordre musculaire

Le système ubiquitine est connu pour jouer un rôle important dans l'atrophie musculaire, symptôme secondaire sévère de nombreuses maladies chroniques.

Une protéine F-Box nommée MAFbx ou Atrogine1 a été identifiée par un profil de transcrits de cellules de muscles atrophiés. Elle est nécessaire et suffisante pour induire l'atrophie musculaire, probablement à travers l'ubiquitination d'une ou plusieurs protéines¹⁸⁹⁻¹⁹⁰.

Une mutation connue dans une protéine F-box nommée Fbxw4 ou Dactyline provoque une maladie autosomale dominante de malformation pied/main (SHFM3, *Split-Hand/Foot Malformation 3*)¹⁹¹.

V.3.2. Désordre neurologique

Dans la plupart des maladies neuro-dégénératives, des inclusions contenant des protéines ubiquitinées ont été identifiées dans le cerveau, suggérant un rôle de la protéolyse ubiquitine-dépendante des protéines neuronales. Dans le système nerveux central, la dégradation via l'ubiquitine joue un rôle majeur dans l'élimination des protéines endommagées par le stress oxydatif et/ou la privation de glucose ou d'oxygène. L'aggrégation des protéines peut donc être directement due à une défaillance du système ubiquitine-protéasome.

Par exemple, Fbxw7 apparaît impliquée dans la maladie d'Alzheimer, en se liant et en provoquant l'ubiquitination de la preseniline, altérant ainsi la production de peptides A- β ¹⁹².

D'autre part, Staropoli et son équipe ont montré que la protéine parkin (possédant un domaine RING) forme un complexe de type SCF avec la protéine Fbxw7 qui semble provoquer la dégradation de la cycline E dans les neurones post-mitotiques, favorisant ainsi le développement de la maladie de Parkinson ¹⁹³.

D. Objectifs de la thèse

I. Importance de la recherche sur la voie ubiquitine

La voie ubiquitine, quelle que soit sa fonction finale (protéolytique ou régulatrice non dégradante), est impliquée, nous l'avons vu, dans une multitude d'événements cellulaires. Dès lors, il est clair que des perturbations ou des dysfonctionnements dans ce système peuvent avoir des conséquences pathologiques dans un organisme. C'est pourquoi, un grand nombre de chercheurs intéressés par un large éventail de domaines, aussi bien la protéolyse, le développement, l'immunologie, la régulation du cycle cellulaire que l'oncologie, mènent des recherches convergentes et cherchent à comprendre comment une molécule aussi simple que l'ubiquitine peut participer à la régulation d'une telle diversité de processus cruciaux ¹⁹⁴.

D'ailleurs, en 2004, trois scientifiques, Aaron Ciechanover, Avram Hershko (tous deux israéliens) et Irwin Rose (américain) ont reçu le prix Nobel de chimie pour leurs travaux portant sur la dégradation ubiquitine-dépendante des protéines dans la cellule.

De plus, des approches thérapeutiques ciblant la voie de l'ubiquitine (surtout celle liée au protéasome) commencent à apparaître. En effet, les domaines de liaison des protéines F-box pour lesquels la structure est connue constituent des cibles attrayantes pour dessiner des inhibiteurs spécifiques¹⁰²⁻¹⁹⁵. La plasticité du système SCF a été exploitée pour créer des enzymes E3 synthétiques, qui en principe peuvent cibler n'importe quelle protéine désirée pour laquelle existe une protéine partenaire. Dans une de ces stratégies, le domaine F-box a été fusionné à un domaine hétérologue d'interaction pour éliminer avec succès les cibles protéiques désirées dans les cellules¹⁹⁶⁻¹⁹⁸.

II. But du travail de thèse

La première étape était d'identifier un gène candidat à proximité du marqueur microsatellite D17S250 délété au niveau somatique dans différents cancers.

Après l'identification *in silico*, il a fallu valider biologiquement son existence et déterminer ses caractéristiques.

Une fois la structure du gène connue, l'interrogation des bases de données nous a permis d'identifier le produit de notre gène comme un des membres de la famille F-box. Or les protéines les mieux caractérisées de cette famille sont soit des oncogènes (Skp2 et β Trcp) soit des suppresseurs de tumeur (Fbxw7).

La seconde étape consistait alors à déterminer la fonction du gène et de son produit dans la carcinogenèse d'une part (expression différentielle, étude de mutation...), et, d'autre part d'identifier les caractéristiques intrinsèques et extrinsèques de la protéine (localisation, fonctionnalité, partenaires protéiques...).

L'ensemble des résultats devait nous permettre d'émettre, au moins des hypothèses, d'une part sur le rôle éventuel de ce gène dans la cancérogenèse et, d'autre part, de manière plus générale, sur la fonction cellulaire de cette nouvelle protéine F-box.

CHAPITRE 2 MATÉRIELS ET MÉTODES

A. Étude de la région génomique de FBXO47

I. Étude de perte d'hétérozygotie LOH

Les pertes d'hétérozygotie ou LOH (*Loss Of Heterozygosity*) sont des délétions chromosomiques affectant uniquement un des deux chromosomes homologues d'une cellule diploïde. La détection d'un tel phénomène s'appuie sur le polymorphisme génétique des marqueurs microsatellites par PCR, les amorces utilisées encadrant alors le microsatellite dont la répétition peut varier d'un allèle à l'autre. La mise en évidence d'une LOH repose sur la variation de l'intensité des pics obtenus (dénotant l'absence de matériel génétique n'ayant pu être amplifié par PCR sur l'un des allèles) en comparant les profils de l'ADN constitutionnel et de l'ADN tumoral d'un même individu.

Sélection des patients, description du matériel d'étude

Quinze patients ont été sélectionnés pour la présence d'une tumeur papillaire rénale opérée par chirurgie et confirmée par analyse anatomopathologique.

Des coupes de 10 µm ont été préparées à partir de tumeur ou de tissu normal correspondant inclus en bloc de paraffine et fixés en formaline. Les aires comprenant le tissu rénal tumoral et normal ont été délimitées sur la lame.

Extraction de l'ADN génomique

L'ADN a été extrait à partir des coupes de tumeur et de tissu normal correspondant. Le tissu est gratté sur la lame à l'aide d'un scalpel puis incubé à 60°C pendant au moins 5 heures dans 500 µl d'une solution 1M Tris HCl, pH 8 ; 5 M NaCl ; 0,5 M EDTA, pH 8 ; 10% SDS ; 10ng/ml de protéinase K. Un volume de 20 µl de protéinase K à 10ng/ml a ensuite ajouté puis les tubes ont alors été remis à incuber sur la nuit à 60°C. L'ADN a été extrait par ajout de 500µl d'un mélange phénol-chloroforme-acide isoamylique. Les tubes ont ensuite été centrifugés 5 minutes à 10 000 rpm. Les phases supérieures contenant l'ADN sont transférées dans des tubes contenant 500µl de chloroforme. Ceux-ci sont centrifugés 5 minutes à 10 000 rpm. Les phases supérieures sont déposées sur microcons, puis centrifugées 7 minutes à 10 000 rpm. Les filtres ont enfin été lavés avec 300 µl de TE 10 :1, puis séchés par une centrifugation de 10 minutes à 10 000 rpm. L'ADN est alors repris avec 60 µl de TE 10 :1. La

préparation d'ADN ainsi obtenue est dosée sur un spectrophotomètre *NanoDrop* (Nyxor Biotech).

Caractéristiques des marqueurs microsatellites étudiés

Nous avons choisi d'étudier 8 marqueurs microsatellites répartis sur le chromosome 17. Le marqueur D17S969 se trouve sur le bras court, les 7 autres sur le bras long entre la bande 11.2 et la bande 23.2. Le tableau suivant regroupe les caractéristiques des marqueurs sélectionnés (ND signifie non déterminé).

Locus	Position cartographique	Taux de variation allélique maximum	Type de répétition	Numéro d'accèsion GDB
D17S969	17p12	ND	Tétranucléotide	228553
D17S1532	17q11.2	ND	Trinucléotide	455733
D17S933	17q11.2	0,8240	Dinucléotide	199552
D17S250	17q12	0,8270	Dinucléotide	177030
D17S800	17q21.2	0,7433	Dinucléotide	188300
D17S579	17q21.31	0,87	Dinucléotide	191691
D17S588	17q21.33	0,8510	Dinucléotide	185206
D17S1290	17q23.2	ND	Tétranucléotide	364316

PCR multiplex

La PCR multiplex permet d'amplifier, dans un même mélange réactionnel, plusieurs fragments d'ADN correspondants à des couples d'amorces différents. Les amorces utilisées dans l'étude des microsatellites sélectionnés ont été synthétisées par Eurogentec, et toutes les amorces « sens » ont été marquées par un fluorochrome pour permettre une analyse en fluorescence. Le tableau ci-dessous présente les amorces utilisées, le type et la couleur de leur fluorochrome ainsi que la taille du produit attendu.

Locus	Séquence des amorces 5'-3'	Type et couleur du fluorochrome	Taille estimée du produit PCR (pb)
D17S969	ATCTAATCTGTCATTCATTCATCTATCCA-6-Fam/ AACTGCAGTGCTGCATCATA	FAM-Bleu	116-140
D17S1532	CAACGGTAGGCAAGTAAGCT-6-Hex/T GCTGTGTGCAGTTTGATCT	HEX-Vert	165-185
D17S933	ACTCACTGGGGTCCTGG-6-Fam/ TGTGGTTTCCTTATAGACTGTAGA	FAM-Bleu	187-206
D17S250	GGAAGAATCAAATAGACAAT-6-Ned/ GCTGGCCATATATATTTAAACC	NED-Jaune	135-165
D17S800	GGTCTCATCCATCAGGTTTT-6-Hex/ ATAGACTGTGTAAGGGCATTGA	HEX-Vert	168-178
D17S579	AGTCCTGTAGACAAAACCTG-6-Fam/ CAGTTTCATACCAAGTTCCT	FAM-Bleu	111-125
D17S588	CCTGGTCTAGGAAGAGTGTC-6-Hex/ GTGTAAGCATCTGTGTATACTAC	HEX-Vert	140-160
D17S1290	GCCAACAGAGCAAGACTGTC-6-Fam/ GGAAACAGTTAAATGGCCAA	FAM-Bleu	170-210

Le tampon réactionnel comprend 20mM Tris HCl, pH 8 ; 50mM KCl, 2mM MgCl₂, 0,125 U de Taq polymérase platinumium (Invitrogen), 10µM de dNTPs, 5 pmol de chaque amorce et 50 ng d'ADN pour un volume final de 25 µl.

PCR Multiplex 1 (*D17S969-D17S1532-D17S588-D17S1290*)

Dénaturation initiale	95°C	5 minutes	
Dénaturation	94°C	30 secondes	30 cycles
Hybridation	55°C	50 secondes	
Elongation	72°C	90 secondes	
Extension finale	72°C	10 minutes	

PCR Multiplex 2 (*D17S800-D17S933-D17S579-D17S250*)

Dénaturation initiale	95°C	5 minutes	
Dénaturation	94°C	30 secondes	30 cycles
Hybridation	52°C	50 secondes	
Elongation	72°C	90 secondes	
Extension finale	72°C	20 minutes	

Toutes les PCRs ont été réalisées sur un thermocycleur *ICycler* (Biorad).

Visualisation et Analyse

Avant dépôt sur le génotypeur automatique 16 capillaires ABI3100 (Applied Biosystems), 1 µl de chaque produit des deux PCRs multiplex sont mélangés à 20 µl de formamide désionisée (96%) et 0,5 µl de marqueur de taille ROX GS350 (35 à 350 pb). Ce mélange est ensuite dénaturé 3 minutes à 95°C et refroidi sur glace.

Le logiciel GeneScan 2.0.2 (Applied Biosystems) permet l'assignation de la taille de chaque fragment. Le logiciel Genotyper 2.1 (Applied Biosystems) mesure l'intensité des pics obtenus c'est-à-dire indique le nombre de copies de l'allèle amplifié.

Ainsi, nous avons considéré que les échantillons hétérozygotes pour un locus sont informatifs alors que les échantillons homozygotes sont déclarés non-informatifs. Une LOH positive a été

définie comme une absence ou une réduction visible de l'intensité de l'un des allèles dans la tumeur comparé au tissu normal correspondant.

Dans le cas d'une réduction d'intensité, un index de valeur de LOH (LIV) a été calculé :

$$\text{LIV} = (\text{hauteur du pic de l'allèle 2 dans le tissu normal} / \text{hauteur du pic de l'allèle 1 dans le tissu normal}) / (\text{hauteur du pic de l'allèle 2 dans le tissu tumoral} / \text{hauteur du pic de l'allèle 1 dans le tissu tumoral}).$$

Un LIV supérieur à 2 définit la perte de l'allèle le plus grand dans le tissu tumoral, tandis que un LIV inférieur à 0,67 correspond à la perte de l'allèle de plus petite taille dans le tissu tumoral. Les LIV comprises entre ces deux valeurs ne représente pas une LOH.

B. Clonage du gène FBXO47

I. Clonage IN SILICO

I.1. Recherche d'une structure exon / intron et phase ouverte de lecture

La région analysée a d'abord été soumise au programme Repeat Masker (<http://www.repeatmasker.org/>) qui permet de masquer les séquences répétées facilitant ainsi l'analyse suivante.

Les programmes de prédictions d'exons utilisés pour analyser la région génomique (100 kb) d'intérêt sont les suivants:

GENSCAN (<http://genes.mit.edu/GENSCAN.html>)

GRAIL EXP (<http://grail.lsd.ornl.gov/grailexp>)

FGENES (<http://genomic.sanger.ac.uk/gf/gf.shtml>)

Les exons les plus fréquemment prédits ont été retenus et assemblés afin d'identifier une possible phase ouverte de lecture à l'aide du logiciel DNA Strider 1.2.

Les séquences masquées ont été comparées à différentes banques de données type BLAST, ou dbEST (banque *d'Expression Sequence Tag*), disponibles sur le serveur NCBI (<http://www.ncbi.nlm.nih.gov/BLAST/>)¹⁹⁹⁻²⁰⁰.

I.2. Analyse des régions non traduites

La région promotrice a été analysée en recherchant les sites potentiels de liaison de facteurs de transcription connus ainsi que la boîte TATA, à l'aide de programmes disponibles sur le serveur Infobiogen : Fgenes, McPromoteur, Corepromoter, Promoter Inspector, TFSEARCH et PROSCAN (http://www.infobiogen.fr/deambulum/liste.php?page=pgn_predgene).

L'identification d'îlots CpG a été menée avec le programme Grail (http://compbio.ornl.gov/GAT_tmp).

Les sites putatifs de polyadénylation ont été déduits des séquences masquées en combinant les résultats obtenus par différentes méthodes incluant les programmes GenScan, GrailExp, Fgenes et PolyAH (<http://genomic.sanger.ac.uk/gf/gf.shtml>).

II. Clonage IN VITRO

II.1. PCR exon-connection

Cette amplification a été réalisée à partir d'ADN rétrotranscrit à l'aide du kit *ThermoScript RT-PCR System* (Invitrogen) sur de l'ARN total de côlon (Clontech).

Les amorces utilisées sont nombreuses et ont la particularité, pour certaines, d'être situées à cheval sur une jonction de deux exons. Nous ne détaillerons pas ici l'ensemble des combinaisons d'amorces testées, nous listons uniquement dans le tableau ci-dessous les séquences de ces amorces ainsi que les conditions générales d'amplification.

Amorces

Nom	Séquence 5'-3'	Tm
Amorce -1s	GAGGATGGTGGGGCAAACGACTA	70,7
Amorce As	CTCAGGCTTCAACCCATATCAACATT	69,2
Amorce Bs	AGGACTTTCATAACCTTGAGCTGC	61,5
Amorce Cs	TTATGCGAACTGACTAATCTCTGCC	62,2
Amorce Ds	AGTGATTCTGCTTTTTGGTTGACGC	67,1
Amorce Es	ACTAAAGAGTGGACAGCAGAT	63,8
Amorce Fs	CAGTGGTTCCTCCGTGAGTGGCTTCTA	73
Amorce Gs	AGACCGTGATGAAGATGACAGAAGCT	68
Amorce Cas	ATCTGCTGTCCACTCTTTAGT	63,8
Amorce Das	GCGTCAACCAAAAAGCAGAATCACT	67,1
Amorce Eas	ATCTGCTGTCCACTTTAG	63,8
Amorce Fas	TAGAAGCCACTCACGGGAACCACTG	73
Amorce Gas	GCTTCTGTCATCTTCATCACGGTCTC	68
Amorce G'as	GACCTCCTTATGGAAGTTAGCCT	61,7
Amorce 3'	CATACATACAGGTTGGCTGGTCCCAGA	75

Mélange réactionnel

Réactifs	Volume (µl)
Tampon Réaction PCR (10X) Invitrogen	5
50 mM Mg(Cl) ₂ Invitrogen	1,5
MIX dNTP 50X (2,5mM chaque) Promega	4
amorces (10 pmol/µl chaque) Sigma	1*2
Taq Platinum Invitrogen	0,3
H ₂ O stérile qsp 50µl	35,2
ADNc (100 ng)	2

Paramètres d'amplification (*thermocycler* : ICycler Biorad):

Dénaturation initiale 94°C 5 minutes

Amplification 35 cycles

- dénaturation 94°C 30 secondes
- hybridation des amorces Tm-5°C 30 secondes
- élongation 72°C 2 minutes

Élongation finale 72°C 10 minutes

Visualisation

La migration des produits PCR est faite par électrophorèse en champ constant, qui permet de faire migrer des fragments d'ADN en fonction de leur taille par rapport à la migration d'un marqueur de taille.

10 µl de chaque produit de PCR sont déposés par puit sur un gel d'agarose à 1,2% préparé en TBE et 0,2 µg/ml de bromure d'éthidium. Le standard de poids moléculaire utilisé est le Smart Ladder (Eurogentec) qui représente une échelle de 100 pb à 10 kb.

La migration est ensuite réalisée à 120 volts pendant environ 40 minutes. Une exposition aux UV (312 nm) excite ensuite le bromure d'éthidium intercalé entre les bases de l'ADN, qui est alors visualisé par fluorescence. Le gel est photographié.

II.2. Race-PCR

Cette technique ²⁰¹ permet d'obtenir l'amplification rapide des extrémités 5' et 3' d'un ADNc par PCR : *Rapid Amplification of cDNA Ends* ou *RACE-PCR*.

Nous avons utilisé le kit « RT SMART Race cDNA synthesis » (Clontech), qui comprend trois étapes (Figure 17).

1) Synthèse d'ADNc simple brin

À partir de deux fois 1 µg d'ARN poly A+ de testicule (Clontech), nous avons produit séparément un ADNc 5'RACE, et un ADNc 3'RACE suivant les instructions du fournisseur.

L'ADNc pour la RACE-5' a été synthétisé à l'aide d'une amorce oligo(dT) modifiée et d'une amorce SMART II A. L'amorce modifiée oligo(dT), nommée 5'-CDS possède deux positions de nucléotides dégénérés en 3'. Ceci permet d'éliminer l'hétérogénéité 3' inhérente à l'utilisation d'une amorce oligo(dT) conventionnel²⁰².

L'ADNc pour la RACE-3' a été synthétisé selon la procédure habituelle d'une PCR inverse mais avec une amorce oligo(dT) spéciale. Ce 3'-CDS inclut des positions nucléotidiques bloquées modifiés comme dans le 5'-CDS et il possède également une partie de la séquence SMART dans sa partie 5'.

Ces incorporations de séquence SMART dans les deux populations d'ADNc nous permettent d'utiliser le mélange d'amorces universelles (UPM), qui reconnaissent la séquence SMART, en parallèle avec des amorces distinctes gène-spécifique.

2) Amplification rapide des extrémités de l'ADNc (RACE)

L'amplification des extrémités 5'et 3' se déroule en deux étapes. La première réaction de PCR a été réalisée avec plusieurs amorces anti-sens spécifiques de l'ADNc pour la RACE 5' (Das; Fas; 3' séquences données pour la PCR exon-connection et F'as 5'-ACCAAAGCCAAAAGACTACGAGCCTT-3' ; 3') et l'amorce UPM, et avec plusieurs amorces sens spécifiques de l'ADNc pour la RACE 3' (-1s,As, Bs et Fs) et également l'amorce UMP. La seconde réaction de PCR (*Nested-PCR*) a été réalisée sur les produits de la première amplification avec des amorces internes à celles décrites précédemment. La seconde réaction permet d'accroître la spécificité de la première.

Les séquences des amorces utilisées en Nested- PCR sont les suivantes :

- NUP pour UMP (Clontech)
- RACE 5' : Cas , Eas, Fas et Has 5'-GGCAAGGCTTTCTGAGGATTTAGGTAGA-3'
- RACE 3' : As, Fs

Ces réactions ont été réalisées suivant les conditions opératoires indiquées par le fournisseur en utilisant la Taq Polymérase « *Advantage 2* » (Clontech).

3) Caractérisation des produits de RACE

Afin d'analyser les fragments obtenus par cette technique, nous avons isolé et purifié les bandes d'intérêt sur gel d'agarose grâce au kit *Wizard SV Gel Extraction* (Proméga). Ces fragments ont ensuite été clonés dans le vecteur pGEM T Easy (Proméga), qui permet une ligation directe puisqu'il bénéficie de la technologie T/A ; les clones obtenus ont été

séquencés par électrophorèse capillaire sur un appareil ABI 3100 (Applied Biosystems). Les séquences ont été analysées à l'aide du logiciel Auto Assembler (Perkin Elmer) et comparées aux banques de données présentes en ligne.

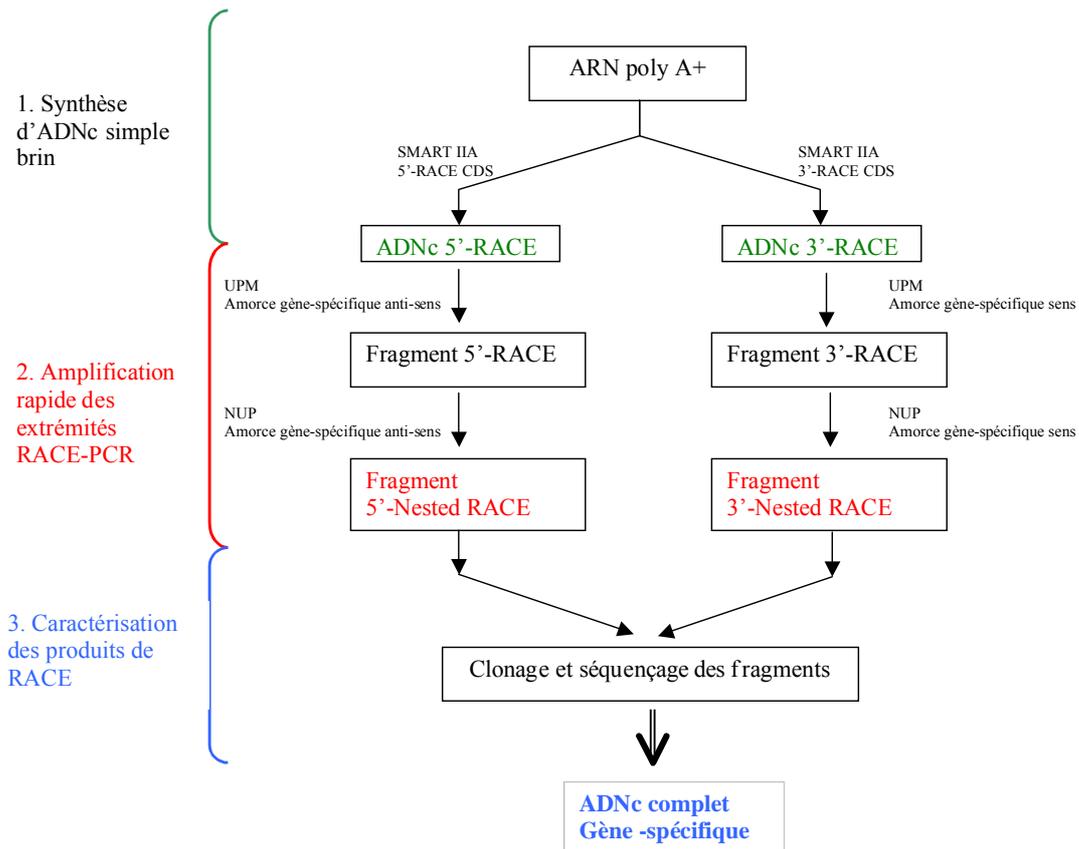


Figure 17. Principe méthodologique de la RACE-PCR.

C. Etude de l'expression du gène FBXO47

I. Northern Blot

Les Northern Blots sont utilisés pour étudier l'expression des ARNm dans certaines cellules ou tissus.

II.1. Membranes utilisées

Les membranes utilisées sont des MTN (*Multiple Tissue Northern*) Blots commercialisés par Clontech (BD Biosciences). Chaque échantillon d'ARN total est préparé grâce à la technique de thiocyanate de guanidium modifié, et chaque échantillon d'ARN poly A+ est enrichi par trois cycles de purification oligo(dT)-cellulose. Chaque ligne d'un blot MNT contient 10 µg d'ARN total ou 2 µg d'ARN poly A+ purifié. L'ARN est fractionné par électrophorèse en gel d'agarose 1%, en condition dénaturante, puis transféré sur une membrane de nylon chargée positivement, et enfin fixé par irradiation UV.

Les deux membranes que nous avons utilisées sont :

- *Human 12-Lane MTN Blot*, contenant de l'ARN de cerveau, de cœur, de muscle squelettique, de côlon, de thymus, de rate, de rein, de foie, d'intestin, de placenta, de poumon, et de leucocytes de sang périphérique.
- *Human Tumor MTN Blot*, contenant de l'ARN de tumeur de sein, de côlon, de rein, de poumon, d'ovaire, de rectum, d'estomac et d'utérus.

I.2. Choix et marquage de la Sonde

L'ADNc utilisé comme sonde a été généré par PCR à partir d'ADN rétrotranscrit de testicule à l'aide du couple d'oligonucléotides suivant : 5'-AGTGGTTCCCCGTGAGTGGCTTCTA-3' et 5'-CATAACATACAGGTTGGCTGGTCCCAGA-3'. Le fragment généré fait une longueur de 750 pb et correspond à la partie 3' de l'ARNm de FBXO47.

La sonde a été marquée au phosphore 32 ($[^{32}\text{P}]\text{dCTP}$) par la technique d'extension d'amorces aléatoires ou *random priming* avec le kit «*rediprime II random prime labelling system*» (Amersham Biosciences) ²⁰³. L'ADN marqué a été purifié sur une colonne NICK

(Amersham biosciences) contenant un gradient de Sephadex G-50, pour un volume final de 400 μl en TE 20 :1. Le taux de marquage de la sonde (en coups par minute ou cpm) a été déterminé sur un aliquot de 2 μl à l'aide d'un compteur à scintillation. La sonde est enfin dénaturée à 100°C durant 5 minutes et refroidie rapidement sur glace juste avant l'hybridation.

I.3. Conditions d'hybridation

Afin de saturer les sites de fixation potentiels de l'ADN, les membranes sont préhybridées pendant 30 minutes à 68°C dans une solution contenant le tampon d'hybridation ExpressHyb (Clontech) et 1,5 mg d'ADN de testicule de saumon dénaturé.

Les membranes ont alors été hybridées sur la nuit à 68°C dans une solution contenant le tampon d'hybridation ExpressHyb, la sonde marquée dans une quantité correspondant à 1 à 2 millions de cpm ainsi que 150 μg d'ADN de testicule de saumon dénaturé.

Les membranes ont ensuite été lavées sous agitation, d'abord trois fois 15 minutes à température ambiante par une solution 2X SSC-0,05%SDS, puis 2 fois 40 minutes à 50°C par une solution 0,2X SSC-0,1% SDS.

I.4. Visualisation

La visualisation de l'hybridation a été faite par autoradiographie sur films (Kodak) dans une cassette et en présence de deux écrans amplificateurs. L'exposition a été réalisée à -80°C pendant une durée variable de 3, 6 ou 10 jours.

II. RT-PCR Panel

II.1. Panels utilisés

Les Panels MTC (Clontech) sont des ensembles d'ADN rétrotranscrits à partir d'ARNs polyA+ extraits de cellules et / ou de tissus. Ces ADNc sont simples brins et normalisés c'est-à-dire fournis à une concentration déterminée (ici environ 1ng/ μL) pour chaque tissu. Cette normalisation a été faite en fonction de la concentration dans les échantillons d'ADNc des gènes suivants : G3PDH, beta-actine, alpha-tubuline, et la phospholipase A2.

Les ADNc inclus dans les panels MTC peuvent être ensuite utilisés comme matrice dans une PCR avec des amorces nucléotidiques gène-spécifique afin d'étudier le profil d'expression dudit gène.

Nous avons utilisé trois types de panel MTC comprenant chacun 8 tissus:

- *Human MTC Panel I*: cœur, cerveau, placenta, poumon, foie, muscle squelettique, rein, pancréas.
- *Human MTC Panel II*: rate, thymus, prostate, testicule, ovaire, intestin, côlon, leucocytes de sang périphérique.
- *Human Fetal MTC Panel*: cerveau fœtal, poumon fœtal, foie fœtal, rein fœtal, cœur fœtal, rate fœtale, thymus fœtal, et muscle squelettique fœtal.

II.2. Conditions d'amplification (PCR)

Nous avons utilisé deux couples d'amorces, le premier permettant d'amplifier toute la partie codante de FBXO47 c'est-à-dire situé d'une part en aval du codon d'initiation et d'autre part en amont du codon stop : amorce-F (5'-GGCAAACGACTAACTCCCCTTCCTGT-3') et amorce R (5'- CATAACATACAGGTTGGCTGGTCCCAGA-3'), et l'autre (fourni par Clontech) permettant d'amplifier le gène G3PDH comme contrôle interne.

Les conditions d'amplification sont les suivantes :

Mélange réactionnel FBXO47

	Volume par réaction (µl)
Tampon PCR Advantage 2 10X (Clontech)	5
dNTPs 10mM (Amersham)	1
Amorce F 10µM (Sigma)	1
Amorce R 10µM (Sigma)	1
Advantage 2 polymérase 50X (Clontech)	1
ADNc (1ng/µl)	5
Eau stérile qsp 50 µl	36

Paramètres d'amplification (*thermocycler* : ICycler Biorad):

Dénaturation initiale 94°C 5 minutes

Amplification 40 cycles

- dénaturation 94°C 30 secondes
- hybridation des amorces/élongation 68°C 3 minutes

Élongation finale 68°C 5 minutes

Mélange réactionnel G3PDH

	Volume par réaction (µl)
Tampon PCR Advantage 2 10X (Clontech)	5

dNTPs 10mM (Amersham)	1
Amorces G3PDH 10µM chacun (Clontech)	2
Advantage 2 polymérase 50X (Clontech)	1
ADNc (1ng/µl)	5
Eau stérile qsp 50 µl	36

Paramètres d'amplification (*thermocycler* : *ICycler Biorad*):

Dénaturation initiale 94°C 5 minutes

Amplification 28 cycles

- dénaturation 94°C 30 secondes
- hybridation des amorces/élongation 68°C 2 minutes

Élongation finale 68°C 5 minutes

II.3. Visualisation

La migration est réalisée par électrophorèse en champ constant dans laquelle 10 µl de chaque produit de PCR sont déposés sur un gel d'agarose 1,2% 0,2 µg/ml BET. Le standard de poids moléculaire utilisé est le Smart Ladder (Eurogentec). Une exposition aux UV (312 nm) excite ensuite le bromure d'éthidium intercalé entre les bases de l'ADN, qui est alors visualisé par fluorescence.

III. Expression Différentielle : *Cancer Profiling Array* (CPA)

III.1. Principe

Le « Disease Profiling Array » DPA (Clontech) est une membrane de nylon sur laquelle est déposé, sous forme de spots, des ADNc amplifiés par la technologie BD SMART (*Switching Mechanism At the 5'end of RNA Transcript*). Cette dernière permet d'obtenir des ADNc entiers représentant ainsi la totalité du message exprimé (c'est-à-dire tous les ARNm) dans un tissu donné. Elle garantit également la présence des transcrits rares, et l'absence d'ADN génomique²⁰⁴⁻²⁰⁵.

Pour notre étude, nous avons utilisé le DPA contenant des paires d'échantillons d'ADNc générées à partir d'ARN total de plusieurs types tissulaires (CPA). Chaque paire consiste en un échantillon tumoral et en un échantillon de tissu normal correspondant obtenu à partir du même individu. Ceci assure qu'une expression différentielle est due à des différences réelles entre les tissus normaux et tumoraux. Une base de données contenant l'histoire détaillée de

chaque patient pour chaque échantillon présent sur la membrane est disponible en ligne (<http://www.clontech.com/clontech/techinfo/manuals/PDF/7847-1.pdf>).

La membrane contenant les ADNc est normalisée sur la base de l'abondance relative de plusieurs gènes de ménage (β -actine, glutamate déshydrogénase et le gène de la protéine hautement basique 23-kDa).

Nous avons utilisé la membrane Cancer Profiling Array II pour étudier l'expression différentielle de FBXO47. Elle contient 154 paires d'échantillons provenant de 19 tissus. Le diagramme complet est présenté en annexe A.

III.2. Choix et Marquage de la sonde

L'ADN utilisé comme sonde est identique à celui utilisé pour les Northern Blot et le marquage a été réalisé de la même façon.

III.3. Hybridation

Afin de saturer les sites de fixation potentiels de l'ADN, les membranes ont été préhybridées pendant 50 minutes à 68°C dans une solution contenant le tampon d'hybridation ExpressHyb (Clontech) et 1,5 mg d'ADN de testicule de saumon dénaturé.

Les membranes sont ensuite hybridées sur la nuit à 68°C dans une solution contenant le tampon d'hybridation ExpressHyb, 20 ng de sonde marquée ainsi que 150 μ g d'ADN de testicule de saumon dénaturé.

Les membranes ont été lavées sous agitation, d'abord trois fois 30 minutes à 68°C sous agitation par une solution 2X SSC-0,5%SDS, puis une fois 40 minutes à 68°C sous agitation par une solution 0,2X SSC-0,5% SDS, et enfin une fois 5 minutes à température ambiante sous agitation par une solution 2X SSC.

III.4. Visualisation et Quantification

La visualisation de l'hybridation a été réalisée par exposition de la membrane dans une cassette de PhosphorImager (Amersham Pharmacia Biotech). L'exposition a été effectuée à température ambiante pendant une durée de 4, 24 ou 48 heures. L'image a ensuite été scannée sur le PhosphorImager et analysée de façon semi-quantitative à l'aide du logiciel IPLabGel.

D. Étude de la protéine Fbxo47

I. Analyse bioinformatique des motifs protéiques et caractéristiques physico-chimiques

Nous avons utilisé un ensemble de programmes en libre accès sur internet, dont nous ne détaillerons pas le fonctionnement puisque celui-ci est disponible en ligne.

La séquence protéique soumise à l'interrogation dans ces programmes de recherche a été traduite à partir de la séquence nucléique à l'aide du logiciel DNA Strider1.2.

I.1. Caractéristiques physico-chimiques

Le programme ProtParam (<http://www.expasy.org/tools/protparam.html>) nous a permis de déterminer plusieurs propriétés physico-chimiques qui peuvent être déduites à partir d'une séquence protéique. Les paramètres déduits sont le poids moléculaire, le point isoélectrique théorique, la composition en acides aminés, la composition atomique, le coefficient d'extinction, le temps estimé de demi-vie, l'index d'instabilité, l'index aliphatique et la GRAVY (*grand average of hydropahicity*).

I.2. Caractéristiques protéiques

I.2.1 Recherche d'analogie de séquence

Nous avons effectué une recherche par itération à l'aide du programme PSI-Blast proposé par le serveur NCBI (<http://www.ncbi.nlm.nih.gov/BLAST>) contre l'ensemble des bases de données protéiques suivantes : toutes les translations CDS de GenBank, PDB (*Protein dataBank*), SwissProt. La matrice utilisée est celle qui est proposée par défaut, BLOSUM62 (*Blocks Substitution Matrix*), qui permet de rechercher des homologies avec un seuil d'identité de 62 %.

I.2.2. Alignement de séquence

L'alignement de plusieurs séquences protéiques d'intérêt a été mené à l'aide de deux logiciels : CLUSTALX (http://www.infobiogen.fr/services/analyseq/cgi-bin/clustalw_in.pl), qui réalise un alignement multiple²⁰⁶ et du logiciel, éditeur d'alignement, SeqVu (Garvan Institute), qui permet de corriger manuellement les alignements et de les analyser.

I.2.3. Identifications de motifs protéiques

Nous avons comparé la séquence de notre protéine au contenu des bases de données PROSITE (<http://www.expasy.org/prosite>) et BLOCKS, dont les motifs sont plus courts (http://blocks.fhrc.org/blocks-bin/blocks_search). Celles-ci fonctionnent comme des méthodes de détermination de la fonction de protéines non connues traduites à partir de séquences génomiques. Cela consiste en une base de données de sites biologiquement significatifs et de profils pouvant être facilement reliés à des familles protéiques.

I.2.4. Prédiction de signaux et localisation de sites

Afin d'identifier d'éventuels peptides signaux dans la protéine Fbxo47, nous avons soumis sa séquence à quatre types de programme :

- iPSORT, prédiction de signaux N-terminaux de sortie (<http://hc.ims.u-tokyo.ac.jp/iPSORT>)
- PSORT II, regroupant une dizaine de programmes (<http://psort.nibb.ac.jp/cgi-bin/runpsort.pl>)
- TargetP v1.1, qui donne une indication sur la localisation sub-cellulaire (<http://www.cbs.dtu.dk/services/TargetP>)
- MitoProt II, qui identifie une séquence d'adressage à la mitochondrie (<http://ihg.gsf.de/ihg/mitoprot.html>).

I.2.5. Identification de sites de modifications post-traductionnelles

Le serveur NetPhos 2.0 (<http://www.cbs.dtu.dk/services/NetPhos/>) fournit un réseau de prédiction pour les sites de phosphorylation des résidus sérine, thréonine et tyrosine. Les prédictions de phosphorylation spécifique des kinases est disponible sur le site : <http://www.cbs.dtu.dk/services/NetPhosK> .

Les sites de N-glycosylation sont fournis par le programme NetNGlyc (<http://www.cbs.dtu.dk/services/NetNGlyc>) et ceux d'O-glycosylation par NetOGlyc (<http://www.cbs.dtu.dk/services/NetOGlyc>)

I.2.6. Structure 2-D

La présence et la localisation d'hélices et de feuillets dans la protéine Fbxo47 ont été recherchées à l'aide du site « *consensus secondary structure prediction* » (http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_seccons.html) qui combine plusieurs méthodes de prédiction et permet ainsi de repérer les structures les plus prédites et donc significatives.

II. Localisation cellulaire

II.1. Clonage des protéines d'intérêt dans des vecteurs d'expression

Les deux protéines d'intérêt sont Fbxo47 et $\delta[1-23]$ Fbxo47 qui est une forme délétée pour les 23 premiers acides aminés (correspondant à l'adressage mitochondrial) de la protéine Fbxo47.

L'ADNc de FBXO47 a été amplifié à partir d'ADNc de testicules (Clontech) à l'aide du couple d'amorce suivant : 5'-TGAGACAAGAAGAAGAATGGCATCCAG-3' et 5'-GGCAAGGCTTTCTGAGGATTTAGGTAG-3'. Le produit obtenu a ensuite été inséré dans le vecteur pGEMT easy (Promega). Cette construction nous a permis ensuite d'amplifier à la fois la forme complète et la forme délétée du gène FBXO47 en y insérant des sites de restriction enzymatiques afin de pouvoir sous-cloner ces deux produits dans le plasmide pEGFP-N3 (Clontech).

Les deux séquences FBXO47 et $\Delta[1-69]$ FBXO47 ont donc été amplifiées à partir de la construction pGEMT-FBXO47 avec la Taq Polymérase Herculase (Stratagène), avec les couples d'amorces suivants : 5'-CCCTCGAGGGAGAAGAAGAATGGCATC-3' (site XhoI souligné) et 5'-CGCGGATCCGCGGGTAGACAGAGGAGTAT-3' (site BamHI souligné) pour la forme complète de Fbxo47, 5'-CCGCTCGAGCGGATGACCAGCTGTTATTCC-3' (site XhoI souligné) et 5'-CGCGGATCCGCGGGTAGACAGAGGAGTAT-3' (site BamHI souligné) pour la forme délétée de Fbxo47.

Ces produits de PCR ont été doublement digérés par XhoI et BamHI et ligués dans le plasmide pEGFP-N3, également digéré par XhoI et BamHI. Les vecteurs d'expression ainsi obtenus produisent donc des protéines de fusion Fbx047-GFP et δ [1-23]Fbxo47-GFP. Toutes les constructions ont été séquencées (ABI 3100, Applied Biosystems) afin de confirmer la phase ouverte de lecture entre la protéine d'intérêt et la protéine EGFP (*Enhancer Green Fluorescent Protein*).

II.2. Culture cellulaire et transfection transitoire

Des cellules adhérentes Cos-7 (*African green monkey kidney fibroblasts*), ont été maintenues en DMEM (*Dubelcco's Modified Eagle's Medium*) supplémenté avec du SVF 10% , pénicilline et streptomycine 1%, glutamine 1%.

Un million de cellules, par type de plasmide, ont ensuite été transfectées transitoirement avec 3 μ g d'ADN plasmidique (soit Fbxo47-EGFP, soit δ [1-23]Fbxo47-EGFP, soit EGFP seule) avec le kit de transfection « *Nucleofactor Kit V* » (Amaya), suivant les recommandations du fournisseur. Les cellules sont ensuite incubées 24 heures à 37°C sous CO₂ dans une plaque 24 puits contenant une fine lamelle de plastique dans chaque puits.

II.3. Marquage du noyau et des mitochondries

Un jour après la transfection, les cellules Cos-7 qui ont adhéré sur la lamelle sont lavées avec du PBS et incubées avec 100nM de MitoTracker Red CMXRos (Molecular Probes) dans du DMEM pendant 45 minutes. Les cellules sont lavées trois fois avec du PBS, puis fixées avec 4 % de formaldéhyde en PBS pendant 15 minutes à 37°C. Les cellules sont ensuite lavées en PBS puis incubées avec 1 μ M de TO-PRO-3 (Molecular Probes) pendant 30 minutes à température ambiante et dans l'obscurité. Enfin, les lamelles sont enfin lavées une dernière fois avec du PBS, puis retournées sur une goutte de « *ProLong antifade kit* » (Molecular Probes) posée sur une lame de verre.

II.4. Lecture des lames et Analyse

Les lames ainsi montées sont lues et analysées par microscopie confocale sur un appareil LEICA TCS-SP1.

Les images obtenues sont traitées par le logiciel MetaMorph qui permet de calculer le pourcentage de co-localisation de marqueurs fluorescents.

III. Identification d'interactions protéiques

III.1. Co-précipitation / Interaction *INVITRO*

Cette technique est basée sur la formation d'un complexe protéique dans lequel une des protéine est marquée par des molécules d'histidine qui permettent sa rétention sur une colonne formée de billes de nickel. Celle-ci retient alors l'ensemble du complexe s'il existe. Le décrochage puis l'analyse sur gel de protéines permettent alors de déterminer les composants dudit complexe.

III.1.1. Clonage des protéine d'intérêt dans des vecteurs d'expression

Pour cette expérience, les trois protéines d'intérêt sont Skp1, Skp2 (contrôle positif d'interaction) et Fbxo47.

L'ADNc de SKP1 a d'abord été amplifié par PCR à partir d'ADNc de placenta avec les amorces suivantes : 5'-GAACACCATGCCTTCAATTAAG-3' et 5'-

CTTCTCTTCACACCACTG-3'. Le produit obtenu a ensuite été inséré en phase dans le vecteur pcDNA3.1 V5/His (Invitrogen). L'ADNc de SKP2 a été obtenu par PCR à partir

d'ADNc de tumeur de sein avec les amorces suivantes : 5'-

CGGACGCTATGCACAGGAAG-3' et 5'-GCAATAAATACTTCATACACA ACTGG-3'.

L'ADNc de FBXO47 a été amplifié à partir d'ADNc de testicules (Clontech) à l'aide du couple d'amorce suivant : 5'-TGAGACAAGAAGAAGAATGGCATCCAG-3' et 5'-

GGCAAGGCTTTCTGAGGATTTAGGTAG-3'. Les ADNc SKP2 et FBXO47 ont ensuite

été sous clonés dans le vecteur pCR3.1 à l'aide du kit « *Eukaryotic TA Expression*

Bidirectional » (Invitrogen). Toutes les PCRs d'amplification d'ADNc ont été réalisées avec

la Taq polymérase Advantage 2 (Clontech). Toutes les constructions plasmidiques ont été

réalisées suivant les instructions des fournisseurs et toutes ont été vérifiées par séquençage

automatique sur ABI 3100 (Applied Biosystems).

III.1.2. Production de protéines marquées au ³⁵S

Afin d'exprimer les protéines Fbxo47, Skp2 et Skp1-His et de les marquer au ³⁵S pour l'analyse, nous avons choisi d'utiliser le système d'expression acellulaire « *TNT Coupled Reticulocyte Lysate system* » (Proméga). Ce système permet de coupler, dans un même tube, toutes les étapes de la transcription et de la traduction, diminuant ainsi la durée d'obtention des protéines. De plus la quantité de protéine d'une part et d'incorporation de ³⁵S d'autre part sont plus importantes que dans les systèmes classiques.

Le mélange réactionnel contient 25 µl de lysat de réticulocytes de lapin, 2 µl de tampon de réaction, 1 µl d'ARN polymérase T7, 1 µl de mélange d'acides aminés privé de méthionine à 1 mM, 2 µl de méthionine [³⁵S] (>1000 Ci/mmol à 10mCi/ml), 1µl d'inhibiteur de ribonucléase RNasin, 2 µl d'ADN plasmidique à 0,5 µg/µl et de l'eau stérile pour un volume final de 50 µl. Ce mélange est incubé 90 minutes à 30°C puis placé sur glace.

Les protéines produites ont été dosées. Ce dosage se fait par précipitation de 2 µl de produit de traduction à l'aide d'1 ml de TCA 25 % en présence de 250 µl de NaOH 1M et incubation 30 minutes sur glace. L'ensemble a ensuite été déposé sur du papier Whatman préalablement imbibé de 3 ml de TCA 5 %, et filtré sous vide. Trois lavages de 3 ml de TCA 5% ont été effectués, puis le filtre (filtre humide= cpm spécifiques) a été mis à sécher quelques minutes à température ambiante. En parallèle, sur un autre filtre de papier Wathman (filtre sec= cpm totaux), 2 µl de produit de traduction non précipité ont été déposés. Les deux filtres ont ensuite été placés dans des fioles contenant 4 ml de liquide de scintillation et la radioactivité contenue est comptée sur un compteur Beckman.

La quantité de protéine produite s'obtient alors par le calcul suivant :

$$\text{Quantité de protéine marquée (fmol/}\mu\text{l)} = [((\text{cpm spécifiques} \times 20 / \text{cpm totaux}) \times \text{nombre de méthionine dans la protéine}) / 1000] \times 1.10^7$$

III.1.3. Formation des complexes protéiques

La formation des complexes protéiques se fait en deux étapes.

La première étape consiste à coupler la protéine marquée à l'histidine sur une colonne de billes de nickel qui fixe les histidines par liaison de coordination.

Pour cela, nous avons incubé à 4°C, sur la nuit, 40 µl de résine Ni-nTA (Quiagen) reprise dans du PBS 1X avec 16 fmol de protéine His-marquée (ici Skp1). Le mélange a ensuite été centrifugé 10 minutes à 10 000rpm à 4°C. Le surnageant est conservé pour analyse (Eluat 1).

La seconde étape consiste à ajouter la protéine à complexer sur la première. Nous avons ajouté 13 fmol de la protéine (ici Skp2 ou Fbxo47) dans un volume final de 80 µl en PBS 1X sur le culot contenant la résine de nickel et la protéine His-marquée. Ce mélange est incubé 4h à 30°C, puis centrifugé 10 minutes à 6000 rpm à 4°C. Le surnageant est conservé pour analyse (Eluat 2). Deux lavages avec 50 µl de PBS 1X ont ensuite été réalisés avec, à chaque fois, une centrifugation de 10 minutes à 6000 rpm à 4°C. Les surnageants sont alors mélangés et conservés pour analyse (Eluat 3).

La dernière étape consiste en la rupture de l'association nickel-histidine, qui se réalise ici en conditions natives. Le complexe est élué de la colonne de nickel par ajout de 40 µl d'imidazole 250mM, analogue de l'histidine, et centrifugation de 10 minutes à 6000 rpm à 4°C. Le surnageant est transféré dans un tube différent, il contient le complexe protéique à déterminer.

III.1.4. Analyse sur gel de protéine

L'analyse des protéines constituant le complexe a été réalisée par migration sur gel SDS-Page 15% en conditions dénaturantes. Dans chaque puits est déposé un volume équivalent à 5 fmol de protéine ou 10 µl d'éluat ou 10 µl de complexe protéique. Pour chaque échantillon, 3,3 µl de bleu de charge β-mercaptoéthanol 4X ont été ajoutés puis l'ensemble est dénaturé 10 minutes à 95°C et placé sur glace avant le dépôt. La migration a été conduite à 200V, 50 mA durant 45 minutes environ. Le gel a ensuite été plongé dans une solution de bleu de Coomassie pendant 15 minutes sous agitation, lavé deux fois 15 minutes sous agitation dans une solution de décoloration. Enfin, le gel est enduit de glycérol 50%, mis à sécher sous vide 15 minutes et exposé en cassette de PhosphorImager pour 24 heures. L'écran a ensuite été lu sur le PhosphorImager (Amersham pharmacia Biotech) et analysé à l'aide du logiciel IPLabGel.

III.2. Double Hybride / Interaction *IN VIVO*

III.2.1. Principe

Le système double-hybride est une méthode efficace pour détecter des interactions protéiques *in vivo*. Cette technique est basée sur la technologie développée par Dove, Joung et Hochschild²⁰⁷⁻²⁰⁸ puis affinée par Joung et Pabo²⁰⁹. La méthodologie est basée sur l'utilisation d'un facteur de transcription dont les domaines de liaison à l'ADN et d'activation peuvent être dissociés. Le principe du double hybride consiste à exprimer dans un organisme vivant (type levure ou bactérie), deux protéines dont on veut tester l'interaction sous la forme de protéines hybrides fusionnées l'une avec le domaine de liaison à l'ADN, l'autre avec le domaine d'activation de la transcription. Si les deux protéines interagissent, les deux domaines sont de nouveau associés, permettant ainsi la transcription d'un gène rapporteur. Nous avons utilisé le kit « *BacterioMatch II Two-hybrid System* » de chez Stratagène.

Dans ce système, une protéine d'intérêt (protéine test) est fusionnée à la protéine répresseur de bactériophage (cI, 237 acides aminés), contenant le domaine N-terminal de liaison à l'ADN et le domaine C-terminal de dimérisation. La protéine cible correspondante est fusionnée au domaine N-terminal de la sous-unité α de l'ARN polymérase (248 acides aminés). La protéine test est fixée sur la séquence de l'opérateur en amont du promoteur du gène reporter grâce au domaine de liaison à l'ADN de λ cI. Lorsque la protéine test et la protéine cible interagissent, cela recrute et stabilise la liaison de l'ARN polymérase au promoteur et active la transcription du gène reporter *HIS3*.

Le gène *HIS3* code pour un composant de la voie de biosynthèse de l'histidine qui complète une mutation *hisB* dans la souche reporter. Le produit du gène *HIS3* est produit à un faible niveau à partir de la cassette gène reporter en l'absence d'activation transcriptionnelle, ceci obligeant la souche reporter à croître sur un milieu minimum dépourvu d'histidine. Le 3-amino-1,2,4-triazole (3-AT) agit comme un inhibiteur compétitif du produit du gène *HIS3*. En présence de 3-AT 5mM, la souche reporter est incapable de pousser sur un milieu ne contenant pas d'histidine.

Lorsque la souche reporter est co-transformée avec la protéine hybride test et les protéines cibles, l'ARN polymérase est recrutée au promoteur, activant alors la transcription de *HIS3*. La croissance de la souche reporter sur un milieu sans histidine et contenant du 3-AT ne survient donc que lorsque l'activation transcriptionnelle augmente l'expression de *HIS3* à un niveau suffisant pour surpasser l'inhibition compétitive du 3-AT.

Un second gène reporter, *aadA*, codant une protéine conférant une résistance à la streptomycine, fournit un mécanisme additionnel pour valider une interaction entre la protéine test et la protéine cible (Figure 18).

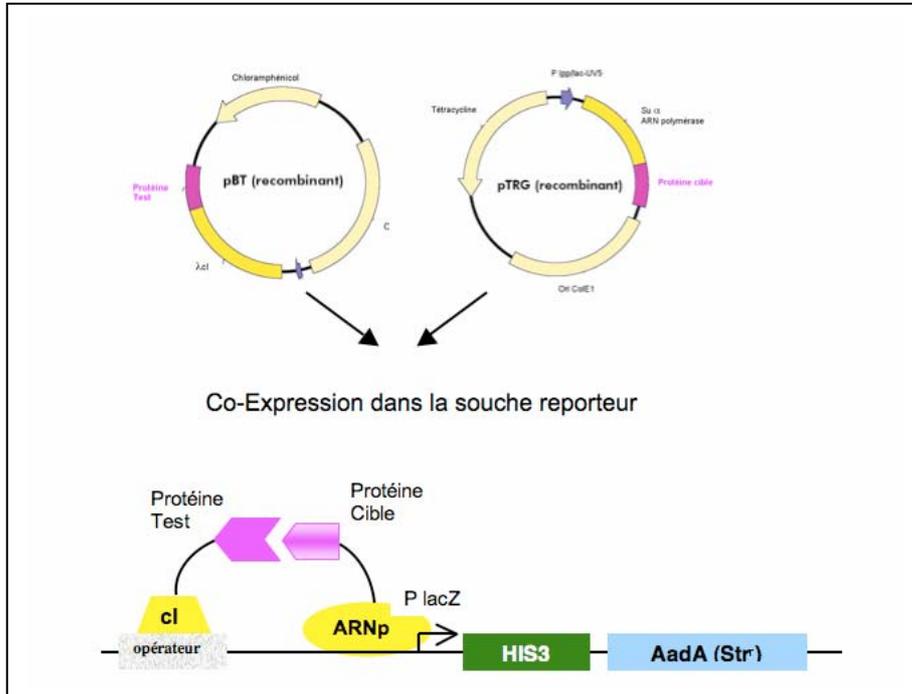


Figure 18. Système Double-Hybride du *BacterioMatch System II*.

III.2.2. Constructions plasmidiques

Nous avons réalisé quatre constructions plasmidiques, contenant les protéines de fusion suivantes : λcI -Fbxo47, λcI - δ [1-155]-Fbxo47 qui est une forme sans domaine F-box, λcI -Skp2, α ARNp-Skp1.

Les deux séquences FBXO47 et Δ [1-465]FBXO47 ont été amplifiées à partir du plasmide pré-existant pGEMT-FBXO47 avec la Taq Polymérase Herculese (Stratagène), avec les couples d'amorces suivants : 5'-CGGGATCCCGAAGAAGAATGGCATC-3' (site BamHI souligné) et 5'-CCCTCGAGGGGGTAGACAGAGGAGTATTCA-3' (site XhoI souligné) pour la forme complète de Fbxo47, 5'-CGGGATCCCGAGCTCCTATGCAGTGT-3' (site BamHI souligné) et 5'-CCCTCGAGGGGGTAGACAGAGGAGTATTCA-3' (site XhoI souligné) pour la forme sans F-Box de Fbxo47. Ces produits de PCR ont été doublement digérés par BamHI et XhoI et ligués dans le plasmide pBT, également digéré par BamHI et XhoI.

La séquence SKP2 a été amplifiée à partir du plasmide pré-existant pCR3.1-Skp2 avec la Taq Polymérase Herculase (Stratagène), avec le couple d'amorces suivant : 5'-CCCTCGAGGGCTGCGGACGCTATGCAC-3' (site XhoI souligné) et 5'-CCCTCGAGGGTCATAGACAACTGGGCTTTTG-3' (site XhoI souligné). Ce produit de PCR a été digéré par XhoI et ligué dans le plasmide pBT, également digéré par XhoI.

Enfin, la séquence SKP1 a été amplifiée à partir du plasmide pré-existant pCR3.1-Skp1 avec la Taq Polymérase Herculase (Stratagène), avec le couple d'amorces suivant : 5'-CGGGATCCCGGAACACCATGCCTTCA-3' (site BamHI souligné) et 5'-CCCTCGAGGGCTTCTCTTCACACCACTG-3' (site XhoI souligné). Ce produit de PCR a été doublement digéré par BamHI et XhoI et ligué dans le plasmide pTRG, également digéré par BamHI et XhoI.

Les vecteurs d'expression ainsi obtenus produisent donc des protéines de fusion λ Ci-Fbxo47 et λ Ci- δ [1-155]-Fbxo47, λ Ci-Skp2 et α ARNp-Skp1. Toutes les constructions ont été séquencées (ABI 3100, Applied Biosystems) afin de confirmer la phase ouverte de lecture entre la protéine d'intérêt et la protéine λ Ci ou la protéine α ARNp.

III.2.3. Amplification et Préparation de la banque d'ADNc

Nous avons utilisé une banque commerciale d'ADNc de rein (Stratagène), dont les ADNc sont déjà sous-clonés dans le vecteur pTRG. Cette banque est fournie dans la souche bactérienne XL1-Blue MRF^rKan qui possède le génotype suivant : (mcrA)183 (mcrCB-hsdSMR-mrr)173endA1 supE44 thi-1 recA1 gyrA96 relA1 lac[F^r proAB lacI^qZ M15 Tn5(Kan^r)].

Cette banque, fournie en 1 ml de stock glycérol avec un titre d'environ 75.10^8 cfu/ml, devait être amplifiée sur des boîtes LB tétracycline (25 cm* 25 cm), chacune devant contenir environ 10^5 cfu dans un volume d'un ml. Étant donné le titre primaire de cette banque ($3,76.10^6$ cfu), nous avons étalé 5 μ l repris dans 38 ml de SOC sur 38 boîtes LB tétracycline 10 μ g/ml. L'incubation dure 24 heures à 30°C et à l'abri de la lumière. Toutes les colonies ayant poussé sur l'ensemble des boîtes ont ensuite été récoltées avec un râteau et reprises dans un même tube avec un volume minimum de milieu SOC. Ceci nous a permis de constituer la banque amplifiée. L'étalement de 2 et 5 μ l de cette banque sur des boîtes LB tétracycline 10 μ g/ml nous a permis de titrer la banque amplifiée en cfu/ml.

Enfin, la moitié des colonies récoltées a été utilisée afin d'extraire les ADN_s plasmidiques nécessaires pour le criblage de la banque. Cette préparation d'ADN a été réalisée avec le kit « HiSpeed Plasmid Maxi Kit » (Quiagen) selon les recommandations du fournisseur.

L'autre moitié de la banque amplifiée a été mélangée à 0,2 volume de glycérol 80%, aliquotée en tube de 1,5 ml, puis congelée à -80°C .

III.2.4. Criblage de la banque

L'expérience de criblage consiste à étaler des cotransformants pBT-recombinant + pTRG-banque d'ADN_c sur un milieu sélectif dans un nombre suffisant pour couvrir la complexité de la banque d'ADN_c. Le nombre de réactions de co-transformation requises a été déterminé par une expérience préliminaire nommée « co-transformation pilote ».

- Co-transformation pilote et Co-transformation contrôle positif

Ces transformations consistent à transformer 50 ng de pBT recombinant (ou pBT-LGF2) avec 50 ng de pTRG-banque d'ADN_c (ou pTRG-Gal11^P) dans 100 μl de cellules compétentes « BacterioMatch II Screening Reporter » (Stratagène).

La réaction de contrôle positif a ensuite été étalée sur deux types de milieux : un milieu de criblage non sélectif (no 3-AT) et un milieu sélectif (5 mM 3-AT), tandis que la co-transformation pilote (100 μl d'une dilution au $100^{\text{ème}}$) a été étalée uniquement sur du milieu non sélectif. Ces boîtes sont incubées 24 heures à 37°C .

Le nombre de colonies obtenu pour la cotransformation pilote, multiplié par 5000 nous a permis ainsi d'estimer le nombre de cotransformations (en 500 μl) nécessaire pour couvrir toute la banque lors du criblage.

Les compositions des milieux et les protocoles détaillés des co-transformations sont disponibles en annexe B.

- Criblage de la banque

Il nécessite trois étapes : un criblage initial, un enrichissement des étalements, un criblage secondaire (Figure 19).

La première étape consiste à effectuer le nombre calculé précédemment de co-transformations. Ces réactions sont réalisées dans un volume de 500 μl de cellules compétentes « *BacterioMatchII Screening Reporter* » (Stratagène) avec 200ng de pBT recombinant (contenant la protéine d'intérêt) et 200 ng de pTRG contenant la banque. La

totalité du volume de chaque co-transformation est étalée par aliquots de 300 μ l sur du milieu sélectif (5 mM 3-AT). Les boîtes sont incubées à 37°C pendant 24 heures.

Les colonies ayant poussées sont alors repiquées en stries sur d'autres boîtes 5 mM 3-AT (Boîte d'enrichissement 1). Les boîtes du criblage initial sont maintenues à température ambiante dans le noir pour 15 à 16 heures de plus, ce qui permet aux interacteurs plus faibles de s'exprimer. Après cette période additionnelle de pousse, les colonies supplémentaires ont également été repiquées en stries sur d'autres boîtes 5 mM 3-AT (Boîte d'enrichissement 2). Enfin, le criblage secondaire permet de vérifier les interacteurs positifs en utilisant le second gène reporteur. Les colonies présentes sur les boîtes d'enrichissement sont repiquées d'une part sur un milieu doublement sélectif, 5 mM 3-AT / 15 μ g streptomycine, et d'autre part sur un milieu de maintenance, 12,5 μ g chloramphénicol / 6,25 μ g tétracycline (Boîte de maintenance). L'ensemble des boîtes a été mis à incuber 24 heures dans une étuve à 37°C, puis 16 heures supplémentaires à température ambiante et à l'abri de la lumière.

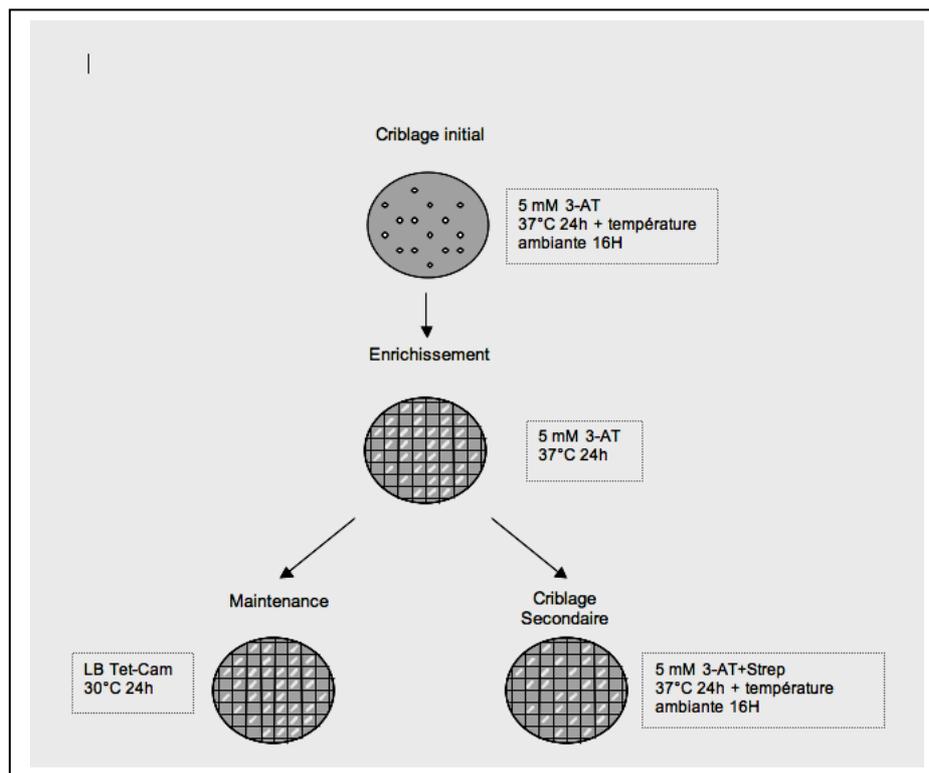


Figure 19. Criblage de la banque d'ADNc dans la technique de double-hybride.

III.2.5. Validation des interactions protéine-protéine détectées (Figure 20)

L'ADN plasmidique des colonies positives lors du criblage secondaire a été extrait à l'aide du kit *FastPlamid MiniPrep* (Ependorf) à partir des stries bactériennes correspondantes de la boîte de maintenance. Cet ADN purifié a été utilisé pour transformer la souche *XLI-blue MRF'Kan* (Stratagène). La réaction est étalée sur LB tétracycline de manière à éliminer par pression de sélection le plasmide pBT recombinant de la colonie. Les colonies obtenues ont alors été repiquées à la fois sur LB tétracycline et sur LB chloramphénicol, de manière à identifier les colonies tet^R cam^S . Les bactéries possédant ce phénotype ont été alors cultivées afin d'en extraire leur ADN plasmidique correspondant à un pTRG contenant un ADNc d'interacteur protéique de la protéine d'intérêt (Figure 20).

L'ADN extrait et purifié, selon des techniques standardisées, a été à nouveau co-transformé avec le pBT contenant la protéine d'intérêt d'une part (Réaction A), et avec le pBT vide d'autre part (Réaction B) dans des cellules compétentes « *BacterioMatch II Validation Reporter* » (Stratagène). Ces deux réactions sont testées à la fois sur milieu sélectif (5 mM 3-AT) et sur milieu non sélectif (no 3-AT). Un ADNc code pour un partenaire protéique validé uniquement si la réaction A pousse sur les deux milieux et si la réaction B ne pousse pas sur le milieu sélectif.

Dans ce cas, l'ADNc du clone a été séquencé (ABI 3100, Applied Biosystems) et les résultats de séquence comparés aux banques nucléiques et protéiques disponibles en ligne.

III.2.6. Double Hybride dirigé

Le double-hybride dirigé consiste à co-transformer, dans une souche bactérienne reporter adaptée (ici « *BacterioMatch II Validation Reporter* »), un plasmide pBT contenant une protéine X connue et un pTRG contenant une protéine Y connue également.

La technique est donc la même que celle qui a été utilisée pour valider définitivement les interacteurs positifs obtenus lors du criblage de la banque d'ADNc.

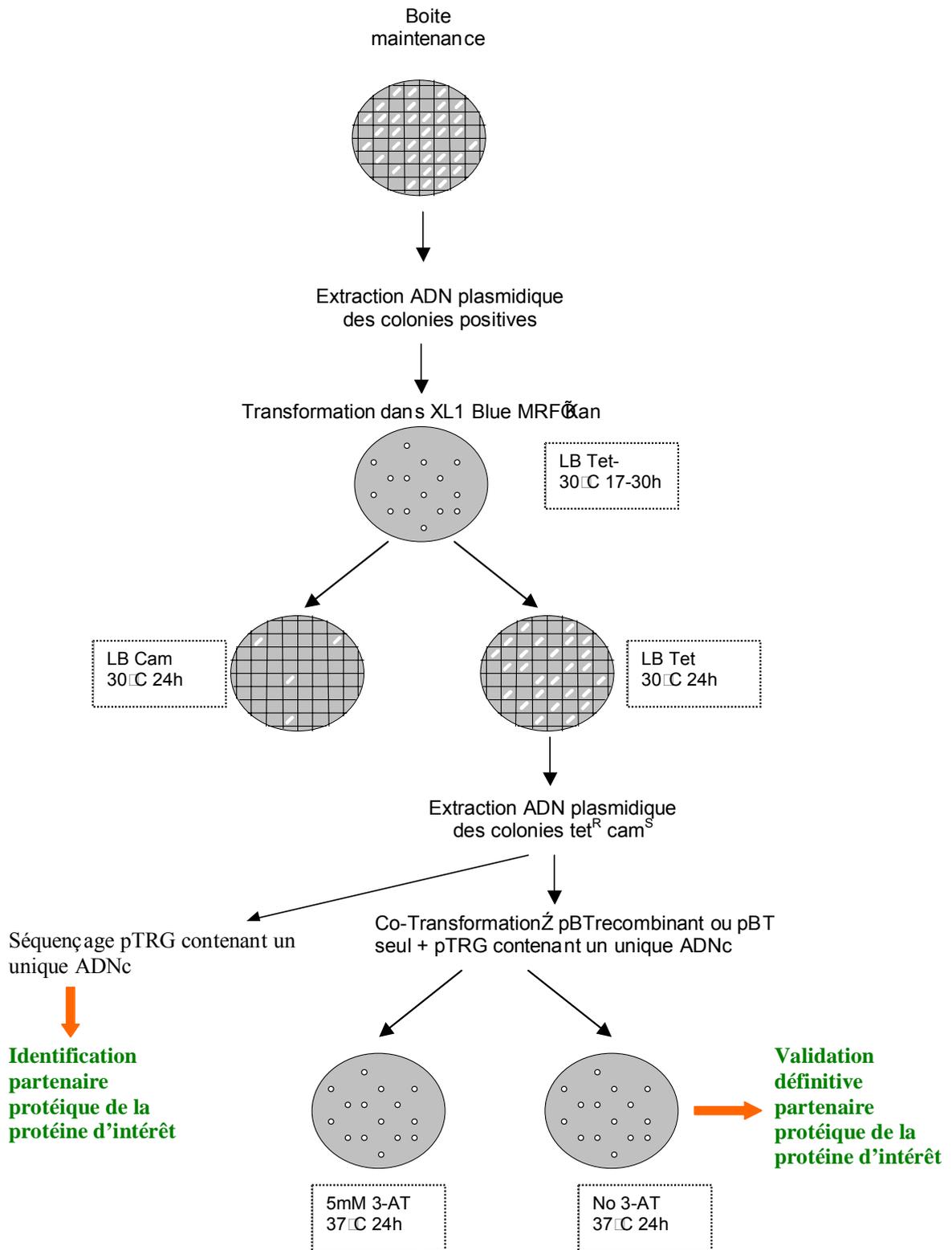


Figure 20. Validation des interacteurs positifs dans la technique de double hybride.

CHAPITRE 3

RESULTATS

A. Identification d'un nouveau gène localisé en 17q12

I. Étude de réarrangements génomiques dans la région 17q12

Étude de perte d'hétérozygotie et Cancers papillaires rénaux

L'approche par LOH consiste à étudier des microsatellites chez des sujets atteints de cancer en comparant leur ADN tumoral à leur ADN constitutionnel²¹⁰. Si l'un des allèles a disparu (hétérozygotie somatique), cela signifie que la région chromosomique correspondante a été perdue dans les cellules tumorales.

Nous avons déterminé la perte d'hétérozygotie à l'aide de huit marqueurs microsatellites (sept répartis sur le bras long du chromosome 17 et un sur le bras court) de 15 cas de cancers papillaires rénaux (pRCC) (Figure 21).

Huit patients présentent une altération allélique (LOH) pour au moins un des huit locus chromosomiques étudiés (patients 4 à 11). Le tableau 2 résume la fréquence des cas informatifs pour chaque marqueur, c'est-à-dire hétérozygote dans le tissu sain. Des altérations alléliques sont présentes dans sept loci étudiés sur huit, avec une fréquence variant de 8,3 % pour le marqueur D17S933 situé en 17q11.2 à 50 % pour le marqueur D17S250 situé en 17q12.

Pour les cas informatifs, une perte d'hétérozygotie est présente à 50 % (7/14) au locus D17S250 et à 44 % (4/9) au locus D17S800. Seulement 8,3 %, 33%, 37,5% et 30% des cas informatifs se trouvent être délétés pour les marqueurs D17S933, D17S579, D17S588 et D17S1290.

De façon intéressante, pour deux patients (n°4 et 10, Figure 21), une perte d'hétérozygotie est observée uniquement au niveau du marqueur D17S250.

La perte d'hétérozygotie au locus D17S250 étant de plus la plus fréquente, nous permet de définir une zone minimale de délétion, autour du marqueur D17S250, qui réduit la région d'intérêt pour la recherche du gène candidat.

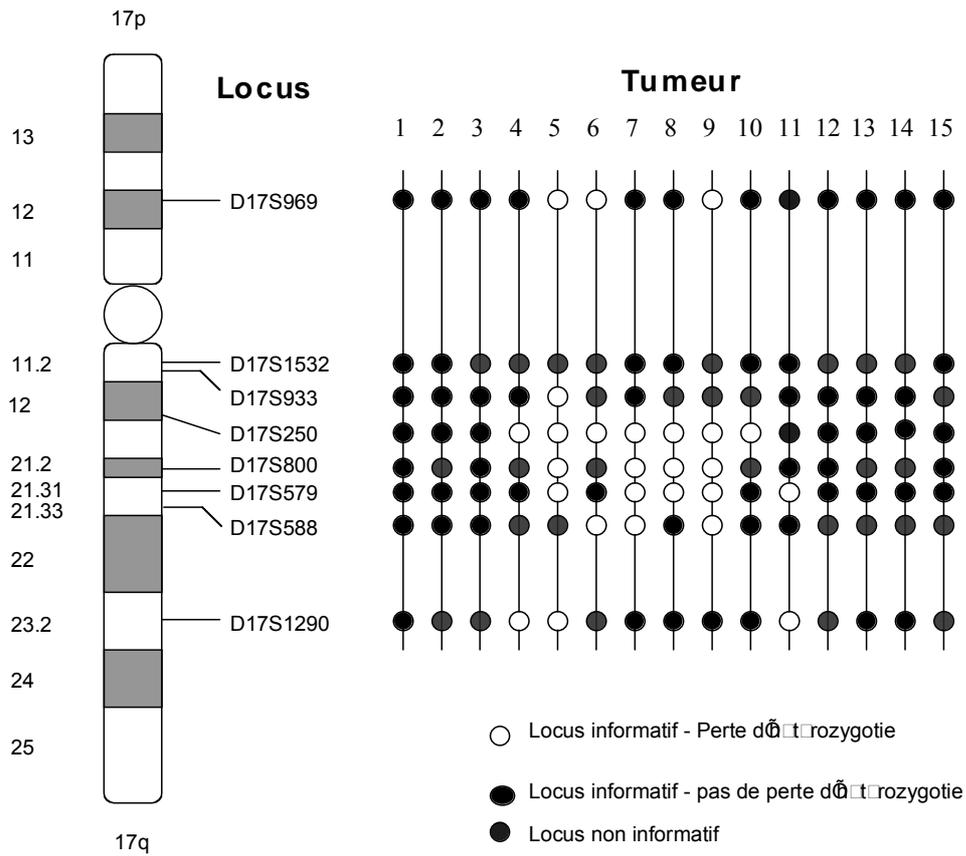


Figure 21. Résultats de l'analyse LOH des 15 cas de pRCC à l'aide de huit marqueurs microsatellites répartis sur le chromosome 17.

Locus	Localisation	LOH/Cas informatif	% LOH
D17S969	17p12	3/14	21,4 %
D17S1532	17p11.2	0/7	0%
D17S933	17q11.2	1/12	8,3%
D17S250	17q12	7/14	50%
D17S800	17q21.2	4/9	44%
D17S579	17q21.31	5/15	33,3%
D17S588	17q21.33	3/8	37,5 %
D17S1290	17q23.2	3/10	30%

Tableau 2. Fréquences de perte d'hétérozygotie des huit microsatellites étudiés sur les 15 cas de cancers papillaires rénaux.

II. Analyse bioinformatique de la région encadrant le marqueur D17S250

II.1. Définition de la zone d'étude

Nous avons défini une région d'analyse d'une taille de 100 kb à l'intérieur du contig AC006441, correspondant à la séquence située entre les bases 34330400 et 34430400 du chromosome 17, définie d'après les deux patients conservant les deux marqueurs adjacents à D17S250 (Figure 22).

Le marqueur D17S250 se situe entre les bases 34405618 et 34405769 sur cette séquence.

Comme le montre la figure 22, il y a peu de gènes connus dans cette région à l'exception de la partie 3' du gène *LASP1*, dans l'extrémité centromérique.

II.2 Annotation des séquences répétées

L'identification et le masquage des séquences répétées sont un préalable indispensable à la recherche de séquences codantes. Nous avons utilisé le logiciel *RepeatMasker* et les paramètres par défaut. Ce programme recherche les séquences répétées, les différents types de répétition nucléotidiques, ainsi que des régions (stretch) extrêmement riches en AT (>84 %) ou GC (>87 %). Parmi les séquences répétées, on trouve les SINE (*Short Interspersed Nucleotide Element*) de type séquence Alu, fragment d'ADN d'environ 300 pb présent en moyenne tous les 3 à 10 kb²¹¹⁻²¹², les SINE de type MIR²¹³, les LINE (*Long Interspersed Nucleotide Element*)²¹⁴, et les LTR (*Long Terminal Repeats*)²¹⁵. En moyenne, 40 % de l'ADN génomique se trouve masqué par ce programme.

Le tableau 3 présente la distribution globale des séquences répétées dans notre région d'intérêt. Le pourcentage total (63,5 %) est plus élevé que la moyenne en séquences répétées que le chiffre généralement observé dans les régions codantes du génome.

Une autre information apportée par le programme *RepeatMasker* concerne la richesse en bases GC de la séquence, qui permet de déduire l'isochore de la région concernée souvent corrélée avec la richesse en gènes²¹⁶. Notre région possède un pourcentage en GC moyen (43 %), elle est donc, a priori peu riche en séquences codantes.

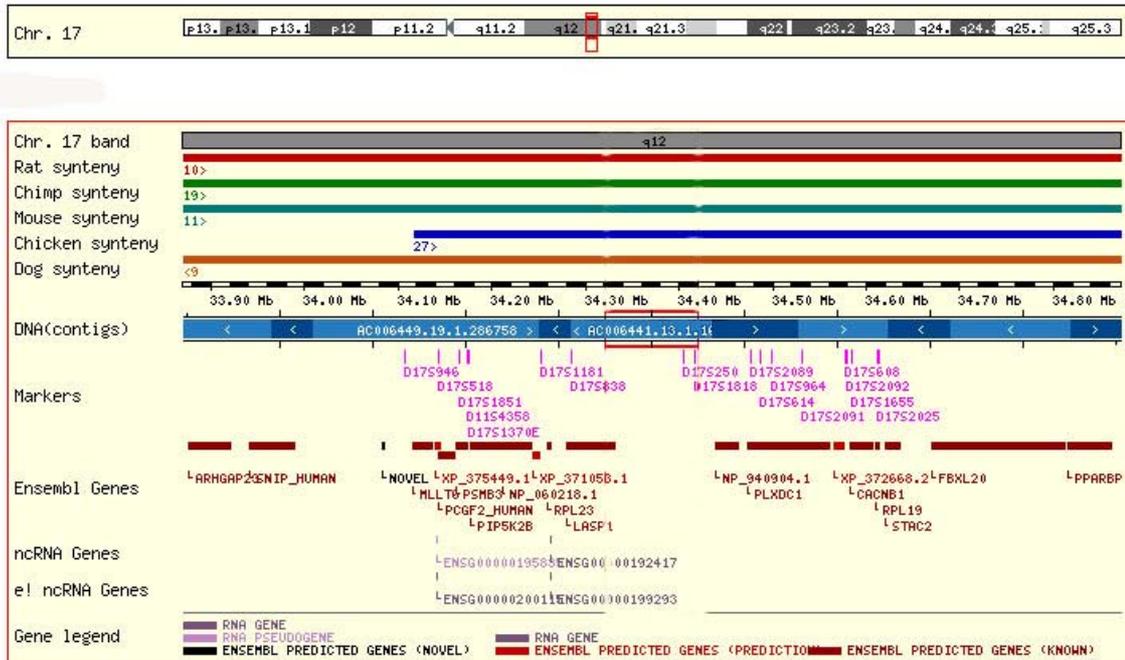


Figure 22. Représentation de la région génomique d'intérêt pour notre recherche de gène candidat. Les 100 kilobases concernées sont encadrées en rouge, à l'extrémité droite se trouve le marqueur D17S250, et à gauche la partie 3' du gène LASP1.

	Nombre d'éléments	% de la séquence
SINE	167	44,248
ALU	153	43,191
MIR	14	1,057
LINE	62	12,067
LINE 1	44	8,046
LINE 2	15	3,716
CR1	3	0,305
LTR	15	4,107
MaLR	10	1,118
ERV1	5	2,989
DNA	10	1,77
MER1_type	6	0,957
MER2_type	1	0,743
Mariners	1	0,07
AcHobo	2	0,108
Autres	1	1,272

Tableau 3. Répartition globale des séquences répétées dans la séquence analysée de 100 kb.

II.3. Prédiction d'exons et de gènes pour la région d'intérêt

II.3.1 Prédiction d'exons, recherche de phase ouverte de lecture

Nous avons utilisé en parallèle différents programmes sur notre séquence d'intérêt. Chaque programme possédant ses propres caractéristiques, nous les résumons dans le tableau ci-dessous :

Programme	Type de prédiction	Longueur maximale de la séquence analysée	Qualité ou score affecté aux exons prédits
GRAIL EXP 3.3	Structure de gènes	200 kb	<i>Excellent</i> <i>Bon</i> <i>Pauvre</i> <i>Marginal</i>
GENSCAN	Structure de gènes	500 kb	Bon score >5
FGENES 1.6	Structure de gènes	500 kb	Bon score >5

Les résultats des programmes de prédiction d'exons pour notre séquence sont résumés dans les tableaux 4 et 5.

Pour l'analyse correspondant au brin «*forward*» de la séquence (5'→3'), nous observons 11 exons prédits au total, chacun n'étant prédit que par un seul programme. Si l'on s'intéresse aux scores attribués à chaque exon, on constate que seulement un exon (exon 8F) évalué par *GENSCAN* présente un bon score (>5), 2 exons (exons 4F et 6F) sont considérés comme « bon » par *GrailExp*, et aucun des exons prédits par *FGenes* n'obtient de bon score.

Sur le brin «*reverse*» de cette même séquence (3'→5'), les logiciels localisent 21 exons, dont 2 prédits par 2 programmes et 4 prédits par les trois systèmes de recherche d'exons.

Sur ces 6 exons, 5 ont des bornes 5' et 3' identiques, et l'une des bornes en commun. Par l'intermédiaire de *GrailExp*, on trouve 4 exons excellents (exons 6R, 13R, 15R et 17R) et un bon (exon 18R). Les exons avec un bon score pour *GENSCAN* sont les exons 13R, 15R, 16R et 21R. Pour les 2 exons prédits par deux programmes, l'un possède les mêmes bornes 5' et 3' (exon 16R), l'autre a une extrémité 3' différente (exon 7R). Pour les exons prédits par les trois programmes (exons 13R, 14R, 15R et 17R), les bornes sont toujours identiques.

Exons prédicts	GrailExp Position sur la séquence	Q	FGENES 1.6 Position sur la séquence	Score	GENSCAN Position sur la séquence	Score
Exon1 F			2419 2613	3,46 ini		
Exon2 F			5521 5649	0,77 int		
Exon3 F			6204 6329	2,90 ter		
Exon4 F	6567 6667	g int				
Exon5 F					41034 41114	1,68 ini
Exon6 F	41917 42012	g ter				
Exon7 F	76034 76320	p ini				
Exon8 F					76279 76418	9,93 int
Exon9 F	85073 85320	m ter				
Exon10 F	89765 89873	m ini				
Exon11 F					89896 90175	4,46 ter

Tableau 4. Prédiction des exons sur la séquence de 100 kb (brin « forward »).

Les exons sont classés selon leur ordre du centromère au télomère. Pour le logiciel GrailExp, la qualité de la prédiction (Q) est de 4 niveaux : m (marginale), p (pauvre), g (bonne), e (excellente). Pour les autres programmes, un score est indiqué. Dans le tableau se trouve également indiqué s'il s'agit d'un exon 5' du gène (ini), 3' (ter), ou interne (int).

Exons prédicts	GrailExp Position sur la séquence	Q	FGENES 1.6 Position sur la séquence	Score	GENSCAN Position sur la séquence	Score
Exon1 R			9719 9734	4,39 ini		
Exon2 R			10511 10574	0,64 int		
Exon3 R			10728 10755	2,59 ter		
Exon4 R			20442 20474	5,05 ini		
Exon5 R			30904 30981	1,01 int		
Exon6 R	34393 34432	e ter				
Exon7 R	44949 45030	m ini	44948 44980	3,55 ter		
Exon8 R	49263 49386	m int				
Exon9 R			57188 57368	2,99 ini		
Exon10 R			58166 58336	0,83 int		
Exon11 R			62690 63036	2,11 int		
Exon12 R					67518 67523	-1,30 ini
Exon13 R	68524 68632	e int	68624 68632	2,70 int	68624 68632	11,24 int
Exon14 R	75196 75253	m ini	75077 75253	1,30 int	75077 75253	3,47 int
Exon15 R	76477 76607	e int	76477 76623	2,32 int	76477 76623	13,89 int
Exon16 R			77293 77435	2,29 int	77293 77435	6,95 int
Exon17 R	81481 81640	e int	81481 81640	1,64 int	81481 81640	2,64 int
Exon18 R	82923 83038	g ter				
Exon19 R			88900 89033	4,47 ter		
Exon20 R					94419 94562	3,63 int
Exon21 R					97057 97409	6,06 ter

Tableau 5. Prédiction des exons sur la séquence de 100 kb (brin « reverse »).

Les exons sont classés selon leur ordre du télomère au centromère. Pour le logiciel GrailExp, la qualité de la prédiction (Q) est de 4 niveaux : m (marginale), p (pauvre), g (bonne), e (excellente). Pour les autres programmes, un score est indiqué. Dans le tableau se trouve également indiqué s'il s'agit d'un exon 5' du gène (ini), 3' (ter), ou interne (int). En vert, sont indiqués les exons prédits par 2 programmes et en rouge les exons prédits par les trois programmes utilisés. En gras, les exons retenus pour former une ORF candidate.

L'analyse de l'ensemble de ces résultats a permis de déterminer une ORF potentielle identifiée sur le brin reverse de la région d'intérêt à l'aide du logiciel *DNA Strider*. Le cadre de lecture ouvert obtenu (couvrant environ 40 kb) englobe les exons 9R, 10R, 11R, 13R, 14R, 15R, 16R, 17R et 18R (Tableau 5).

II.3.2. Recherche d'homologie chez la Souris

Étant donné la conservation des gènes entre les espèces et l'existence d'une région de synténie entre la région 17q et le chromosome 11 de la souris, nous avons comparé la région de 40 kb (contenant l'ORF potentielle) au génome de la souris à l'aide du logiciel BLAST. Il est apparu de nombreuses régions d'homologie dues à la synténie des chromosomes. Néanmoins, lorsqu'on s'intéresse aux exons déjà retenus à l'aide des logiciels de prédiction de gènes, on s'aperçoit que les 9 exons retenus sur le chromosome 17 humain s'alignent sur le brin sens du chromosome 11 avec une moyenne de 87,44 % d'identité des acides nucléiques (Tableau 6). Ceci renforce la prédiction des exons obtenus en II.3.1.

Il existe également une séquence d'homologie sur le chromosome 11 murin avant le premier exon et entre les exons 3 et 4. Ces deux séquences peuvent correspondre à des exons supplémentaires de l'ORF bien qu'ils ne soient pas prédits par les programmes utilisés précédemment sur le génome humain. En effet, ils contiennent tous deux des caractéristiques d'exons, c'est-à-dire qu'ils concordent avec la séquence d'épissage consensus AT/GT sur notre région d'intérêt, enfin ils forment une phase ouverte de lecture cohérente avec notre ORF candidate.

Notons également que l'exon terminal, qui n'était prédit que par un seul programme (*GrailExp*) est également présent chez la souris avec une homologie de 87 %.

N° exon prédit chez la souris	Position sur le chromosome 11 murin (pb)- taille exon (pb)	Position sur le chromosome 17 humain (pb)- taille exon (pb)	% Identité Homme/Souris
Exon 1	9284399 9284480 81	52904 52985 81	90
Exon 2	9279861 9280041 180	57188 57368 180	81
Exon 3	9278432 9278606 174	58166 58336 170	85
Exon 4	9273333 9273413 80	62952 63037 85	91
Exon 5	9269069 9269143 74	65280 65358 78	92
Exon 6	9265535 9265647 112	68524 68632 108	87
Exon 7	9263835 9264014 179	75077 75253 176	88
Exon 8	9258145 9258289 144	76477 76623 146	89
Exon 9	9256631 9256777 146	77293 77435 142	89
Exon 10	9255844 9255990 146	81481 81640 159	85
Exon 11	9254869 9254966 97	82923 83019 96	87

Tableau 6. Comparaison entre la séquence génomique contenant les exons candidats sur le chromosome 17 humain et le chromosome 11 de la Souris.

En bleu, les exons retenus, prédits par plusieurs programmes et présents chez la souris, en vert les exons retenus, prédits par un seul programme et présents chez la souris, en rouge et en gras, les exons non prédits, présents chez la souris et compris dans la cadre de lecture candidat du gène candidat humain.

II.4. Synthèse des résultats : définition d'une ORF potentielle

La figure 23 résume la structure potentielle du gène, à partir de laquelle nous avons biologiquement validé son existence.

Ce gène s'étend sur environ 31 kb, comporte 11 exons dont 10 forment une phase de lecture ouverte. Le premier exon ne serait donc pas codant, l'ATG se trouvant dans l'exon 2. Par contre le dernier exon, confirmé par comparaison d'homologie, contient bien le codon stop putatif.

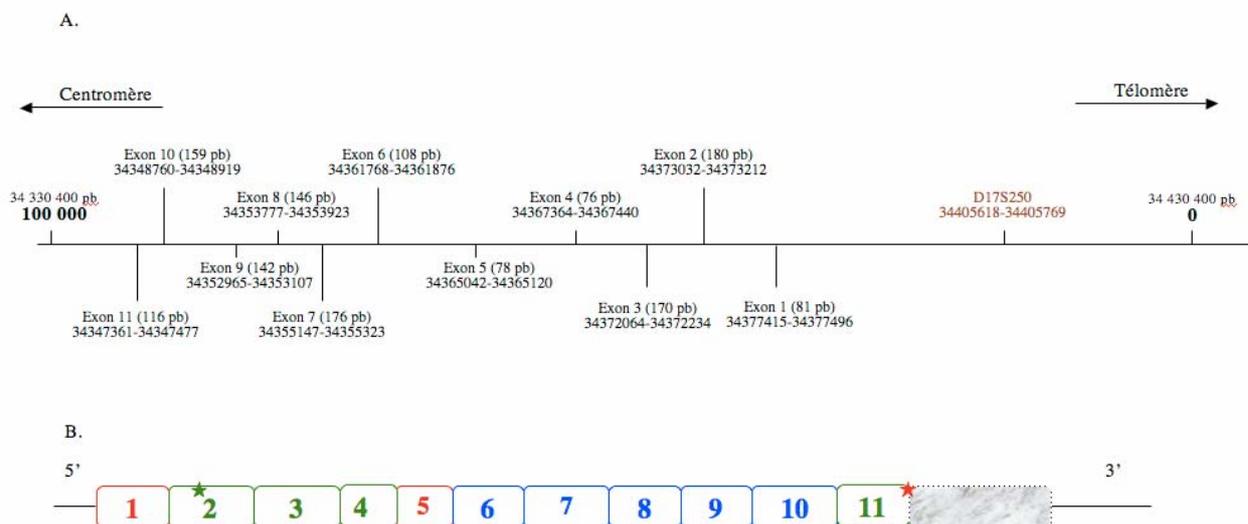


Figure 23. Définition *IN SILICO* de la structure du gène candidat (31kb, 11 exons).

A. Répartition et taille des exons formant l'ORF candidate sur la région d'intérêt analysée.

B. Représentation schématique du gène cloné *in silico* par rapport au tableau 6.

L'étoile verte représente le codon START (ATG) et l'étoile rouge le codon STOP (TAA), la partie grisée représentant la partie 3' du gène.

III. Validation biologique des transcrits potentiels

Cette étape a nécessité deux techniques. Premièrement, la PCR exon-connection est employée pour valider la structure prédite *in silico*. Il s'agit en fait de désigner des amorces PCR soit à l'intérieur des exons soit à cheval sur deux exons, ceci permettant d'amplifier un produit à partir d'ADNc uniquement lorsque la construction prédite est la bonne et de s'affranchir de toute contamination génomique. Cette technique doit nous permettre d'obtenir l'ensemble de la phase ouverte de lecture. Deuxièmement, la RACE-PCR, dont le but est de caractériser les régions non traduites de l'ARNm, c'est-à-dire les régions 5' et 3'.

La validation en biologie moléculaire implique le séquençage des produits PCR et est donc soumise à des limites en terme de quantité de matériel. Pour les deux techniques, afin de pallier ces problèmes de quantité, nous avons dû utiliser des PCR ancrées (*Nested-PCR*), c'est-à-dire ré-amplifier un produit PCR avec des amorces internes. Cette technique augmente la sensibilité et la spécificité de la réaction.

III.1. PCR exon-connection : structure de la phase ouverte de lecture

Nous avons désigné des oligonucléotides répartis sur la séquence formant l'ORF candidate (Figure 24). Ils nous ont permis d'effectuer diverses PCRs afin de valider le gène candidat sur de l'ADNc de lignée de côlon.

Dans un premier temps, nous avons cherché à amplifier les exons les plus prédits à l'aide des amorces As et Fas. Nous avons obtenu un produit PCR d'environ 800 pb. Cette bande étant trop faible pour être analysée directement, nous avons réalisé des PCR ancrées sur ce produit avec les deux couples d'amorces suivants, As/Cas et Cs/Fas. Les produits obtenus sont de la taille attendue. Ceux-ci ont alors été purifiés et séquencés.

Ce premier résultat nous a permis la validation des exons 3, 4, 5, 6, 7 et 8.

Dans un deuxième temps, les amorces Cs et 3' nous ont fourni un produit PCR d'une taille de 850 pb à partir duquel nous avons ensuite réalisé 4 PCRs ancrées avec les couples d'amorces Ds/Fas, Ds/Gas, Es/Gas et Fs/Gas. Elles ont amplifié, respectivement, des produits de 274, 551, 352 et 283 pb, qui ont directement été séquencés et nous ont permis la validation de la structure des exons 9, 10 et 11.

L'ensemble des résultats obtenus nous a permis de désigner les oligonucléotides –1s/3' couvrant l'ensemble de l'ORF. Ainsi, dans une troisième et dernière expérience d'exon-

connection, nous avons amplifié une bande d'environ 1400 pb, qui après ré-amplification, purification, sous-clonage et séquençage, a permis de confirmer la structure entière de la phase ouverte de lecture du gène candidat de l'exon 2 à l'exon 11 avec 300 pb supplémentaires (après le codon stop) dans la partie 3'.

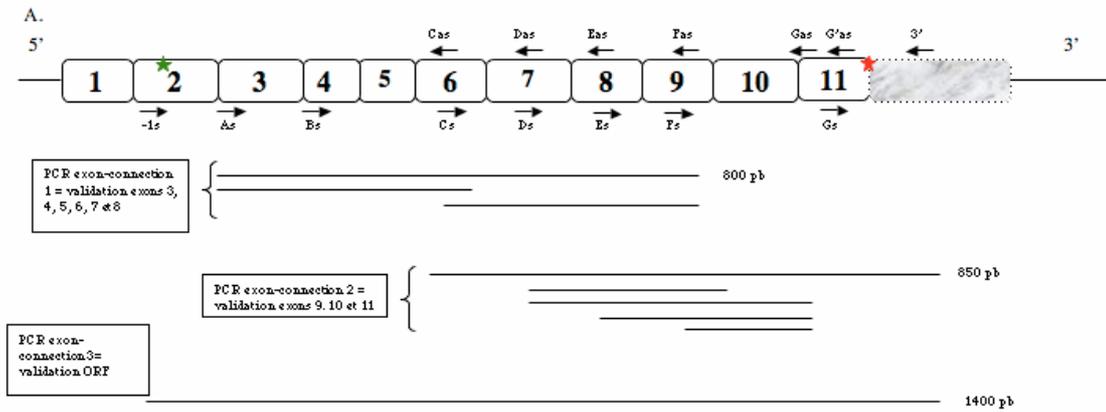


Figure 24. PCR exon-connection.

Répartition et localisation des amorces utilisées pour la PCR exon-connection et schématisation des produits PCRs obtenus (lignes). L'étoile verte représente le codon START (ATG) et l'étoile rouge le codon STOP (TAA).

III.2. Structure des extrémités 5' et 3' non codantes

III.2.1. RACE-PCR

Ce type de PCR nous a permis d'obtenir les parties 5' et 3' du gène à partir d'ADNc de testicule commercialisé par Clontech.

Pour ce faire, nous avons désigné 4 oligonucléotides anti-sens spécifiques situés dans les exons 7, 9 et 11 afin d'élonger vers la partie 5' du gène (Figure 25). Nous avons obtenu par cette première amplification 4 produits dont la taille, concorde avec la description de l'ORF potentielle (Figure 26A). Étant donné la faible amplification pour trois des réactions, nous avons ensuite réalisé des PCR ancrées sur ces produits de PCR à l'aide d'oligonucléotides spécifiques internes à la séquence pré amplifiée, situés respectivement dans les exons 6, 8, 9 et 11 (Figure 25). Les résultats sont présentés sur la figure 26C.

L'amplification secondaire, plus spécifique, nous apporte cette fois la quantité suffisante de produits PCR exploitables, c'est-à-dire, que nous avons purifiés, sous-clonés et séquencés. Cette expérience nous a confirmé l'épissage de l'exon 1 non codant avec le reste de la phase ouverte de lecture.

Nous avons, de la même façon, effectué une RACE 3', à l'aide de 4 oligonucléotides sens situés dans les exons 1, 2, 3 et 9 (Figure 25).

Cette première amplification n'a fonctionné que pour deux des 4 amorces utilisées (amorce située dans l'exon 1 et amorce située dans l'exon 9) (Figure 26B). Notons, qu'il semble exister plusieurs transcrits alternatifs en 3' puisqu'il y a une double bande pour un des produits obtenus.

Comme pour la RACE 5', nous avons ré-amplifié les produits obtenus avec des amorces internes, situées respectivement dans les exons 2 et 9 (Figure 26D). Pour les deux réactions, nous avons obtenu une amplification de plusieurs bandes, ce qui semble confirmer l'existence de transcrits alternatifs. Néanmoins, il existe, pour chacune, une bande majoritaire. Nous avons purifié sur gel d'agarose cette bande majoritaire ainsi que les deux bandes les plus proches. L'ensemble a été sous-cloné puis séquencé. Seuls les clones contenant les bandes majoritaires étaient exploitables. Ils nous ont permis de définir la région 3' non traduite du gène, qui comporte au moins 710 pb après le codon stop TAA situé dans l'exon 11.

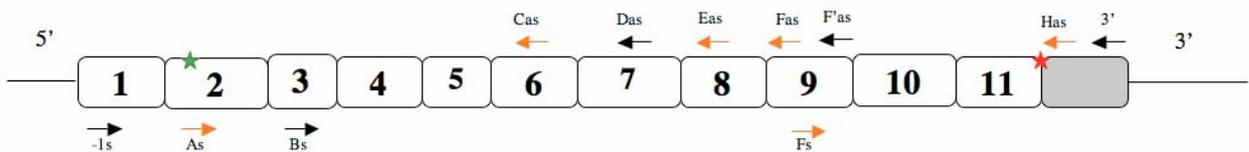


Figure 25. Localisation des oligonucléotides utilisés pour la RACE-PCR.

Répartition et localisation des amorces utilisées pour la RACE-PCR en noir, en orange les amorces utilisées en Nested RACE-PCR.

L'étoile verte représente le codon START (ATG) et l'étoile rouge le codon STOP (TAA).

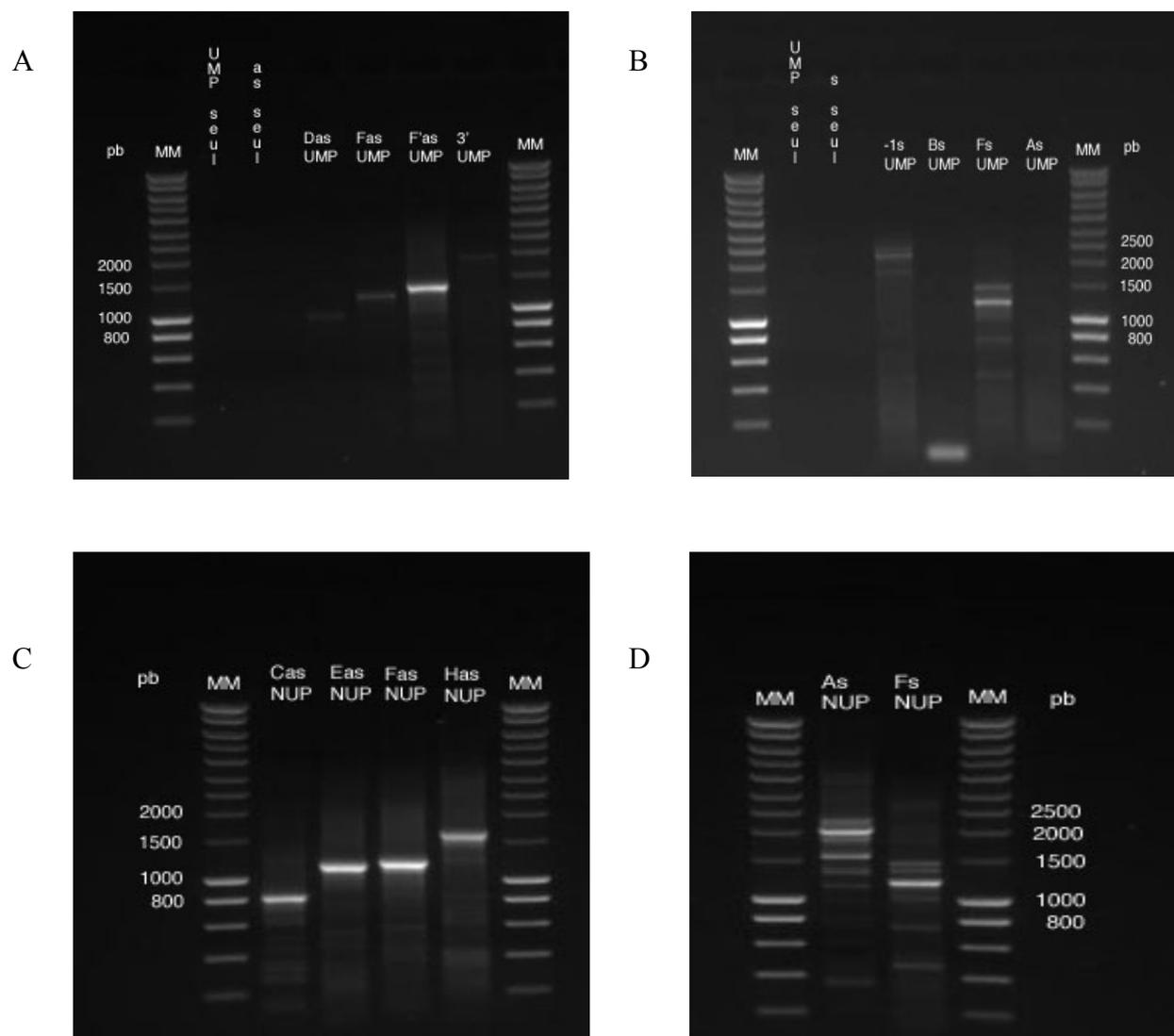


Figure 26. Résultats obtenus par RACE-PCR.

A. RACE-PCR 5'

B. RACE-PCR 3'

C. Nested-RACE 5'

D. Nested-RACE 3'

MM = marqueur moléculaire (Smart Ladder Eurogentec). UMP, NUP voir matériels et méthodes

III.2.2. Recherche de motifs caractéristiques de la structure de gènes

Nous avons recherché les caractéristiques des parties non codantes du gène par bioinformatique (régions 5' et 3' définies par RACE-PCR).

La région promotrice a été analysée pour des sites de liaison putatifs aux facteurs de transcription connus et pour la présence d'une boîte TATA, en utilisant les programmes suivants disponibles sur le serveur Infobiogen : *Fgenes*, *McPromoter*, *CorePromoter*, *Promoteur Inspector*, *TFSEARCH* et *PROSCAN*. Nous avons élargi l'analyse aux 3000 pb en amont de l'exon 1.

Nous avons détecté une zone promotrice candidate contenant une boîte TATA, localisée à 200 pb en amont de l'exon 1 (53673). Des sites putatifs de liaison aux facteurs de transcription Sp1 et UCE.2 ont également été mis en évidence aux positions suivantes : Sp1 (53857) et UCE.2 (53659, 53755). De plus, nous avons identifié un îlot CpG entre les exons 1 et 2 (l'exon 2 contenant l'ATG). Cette zone riche en GC remplit les conditions définies par Gardiner-Garden et Frommer²¹⁷, à savoir qu'elle possède une longueur de plus de 200 pb et une richesse en G+C supérieure à 50% (ici 54,88) avec un ratio CpGo/e supérieur à 0,6 (0,75 dans notre cas).

La région 3' non traduite contient un site de polyadénylation (AATAAA) en position +2057 par rapport à l'ATG. Cette séquence consensus est fréquemment utilisée comme signal de polyadénylation pour les ARNm eucaryotes.

La Figure 27 représente schématiquement la structure et les caractéristiques du gène cloné.

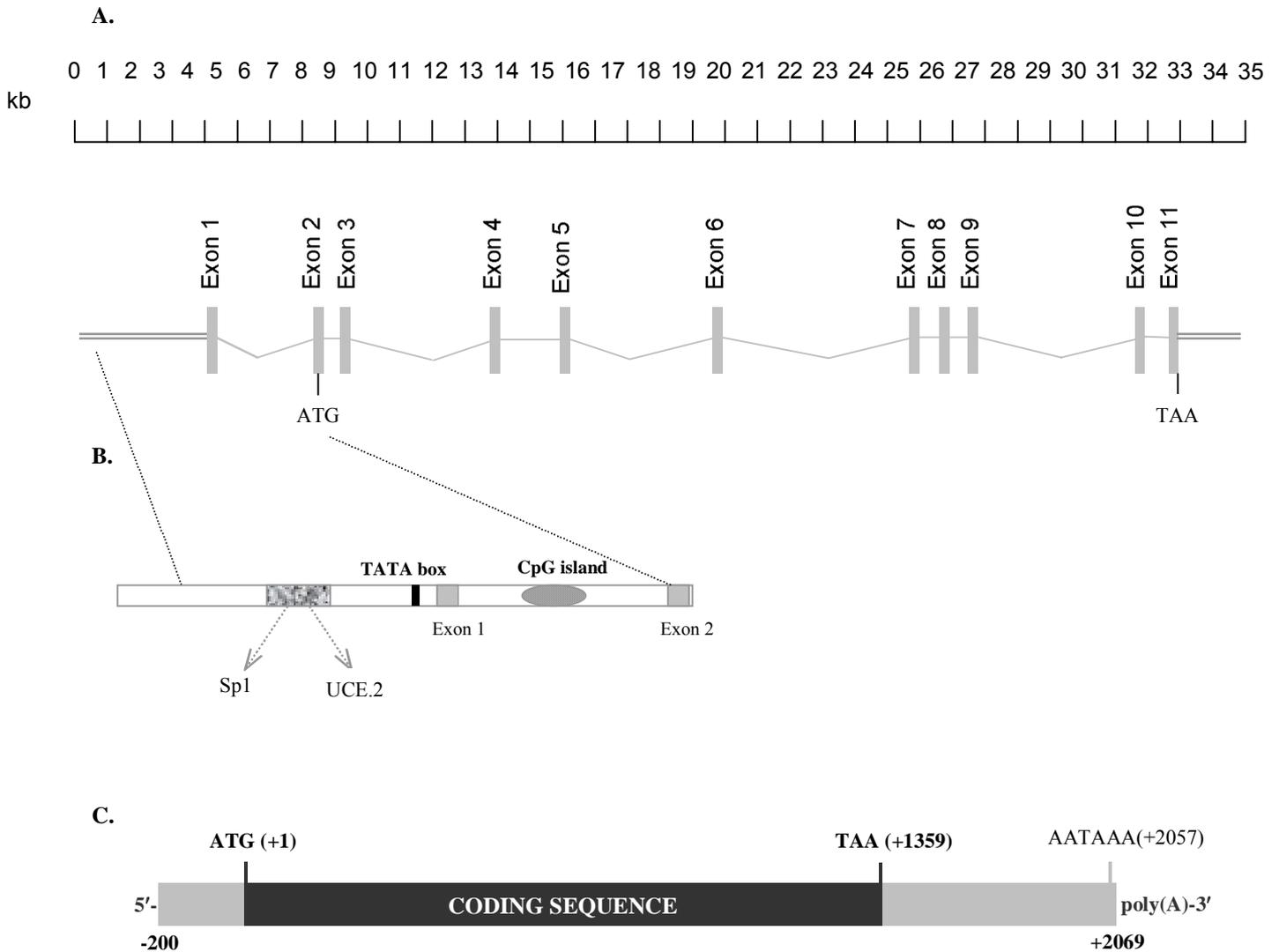


Figure 27. Représentation de la structure génomique du gène FBXO47.

- A. Ce gène couvre 31 kb et comprend 11 exons, avec le premier non codant (barres grises = exons, lignes = introns).
- B. La région promotrice potentielle contient une boîte TATA et des sites putatifs pour Sp1 et UCE.2 et un îlot CpG.
- C. ARNm. La traduction débute 26 pb après le début de l'exon 2. Les positions sont notées à partir de l'ATG.

B. Caractérisation fonctionnelle du transcrit

I. Comparaison aux banques d'EST

Les EST (*Expressed Sequence Tag*) sont des séquences exprimées issues de banques d'ADNc dont on connaît une partie de la séquence. Chacun d'entre eux est défini par cette séquence partielle qui a été réalisée en 5', en 3' ou les deux selon les banques où ils sont disponibles. Il existe une banque de données rassemblant un grand nombre de ces séquences, nommée *dbEST* qui, en novembre 2005, contenait 7 057 754 séquences humaines.

La séquence génomique contenant notre gène (c'est-à-dire environ 40 kb) a été comparée aux banques d'EST par une interrogation directe de *dbEST* par le programme BLAST. Nous n'avons obtenu aucun résultat, ce qui signifie que cette région est vierge de tout ADNc connu.

Au vu de cette interrogation, nous avons comparé l'ensemble de la région d'intérêt (100 kb) à la banque *dbEST*. Nous avons détecté 90 ESTs dont, seulement 8 sont différents. Tous les autres se sont révélés correspondre à un même cluster d'ADNc codant pour le gène *LASP1* situé dans l'extrémité centromérique de la zone étudiée. Les huit autres clones correspondaient à une région ne contenant pas notre gène et ne présentaient donc pas d'intérêt pour notre étude. L'absence d'EST sera discutée dans le chapitre suivant.

II. Expression de l'ARNm

Le profil d'expression sera étudié soit en cherchant à révéler l'ARNm à l'aide d'une sonde spécifique (Northern Blot et CPA) ou à amplifier le gène à partir d'ADNc (RT-PCR).

II.1. Northern Blot

Nous avons utilisé deux Northern Blots commerciaux (*Human 12-Lane MTN Blot* et *Human Tumor MTN Blot*). Pour réaliser ces hybridations, nous avons choisi de générer par PCR une sonde correspondant à la partie 3' de l'ARNm du gène candidat identifié, ce qui permet de détecter l'ensemble des transcrits potentiels. La partie 3' d'un gène possède en plus

une plus grande spécificité que les parties codantes notamment dans les familles de gènes. Cette sonde a été purifiée et marquée au ^{32}P . Sa spécificité a été testée par hybridation sur une membrane obtenue par Southern Blot contenant un clone comprenant le gène en entier.

L'hybridation de la sonde spécifique de la partie 3' sur les membranes testées n'a pas permis de mettre en évidence de signal et ce même après exposition prolongée (jusqu'à 10 jours) (résultats non montrés).

II.2. RT-PCR - banque de tissus

Etant donné l'absence de résultats observés en Northern Blot, nous avons voulu étudier l'expression du gène d'intérêt par une technique plus sensible, la RT-PCR, notamment pour évaluer les transcrits rares.

Nous avons utilisé trois panels commerciaux contenant de l'ADN rétrotranscrit d'un certain nombre de tissus : *Human MTC Panel I*, *Human MTC Panel II* et *Human Fetal MTC Panel I* (Clontech). Nous avons amplifié le gène candidat par PCR avec des amorces situées dans les parties 5' et 3' non codantes de l'ARNm pour le MTC de tissus fœtaux et pour les MTC I et II.

Les résultats sont présentés sur la Figure 28.

Nous n'avons obtenu aucune amplification du gène dans les tissus fœtaux, et ceci avec 40 cycles d'amplification (taille attendue 1750 pb).

Les résultats de PCR sur les tissus adultes nous ont fourni deux signaux pour le pancréas et le testicule à la taille attendue de 1750 pb. Le signal visualisé dans le pancréas est faible tandis que celui observé dans le testicule a une intensité comparable à celle du contrôle constitué par l'amplification du gène G3PDH qui possède une expression ubiquitaire forte.

Ce profil d'expression est aussi observé lorsque l'on diminue le nombre de cycles de PCR, même si la bande présente dans le pancréas est difficile à détecter pour 25 cycles d'amplification, contrairement à celle dans le tissu testiculaire qui reste forte (résultats non présentés).

Notre gène d'intérêt semble donc avoir une expression spécifique et forte dans le tissu testiculaire.

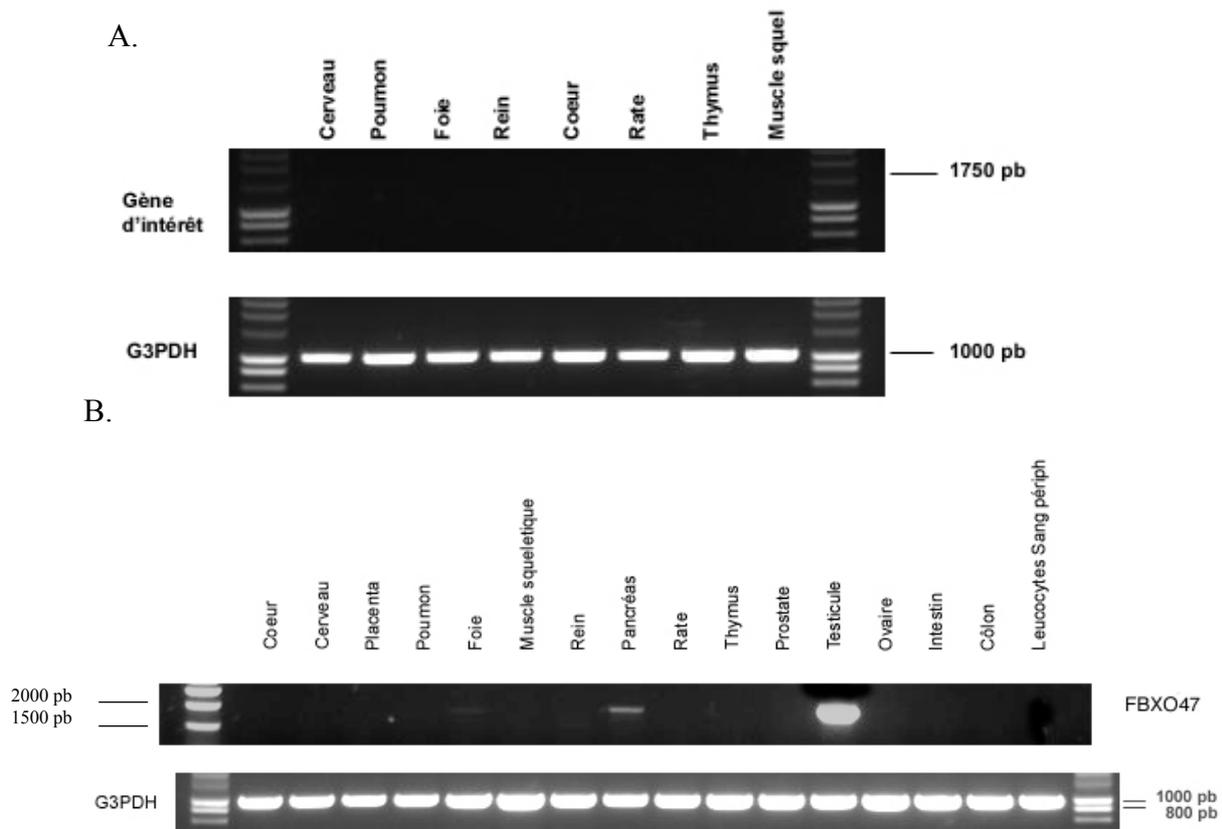


Figure 28. Expression tissulaire du gène d'intérêt.

- A. Amplification de l'ADNc entier sur des tissus fœtaux (produit PCR de 1750 pb) 40 cycles d'amplification.
- B. Amplification de l'ADNc entier sur des tissus adultes (produit PCR de 1750 pb) 40 cycles d'amplification.

G3PDH (produit PCR 900 pb) permet de contrôler la qualité et la quantité des ADNc servant de matrice à l'amplification.

II.3. Expression différentielle

Dans notre démarche de caractérisation d'un gène impliqué dans la cancérogenèse, nous avons cherché à étudier l'expression différentielle de notre gène d'intérêt entre des tissus tumoraux et les tissus normaux correspondants.

Nous avons utilisé pour cela une membrane commerciale nommée *Cancer Profiling Array II* (Clontech). Cette membrane de nylon contient des paires d'échantillons (normal/tumoral) d'ADNc générées à partir d'ARN total de plusieurs tissus. Ici, nous avons étudié sur une

même membrane 154 paires d'échantillons provenant de 19 tissus différents (de 2 à 8 patients par tissu) (Figure 29).

Le diagramme complet et les détails concernant les prélèvements sont présentés en annexe A. Pour hybrider cette membrane, nous avons utilisé la même sonde que celle employée pour les Northern Blot c'est-à-dire celle qui correspond à la fin de la phase ouverte de lecture et à la partie 3' spécifique du gène.

Du fait de la grande variation d'expression des ARNm entre les tissus, nous avons exposé la membrane pour des temps variables: 4 h, 24 h et 48 h. Ceci nous a permis de valider la présence ou l'absence du transcrit spécifique. Ces expériences ont, en parallèle, été réalisées avec une sonde radiomarquée spécifique du gène ubiquitine comme contrôle interne.

La figure 29 présente les résultats obtenus après 4 heures d'exposition en cassette de PhosphorImager. C'est la durée que nous avons retenue comme étant la plus pertinente. En effet, les images obtenues avec des expositions plus longues n'apportaient pas, d'une part de nouvelles informations sur le transcrit, et d'autre part, elles étaient moins informatives pour la quantification de signal (résultats non montrés).

La première observation est que notre gène semble exprimé à un niveau faible par rapport à l'expression de l'ubiquitine dans tous les tissus présents, normaux comme tumoraux, ainsi que dans les neuf lignées cellulaires. Il apparaît que les spots les plus forts sont observés dans les tissus normaux. De façon étonnante, l'expression la plus forte n'apparaît pas être dans le testicule mais dans le rein contrairement aux résultats obtenus en RT-PCR.

Nous avons réalisé une semi-quantification des signaux d'hybridation à l'aide du logiciel IPLabgel qui permet de mesurer l'intensité du signal. Le graphique (Figure 30A) représente la moyenne des intensités de signal obtenues sur tous les spots normaux et tumoraux d'un même tissu (moyenne de tous les patients pour un même tissu). Les tissus normaux exprimant le plus le gène d'intérêt sont le rein, la vulve, le foie, le rectum, la thyroïde, le testicule et le pancréas. Nous avons également observé que l'expression moyenne est plus forte dans les tissus normaux que dans les tissus tumoraux correspondants pour la majorité des tissus testés (sauf sein et ovaire).

Le graphe B (Figure 30B) présente le ratio Normal/Tumoral des moyennes d'intensité de signal pour l'ensemble des tissus.

Nous avons pu distinguer quatre groupes d'expression :

- Le premier pour lequel l'expression différentielle est nulle ou presque (ratio proche de 1) comprenant le sein, l'ovaire, le côlon.
- Le second caractérisé par un ratio compris entre 1,2 et 1,4 comprenant le testicule, le poumon, l'utérus, l'estomac, la prostate, la vessie, le col de l'utérus, la trachée et le foie.
- Le troisième (ratio proche de 1,5) rassemblant le rectum, le rein, la peau et l'intestin.
- Le quatrième pour lequel le gène d'intérêt est deux fois plus exprimé dans les tissus normaux que tumoraux, comprenant la thyroïde, le pancréas et la vulve.

Ces résultats montrent que l'expression différentielle entre tissu normal et tumoral du gène d'intérêt est cohérente avec un rôle possible de ce gène dans la cancérogenèse, au moins pour les tissus des groupes II, III et IV.

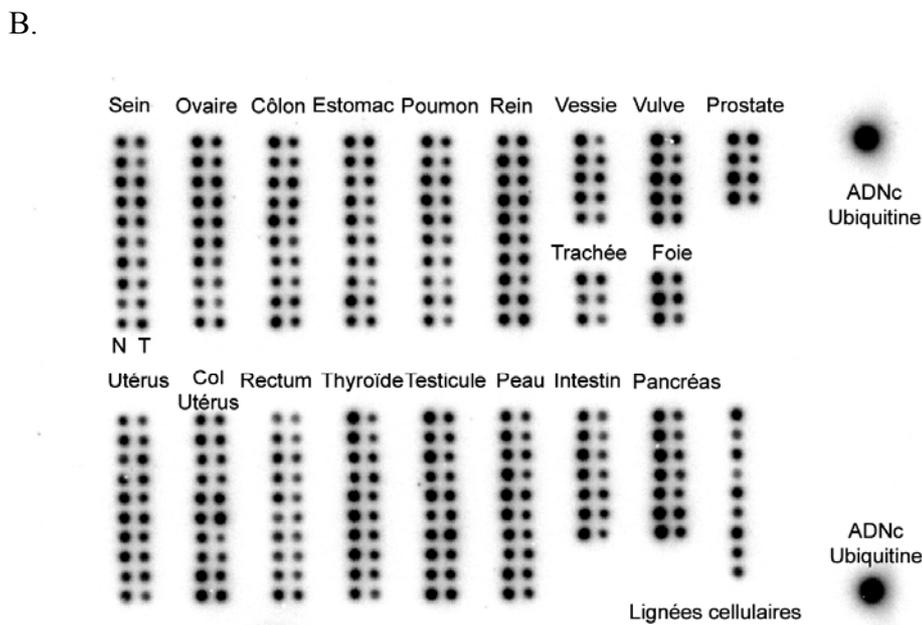
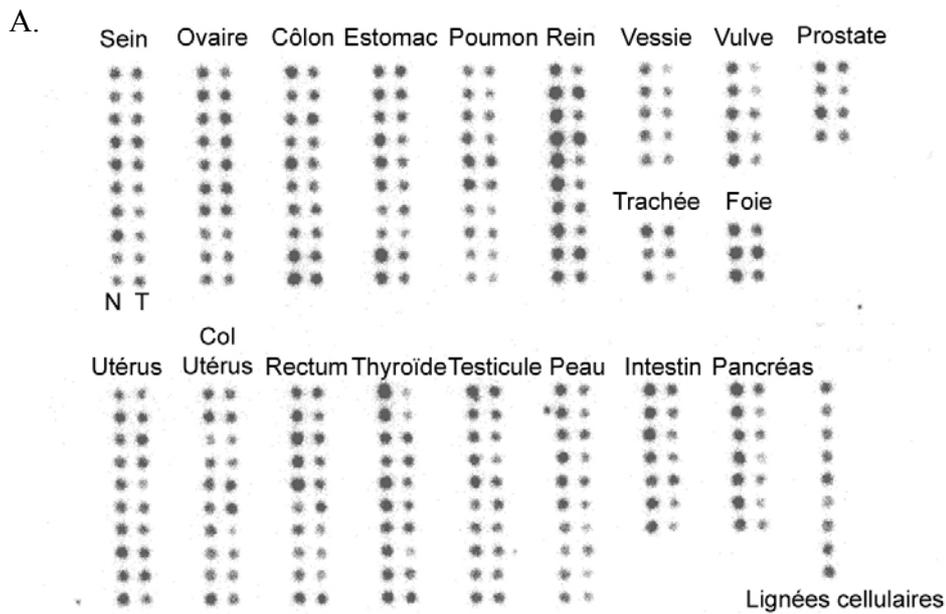


Figure 29. Profil d'expression différentielle sur des ADNc de tissus sains et tumoraux (N et T) et sur des lignées cellulaires.

- A. Cette membrane a été hybridée avec la sonde radiomarquée spécifique du gène d'intérêt et exposée 4 heures en cassette de PhosphorImager.
- B. Cette membrane contrôle a été hybridée dans les mêmes conditions avec une sonde spécifique de l'ubiquitine.

Un diagramme représentant les positions et les caractéristiques exactes de chaque spot est disponible en annexe A.

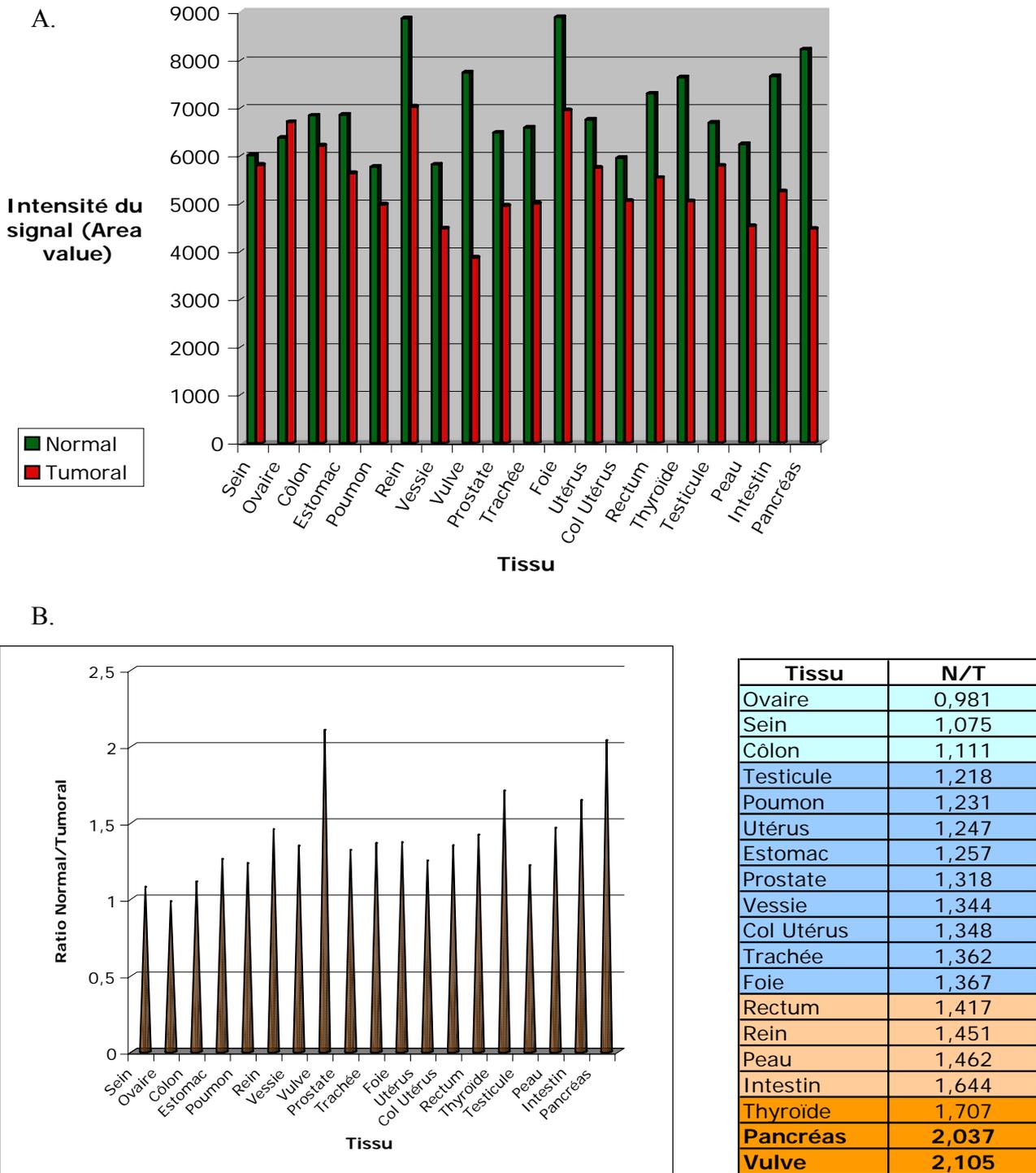


Figure 30. Semi-Quantification des résultats obtenus au PhosphorImager sur Cancer Profiling Array II avec la sonde gène-spécifique.

- A. Graphe représentant la moyenne des intensités de signal pour l'ensemble des spots normaux ou tumoraux d'un même tissu.
- B. Graphe représentant le ratio (N/T) des moyennes et tableau représentant les quatre groupes de ratio différents.

III. Étude bioinformatique du produit codé par le transcrit

Nous avons donc identifié un nouveau gène localisé en 17q12 dont l'expression semble assez faible et plutôt favorisée dans les tissus normaux. Afin d'approfondir la caractérisation du gène mais également d'identifier son rôle dans la cellule, nous nous sommes ensuite intéressé à la protéine codée par le transcrit.

Le logiciel *DNA Strider* nous a permis de traduire la séquence nucléique de l'ATG au TAA en séquence protéique.

III.1. Caractérisation bioinformatique de propriétés physico-chimiques

Nous avons soumis la séquence protéique déduite du transcrit au logiciel *ProtParam*, qui permet de déterminer les paramètres suivants : nombre d'acides aminés, poids moléculaire, point isoélectrique, composition en acide aminé, demi-vie estimée, index d'instabilité, et index aliphatique.

La protéine codée est composée de 453 acides aminés pour une masse moléculaire de 52124,65 Da et un point isoélectrique théorique de 8,77.

Le tableau suivant fournit la composition détaillée en acide aminé de la protéine :

Acide aminé	Nombre dans la séquence	% de la séquence
Ala (A)	24	5,3
Arg (R)	26	5,7
Asn (N)	23	5,1
Asp (D)	19	4,2
Cys (C)	16	3,5
Gln (Q)	18	4
Glu (E)	24	5,3
Gly (G)	16	3,5
His (H)	10	2,2
Ile (I)	26	5,7
Leu (L)	63	13,9
Lys (K)	26	5,7
Met (M)	17	3,8
Phe (F)	24	5,3
Pro (P)	15	3,3
Ser (S)	35	7,7
Thr (T)	25	5,5
Trp (W)	8	1,8
Tyr (Y)	11	2,4
Val (V)	26	5,7
Asx (B)	0	0
Glx (Z)	1	0,22

Le nombre total de résidus chargés négativement (Asp + Glu) est de 43, et celui de résidus chargés positivement (Arg + Lys) est de 52.

Le temps de demi-vie est estimé en prenant la méthionine comme partie N-terminale de la séquence et à partir de la composition en acides aminés. La demi-vie de la protéine est estimée à 30 h dans des réticulocytes de Mammifères *in vitro*, 20 h *in vivo* chez la levure, et supérieure à 10 h *in vivo* chez *Escherichia coli*.

L'index d'instabilité (II) est de 39,23 ce qui selon la classification de Guruprasad ²¹⁸ qualifie cette protéine de stable. Remarquons tout de même qu'elle se situe à la limite entre les deux catégories, stable et instable, qui respectivement se caractérisent par un II inférieur à 40 ou supérieur à 40.

III.2. Identification d'un domaine F-Box

III.2.1. Recherche d'analogie de séquence

Afin d'identifier la famille protéique à laquelle pourrait appartenir notre protéine, nous avons effectué une confrontation de notre séquence protéique contre plusieurs banques de données (comprenant 1 541 613 séquences en tout) à l'aide du logiciel *PSI-Blast*.

Nous avons réalisé trois itérations, car plus de cycles ne donnait pas plus de résultats interprétables (Tableau 7).

La première itération nous a donné 10 résultats dont seulement 3 ont un bon score c'est-à-dire avec une *e-value* inférieure à 0,001. La première séquence (hypothetical protein *Rattus norvegicus*) s'aligne sur les 100 premiers acides aminés, la seconde (hypothetical protein *Mus musculus*) sur les 170 derniers, et la troisième (cyclin-like F-box *Caenorhabditis elegans*) sur 50 acides aminés situés entre les acides aminés 50 et 100 de notre séquence protéique. Cette dernière protéine a retenu notre attention car dans les autres résultats (avec une *e-value* > 0,001), nous retrouvons deux protéines F-box qui s'alignent sur les mêmes acides aminés.

La seconde itération n'a produit que quatre résultats dont trois ont déjà été fournis par la première recherche.

Enfin, la troisième itération a apporté trente et un résultats dont seulement quatre présentaient un score correct.

La synthèse des résultats fait apparaître fréquemment que la région de notre protéine la plus retrouvée dans les bases de données correspond à un domaine caractéristique des protéines de la famille F-box. Les protéines de cette famille sont définies par la présence d'un domaine F-Box, initialement identifié dans la cycline F⁹⁹.

La petite taille de la région (40 résidus environ) qui s'aligne peut expliquer les mauvais scores observés.

L'existence du domaine F-box au sein de notre protéine a été confirmée à l'aide du programme PROSITE. Celui-ci a identifié, avec une haute probabilité, l'existence d'un domaine F-Box entre les acides aminés 41 et 91 de notre protéine.

III.2.2. Alignement de séquences

Le domaine F-box est un motif protéique d'environ 50 acides aminés qui fonctionne comme un site d'interaction protéine-protéine.

Il est caractérisé par une séquence consensus avec quelques positions invariantes (cf. introduction)⁷⁴ (Figure 31).

La famille des protéines F-box est une famille très conservée à travers l'évolution des Eucaryotes et comprend de nombreux membres dont quelques uns seulement sont bien caractérisés.

Nous avons comparé, à l'aide des logiciels CLUSTALX et SeqVu, les domaines F-box de plusieurs protéines et la séquence protéique (41 à 91) de notre protéine afin d'identifier les résidus consensus du motif F-box identifié dans notre séquence protéique (Figure 32).

L'alignement de séquence montre que les positions conservées 8, 9, 16, 20 et 32 sont présentes dans notre motif et qu'elles s'alignent de façon satisfaisante et concordante avec d'autres protéines F-Box d'organismes eucaryotes divers (souris, levure, homme).

Nous avons donc identifié un nouveau membre de la super famille des protéines F-box.

Itération 1			
10 résultats			
référence NCBI	Nom/Organisme	Score	e-value
XP 346902.1	hypothetical protein <i>Rattus norvegicus</i>	318	2,00E-85
XP 358410.1	hypothetical protein <i>Mus musculus</i>	170	4,00E-41
NP 496467.1	cyclin-like F-box <i>Caenorhabditis elegans</i>	54	6,00E-06
CAB56536.1	sensor kinase <i>Pseudomonas stutzeri</i>	36	1,40E+00
XP 232833.2	similar to expressed sequence AI836810 <i>Rattus norvegicus</i>	35	1,90E+00
NP 735338.1	Phenylalanyl-tRNA synthetase su beta <i>Streptococcus</i>	35	2,00E+00
NP 687855.1	Phenylalanyl-tRNA synthetase chain beta <i>Streptococcus</i>	35	2,00E+00
AAF09247.1	F-Box protein FBL5 <i>Homo sapiens</i>	35	3,20E+00
AAF03699.1	F-Box protein FBL4 <i>Homo sapiens</i>	35	3,20E+00
NP 036292.2	F-Box protein FBL4 <i>Homo sapiens</i>	35	3,40E+00
Itération 2			
4 résultats			
référence NCBI	Nom/Organisme	Score	e-value
XP 346902.1	hypothetical protein <i>Rattus norvegicus</i>	348	1,00E-94
NP 496467.1	cyclin-like F-box <i>Caenorhabditis elegans</i>	346	7,00E-94
XP 358410.1	hypothetical protein <i>Mus musculus</i>	212	1,00E-53
NP 492477.1	putative protein with a coiled coil-4 domain <i>C.elegans</i>	51	5,00E-05
Itération 3			
31 résultats			
référence NCBI	Nom/Organisme	Score	e-value
NP 492477.1	putative protein with a coiled coil-4 domain <i>C.elegans</i>	409	1,00E-113
XP 346902.1	hypothetical protein <i>Rattus norvegicus</i>	333	6,00E-90
NP 496467.1	cyclin-like F-box <i>Caenorhabditis elegans</i>	326	4,00E-88
XP 358410.1	hypothetical protein <i>Mus musculus</i>	192	1,00E-47
NP 191441.1	F-box protein family <i>Arabidopsis thaliana</i>	37	6,30E-01
XP 313054.1	ENSANGP00000023140 <i>Anopheles gambiae</i>	37	6,80E-01
NP 201057.1	F-box protein family <i>Arabidopsis thaliana</i>	37	7,30E-01
NP 611647.1	CG6758-PA <i>Drosophila melanogaster</i>	37	7,40E-01
NP 197070.1	hypothetical protein <i>Arabidopsis thaliana</i>	36	1,40E+00
NP 192896.1	F-box protein family <i>Arabidopsis thaliana</i>	36	1,50E+00
NP 194726.1	F-box protein family <i>Arabidopsis thaliana</i>	35	2,40E+00
AAO037188.1	hypothetical protein <i>Arabidopsis thaliana</i>	35	2,50E+00
NP 1800975.1	F-box protein family <i>Arabidopsis thaliana</i>	35	2,50E+00
NP 175150.1	F-box protein-related <i>Arabidopsis thaliana</i>	35	2,70E+00
NP 191693.1	F-box protein family <i>Arabidopsis thaliana</i>	35	3,00E+00
NP 179784.1	F-box protein family <i>Arabidopsis thaliana</i>	35	3,20E+00
AAF09247.1	F-box protein FBL5 <i>Homo sapiens</i>	35	3,30E+00
AAF03699.1	F-box protein FBL4 <i>Homo sapiens</i>	35	3,30E+00
NP 036292.2	F-box and LLR protein 4 <i>Homo sapiens</i>	35	3,40E+00
NP 191459.1	F-box protein family <i>Arabidopsis thaliana</i>	35	3,40E+00
NP 190299.2	F-box protein family <i>Arabidopsis thaliana</i>	35	3,40E+00
NP 197071.1	F-box protein family <i>Arabidopsis thaliana</i>	35	3,50E+00
T45651	hypothetical protein F13112.200 <i>Arabidopsis thaliana</i>	35	4,00E+00
NP 419404.1	sensory box histidine kinase <i>Caulobacter</i>	34	4,20E+00
NP 8611243.1	GTP cyclohydrolase I <i>Helicobacter</i>	34	4,50E+00
T49173	hypothetical protein T20N10.250 <i>Arabidopsis thaliana</i>	34	7,30E+00
NP 567075.1	F-box protein family <i>Arabidopsis thaliana</i>	34	7,50E+00
NP 191445.1	F-box protein family <i>Arabidopsis thaliana</i>	34	7,70E+00
NP 174386.1	F-box protein family <i>Arabidopsis thaliana</i>	34	7,90E+00
AAF98183.1	F17F8.8 <i>Arabidopsis thaliana</i>	34	8,10E+00
NP 180705.1	F-box protein family <i>Arabidopsis thaliana</i>	33	9,80E+00

Tableau 7. Résultat des trois itérations effectuées à l'aide du programme PSI-BLAST.

Les membres de la super famille F-Box sont soulignés en bleu, et les valeurs de e-value (<0,0001) sont indiquées en rouge.

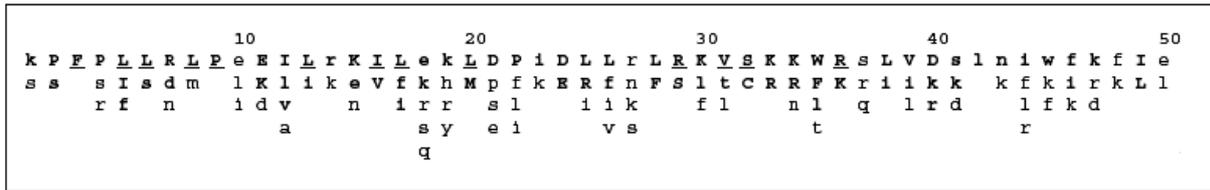


Figure 31. Séquence consensus du motif F-Box.

Celle-ci a été déterminée à partir de l'alignement de 234 séquences utilisées pour créer le profil F-box dans la base Pfam (<http://pfam.wuols.edu/cgi-bin/getdesc?name=F-box>.) Les résidus en majuscules, gras et soulignés sont trouvés dans plus de 40% des séquences F-box, en gras seulement dans 20 à 40%, en minuscules et gras dans 15 à 19%, et les minuscules uniquement dans 10 à 14% des F-Box.

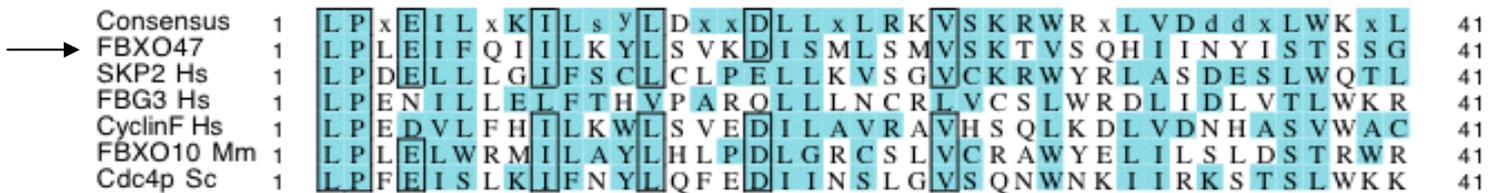


Figure 32. Alignement multiple de séquences F-Box pour quelques membres représentatifs de la superfamille des protéines F-box.

Notre protéine est indiquée par une flèche (FBXO47).

Les résidus homologues sont en bleu, les résidus identiques sont encadrés.

Hs : *Homo sapiens*, Mm : *Mus musculus*, Sc : *Saccharomyces cerevisiae*.

Les numéros d'accèsion GenBank des protéines comparées sont respectivement : Skp2 AAH07441, Fbg3 AAK77940, CyclinF AAH12349, Fbxo10 AAH54731 et Cdc4p NP_116585.

III.3. Nomenclature

La protéine codée par notre gène se nomme donc, selon la nomenclature HUGO, Fbxo47 puisque nous n'avons pas identifié d'autre domaine d'interaction protéique connu dans la partie C-terminale.

Les séquences du gène et de la protéine ont été déposées dans la banque de données GenBank sous le numéro d'accèsion AY700575.

III.4. Identification de signaux/motifs protéiques

Le programme *ScanProsite* en plus de nous confirmer l'existence du motif F-box dans Fbxo47, nous a également apporté d'autres informations quant à la présence de domaines protéiques.

La première est la présence d'un motif *Leucine Zipper* au niveau des acides aminés 120 à 141 (**LfkrctLlPtherLkyihkiL**) avec une probabilité de $0,2311^{e-2}$.

La seconde information concerne la présence de sites de modifications post-traductionnelles, comme des sites candidats de glycosylation de type N (*ASN-Glycosylation*), et des sites de phosphorylation pour les protéines kinases cAMP et cGMP dépendantes (*CAMP-PHOSPHO site*), les protéines kinases C (*PKC-PHOSPHO site*) et les protéines kinases caséine 2 (*CK2- PHOSPHO site*).

Afin d'affiner la pertinence de ces sites de modifications post-traductionnelles, nous avons soumis notre séquence pour des programmes spécialisés : *NetPhos 2.0*, *NetNGlyc* et *NetOGlyc*, pour les deux phénomènes que sont la phosphorylation et la glycosylation.

Les résultats sont résumés dans le tableau 8. Le programme Prosite donne les résultats de phosphorylation selon le type de kinase ciblé, alors que le programme NetPhos donne des résultats en type d'acide aminé phosphorylé. Nous avons croisé ces résultats, certains se retrouvent (sites 20 et 60 par exemple) et d'autres non (sites 65 et 443). De la même façon, certains profils de modification sont prédits avec une grande probabilité par Prosite et une faible probabilité par NetPhos (site 228).

Nous avons donc retenu 9 sites potentiels de phosphorylation prédits par les deux programmes avec une forte probabilité (en rouge dans le tableau 8).

En ce qui concerne la glycosylation, Les programmes Prosite et par NetNGlyc révèlent un site unique de glycosylation (8-11).

Une fois traduite, la protéine peut subir une maturation post-traductionnelle. Au cours de son transfert entre différents compartiments cellulaires, la protéine peut être modulée par une série de modifications biochimiques (glycosylation, hydroxylation, phosphorylation, prénylation, acylation, ubiquitination, protéolyse partielle) la modifiant profondément, de telle sorte que la protéine finale est bien différente de la molécule initialement codée par le gène. Ces modifications chimiques contribuent à la régulation de l'activité de la protéine, ainsi qu'à sa localisation.

Ainsi les résultats obtenus par prédiction de telles modifications (glycosylation, phosphorylation) nous informent sur la potentialité de régulation de la protéine Fbxo47.

N-Glycosylation						
position dans la séquence	Prosite		Score	NetNGlyc		Probabilité
8-11	NFTL		e	NFTL		0,8

Phosphorylation						
position dans la séquence	Prosite	Kinase	Score	NetPhos 2.0	Résidu	Probabilité
20	SnR	PKC	e	KLRR SNROT	sérine	0,992
24				SNROTSCYS	thréonine	0,829
25				NRQTSCYSK	sérine	0,987
60	SvK/SvkD	PKC/CK2	e	LKYL SVKDI	sérine	0,998
65				VKDISMLSM	sérine	0,545
83				INYISTSSG	sérine	0,566
88	SkR	PKC	e	TSSG SKRLL	sérine	0,853
108				RRQDSAILE	sérine	0,933
125	KRcT	CAMP	e	FKRCT LLLP	thréonine	0,082
207	SaR	PKC	e	SKPGSARKL	sérine	0,922
228	ThR	PKC	e	LDHW THRSD	thréonine	0,0431
233				HRSDSAFWL	sérine	0,987
261	SpqD	CK2	e	FGPI SPQDG	sérine	0,99
281	SIK	PKC	e	TDEF SLKGL	sérine	0,933
296	StK/StkE	PKC/CK2	e	LYD ASTKEW	sérine	0,734
297	StK/StkE	PKC/CK2	e	LYD ASTKEW	thréonine	0,673
301	TadD	CK2	e	TKEWT ADDV	thréonine	0,448
307	SlvD	CK2	e	DDVI SLVDE	sérine	0,937
368				EKELYCMDW	tyrosine	0,827
386	TpvE	CK2	e	KVF STPVER	thréonine	0,942
414				QSIMSGDRD	sérine	0,885
443				KEVLYLTMN	tyrosine	0,943
448				LTMNTPLST	thréonine	0,679
451				NTPLST	sérine	0,696

Tableau 8. Compilation des résultats pour les modifications post-traductionnelles de Fbxo47, obtenus par les programmes Prosite, NetPhos 2.0 et NetNGlyc.

En rouge, sont indiquées les modifications les plus pertinentes.

e = excellent

III.5. Recherche d'un peptide-signal et/ou d'un signal d'adressage

Il existe d'autres types de modifications post-traductionnelles que celles étudiées dans le paragraphe précédent, notamment, le clivage protéolytique d'un peptide signal ou d'une pré-séquence.

Le devenir et la fonction d'une protéine sont intimement liés à sa localisation subcellulaire ou à son orientation dans la cellule après traduction (exportation, adressage à la membrane...).

Nous avons d'abord soumis notre séquence protéique au programme *PSORTII*.

Les résultats sont présentés dans la figure 33.

Notre protéine ne contient pas de peptide-signal et sa localisation sub-cellulaire est à 78,3 % mitochondriale. Remarquons que l'on confirme également par ce programme la présence du motif de liaison à l'ADN « *Leucine Zipper* ».

La localisation mitochondriale de Fbxo47 a été renforcée par la soumission de sa séquence à d'autres programmes plus spécialisés comme *TargetP v1.1* et *MitoProII*.

Ces deux logiciels ont confirmé à 83,6 et 93,8 % respectivement, la présence d'une séquence d'adressage mitochondrial.

La grande majorité des protéines mitochondriales est codée par le génome nucléaire, synthétisée dans le cytoplasme puis importée dans les mitochondries. Deux types de signaux d'adressage mitochondrial sont généralement décrits : les pré-séquences et les signaux internes. La pré-séquence consiste en une extension N-terminale de la protéine à importer. Les pré-séquences ne contiennent pas de séquence consensus d'adressage mitochondrial, mais elles possèdent quelques particularités. Au niveau de leur structure primaire, elles sont riches en acides aminés basiques, hydrophobes et hydroxylés et elles ont une taille de 10 à 80 acides aminés.

La protéine Fbxo47 est adressée à la mitochondrie par un signal de type pré-séquence d'une longueur de 23 acides aminés. En effet, les programmes de prédiction fournissent un site de clivage en position 24. Ainsi, sous sa forme mature dans la matrice mitochondriale, Fbxo47 est amputée de ses 23 premiers acides aminés dont 7 sont fortement basiques et hydrophobes (aucun n'est acide).

PSORT II**Results of Subprograms**

PSG: a new signal peptide prediction method

N-region: length 4; pos.chg 1; neg.chg 0
 H-region: length 11; peak value 2.87
 PSG score: -1.53

GvH: von Heijne's method for signal seq. recognition

GvH score (threshold: -2.1): -9.63
 possible cleavage site: between 53 and 54

>>> Seems to have no N-terminal signal peptide

ALOM: Klein et al's method for TM region allocation

Init position for calculation: 1
 Tentative number of TMS(s) for the threshold 0.5: 2
 Number of TMS(s) for threshold 0.5: 0
 PERIPHERAL Likelihood = 2.54 (at 54)
 ALOM score: -0.37 (number of TMSs: 0)

MITDISC: discrimination of mitochondrial targeting seq

R content: 4 Hyd Moment(75): 7.87
 Hyd Moment(95): 7.03 G content: 3
 D/E content: 1 S/T content: 11

Score: 0.90

Gavel: prediction of cleavage sites for mitochondrial preseq

R-2 motif at 100 KRL|LL

NUCDISC: discrimination of nuclear localization signals

pat4: none
 pat7: PNQKLRR (3) at 13
 bipartite: none
 content of basic residues: 11.5%
 NLS Score: -0.22

KDEL: ER retention motif in the C-terminus: none

ER Membrane Retention Signals:

XXRR-like motif in the N-terminus: ASRI none

SKL: peroxisomal targeting signal in the C-terminus: none

SKL2: 2nd peroxisomal targeting signal: none

VAC: possible vacuolar targeting motif: none

RNA-binding motif: none

Actinin-type actin-binding motif:

type 1: none
 type 2: none

NMYR: N-myristoylation pattern : none

Prenylation motif: none

memYQRL: transport motif from cell surface to Golgi: none

Tyrosines in the tail: none

Dileucine motif in the tail: none

checking 63 PROSITE DNA binding motifs:

Leucine zipper pattern (PS00029): *** found ***
 LFKRCTLLLPTKERLKYIHKIL at 120

checking 71 PROSITE ribosomal protein motifs: none

checking 33 PROSITE prokaryotic DNA binding motifs: none

NNCN: Reinhardt's method for Cytoplasmic/Nuclear discrimination

Prediction: cytoplasmic
 Reliability: 76.7

COIL: Lupas's algorithm to detect coiled-coil regions

total: 0 residues

Results of the k-NN Prediction

k = 9/23

78.3 %: mitochondrial

13.0 %: cytoplasmic

4.3 %: plasma membrane

4.3 %: nuclear

Figure 33. Résultat de la soumission de notre séquence protéique au programme PSORT II.

III.6. Structure secondaire de Fbxo47

La structure secondaire d'une protéine correspond à la conformation locale de son squelette. Ceci conduit à définir les motifs issus du repliement régulier de la séquence polypeptidique : hélice α , feuillet β et coude.

Nous avons voulu déterminer la structure secondaire de Fbxo47 de manière à confirmer certaines caractéristiques structurales de 2 des motifs identifiés dans notre protéine : le domaine F-box et la pré-séquence d'adressage mitochondrial. Ces deux motifs possèdent en plus des caractéristiques de leur structure primaire, des particularités de structure secondaire. Le motif F-Box consiste en trois hélices principales. L'hélice H1 s'oriente orthogonalement à la paire anti-parallèle formée par les deux autres hélices H1/H2 ²¹⁹.

La pré-séquence d'adressage mitochondrial se replie en une hélice α amphiphile possédant un côté chargé positivement et un côté hydrophobe ²²⁰. La structure secondaire consensus obtenue, à l'aide du site « *consensus secondary structure prediction* », pour Fbxo47 confirme les caractéristiques des deux motifs préalablement identifiés : il existe une hélice α dans les 23 premiers acides aminés de la séquence correspondant à l'adressage mitochondrial, et les trois hélices H1 (48-59), H2 (64-77) et H3 (88-95) du domaine F-box se répartissent entre les acides aminés 50 et 100 de Fbxo47 (Figure 34).

La figure 35 présente un résumé des caractéristiques de l'ARNm et de la protéine codée par le gène FBXO47.

L'ensemble des résultats présentés dans les parties A et B de ce chapitre, ont fait l'objet d'une publication dans le journal *Genes, Chromosomes & Cancer*, en mai 2005, sous le titre « *Molecular cloning and characterisation of FBXO47, a novel gene containing an F-box domain, located in the 17q12 band deleted in papillary renal cell carcinoma* » ²²¹. (Annexe C)

Consensus prediction result for : Fbxo47

View Consensus in: [\[MPSA \(Mac, UNIX\) , About...\]](#) [\[AnTheProt \(PC\) , Download...\]](#) [\[HELP\]](#)

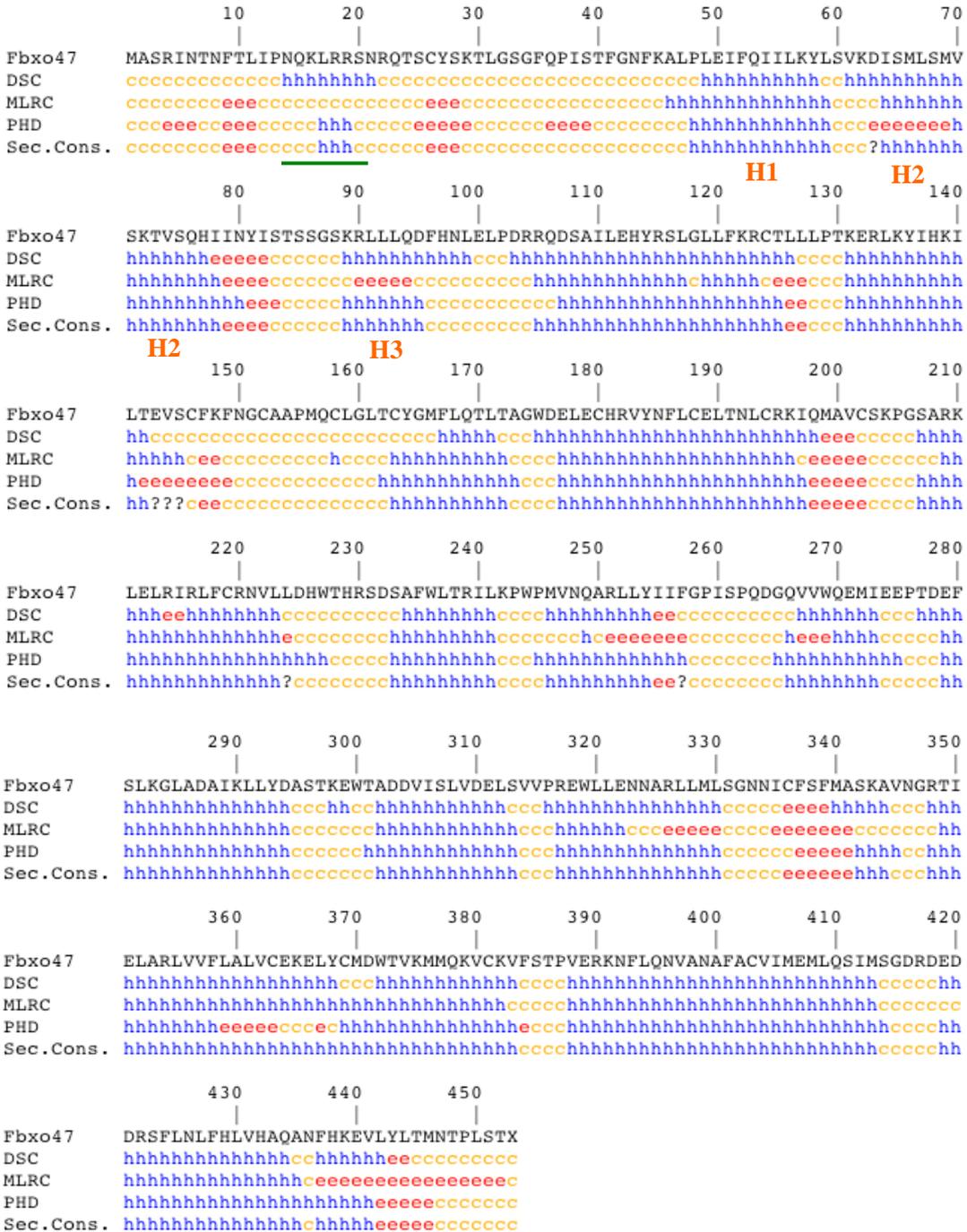


Figure 34. Prédiction de structures secondaires pour la protéine Fbxo47.

L'hélice de la préséquence d'adressage mitochondriale est soulignée en vert. Les trois hélices H1, H2 et H3 du motif F-box sont indiquées en orange.

gggagggtgcagctccggagcctacgggagtgtgacctctcgctcctggcgagaagctcttgaggaga	71	
tattttcttttctaaacactaagaatttgaggatggtggggcaaacgactaactccactttcctgtgac	143	
tctgtagccggaaaaactgtcaaaatctgtggataaaatttgtgagacaagaaga	209	
		<i>M A S</i>
	3	
AGA ATA AAT ACA AAT TTC ACT TTG ATT CCC AAC CAG AAA CTT AGA CGT AGT AAT	263	
R I N T N F T L I P N Q K L R R S N	21	
CGT CAA ACC AGC TGT TAT TCC AAG ACC CTT GGC TCA GGC TTT CAA CCC ATA TCA	317	
R Q T S C Y S K T L G S G F Q P I S	39	
ACA TTT GGA AAT TTT AAA GCC TTA CCA TTG GAA ATA TTC CAG ATA ATT TTA AAA	371	
T F G N F K A L P L E I F Q I I L K	57	Domaine F-box
TAT TTG TCA GTG AAG GAT ATC AGC ATG CTA AGC ATG GTG TCC AAA ACA GTC AGC	425	
Y L S V K D I S M L S M V S K T V S	75	
CAA CAC ATT AIT AAT TAT ATC TCA ACC TCA TCA GGA AGC AAA AGA CTT TTA CTA	479	
Q H I I N Y I S T S S G S K R L L L	93	
CAG GAC TTT CAT AAC CTT GAG CTG CCT GAC AGG AGA CAA GAC TCT GCT ATA CTG	533	
Q D F H N L E L P D R R Q D S A I L	111	
GAG CAC TAC AGA TCT CTA GGT CTA CTG TTT AAA AGA TGC ACA TTG CTG CTA CCC	587	
E H Y R S L G L L F K R C T L I L P	129	Motif Leucine-Zipper
ACC AAG GAA AGG CTA AAA TAC AIT CAC AAG ATA CTC ACA GAA GTT TCC TGC TTT	641	
T K E R L K Y I H K I L T E V S C F	147	
AAA TTC AAT GGC TGT GCA GCT CCT ATG CAG TGT TTA GGA TTA ACA TGT TAT GGC	695	
K F N N G C A A P M Q C L G L T C Y G	165	
ATG TTT TTA CAG ACC TTA ACA GCA GGT TGG GAT GAA CIT GAG TGC CAT CGC GGT	749	
M F L Q T L T A G W D E L E C H R V	183	
TAT AAT TTC TTA TGC GAA CTG ACT AAT CTC TGC CGC AAG ATA CAA ATG GCT GTC	803	
Y N F L C E L T N L C R K I Q M A V	201	
TGC AGC AAA CCA GGA AGT GCC GGA AAA CTG GAG TTA AGA ATC AGA CTC TTC TGT	857	
C S K P C G S A R K L E L R I R L F C	219	
AGG AAT GTC CFC CTT GAT CAT TGG ACA CAT CGA AGT CAT TCT GCT TTT TGG TTG	911	
R N V L L D H W T H R S D S A F W L	237	
ACG CGA ATA TTA AAA CCA TGG CCA AIG GTG AAT CAG CCA AGA TTA CTG TAT AIC	965	
T R I L K P W P M V N Q A R L L Y I	255	
ATC TTT GGG CCA ATA TCT CCT CAA GAT GGA CAG GTG GTT TGG CAG GAA ATG ATA	1019	
I F G P I S P Q D G Q V V W Q E M I	273	
GAA GAA CCT ACA GAT GAA TTC AGT CTG AAA GGT TTG GCT GAT GCC ATT AAG TTA	1073	
E E P T D E F S L K G L A D A I K L	291	
CTA TAT GAC GCT AGC ACT AAA GAG TGG ACA GCA GAT CAT GTT ATC AGT CTT GTA	1127	
L Y D A S T K E W T A D D V I S L V	309	
GAT GAA CTA TCA GTG GTT CCC GGT GAG TGG CTT CTA CAG AAT AAT GCA CGT CTC	1181	
D E L S V V P R E W L L E N N A R L	327	
CTA ATG CTA AGT GGA AAC AAC AIC TGT TTC AGT TTC AIG GCT AGT AAA GCT GIG	1235	
L M L S G N N I C F S F M A S K A V	345	
AAT GGA CGC ACC AIT GAA CTG CCA AGG CTC GTA GTC TTT TTG GCT TTG GTG TGT	1289	
N G R T I E L A R L V V F L A L V C	363	
GAG AAA GAA CTG TAC TGC ATG GAT TGG ACA GTT AAA ATG ATG CAA AAA GTC TGC	1343	
E K E L Y C M D W T V K M M Q K V C	381	
AAA GTC TTT AGC ACT CCA GTG GAA AGA AAG AAC TTC CTG CAG AAT GTG GCA AAT	1397	
K V F G S T P V E R K N F L Q N V A N	399	
GCA TTT GCA TGT GTT ATA ATG GAA AIG CTG CAA TCA AIT ATG TCT GGA GAC GGT	1451	
A F A C V I M E M L Q S I M S G D R	417	
GAT GAA GAT CAC AGA AGC TTT TTG AAT TTG TTC CAT CTT GTA CAT GCT CAG GCT	1505	
D E D D R S F L N L F H L V H A Q A	435	
AAC TTC CAT AAG GAG GTC CTG TAT TTG ACC ATG AAT ACT CCT CTG TCT ACC TAA	1559	
N F H K E V L Y L T M N T P L S T *	453	
atc ctc aga aag cct tgc ctt tag aga acg tca ctg aac gaa tca tta gag gag	1613	
	471	
tgc	1616	
	472	

Figure 35. ADNc et séquence protéique déduite du gène humain FBXO47.

Les nucléotides constituant l'ORF sont désignés en lettres capitales. L'identité et la position des acides aminés sont situés sous les codons. Les résidus formant le domaine F-Box sont surlignés en bleu et les acides aminés les plus conservés sont en gras. Le motif Leucine Zipper est souligné avec les Leucines conservées en gras et en jaune. La préséquence d'adressage mitochondrial est constituée des résidus en italique.

C. Recherche de la fonction biologique de la protéine Fbxo47

I. Etude de la fonctionnalité du domaine F-Box

Il a été démontré que les protéines F-box interagissent à travers leur domaine F-box avec Skp1 (et indirectement avec Cul-1)⁷⁰⁻⁹⁹⁻¹⁰³. Afin de montrer la fonctionnalité du domaine F-box de Fbxo47, nous avons étudié son interaction avec la protéine Skp1 par deux types d'expériences. Dans la première, nous avons exprimé les protéines d'intérêt *in vitro* et analysé leur capacité à se complexer. La seconde expérience visait à faire exprimer ces protéines *in vivo* dans une souche bactérienne, et par la technique du double hybride et de son système rapporteur, de confirmer et d'évaluer la force de cette l'interaction.

I.1. Interaction *in vitro*

Pour tester l'interaction *in vitro* de Fbxo47 et de Skp1, nous avons voulu co-précipiter le complexe éventuellement formé par ces deux protéines. Le principe de cette expérience est basé sur la formation d'un complexe protéique dans lequel une des protéine est marquée par des molécules d'histidine qui permettent sa rétention sur une colonne formée de billes de nickel. Celle-ci retient alors l'ensemble du complexe s'il existe.

Nous avons produit les protéines Skp1-His-³⁵S (taille attendue 23,5 kDa), Fbxo47-³⁵S (taille attendue 52,5 kDa) et la protéine Skp2-³⁵S comme contrôle positif de l'expérience (taille attendue 47 kDa), dans un système acellulaire à l'aide du kit « *TNT Coupled Reticulocyte Lysate System* » (Promega) (Figure 36 ligne 1, 2 et 3). L'ensemble des protéines a été déposé sur la colonne de nickel, puis décroché par compétition à l'aide d'imidazole. Les co-précipitants ont été séparés sur SDS-Page et révélés par autoradiographie (Figure 36).

Le décrochage de la protéine Fbxo47 dans la fraction d'élution montre qu'il y a effectivement interaction entre Fbxo47 et Skp1, de la même façon que pour la protéine F-box Skp2 (lignes 4 et 5 Figure 36).

Il faut noter qu'il existe plusieurs bandes pour les protéines Fbxo47 et Skp2 après production par le système TNT. Il semblerait que ce soit dû à la présence de méthionine interne à la séquence protéique, formant alors des phases de lectures ouvertes alternatives plutôt qu'à une dégradation des protéines.

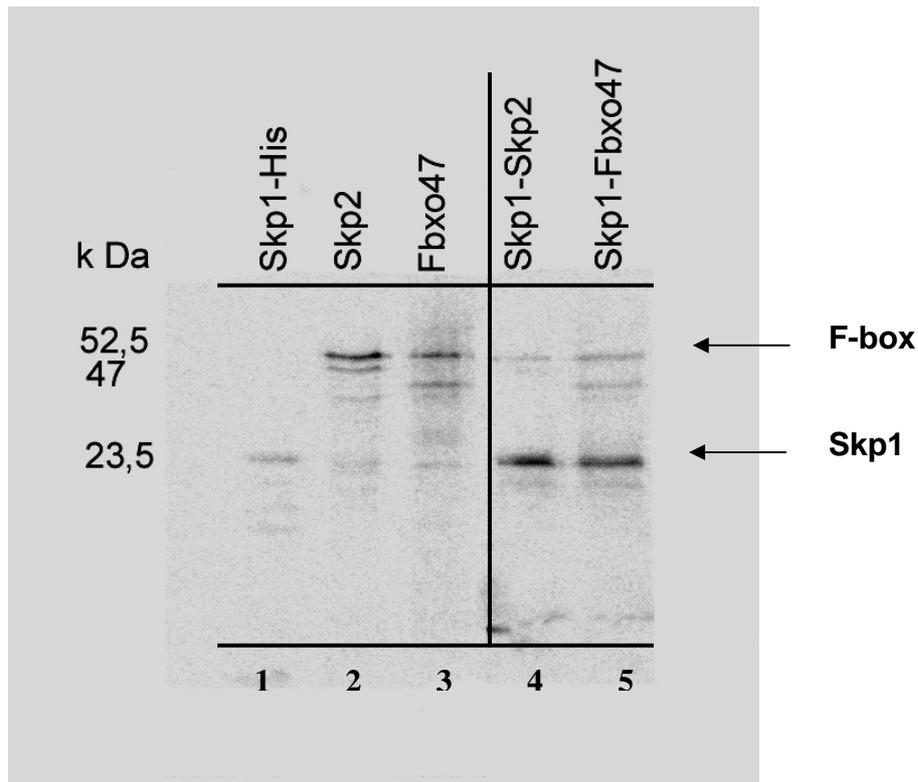


Figure 36. Interaction Fbxo47-Skp1 *in vitro* (SDS-Page et autoradiographie des protéines radiomarquées).

Lignes 1-3, Production de protéines radiomarquées au ^{35}S après transfection avec le système TnT (Promega)

Lignes 4 et 5, complexes élués de la colonne de Nickel par compétition avec l'imidazole.

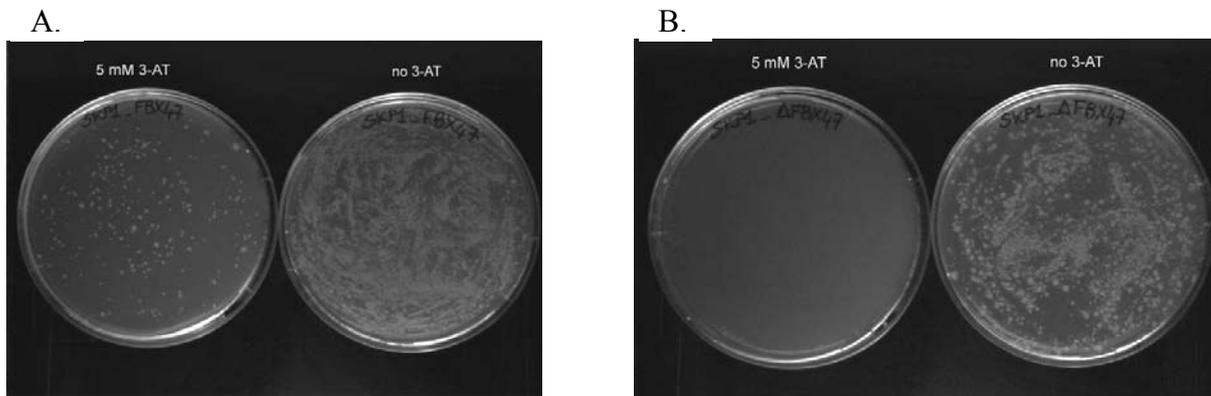


Figure 37. Test d'Interaction Skp1-Fbxo47 (A) et Skp1- Δ [1-155]Fbxo47 (B) par le système double-hybride bactérien.

5 mM 3-AT : milieu sélectif

no 3-AT : milieu non sélectif

Images réalisées après 24 heures d'incubation à 37°C et 24 heures supplémentaires à 18 °C dans l'obscurité.

1.2. Interaction *in vivo*

L'expérience précédente nous a donné une bonne indication quant à la fonctionnalité du domaine F-Box de Fbxo47. Nous avons décidé de valider à nouveau ce domaine mais cette fois ci dans un organisme vivant par le système du double-hybride bactérien. De plus, nous avons voulu également validé le fait que le domaine F-box est indispensable à l'interaction en utilisant une forme tronquée de notre protéine (délétée de la F-box).

La méthodologie est basée sur l'utilisation d'un facteur de transcription dont les domaines de liaison à l'ADN et d'activation peuvent être dissociés. Ainsi les protéines, dont on veut tester l'interaction, peuvent être exprimées comme protéines hybrides fusionnées l'une avec le domaine de liaison à l'ADN, l'autre avec le domaine d'activation de la transcription. Si les deux protéines interagissent, les deux domaines sont de nouveau associés, permettant ainsi la transcription d'un gène rapporteur.

Nous avons utilisé le kit « *BacterioMatch II Two-hybrid System* » de chez Stratagène, pour lequel le gène reporter est HIS3. En l'absence d'activation transcriptionnelle, la souche bactérienne ne peut pas croître sur un milieu contenant du 3-AT, inhibiteur compétitif du produit du gène HIS3 (voir partie Matériels et Méthodes).

Nous avons co-transformé dans la souche reporter le plasmide codant pour la protéine λ CI-Fbxo47 et le plasmide codant α ARNp-Skp1. La figure 37A confirme l'interaction Fbxo47-Skp1 observée *in vitro*. En effet, ayant intégré les deux plasmides, la souche reporter est capable de croître à la fois sur le milieu non sélectif (no 3-AT) et sur le milieu sélectif (5 mM 3-AT). Par contre, la souche reporter co-transformée avec les plasmides codant les protéines α ARNp-Skp1 et λ CI- δ [1-155]-Fbxo47 (forme de Fbxo47 délétée du motif F-box) ne peut pas pousser sur le milieu sélectif (Figure 37B). Ainsi, nous avons démontré que l'interaction Skp1-Fbxo47 impliquait le domaine F-box de Fbxo47.

L'interaction Skp1-Fbxo47 est détectée tardivement (après 24 h d'incubation à 37°C plus 24 h à 18°C), par comparaison avec le contrôle positif (pBT-LGF2 + pTRG-Gal11^P) (24 h d'incubation à 37°C). Remarquons que pour l'interaction, Fbxo47 et Skp1, il y a toujours moins de colonies sur le milieu sélectif que sur le milieu non sélectif (résultats non montrés). Afin de comparer l'intensité de l'interaction, nous avons également co-transfecté Skp1 avec Skp2, protéine F-box dont l'interaction avec Skp1 est bien décrite⁹⁹⁻¹⁰⁷⁻²²²⁻²²³. Le temps de détection de cette interaction reste identique (24 h 37°C + 24 h 18°C) (résultats non montrés).

On peut donc conclure que l'interaction protéique entre un domaine F-Box et Skp1 est plutôt de « type » faible.

Nous avons donc démontré que Fbxo47 possède un domaine F-box fonctionnel, capable de reconnaître et d'interagir avec la protéine Skp1.

II. Localisation mitochondriale de la protéine Fbxo47

L'analyse bioinformatique de la séquence protéique de Fbxo47 nous a indiqué la présence d'une pré-séquence d'adressage mitochondriale avec une probabilité d'environ 90 %. Afin de visualiser la localisation de la protéine dans un système cellulaire, nous avons construit deux protéines de fusion : Fbxo47-EGFP et $\delta[1-23]$ Fbxo47-EGFP, qui est une forme sans adressage mitochondrial de Fbxo47. Les plasmides contenant ces deux protéines, ainsi que le plasmide contrôle contenant l'EGFP seule, ont été transfectés dans des cellules COS, pendant 24 heures. Après cette incubation, les cellules ont subi un double marquage fluorescent noyau/mitochondrie (TOPRO-3/MitoTracker red, respectivement) puis ont été montées sur lame pour être visualisées par microscopie confocale.

Les résultats sont présentés dans la figure 38.

L'image obtenue avec le vecteur contenant l'EGFP seule (Figure 38B) montre un marquage vert diffus dans l'ensemble de la cellule, lié à l'excitation de l'EGFP.

Par contre, le marquage vert produit par la protéine de fusion Fbxo47-EGFP (Figure 38A), est beaucoup plus restreint et prend la forme d'agrégats très localisés. Ces points concentrés se superposent à ceux obtenus par le marquage des mitochondries (marquage rouge). Néanmoins, on peut remarquer que les deux marquages ne sont pas équivalents : l'ensemble du marquage mitochondrial n'est pas retrouvé avec le marquage EGFP (vert).

La protéine de fusion contenant l'EGFP et la forme tronquée (sans la séquence d'adressage mitochondrial prédite) se localise de façon beaucoup moins concentrée que la protéine non tronquée. Cette répartition diffuse confirme la fonctionnalité de la séquence d'adressage mitochondrial dans la protéine sauvage Fbxo47.

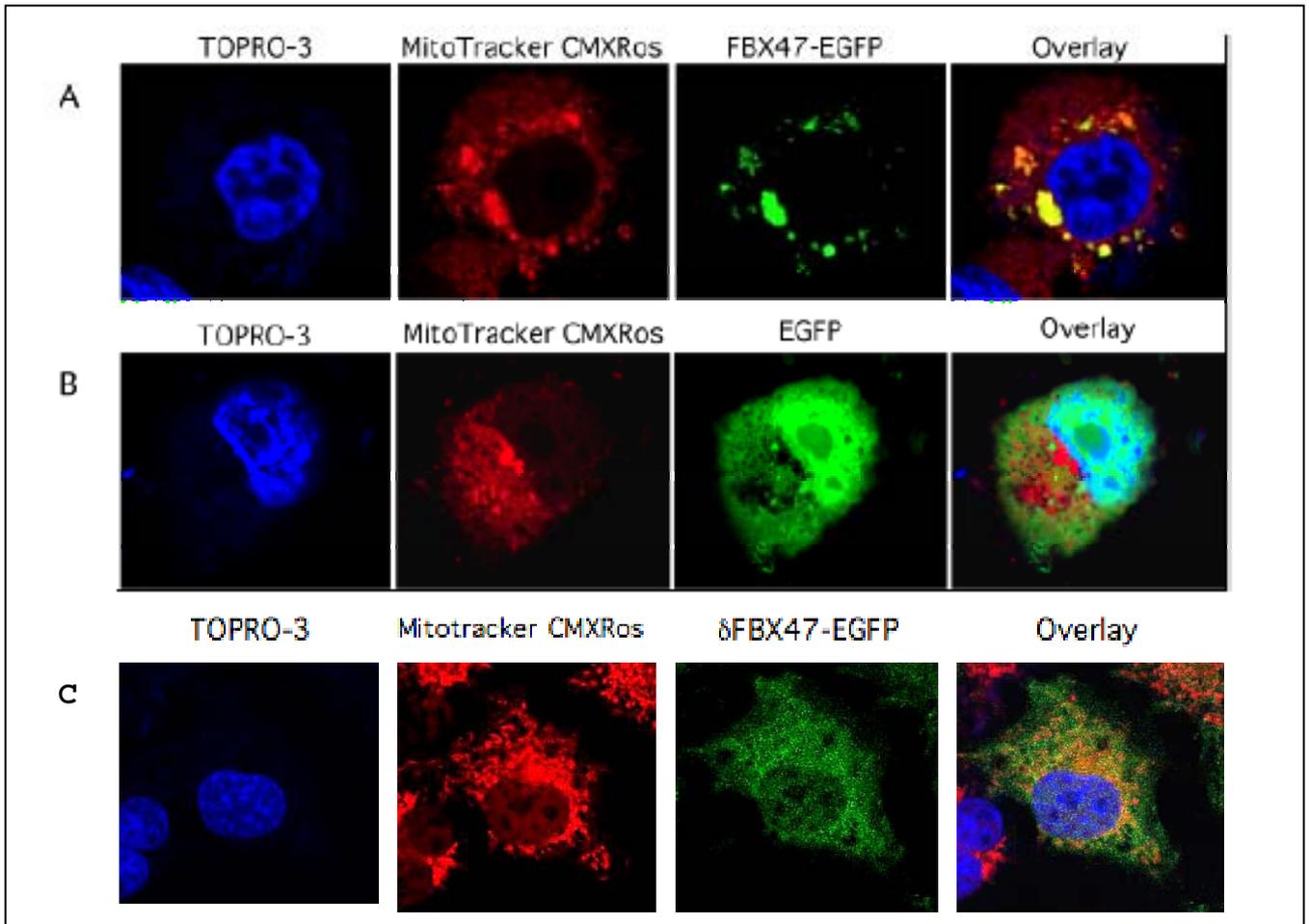


Figure 38. Localisation sub-cellulaire de la protéine Fbxo47 après transfection de cellules COS.

A Transfection des cellules COS avec le vecteur contenant la forme Sauvage (WT) de Fbxo47 fusionnée à l'EGFP

B Transfection des cellules COS avec le vecteur contenant l'EGFP seule (contrôle)

C Transfection des cellules COS avec le vecteur contenant la forme mutée (sans adressage mitochondrial) de Fbxo47 fusionnée à la protéine EGFP.

III. Recherche de partenaires protéiques de Fbxo47

Les protéines F-box possèdent une structure bipartite avec deux domaines de liaison protéiques : l'un se liant à la protéine Skp1 (le motif F-Box), et l'autre capable de reconnaître la protéine cible à ubiquitiner. A travers notre analyse bioinformatique, nous n'avons pas mis en évidence de domaine connu de liaison protéique susceptible de nous orienter vers les substrats potentiels de Fbxo47.

Nous avons donc décidé de réaliser un criblage de banque d'ADNc de rein par la méthode du double-hybride bactérien (*Bactériomatch II*, Stratagène).

Le principe de l'expérience est le même que pour la recherche dirigée d'interaction (comme nous l'avons fait entre Fbxo47 et Skp1), sauf qu'au lieu de co-transformer le plasmide contenant Fbxo47 et un autre plasmide contenant une protéine connue, nous avons co-transformé le plasmide contenant Fbxo47 avec une multitude de plasmides contenant l'ensemble des ADNc (c'est-à-dire la totalité du message exprimé) extraits de cellules de rein.

Cette expérience a comporté plusieurs étapes : l'amplification de la banque, la détermination du nombre de co-transformations nécessaires pour couvrir l'ensemble des ADNc, un criblage primaire, un criblage secondaire, puis la validation des interactions positives et enfin l'identification des interacteurs de Fbxo47.

III.1. Amplification et titrage de la banque d'ADNc de rein

Nous disposions d'une banque commerciale de rein (Stratagène), dont les ADNc sont déjà sous-clonés dans le vecteur pTRG contenu dans la souche bactérienne XL1-Blue MRF^rKan. Le titre primaire était de $3,76.10^6$ cfu et le titre de la solution pré-amplifiée était de 75.10^8 cfu.

Après amplification de cette banque, par étalement et croissance sur milieu LB tétracycline, nous avons titré notre banque à $1,7.10^{10}$ cfu/mL. Nous l'avons donc amplifié environ 2,3 fois.

La moitié du volume nous a ensuite servi pour extraire, par préparation d'ADN plasmidique, les vecteurs contenant les ADNc.

III.2. Détermination du nombre de co-transformations

Pour évaluer le nombre de réactions nécessaires pour couvrir la totalité de la banque d'ADNc, il a fallu réaliser une co-transformation pilote. Elle consiste à transformer d'une part avec une quantité moindre de plasmides et d'autre part dans un volume moins important, le vecteur contenant Fbxo47 et des vecteurs contenant des ADNc de la banque. Ces réactions sont étalées uniquement sur du milieu non sélectif (no 3-AT).

Ceci détermine en fait l'efficacité de la réaction, qui a été de 1.10^5 cfu/100 μ L.

Le nombre de colonies obtenues nous a permis d'estimer à 38 le nombre de co-transformations (de 500 μ L) nécessaire pour couvrir l'ensemble de la banque.

III.3. Criblage primaire

Nous avons donc transformé 38 fois le vecteur contenant Fbxo47 avec des vecteurs contenant la banque d'ADNc. L'efficacité a été déterminée sur le milieu no 3-AT et nous avons obtenu 2.10^5 cfu/mL pour une réaction de 500 μ L. Or le titre primaire de la banque étant de $3,76.10^6$ cfu, nous avons donc couvert, par ce criblage de 40 co-transformations, 2,5 fois la banque. Ces réactions ont toutes été étalées sur du milieu sélectif (5 mM 3-AT) par aliquots de 300 μ L, puis mises à incuber 24 h à 37°C. Ainsi nous avons 200 boîtes à analyser après ce premier criblage.

Après cette première période de croissance, nous avons obtenu, pour l'ensemble des boîtes, 78 colonies qui ont été repiquées nominativement sur d'autres boîtes 5 mM 3-AT (Boîte d'enrichissement 1).

Les 200 boîtes sont placées à température ambiante à l'obscurité pendant 16 h afin de détecter des interactions plus faibles. Nous avons obtenu 705 colonies de plus à la suite de cette seconde incubation. Comme précédemment, elles ont été repiquées sur milieu sélectif (Boîte d'enrichissement 2).

Nous possédons donc 803 interacteurs potentiels à la fin du criblage primaire.

III.4. Criblage secondaire

Les 803 clones sélectionnés à la suite du premier criblage ont été soumis à un deuxième criblage plus restrictif. En effet, le système utilisé (*BacterioMatch II*) possède deux systèmes

rapporteurs en série (c'est-à-dire dépendant du même activateur transcriptionnel) : le gène HIS3 et le gène de résistance à la streptomycine. Ainsi lorsque l'activation transcriptionnelle est réelle et pérenne, l'interaction des deux protéines traduites à partir des plasmides co-transformés permet à la souche bactérienne de pousser sur un milieu doublement sélectif (5 mM 3-AT + 12,5 µg/mL streptomycine).

L'ensemble des clones positifs a donc été repiqué sur milieu doublement sélectif mais également sur LB tétracycline/chloramphénicol (Boîte de maintenance). Après 24 h à 37°C, l'analyse a révélé 113 colonies positives (bonne pousse), 365 colonies semi-positives (pousse moyenne) et 327 négatives (pas de pousse) qui représentent donc des faux positifs du criblage primaire. Les boîtes de criblage secondaires ont été incubées 18 h supplémentaires à température ambiante dans l'obscurité et ré-analysées. Parmi les 365 colonies semi-positives, 181 ont poussées correctement, 184 sont restées semi-positives et les 327 négatives le sont restées.

En conclusion, à la fin de cette étape, nous disposons de 294 interacteurs potentiels de Fbxo47 à valider.

III.5. Validations des interactions positives

Le principe de la validation est de co-transformer le vecteur contenant l'interacteur isolé avec le vecteur contenant Fbxo47 suivant la même méthode que celle utilisée lors de l'expérience de double hybride dirigé.

La première étape a consisté à isoler tous les interacteurs positifs. L'ADN plasmidique est isolé à partir de la boîte de maintenance et utilisé pour transformer la souche bactérienne XL1-Blue MRF^rKan. La réaction est ensuite étalée sur du milieu LB tétracycline de manière à exercer une pression de sélection afin de perdre le plasmide contenant Fbxo47 (qui porte la résistance au chloramphénicol). Les colonies présentes sont ensuite repiquées sur deux milieux : LB tétracycline, LB chloramphénicol. Nous avons ainsi identifié les clones d'intérêt c'est-à-dire de phénotype tet^S cam^R (environ 90 %).

L'ADN plasmidique des bactéries possédant ce phénotype a été extrait. Il correspond à un vecteur pTRG contenant un ADNc d'interacteur protéique de la protéine d'intérêt.

Enfin, la validation elle-même a consisté à co-transformer 294 fois, le vecteur pBT-Fbxo47 avec le vecteur pTRG-ADNc d'intérêt isolé (Réaction A), et avec le pBT vide d'autre part (Réaction B) dans des cellules compétentes « *BacterioMatch II Validation Reporter* »

(Stratagène). Ces réactions ont été étalées à la fois sur milieu sélectif (5 mM 3-AT) et sur milieu non sélectif (no 3-AT) et incubées 24 h à 37°C puis 24 h à 18°C à l'abri de la lumière.

Finalement, cette expérience de validation confirme uniquement 10 clones parmi les 294. Remarquons qu'environ 5 interacteurs potentiels n'avaient pas besoin de la présence de Fbxo47 pour permettre une activation transcriptionnelle (réaction B positive sur les deux milieux).

III.6. Analyse et identification d'interacteurs protéique de Fbxo47

Nous avons séquencé l'ADN plasmidique des vecteurs dont l'interaction était validée. Ces séquences ont été ensuite préparées de manière à être en phase de lecture avec la sous-unité α de l'ARN polymérase.

Les séquences ont ensuite été soumises au programme BlastN, qui les compare à l'ensemble des banques de données nucléotidiques suivantes : GenBank, EMBL, DDBJ et PDB (3 535 328 séquences en novembre 2005) (Tableau 9).

Puis ces mêmes séquences ont été soumises au programme TBlastX, qui les compare après traduction dans les 6 phases de lecture possibles aux banques de données protéiques suivantes : GenBank CDS translations, PDB, SwissProt, PIR et PRF (2 905 734 séquences en novembre 2005) (Tableau 10).

L'analyse des séquences des clones révèle qu'il y a 4 clones (C201-C286 et C228-C 89) qui correspondent à 2 ADNc différents. Ceci n'est pas surprenant dans la mesure où nous avons couvert environ 2,3 fois la banque lors du criblage et de plus le message représenté par le titre primaire peut contenir, pour les gènes très exprimés, plusieurs fois le même ARNm. La seconde observation concerne les clones 201 et 286 contenant le même ADNc, qui s'aligne en discontinuité sur la séquence de l'ARNm de NR4A1 (Tableau 9). En effet, les bases 50 à 243 puis les bases 331 à 628 s'alignent sur la séquence de l'ARNm du membre 1 de la sous-famille des récepteurs nucléaires (BT007144.1).

N° Clone	Source ADN	Nb nucléotides	BlastN	N° Genbank	E-value	% identité	Force I
C142	M	526	<i>Hs</i> isolate 13_R21 mitochondrion	AY963584.2	0.0	98	moyen
C170	G	502	<i>Hs</i> PP2CB mRNA transcript 1	NM_004156.2	0.0	92	moyen
C201/C286	G	297-193	<i>Hs</i> Nr4a1 mRNA	BT007144.1	3,00E-153	98-100	forte
C228/C89	G	415	<i>Hs</i> G elongation factor mito mRNA	BC049210.1	0.0	97	moyen
C164	G	75	<i>Hs</i> lecithin-cholesterol acyl transferase gene cds	AY422210.1	7,00E-69	56	moyen
C193	G	527	<i>Hs</i> Cdc42 GTPase activating protein mRNA	NM_020754.1	0.0	95	moyen
C195	G	745	<i>Hs</i> HSP90protein 1 alpha mRNA	BC007989.2	0.0	98	moyen
C250	G	377	<i>Hs</i> Granulin transcript 2 mRNA	NM_001012479.1	0	98	moyen

Tableau 9. Résultats de comparaison des séquences des interacteurs positifs validés aux banques de données nucléotidiques via le programme BlastN.

(PP2CB : isoforme β de la sous-unité catalytique de la protéine phosphatase 2A)

M : ADN mitochondrial

G : ADN nucléaire (génomique)

La force I évalue la force d'interaction par rapport au temps de croissance observé des colonies par rapport au contrôle positif.

e désigne la probabilité d'identité des séquences identifiées par rapport aux séquences soumises

Lorsqu'on analyse plus précisément l'alignement de la séquence des clones C201-C286 (662 nucléotides), il apparaît qu'elle s'aligne en continu (bases 50 à 628) sur le BAC *Hs* 12 BAC RP11-1100L3 (AC025259.49) mais également (bases 193-628) sur le clone *Hs* cDNA FLJ39497 *fis* clone PROST 2016582 (AK 096816.1).

Ces deux séquences ne représentent pas des transcrits variants connus du gène NR4A1. Il existe 3 transcrits variants connus codant pour 2 isoformes protéiques de Nr4a1. La figure 39 présente ces 3 transcrits et l'ADNc séquencé à partir des clones 201 et 286. Dans cet alignement, on s'aperçoit que la séquence d'intérêt rassemble l'exon 6, l'intron situé entre les exons 6 et 7, l'exon 7 et l'exon 8 du gène NR4A1.

Toutes les autres séquences analysées et soumises au programme BLAST nous ont fourni un alignement continu par rapport aux séquences des bases de données.

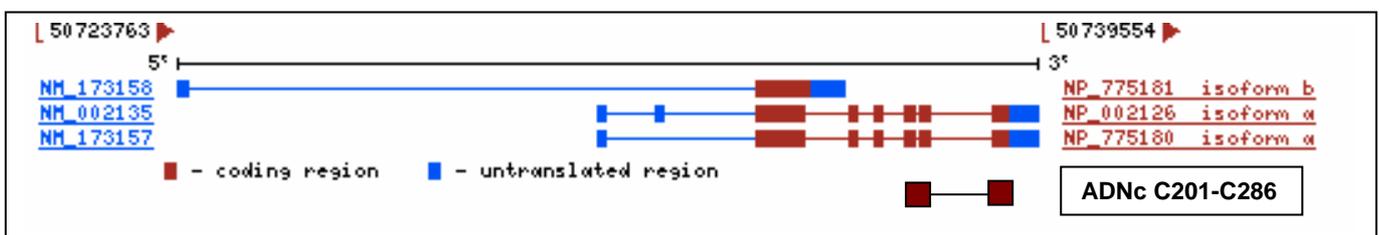


Figure 39. Structure exons / introns des différents ARNm variants de NR4A1 et alignement avec l'ADNc contenu dans les clones C201 et C286.

Le tableau 10 présente les résultats obtenus par le programme TBLASTX.

N° Clone	ORF-RNAP	Nb AA	TBlastX	N° Genbank	E-value	% identité	Force I
C142	phase 2	150	<i>Hs</i> cytochrome c oxydase subunit 3	AAU08540.1	1,00E-69	99	moyen
C170	phase 2	178	<i>Hs</i> PPP2CB	AAV38333.1	8,00E-69	78	moyen
C201/C286	Phase 1	99-64	<i>Hs</i> NR4a1 (HOL I)	CAG32985.1	3,00E-29	81-95	forte
C228/C89	phase 1	124	<i>Hs</i> G elongation factor G1	AAK53402.1	2,00E-66	100	moyen
C164							moyen
C193							moyen
C195	Phase 3	30	<i>Hs</i> HSPCA90 protein	AAH07989.2	5,00E-09	100	moyen
C250	Phase -3	45	<i>Hs</i> granulin isoform 2 precursor	AAH00324.1	4,00E-17	84	moyen

Tableau 10. Résultats de comparaison des séquences des interacteurs positifs validés aux banques de données protéiques via le programme TBlastX.

La force I évalue la force d'interaction par rapport au temps de croissance observé des colonies par rapport au contrôle positif.
e désigne la probabilité d'identité des séquences identifiées par rapport aux séquences soumises

Pour les clones C201-C286, nous retrouvons un alignement discontinu sur la séquence protéique de Nr4a1. En fait, seules les bases 4 (18) à 246 s'alignent sur cette séquence protéique en phase 1, c'est-à-dire forment une protéine de fusion avec la sous-unité α de l'ARN polymérase. Ces 64 acides aminés correspondent au domaine HOL1 de la protéine Nr4a1. Ce motif protéique est un domaine de liaison au ligand présent dans les récepteurs aux hormones. C'est donc plus probablement ce motif protéique qu'est capable de lier FBXO47. Pour le confirmer, nous avons réalisé un double hybride dirigé en cotransformant pBT-FBXO47 et pTRG-Nr4a1 (ADNc entier codant pour l'isoforme 1) et l'interaction n'était pas positive. Nous ne pouvons donc pas considérer Nr4a1 comme cible pour Fbxo47.

Parmi tous les autres clones, seuls les clones C228-C89 (contenant le même ADNc) qui fournissent une séquence protéique en phase avec la sous-unité α de l'ARN polymérase. En effet, les clones C142 et C170 codent, respectivement, la sous-unité 3 de la cytochrome c oxydase et l'isoforme β de la sous-unité catalytique de la protéine phosphatase 2A. Ils sont en phase 2 par rapport à la sous-unité α de l'ARN polymérase. Le clone C195 contient, en phase 3, 30 acides aminés de la protéine Hsp90 α 2 et le clone C250 contient, en phase -3, 45 acides aminés de la forme précurseur de la granuline. Les clones C164 et C193 n'ont pas de séquences protéiques connues. Nous considérons que les clones positifs et validés comme interacteurs potentiels de FBXO47 sont ceux dont la séquence d'ADNc de la banque est en phase (ici phase 1) avec la sous-unité α de l'ARN polymérase. En effet, in vivo, l'interaction doit rassembler la sous-unité α de l'ARN polymérase et la protéine répresseur de bactériophage λ cI.

Le facteur d'élongation mitochondrial G1 (GFM1 ou mEF1) serait donc le seul partenaire protéique de Fbxo47 identifié par cette technique.

Nous avons également testé l'interaction entre le clone C228, contenant l'ADNc du facteur d'élongation mitochondrial G1, et la forme sans F-box de la protéine Fbxo47. L'interaction est toujours positive, et de même « affinité » (temps de pousse) qu'avec la forme sauvage (Figure 39). Cette interaction est donc indépendante du domaine F-box de notre protéine, et nous permet donc de supposer l'existence d'un autre domaine d'interaction protéique entre les acides aminés 156 et 453 de Fbxo47.

La comparaison de séquence protéique entre le clone 228 et la séquence de référence (AKK53402.1) de GFM1, montre que Fbxo47 interagit avec la partie comprise entre les acides aminés 609 à 732 de celle-ci. Ces résidus correspondent au domaine nommé EFG_C (pfam00679.11), qui est retrouvé dans la région C-terminale des facteurs d'élongation G et dans certaines protéines de résistance à la tétracycline. Il adopte une conformation « ferredoxin-like », dont la fonction n'est pas connue dans Gfm1. Néanmoins, dans d'autres protéines où cette conformation existe, la structure suggère un rôle dans les interactions protéine-protéine, peut-être régulées par la liaison de petites molécules ligands.

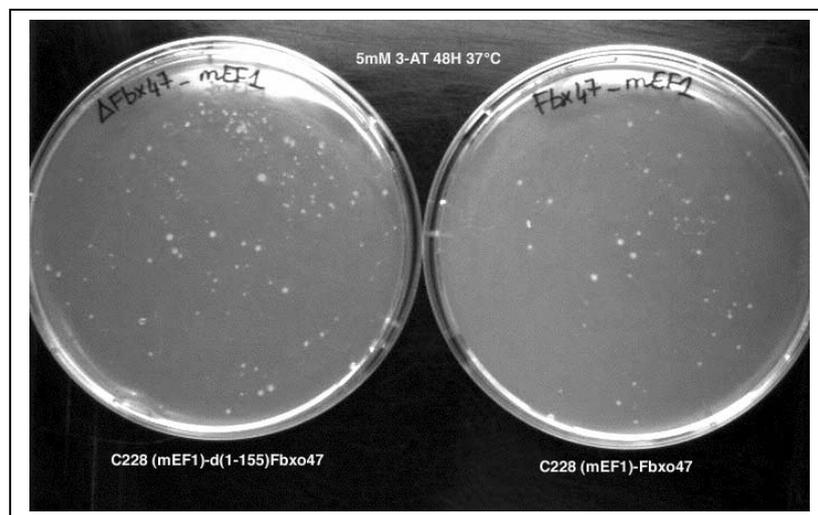


Figure 40. Visualisation de l'interaction Gfm1 et formes avec et sans F-box de la protéine Fbxo47.

CHAPITRE 4

DISCUSSION

A. Clonage du gène FBXO47

I. De la présence d'un gène suppresseur de tumeur en 17q12

I.1. Perte d'hétérozygotie en 17q12 et cancer papillaire rénal

L'inactivation de gènes suppresseurs de tumeur via la perte d'un allèle au niveau somatique (LOH, hyperméthylation) et l'inactivation de l'autre allèle constitutionnel par mécanismes mutationnels variés, constitue l'un des processus les plus courants dans le développement d'un certain nombre de cancers²²⁴⁻²²⁵. Des pertes d'hétérozygotie sur le bras long du chromosome 17 ont été observées dans certains cancers comme le mélanome, le cancer du sein, de l'ovaire, de la prostate, de la thyroïde ou encore dans les carcinomes pancréatiques²²⁻²²⁶⁻²²⁹. L'ensemble de ces données converge vers l'existence d'une région candidate contenant un gène suppresseur de tumeur dans la région 17q12.

Nous nous sommes plus précisément intéressés aux délétions de la bande q12-q21 du chromosome 17 dans 15 cas de cancers papillaires rénaux. Les publications récentes concernant les anomalies moléculaires dans les carcinomes papillaires rénaux sont peu claires et contradictoires. En effet, les pRCC sont normalement caractérisés par des trisomies ou des tétrasomies des chromosomes 7 et 17 et la perte du chromosome Y⁵⁻²³⁰⁻²³¹. Or des délétions sur plusieurs locis ont été également décrites dans des cas de pRCC⁷⁻²²⁹⁻²³²⁻²³⁴. Pour définir les altérations chromosomiques en 17q12-q21 sur nos patients, nous avons réalisé une étude LOH à l'aide de 8 marqueurs microsatellites. La zone minimale de délétion obtenue se situe autour du marqueur D17S250 (17q12) puisque d'une part, deux des patients analysés perdent uniquement ce marqueur, et d'autre part, il s'agit du marqueur le plus fréquemment perdu (50% des cas informatifs de notre étude). Remarquons également que par rapport au marqueur D17S250, les fréquences de pertes des autres marqueurs (% LOH) sont plus importantes vers le télomère (>30%) que vers le centromère (8%).

La figure 22 (Chapitre 3 : Résultats) montre les gènes connus présents à proximité de D17S250. Ce marqueur n'est pas intragénique, c'est-à-dire qu'il ne se situe pas dans une région codante du génome. Nous remarquons, de façon très intéressante, que dans cette région plutôt riche en gènes, il n'y a pas de gène identifié à proximité directe du marqueur D17S250. Les gènes les plus proches sont : LASP1, codant pour une protéine de type LIM SH3,

(fusionné avec MLL dans certaines leucémies myéloïdes ²⁸) et PLXD1 (ou TEM7), codant une protéine transmembranaire à domaine « plexin-like », qui montre une expression très élevée lors de l'angiogenèse tumorale ²³⁵. A priori, ces deux gènes peuvent être plutôt considérés, au vu de leurs caractéristiques, comme des oncogènes, ce qui ne conforte pas notre hypothèse de gène candidat.

Néanmoins, il existe dans cette région à la fois des oncogènes (CCR7) et des gènes suppresseurs de tumeurs (BRCA1), dans une région chromosomique restreinte.

1.2. Expression différentielle tissu normal/tissu tumoral

Une autre caractéristique qui qualifie un gène impliqué dans le processus oncogénique est son niveau d'expression. Un oncogène est surexprimé dans une cellule tumorale par rapport à une cellule normale tandis qu'un gène suppresseur de tumeur est souvent sous exprimé voir non exprimé. La plupart des études sur ce type de gènes utilisent des techniques d'expression différentielle comparant des échantillons sains et tumoraux afin d'identifier des gènes sur ou sous-régulés.

1.2.1. Pertinence de la technique

Ces techniques nécessitent, pour être performantes, de cribler un grand nombre d'échantillons. La membrane que nous avons choisie et utilisée contient 154 paires d'échantillons provenant de 19 tissus différents (*Cancer Profiling Array II* Clontech), assurant ainsi une représentativité pertinente.

Ces échantillons sont des ADNc, normalisés quantitativement et fixés sur une membrane de nylon. Ceci assure une grande sensibilité, mais également une bonne représentation des transcrits rares.

La sonde que nous avons désignée correspond à la partie 3' de l'ADNc de FBXO47, elle est donc spécifique de l'ensemble des transcrits potentiels du gène.

1.2.2. Conclusions

Les résultats d'hybridation obtenus avec notre sonde montrent sur l'ensemble des tissus testés, que le gène FBXO47 est plus exprimé dans le contexte sain que tumoral. On le constate aussi dans le rein, où sur l'ensemble des intensités de signaux quantifiées le ratio

normal sur tumoral est proche de 1,5. De façon remarquable, FBX047 est également deux fois moins exprimé dans les cancers du pancréas et de la vulve que dans la composante saine. Notons également que l'on n'observe pas d'extinction totale de l'expression des messagers de FBX047 en contexte tumoral.

I.3. FBX047 est-il un gène suppresseur de tumeur?

Nous avons identifié un nouveau gène localisé en 17q12, région fréquemment délétée dans un certain nombre de cancers et dans la moitié des cas de pRCC que nous avons étudiés. De plus, son expression apparaît réduite dans les échantillons tumoraux par rapport aux échantillons normaux correspondants dans un certain nombre de tissus testés (et notamment le rein). Ces résultats confortent notre hypothèse d'un rôle de gène suppresseur de tumeur pour FBX047.

Néanmoins, en accord avec la théorie de Knudson, les gènes suppresseurs de tumeur doivent être inactivés sur les deux allèles dans la cellule tumorale⁴⁸. La délétion génomique reste toujours l'inactivation d'un des deux allèles la plus fréquente. La perte du premier allèle est sujette à des phénomènes plus variés. En effet, dans les cancers de type familiaux, c'est souvent une mutation germinale et le processus est alors simple. Dans les cancers sporadiques, à l'exception de certains gènes comme le gène du rétinoblastome (RB) et le gène APC dans le côlon, la probabilité de trouver une mutation somatique dans un gène suppresseur de tumeur localisé dans une région avec une fréquence de LOH forte a été montrée comme étant très faible²³⁶. Ceci est particulièrement vrai pour les gènes suppresseurs de tumeur de type « *caretaker* » comme le gène BRCA1 dont on ne connaît pas de point de mutation dans les cancers sporadiques du sein²³⁷.

Cette inconsistance entre le modèle classique des « deux évènements » et les observations faites dans les cancers sporadiques repose sur deux explications possibles : le phénomène d'haploinsuffisance ou les altérations épigénétiques, se traduisant par l'hyperméthylation du promoteur ou des mutations dans la région promotrice.

Premièrement, le mécanisme d'haploinsuffisance montre que des mutations inactivatrices et la perte totale de fonction ne sont pas toujours nécessaires pour annuler le rôle de suppresseur de tumeur d'un gène²³⁸. Un exemple typique d'haploinsuffisance a été démontré avec le gène p27, une protéine suppresseur de tumeur qui inhibe les kinases dépendantes des cyclines et bloque la prolifération cellulaire²³⁹.

Deuxièmement, la seconde hypothèse repose sur les altérations épigénétiques comme l'hyperméthylation du promoteur ou des mutations dans les régions régulatrices. Ceci survient à un niveau soit transcriptionnel soit traductionnel pour affecter l'expression normale du gène suppresseur de tumeur plutôt que de l'abolir totalement. Des études analysant la méthylation en parallèle d'une étude LOH ont montré que la fréquence avec laquelle la méthylation du promoteur contribuait à l'inactivation de gène était de 33% pour le gène VHL dans le syndrome Von Hippel Lindau²⁴⁰ et de 84% pour le gène MLH1 dans les tumeurs colorectales avec une instabilité microsatellitaire²⁴¹.

Nous avons cherché à étudier à la fois, les mutations dans le gène FBXO47 et la présence ou non d'une hyperméthylation du promoteur dans l'îlot CpG identifié, *in silico*, pour FBXO47. Nous n'avons pu mener ni l'une ni l'autre de ces expériences en raison d'un manque de matériel génomique. En effet, nous avons réalisé notre étude LOH à partir d'ADN extrait sur des lames d'histopathologie. Le tissu est inclus en paraffine puis monté sur lame. Les techniques actuelles nous ont permis d'extraire de l'ADN des patients à partir de ces lames, puis de réaliser des PCRs amplifiant les microsatellites. Ces produits d'amplification sont de très petite taille (< 200 pb). Par contre, lorsque l'on tente d'amplifier des produit plus grands, comme des exons entiers ou une région promotrice, la qualité et/ou la concentration de l'ADN extrait sur lames ne le permet plus. En effet, l'ADN extrait de tissu qui a été exposé à la formaline pendant une longue période est fortement fragmenté et l'efficacité de la PCR diminue proportionnellement avec l'augmentation de la taille du fragment à amplifier²⁴²⁻²⁴³. Il n'a pas été possible par ailleurs d'obtenir, du tissu frais des patients testés, afin d'en ré-extraire un ADN de bonne qualité, utilisable pour la recherche de mutations.

Le rôle de gène suppresseur de tumeur pour FBXO47 reste donc à prouver par des analyses complémentaires.

II. Profil d'expression du gène FBXO47.

Pour valider biologiquement l'existence d'un gène, il faut identifier son ARN messager dans un ou plusieurs tissus. Par la technique de PCR exon-connection, nous avons validé la structure intron/exon prédite *in silico* de notre gène d'intérêt. Ensuite, des techniques d'élongation de l'ADNc en 5' et 3' nous ont permis d'obtenir les régions non traduites de FBXO47.

II.1. Pertinence des techniques d'étude d'expression de gènes

L'étape suivante a consisté à examiner le profil d'expression tissulaire de ce gène. Il existe différentes techniques dont la plus commune est l'hybridation d'une sonde spécifique du gène sur Northern Blot.

II.1.1. Northern Blot

Les résultats obtenus sur les deux membranes commerciales (*MNT Blot*) se sont révélés très décevants. En effet, nous n'avons obtenu aucun signal d'hybridation avec la sonde 3' spécifique de FBXO47, et ce même après de longues périodes d'exposition. Plusieurs raisons peuvent être avancées :

- Les tissus exprimant le gène ne sont pas représentés sur la membrane d'hybridation.
- Cette technique connaît des limites lorsque le transcrit est peu exprimé²⁴⁴.
- Le gène peut avoir une expression spatio-temporelle très limitée et les Northern Blots utilisés ici ne comportent pas d'ARN de ces cellules au stade de développement où s'exprime le gène recherché.

En réalité, ce résultat n'est pas si surprenant. En effet, l'étude bioinformatique des ESTs de la région contenant notre gène candidat n'avait fourni aucun résultat. Or les EST sont représentatifs de gène dont l'expression relativement forte.

Il est donc fort probable que l'absence de détection en Northern Blot soit corrélée à la faible expression du gène FBXO47.

II.1.2. RT-PCR

Afin d'améliorer la sensibilité de la recherche d'expression de FBXO47, nous avons réalisé une PCR sur des panels d'ADNc commerciaux, rassemblant 16 tissus normaux et 8 tissus fœtaux. L'utilisation d'ADNc « prêt à l'emploi » garantit une qualité et une normalisation des échantillons de départ et ainsi des résultats observés. Notons que le produit d'amplification obtenu correspond à la sonde utilisée pour hybrider les membranes de Northern Blot, ce qui nous permet de comparer à la fois la sensibilité des techniques et les résultats.

Les résultats obtenus sur les tissus adultes sont plus probants que par la technique de Northern Blot. En effet, nous observons une expression forte dans le tissu testiculaire et plus faible dans le pancréas, or ces tissus n'étaient pas testés sur les membranes de Northern Blot utilisées.

II.1.3. Hybridation sur membrane ADNc (CPA II)

L'expérience d'expression différentielle, réalisée par hybridation de la sonde FBXO47 spécifique, nous informe également sur le profil de transcription de notre gène puisqu'elle contient 154 ADNc normaux provenant de 19 tissus différents. Cette technique est très sensible par rapport au Northern Blot. En revanche, elle ne donne aucune information quant à la taille du transcrit ou à l'existence de transcrit alternatifs.

Les résultats obtenus tranchent avec ceux obtenus par les autres techniques car FBXO47 apparaît dans tous les tissus testés. Néanmoins, son expression reste faible comparée au témoin ubiquitine. Les tissus exprimant le plus notre gène sont le rein, le foie, la vulve, et le pancréas. L'expression dans le pancréas, détectée en RT-PCR est confirmée. En revanche, la vulve n'était pas testée ni sur le Northern Blot ni en RT-PCR, par contre le foie et le rein qui ne montraient pas d'expression avec les deux techniques précédentes, sont positifs pour l'expression de FBXO47.

II.2. Synthèse : Profil d'expression du gène FBXO47.

La recherche du profil d'expression du gène FBXO47, dans l'ensemble des tissus testés par les différentes techniques, est résumée dans le tableau 11.

Les résultats ne sont pas concordants dans les trois expériences, pourtant la sonde utilisée pour l'hybridation des deux types de membranes, concorde avec le produit de PCR obtenu par les amorces utilisées pour amplifier le transcrit spécifique en RT-PCR.

	Northern Blot	RT-PCR	CPA II
Cerveau			
Cœur			
Muscle			
Côlon			
Thymus			
Rate			
Rein			
Foie			
Intestin			
Placenta			
Poumon			
Sang			
Pancréas			
Prostate			
Testicule			
Ovaire			
Sein			
Estomac			
Vessie			
Vulve			
Trachée			
Utérus			
Col utérus			
Rectum			
Thyroïde			
Peau			

Non testés

Absence d'expression

Expression faible

Expression moyenne

Expression forte

Tableau 11. Comparaison des résultats obtenus pour l'étude de l'expression de FBXO47 dans des tissus normaux adultes par les trois techniques employées : Northern Blot, RT-PCR, et CPA II (hybridation sur ADNc).

Plusieurs hypothèses peuvent être proposées :

Premièrement, en ce qui concerne les deux types d'hybridation, Northern Blot et CPA, la sensibilité de la technique peut être à l'origine des différences observées (rein et foie). En effet, dans le cas du Northern Blot, l'hybridation s'effectue entre de l'ARN et de l'ADNc, alors que dans le cas de la CPA, elle a lieu entre deux ADNc. De plus, le transcrite de FBXO47 reste rare comparé au témoin ubiquitine, ne favorisant pas un signal fort en Northern Blot.

Deuxièmement, la RT-PCR est une technique qui amplifie le signal de base tout en respectant la proportion de transcrite présent dans le tissu testé. Elle a été montrée comme étant 1000 à 10000 fois plus sensible que d'autres techniques traditionnelles d'hybridation ARN²⁴⁵⁻²⁴⁶. Cette hypothèse reste valable par l'utilisation d'un témoin interne pour les techniques de RT-PCR et de CPA. Il est donc difficile d'expliquer pourquoi le rein, le foie et le côlon semblent exprimer fortement FBXO47 dans l'expérience d'hybridation ADNc alors qu'aucun

signal pour ces mêmes tissus n'est visualisé en RT-PCR. De la même façon, pourquoi le tissu testiculaire n'exprime pas plus FBXO47 sur la CPA alors qu'en RT-PCR l'expression est forte et comparable à celle de l'ubiquitine ?

Troisièmement, on pourrait remettre en cause la nature des échantillons testés (nombre d'individus, type cellulaire) dans les techniques reposant sur de l'ADNc. En effet, en RT-PCR chaque tissu est représenté par de l'ADNc obtenu par rétro-transcription puis normalisation des taux d'expression d'ARNm par au moins 4 gènes de « ménages ». L'ARN source est de l'ARN polyA+ issu d'au moins deux individus différents pour chaque tissu. Néanmoins, il n'y a pas, dans un même tissu, de différenciation des différents types cellulaires pouvant être présents. En ce qui concerne la CPA, il y a plusieurs échantillons par tissu, chaque échantillon tumoral ayant été obtenu à partir d'un individu est caractérisé par le type cellulaire, le stade, la classification néoplasique et la morphologie néoplasique (Annexe A). L'échantillon normal correspondant est issu du même tissu chez le même patient.

Le testicule est un organe particulier puisqu'il est formé de couches multiples composées de cellules à des stades de différenciation variés ²⁴⁷⁻²⁴⁸. Il aurait donc été intéressant d'évaluer l'expression de FBXO47, selon les différents types cellulaires. Nous n'avons pas réussi à nous procurer de l'ARN correspondant à cette recherche. Nous ne pouvons donc pas savoir si FBXO47 a un profil d'expression particulier dans le testicule, ce qui nous aurait peut-être permis d'apporter des informations sur le rôle de la protéine codée.

B. Caractéristiques de la protéine Fbxo47

I. Fbxo47, un nouveau membre de la famille des F-box

Le gène que nous avons identifié en 17q12 code pour une protéine de la famille des F-box. Cette famille regroupe toutes les protéines contenant le motif F-box, consensus dégénéré, d'environ 40 paires de bases.

En mars 2006, cette famille regroupait chez les eucaryotes, 4459 protéines (base de donnée *EntrezProtein NCBI*), dont 452 chez les végétaux, 91 chez les levures, 107 chez la Drosophile, 514 chez la Souris et 498 chez l'Homme. C'est donc une famille protéique très grande et très conservée dont les membres possèdent tous en commun ce domaine F-box, caractérisé comme domaine de liaison à la protéine Skp1 ⁹⁹.

La figure 41 présente la distribution taxonomique des protéines F-Box selon la base de données InterPro (<http://www.ebi.ac.uk/interpro/DisplayIproEntry?ac=IPR001810>).

La distribution évolutive des protéines F-box reflète leur rôle logistique de protéines interchangeables dans le cœur invariant du système ubiquitine ligase SCF. Quelque soit l'espèce donnée, on trouve une petite variété de culline-1 et de protéines Skp1 tandis qu'il existe un très grand nombre de protéines F-box correspondantes (par exemple, 1 SKP1 et 117 protéines F-Box chez l'homme contre 10 à 20 SKP1 et 350 F-box chez *Caenorhabditis elegans*)²⁴⁹. Selon les bases de données consultées, les chiffres varient mais les proportions entre les espèces restent les mêmes.

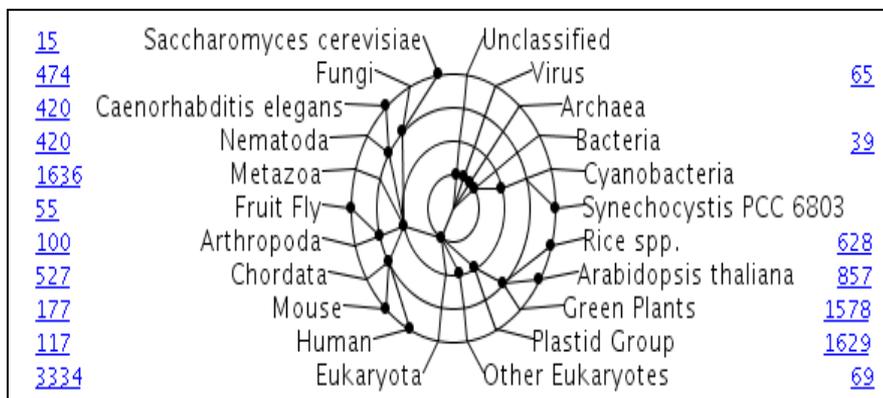


Figure 41. Distribution taxonomique des protéines F-Box selon la base de données InterPro. (<http://www.ebi.ac.uk/interpro/DisplayIproEntry?ac=IPR001810>)

Les grands cercles représentent les branches distales de l'arbre phylogénétique, qui sont le genre et l'espèce, opposés aux petits cercles pleins (noirs), qui représentent l'ordre et le règne.

Le nombre en bleu représente le nombre de protéines F-Box prédites dans l'organisme correspondant.

I.1. Fonctionnalité du domaine F-Box

Le domaine F-box est considéré fonctionnel lorsqu'il fournit la capacité de se lier à la protéine Skp1. Après avoir identifié par bioinformatique l'existence d'un tel motif dans Fbxo47, nous avons démontré sa fonctionnalité par deux méthodes : une complexion protéique *in vitro* et une interaction protéique *in vivo* par le système de double-hybride dirigé en système bactérien. Ces deux techniques fournissent le même résultat : Fbxo47 n'est capable de se lier à la protéine Skp1 qu'en présence du domaine F-box, celui-ci est donc fonctionnel.

Ce type d'interaction est d'intensité plutôt moyenne, c'est pourquoi la pousse des colonies positives dans le système double-hybride est plus lente (24 heures supplémentaires) qu'avec le couple témoin positif.

Les interactions intramoléculaires sont responsables de la conformation spatiale des protéines, et dans une moindre mesure de celle des peptides. Les principales interactions non-covalentes sont soit de nature électrostatique, comme les liaisons ioniques, les liens hydrogène et les interactions de type van der Waals, soit de nature hydrophobe, qui implique davantage des interactions avec le solvant. Les interactions covalentes se limitent quant à elle à l'établissement de ponts disulfures entre deux résidus cystéine. Dans le cas de l'association Skp1/F-box, l'interaction est de type électrostatique.

Les interactions électrostatiques de type Van der Waals sont des liens d'assez faible énergie, entre $1-7 \text{ kcal mol}^{-1}$ ²⁵⁰, et la force de l'interaction est alors proportionnelle à l'électronégativité des molécules.

1.2. Le second domaine d'interaction protéique

Nous avons identifié, bioinformatiquement par le programme *ScanProsite*, un domaine Leucine-Zipper (**L**fkrct**L**lpther**L**kyihki**L**) entre les acides aminés 120 à 141 avec une *e-value* de $0,2311 \times 10^{-2}$. Cette probabilité reste moyenne, et donc sa fonctionnalité *in vitro* et *in vivo* reste à prouver.

Le motif Leucine Zipper n'est pas à proprement parler un motif de fixation à l'ADN. Il est cependant important de l'évoquer car c'est un motif structural qui va permettre le positionnement correct d'une région de la protéine se fixant à l'ADN. Ce motif a été identifié à l'origine dans une protéine enhanceur. Il est caractérisé par une répétition de leucine tous les 7 résidus, elles sont ainsi toutes situées du même côté d'une hélice α . En général cette hélice est très longue avec un minimum de 6 tours d'hélice. De plus cette hélice est amphipatique, ce qui va permettre sa stabilisation vis à vis du corps plus ou moins globulaire de la protéine. Cette hélice va interagir fortement avec la même hélice d'un autre monomère et former ainsi le zip.

I.3. L'appartenance à une ligase E3 ?

La grande majorité des protéines F-box connues appartiennent au système SCF. Ce complexe multi-protéique fonctionne comme une ligase E3 dans le système d'ubiquitination. Il est composé de quatre sous-unités principales : trois invariantes, la culline 1, Skp1 et Roc1, et une sous-unité variable, une protéine F-box.

Nous avons montré que Fbxo47 était capable d'interagir avec Skp1, mais nous ne l'avons pas isolée dans un tel complexe multi-protéique. Ainsi, nous ne pouvons pas conclure catégoriquement quant à son appartenance à une ligase E3. Nous ne pouvons pas non plus écarter la possibilité qu'elle intervienne dans un autre complexe puisqu'il existe des protéines F-box s'associant à d'autres cullines, comme Cul7 et Fbxw6 par exemple ⁷⁷, ou n'appartenant pas à une ligase E3, comme la protéine F-box Emi-1 ⁹³.

II. Fbxo47 : une protéine mitochondriale.

L'étude bioinformatique de la séquence protéique de Fbxo47 nous a révélé l'existence d'une séquence d'adressage mitochondrial de type pré-séquence.

La mitochondrie est un organe contenant plusieurs centaines de protéines différentes. L'immense majorité des protéines mitochondriales est codée par le génome nucléaire, synthétisée dans le cytoplasme puis importée dans les mitochondries. Ces protéines possèdent un signal d'adressage qui leur permet d'être importées dans les mitochondries ²⁵¹. Deux types de signaux d'adressage mitochondrial co-existent : les pré-séquences et les signaux internes. L'importation des protéines dans les mitochondries a été essentiellement étudiée dans la levure mais, les mécanismes mis en oeuvre semblent bien conservés dans le règne des eucaryotes ²⁵².

II.1. Le signal d'adressage mitochondrial de type pré-séquence

La pré-séquence consiste en une extension N-terminale de la protéine à importer. Lors de l'importation d'une protéine contenant une pré-séquence dans une mitochondrie (Figure 42), la protéine est préalablement dénaturée par une chaperonne cytosolique de type Hsp70 puis la pré-séquence est reconnue par les récepteurs Tom20 et Tom22 qui sont localisés au niveau de la membrane mitochondriale externe. Ces deux récepteurs font partie du complexe TOM (*Translocase Outer Membrane*). La protéine traverse la membrane externe en

empruntant le pore du complexe TOM formé par Tom40. Puis elle emprunte le canal formé par Tim23 au niveau du complexe TIM23 (*Translocase Internal Membrane* contenant Tim23). Le domaine N-terminal de Tim23 s'accroche à la membrane externe et ainsi les complexes TOM et TIM23 se retrouvent proches l'un de l'autre²⁵³.

Un potentiel transmembranaire au niveau de la membrane interne est indispensable pour l'importation des protéines contenant une pré-séquence. Il permet d'attirer les préséquences chargées positivement dans le pore du complexe TIM23 (le côté matriciel de la membrane interne est chargé négativement)²⁵⁴. Une chaperonne mitochondriale associée au complexe TIM23, mtHsp70, tire la protéine dans la matrice. Ce processus consomme de l'ATP. Lorsque la pré-séquence a pénétré dans la matrice, elle est clivée par une protéase MPP (*Mitochondrial Processing Peptidase*). La protéine importée se trouve alors dans la matrice mitochondriale.

Cette voie d'importation est appelée voie générale d'importation car elle concerne un très grand nombre de protéines mitochondriales. Certaines protéines importées par cette voie sont réexportées dans la membrane interne par la machinerie d'exportation des protéines codées par le génome mitochondrial²⁵⁵. D'autres sont arrêtées en cours d'importation et transférées dans la membrane interne²⁵⁶ ou l'espace intermembranaire (Figure 42).

II.2. Localisation cellulaire de Fbxo47.

La localisation cellulaire d'une protéine permet d'en éclairer davantage la fonction. Puisque Fbxo47 semble posséder une pré-séquence caractéristique des protéines de la matrice mitochondriale, sa localisation doit être mitochondriale et ce quelque soit le modèle cellulaire étudié.

Le premier point que nous voulons discuter dans cette partie est la faible expression du plasmide contenant la protéine de fusion Fbxo47 par rapport au contrôle (EGFP seule) ou encore à la construction mutante malgré la présence d'un seul et même promoteur fort sur le plasmide. De plus, le taux de transfection des cellules COS avec les plasmides contenant les différentes formes de Fbxo47 était inférieur à celui observé avec le plasmide EGFP seul, bien que la conservation du rapport taille de l'insert sur quantité soit respectée dans les trois types de transfection.

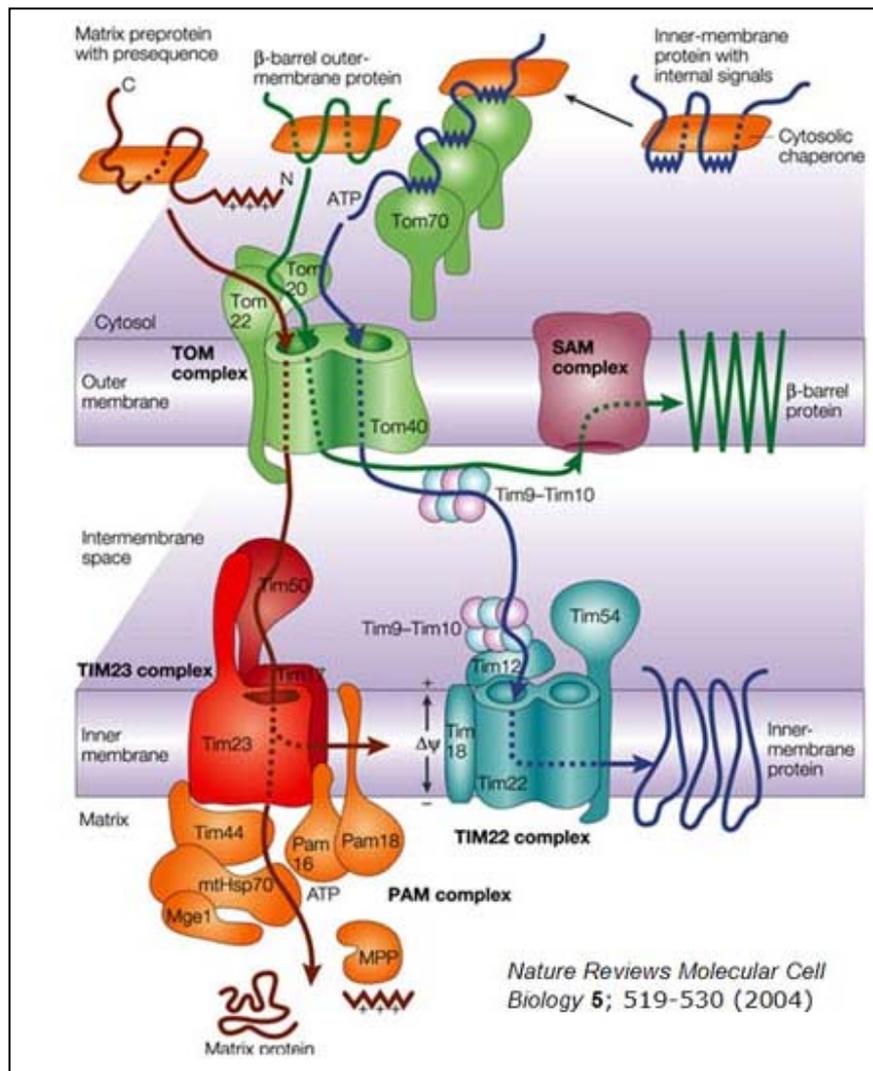


Figure 42. L'importation des protéines à la mitochondrie.

Les protéines synthétisées sous forme de précurseurs dans le cytosol sont maintenues dans un état conformationnel pour l'importation, par des chaperonnes qui hydrolysent l'ATP. La reconnaissance d'une pré-séquence par un récepteur de la membrane externe (TOM) est suivie par la translocation du précurseur à travers les deux membranes mitochondriales, la dernière translocation nécessitant un certain potentiel de membrane. La protéine précurseur est ensuite extraite activement des membranes via Hsp70. Enfin, lorsque la protéine se trouve dans la matrice, les peptidases mitochondriales de process clivent la pré-séquence : la protéine est alors sous sa forme mature.

Les hypothèses que nous pouvons proposer par rapport à ces observations sont multiples. Tout d'abord, on ne peut exclure une certaine toxicité de la protéine, ou son interaction avec d'autres protéines, ce qui provoquerait dans les deux cas une visualisation de l'EGFP dans de moins nombreuses cellules. En effet, le taux de transfection s'observe par le nombre de cellules « vertes » observées, alors qu'il correspond en réalité au nombre de cellules ayant intégré le plasmide. Ensuite, la technologie de transfection employée est une méthode non-virale, les conditions optimales de transfection ne dépendent donc pas de la taille du plasmide mais du type cellulaire à transfecter, et nous avons employé la méthodologie recommandée pour les cellules COS.

En dépit de ces faibles rendements de transfection, la construction de Fbxo47 avec l'EGFP nous a permis d'observer que Fbxo47 semble bien située dans la mitochondrie ou du moins à proximité très immédiate. Par contre, on ne la visualise pas dans l'ensemble du réseau mitochondrial mais seulement dans les fortes concentrations de mitochondries. Or dans d'autres exemples de protéines mitochondriales, le marquage observé correspond en totalité à celui visualisé avec le MitoTracker, celui cependant ne s'accumule que dans les mitochondries actives (c'est-à-dire avec un potentiel de membrane négatif). La construction contenant la forme amputée de l'adressage mitochondrial de Fbxo47 en phase avec l'EGFP, nous a prouvé qu'en l'absence de la pré-séquence, notre protéine perdait sa localisation mitochondriale et devenait alors une protéine cytoplasmique.

La localisation en foci pour Fbxo47 peut avoir plusieurs explications.

La première hypothèse repose sur le fait que nous avons réalisé une expression transitoire de la protéine de fusion (48 h maximum) et que cela reste insuffisant pour couvrir le réseau mitochondrial, c'est-à-dire que la protéine Fbxo47 n'est pas suffisamment exprimée. Nous n'avons pas pu, faute de moyens techniques réaliser une construction stable Fbxo47-EGFP.

Enfin, la seconde hypothèse est que le tag EGFP, empêche partiellement l'importation dans la mitochondrie, pour des raisons conformationnelles.

II.3. Le système F-box et la mitochondrie

Nous avons identifié un gène codant pour une protéine F-box localisée dans la mitochondrie. Quelles en sont les implications ?

La grande majorité des protéines F-box connues appartient au système SCF. Ce type d'ubiquitine ligases E3 fonctionne en partenariat avec le protéasome 26S. Découvert au début des années 80²⁵⁷, le protéasome est un complexe cytosolique et nucléaire. En microscopie électronique, il apparaît formé d'une partie centrale cylindrique, correspondant au protéasome 20S, entourée de deux complexes orientés en direction opposée et représentant chacun le complexe 19S. Le protéasome 26S est impliqué dans la dégradation des protéines marquées à l'ubiquitine. Il serait responsable de la dégradation de la plupart des protéines à courte demi-vie et certaines à longue demi-vie dont la localisation peut être cytosolique, nucléaire ou membranaire²⁵⁸. Il n'a jamais été décrit de protéasome dans la matrice mitochondriale.

Pourrait-il exister un complexe ligase E3 matriciel dans la mitochondrie en l'absence de protéasome ? En effet, cela implique que les protéines poly-ubiquitinées traversent la double membrane mitochondriale pour être dégradées dans le protéasome cytosolique. Cette voie semble assez peu probable malgré la richesse de ces membranes en systèmes transporteurs, car elle nécessiterait une dépense d'énergie très importante (transport actif), mais on ne peut pas totalement l'exclure.

Les voies de dégradation des protéines mitochondriales sont, par ailleurs connues. Comme pour les protéines cytosoliques à longue durée de vie, la protéolyse des protéines mitochondriales peut être effectuée dans le lysosome par autophagie de l'organite entier²⁵⁹. Alors que ce processus est prédominant en condition de quiescence chez la levure et résulte en une dégradation non spécifique, la protéolyse spécifique des protéines mitochondriales est effectuée par des peptidases présentes dans la matrice mitochondriale²⁶⁰⁻²⁶¹. Elles sont très conservées et pour la plupart, semblent être présentes de façon ubiquitaire dans les mitochondries des cellules eucaryotes. Il existe trois types de peptidases mitochondriales, ayant chacune une fonction donnée. Tout d'abord, les peptidases de process assurant le clivage des pré-séquences d'adressage à la mitochondrie des protéines codées par le noyau²⁶²; ensuite les oligopeptidases, et enfin les protéases ATP-dépendantes, qui dégradent entièrement les protéines mitochondriales assurant ainsi un double rôle de contrôle qualité²⁶³ et de protéolyse sélective²⁶⁴.

Il existe une protéine pro-apoptotique de la matrice mitochondriale, ARTS, dont la dégradation est effectuée par le système ubiquitine-protéasome cytosolique. En effet, l'équipe de Lotan propose le modèle suivant : les cellules normales peuvent exprimer ARTS sans être détruites, car la protéine est séquestrée dans la mitochondrie où elle ne peut interagir avec ses cibles apoptotiques, les IAPs (*Inhibitors of Apoptosis Proteins*), qui sont localisées dans le cytosol. Cette hypothèse repose sur le fait que de fortes concentrations d'ARTS cytosoliques suffisent à initier l'apoptose sans autres stimuli²⁶⁵. Leurs travaux concluent que les taux de protéine ARTS sont étroitement régulés via la dégradation protéique ubiquitine-dépendante et que ce système est stoppé en réponse à un stimulus pro-apoptotique. Il existerait donc une balance entre l'importation mitochondriale de la protéine ARTS nouvellement synthétisée et son efflux suivi de son immédiate ubiquitination/dégradation dans le cytosol, ceci permettant de maintenir un taux nul d'ARTS cytosolique en conditions non-apoptotiques.

L'exemple précédent ne repose ni sur un système d'ubiquitination interne à la mitochondrie ni sur une protéine F-box.

Néanmoins, il existe dans la littérature des protéines F-box rapportées à la mitochondrie.

La protéine Mdm30 a été isolée à la fois dans le cytosol et dans la mitochondrie¹⁴⁹. Sa localisation mitochondriale n'a pas été précisément caractérisée c'est-à-dire qu'on ne sait pas si elle se trouve dans la membrane externe, l'espace intermembranaire, la membrane interne ou dans la matrice. Son rôle dans la dégradation de la protéine Fzo1, protéine de la membrane externe et l'absence d'une pré-séquence caractéristique de l'importation matricielle, laisse cependant supposer que ce n'est pas une protéine interne de la mitochondrie. De façon intéressante, lors de notre analyse systématique de la prédiction de localisation cellulaire des protéines F-box humaines présentes dans la base de données NCBI, nous avons identifié une autre protéine (que Fbxo47) présentant une pré-séquence d'adressage à la matrice mitochondriale avec une forte probabilité (>95%). Il s'agit de la protéine Fbxo33, identifiée lors d'une analyse systématique des protéines F-box de mammifères¹⁰⁵. Le rôle de cette protéine n'est pas connu, mais son gène homologue murin est surexprimé dans le cerveau après une crise d'épilepsie²⁶⁶.

Quels peuvent donc être le ou les rôles de telles protéines F-box mitochondriales en dehors de d'une protéolyse ubiquitine-dépendante ?

Une des réponses pourrait être apportée par la compréhension complète du phénomène de l'hérédité mitochondriale. La transmission uniparentale de l'ADN mitochondrial chez les mammifères est un processus connu. En effet, les œufs fertilisés et les embryons en développement ne contiennent pas d'ADN mitochondrial paternel, il existe une dégradation spécifique des mitochondries du sperme dans l'ovocyte fécondé. L'équipe de Sutovsky a montré par plusieurs études l'implication de l'ubiquitine et de son système protéolytique associé, dans la reconnaissance et la destruction spécifique des mitochondries d'origine paternelle contenues dans les spermatozoïdes²⁶⁷⁻²⁶⁸. À l'intérieur du cytoplasme d'œufs de mammifères (singe, vache), le sperme fertilisant a été trouvé ubiquitiné sur ses mitochondries. Cette observation a été effectuée à l'aide d'un anticorps monoclonal anti-ubiquitine et d'un marqueur spécifique de la mitochondrie paternelle, les deux pouvant ensuite être visualisés par imagerie. Les mitochondries et leur signal ubiquitine diminuaient jusqu'à disparaître entre les stades 4 à 8 cellules du développement embryonnaire²⁶⁸.

Une protéine membranaire mitochondriale, la prohibitine (PHB) serait la principale cible de cette ubiquitination²⁶⁹. Identifiée à l'origine comme des régulateurs négatifs potentiels du cycle cellulaire, les protéines PHB agissent comme des chaperonnes de l'assemblage des sous-unités des complexes de la chaîne respiratoire mitochondriale. Les deux protéines PHB, Phb1p et Phb2p, sont localisées dans la membrane interne de la mitochondrie où elles forment un complexe PHB. Ce complexe lie directement les protéines mitochondriales nouvellement synthétisées et les protège ainsi de la dégradation par les protéases ATP-dépendantes²⁷⁰. PHB est essentielle à un développement mitochondrial normal et une déficience PHB chez *C.elegans* a été associée à une biogenèse mitochondriale déficiente²⁷¹.

Par ailleurs, Thompson et ses collaborateurs proposent que l'ubiquitination de la prohibitine pourrait survenir à trois étapes distinctes de la vie du spermatozoïde : dans le testicule, dans l'épididyme et dans le cytoplasme de l'oocyte²⁶⁹. Les sites ubiquitinés seraient masqués par des ponts disulfures au cours du transit épидидymial et découverts uniquement lors de la décondensation du spermatozoïde dans le cytoplasme ovocytaire après fécondation. Les complexes prohibitine-ubiquitine exposés peuvent ainsi cibler la dégradation des mitochondries paternelles via le protéasome 26S ou faciliter une polyubiquitination d'autres substrats mitochondriaux. En même temps, la dégradation rapide de la prohibitine des spermatozoïdes protège l'embryon de ses effets anti-prolifératifs. Dans les spermatozoïdes défectueux, la surface d'ubiquitination survient dans l'épididyme, en accord avec

l'augmentation du nombre de substrats ubiquitinés dans les fractions défectueuses de sperme, qui est ainsi éliminé.

Il existerait donc deux voies différentes d'ubiquitination mitochondriale du sperme, dépendantes de l'état défectueux ou non des spermatozoïdes.

Dans les spermatozoïdes normaux, la prohibitine est constitutivement ubiquitinée dans le testicule et, peut-être dissimulée par des ponts disulfures afin de servir de signal protéolytique après fécondation. Dans les spermatozoïdes défectueux, la prohibitine et d'autres protéines membranaires mitochondriales pourraient être ubiquitinés à un taux plus important, assurant leur dégradation lors du passage épидидymial ²⁶⁹.

Ces phénomènes d'ubiquitination de protéines internes de la mitochondrie laissent penser que des protéines F-box pourraient jouer un rôle dans ce compartiment cellulaire, même en l'absence de protéasome 26S.

C. Quel rôle pour Fbxo47 ?

Nous avons identifié un gène, exprimé principalement dans le tissu testiculaire, codant pour une protéine F-box mitochondriale. Nous avons ensuite cherché quel pouvait être son rôle au sein de la cellule. Pour réaliser ce projet, nous avons choisi de rechercher d'éventuels partenaires protéiques, cibles ou non, par la technique du double-hybride.

I. Technique du Double-Hybride bactérien

I.1. Double-Hybride Bactérie versus Levure

Nous avons choisi la technique double-hybride dans un système bactérien plutôt que de levure pour plusieurs raisons :

- La facilité de mise en œuvre par rapport aux moyens financiers et matériels dont nous disposons : la manipulation et l'extraction d'ADN à partir de bactéries sont plus simples et moins coûteuses ²⁷².
- Le temps d'obtention de résultats ; les bactéries *E.coli* poussent beaucoup plus vite que les levures, permettant à la fois d'obtenir des résultats plus rapidement et de cribler une plus grande banque d'ADNc.

- La détection des transcrits rares : la grande efficacité de transformation obtenue avec les bactéries permet de cribler plus facilement une banque et de détecter les transcrits rares présents (ici le titre primaire de la banque était de $3,76 \cdot 10^6$ cfu).
- L'affranchissement du problème des homologues : l'utilisation d'un système bactérien réduit le risque que les souches contiennent un homologue eucaryote de l'une ou l'autre des protéines partenaires.
- La diminution du nombre de faux positifs : c'est le principal problème du système double-hybride en levure. Le taux de faux-positif dans le système levure est estimé à 20 %. Dans le système bactérien, ce taux est plus faible en raison de l'environnement procaryote, et de l'utilisation d'une double sélection (3 AT et streptomycine)²⁰⁹.

I.2. Choix de la banque testée

Nous avons criblé une banque d'ADNc de rein avec la protéine Fbxo47 comme appât. Etant donné le profil d'expression du gène, il aurait été préférable de cribler plutôt une banque d'ADNc de tissu testiculaire. Malheureusement celle-ci n'était pas disponible. En effet, nous avons utilisé une banque d'ADNc « prête à l'emploi » c'est-à-dire que les ADNc, obtenus par rétro-transcription des ARN poly A totaux d'un tissu, étaient déjà insérés dans le vecteur pTRG. Il nous était effectivement impossible d'obtenir du tissu testiculaire et de mettre en œuvre les moyens techniques nécessaires à la construction d'une banque. De plus, nous n'avons déjà pas pu obtenir de tissu testiculaire pour approfondir le profil d'expression du gène.

Néanmoins, sur la base des résultats d'expression obtenus sur la membrane CPA, l'utilisation de la banque d'ADNc de rein se justifiait pleinement pour identifier les partenaires protéiques de Fbxo47.

I.2. Limites de la technique

II.2.1. Détection de la Sensibilité des interactions

Nous avons détecté des interactions entre Fbxo47 et plusieurs clones de double-hybride. Ces interactions sont définies par la pousse des souches bactériennes sur le milieu sélectif (5 mM 3-AT). Le nombre de colonies et leur vitesse de croissance, sont plus faibles et plus lentes que ce que l'on observe avec la co-transformation des plasmides témoins.

En effet, l'interaction témoin entre les protéines LGF2 (domaine de dimérisation de l'activateur transcriptionnel Gal4, 40 aa) et Gal11p (un domaine de la forme mutante de Gal11, 90 aa) fournit un nombre de colonies réduit de 25% environ sur le milieu sélectif par rapport au milieu non sélectif. Les interactions positives observées dans notre expérience se caractérisent par un nombre beaucoup moins élevé de colonies sur le milieu sélectif que sur le milieu non sélectif. De plus, les colonies contenant les plasmides témoins apparaissent au bout de 24 heures d'incubation à 37°C alors que dans la validation des interacteurs positifs de Fbxo47, la visualisation des colonies n'est possible qu'à partir de 36 heures d'incubation.

Les deux caractères quantitatifs variants, à savoir le nombre et la durée de croissance des colonies, peuvent s'expliquer par la force de l'interaction des protéines testées, qui dépend du type de liaison entre les protéines. L'affinité sera différente selon que la liaison est permanente ou transitoire. Ceci pourrait expliquer que l'activation de la transcription du gène HIS3 soit plus ou moins grande. La taille des protéines de fusion portées par les plasmides reporteurs pourrait également constituer une explication : pour l'expérience témoin, ce sont des domaines (40 et 90 aa) qui interagissent et non pas des protéines entières (Fbxo47). Ceci pourrait faciliter l'interaction, notamment par rapport à la formation de structure tertiaire protéique.

II.2.2. Les faux négatifs

Le système double-hybride ne peut détecter toutes les interactions protéiques. Cela génère ce que l'on nomme communément des « faux négatifs », qui ne sont pas détectés. Dans le système levure, le taux de faux négatifs est estimé à environ 80 %.

Quelles sont les interactions qui ne peuvent pas être détectées ?

Les interactions impliquant des protéines qui présentent des problèmes de détection sont liées à des problèmes structuraux, de stabilité, de toxicité, de mauvaise localisation (protéines membranaires), ou de modification post-traductionnelle.

Les problèmes de repliement et de stabilité des protéines ne peuvent pas être « contrôlés ». Le phénomène de toxicité est, en revanche, minimisé dans le système bactérien par rapport au système levure. En effet, certaines protéines eucaryotes, comme les protéines des points de contrôle du cycle cellulaire et des voies de transduction de signaux, peuvent être toxiques dans la levure car elles interfèrent avec la fonction de leur homologue chez la levure.

Il peut bien sûr exister un phénomène hétérologue semblable chez la bactérie mais sa probabilité est moindre.

La limite la plus importante dans le double-hybride en bactérie reste l'inexistence de modifications post-traductionnelles chez les procaryotes. En effet, la reconnaissance d'une protéine cible par une protéine F-box nécessite, le plus souvent, une modification post-traductionnelle (phosphorylation, glycosylation) du substrat⁸⁵. Nous ne sommes donc pas en conditions optimales pour l'étude des interactions avec des protéines substrats. Néanmoins, on peut penser que ce système, produisant une forte sur-expression des protéines apportées par les plasmides pBT et pTRG, compense grâce à la forte concentration des deux protéines partenaires, permettant l'affranchissement des modifications post-traductionnelles. Seul le domaine d'interaction protéique de la protéine F-box, type LRR ou WD40 par exemple, pourrait suffire à produire une interaction positive.

Ceci ne constitue pas une limite pour les protéines partenaires (non substrats) de Fbxo47, qui, à priori ne doivent pas forcément être modifiées pour former un complexe.

Par conséquent, on ne peut pas distinguer par cette technique une protéine partenaire d'une protéine substrat de Fbxo47.

II.2.3. Qualité de la banque d'ADNc

Certains résultats obtenus par la technique du double-hybride posent le problème de la qualité de la banque d'ADNc clonée dans le vecteur pTRG.

Tout d'abord, nous avons obtenu 8 clones positifs différents c'est-à-dire que les bactéries co-transformées avec chacun de ces clones et Fbxo47 étaient capables de croître sur le milieu sélectif. Or pour la majorité d'entre eux (6/8), la recherche de phase avec la sous-unité α de l'ARN polymérase (pTRG) n'a pas été concluante. En effet, seuls quelques acides aminés (12 au maximum) sont trouvés en phase pour une protéine de fusion capable d'activer la transcription du gène reporteur. De plus dans tous ces cas, la séquence protéique trouvée en phase ne correspond à aucune protéine connue dans les bases de données. La première possibilité serait une erreur de séquençage des clones, mais ceux-ci ont été séquencés au moins deux fois et ce, avec des amorces permettant de vérifier la protéine de fusion (couvrant les deux protéines). La seconde explication pourrait être trouvée dans la construction de la banque. Celle-ci est basée sur une rétro-transcription à l'aide d'un oligonucléotide contenant un site poly(dT), un site XhoI et une séquence « GAGA ». Celle-ci protège le site XhoI d'une

reconnaissance enzymatique précoce. Le second oligonucléotide utilisé contient un site EcoRI, ainsi après digestion enzymatique, l'ADNc double brin peut être insérer de façon dirigée dans le vecteur pTRG. Néanmoins, la longueur de rétro-transcription de l'ARN n'est pas contrôlée, et donc la phase ouverte de lecture avec la sous-unité α de l'ARN polymérase n'est pas assurée. Mais ce type de vecteur ne peut alors pas produire une interaction positive. Nous ne pouvons expliquer de façon rationnelle ce que signifie la pousse tardive de ces clones, plus faible que pour les vrais positifs, mais néanmoins non négligeable par rapport aux interacteurs négatifs. En effet, ces derniers ne poussent pas même après une semaine d'incubation, et la croissance ne peut donc être due à une dégradation des antibiotiques de sélection.

Enfin, le cas du clone C201-C286 suscite également discussion. Il a constitué un des meilleurs candidats du fait de la bonne pousse des bactéries co-transformées avec Fbxo47, et ce clone. Or après séquençage et comparaison aux banques de données nucléiques et protéiques, nous trouvons un alignement discontinu à la fois sur la séquence nucléique (297 puis 193 nucléotides) et sur la séquence protéique (99 puis 64 acides aminés) du gène codant la protéine Nr4a1.

Au niveau nucléotidique, tout se passe comme si il y avait les exons 6, 7 et 8 plus l'intron 6-7 compris dans l'ADNc présent dans le vecteur pTRG (Figure 39 Chapitre 3 :Résultats). Cette séquence ne correspond à aucun transcrit connu de Nr4a1. Cela pourrait donc être un nouveau transcrit du gène NR4A1. En fait, cette hypothèse n'est pas plausible car l'ensemble des exons 6,7 et 8 plus l'intron 6-7 ne forment pas une phase ouverte de lecture. Ceci est confirmé lors de l'interrogation des bases de données protéiques où l'on obtient seulement un alignement pour les bases 4 à 246 en phase 1, c'est-à-dire formant une protéine de fusion avec la sous-unité α de l'ARN polymérase. Ces 64 acides aminés correspondent au domaine HOL1 de la protéine Nr4a1. Ce motif protéique est un domaine de liaison au ligand présent dans les récepteurs aux hormones. D'autre part, le double-hybride dirigé Nr4a1 (forme entière sauvage)/Fbxo47 ne nous a pas confirmé l'interaction entre ces deux protéines. Nous ne pouvons donc pas considérer Nr4a1 comme cible pour Fbxo47. C'est donc plus probablement le motif protéique HOL1 que la protéine Fbxo47 est capable de reconnaître. Nous n'avons pas réalisé de double hybride dirigé avec Fbxo47 et le domaine HOL1 d'une autre protéine que Nr4a1, cette expérience pourrait confirmer ou infirmer le domaine HOL1 comme interacteur de Fbxo47.

Reste en suspens la présence d'une telle séquence nucléotidique dans la banque d'ADNc, qui a été construite à partir d'ARNs polyA+ qui sont, en principe, dépourvus de toute séquence intronique.

II. GFM1, protéine cible de Fbxo47?

Le facteur d'élongation mitochondrial G1, GFM1 (AAK 53402.1), constitue la seule protéine identifiée lors du criblage double hybride avec Fbxo47 comme « appât ». Le fait que ce polypeptide ait été retrouvé deux fois (C89 et C228) renforce sa pertinence puisque la banque contient des séquences redondantes.

Il existe trois nomenclatures différentes pour nommer le facteur d'élongation mitochondrial G1 : GFM1, EFG1 et EF-G1_{mt}. Nous retiendrons la première qui correspond à la nomenclature HUGO : GFM1 (*G elongation Factor, Mitochondrial*).

II.1. Présentation du facteur d'élongation mitochondrial G1

II.1.1. Caractéristiques de GFM1

Le gène humain est localisé en 3q25.1-q26.2, comprend 18 exons et s'étend sur au moins 40 kb. Par la recherche d'homologues du gène GFM de rat dans une base de données EST, suivi par PCR à partir d'une banque d'ADNc de tissu testiculaire, l'équipe de Gao et *al.*²⁷³ en 2001, a obtenu deux ADNc codant les formes humaines et murines de GFM. Le profil d'expression du gène humain, obtenu par Northern Blot, révèle deux transcrits principaux (3,8 et 3,4 kb) largement exprimés et plus abondants dans le cœur, le muscle et les testicules, mais révèle également l'existence d'un transcrit spécifique du tissu testiculaire de 2,9 kb²⁷³.

La protéine déduite comporte 751 acides aminés et montre 84 et 89% d'homologie avec les protéines de rat et de souris, respectivement. La signature de facteur d'élongation liant le GTP, qui est conservé à travers tous les facteurs d'élongation, est présent entre les acides aminés 90 et 105. Gfm1 contient également un domaine consensus de liaison au GTP composé de trois motifs (A/G)XXXXGK(S/T) (53-60 aa), DXXG (120-123 aa) et NKXD (174-177 aa).

Indépendamment, un autre facteur d'élongation mitochondrial, nommé mEfg2 (Gfm2), a été identifié et caractérisé par Hammarsund²⁷⁴. Ce gène est localisé en 5q13 et est exprimé dans tous les tissus (3,2 kb). Il présente également une homologie significative avec plusieurs

protéines bactériennes EFG, mais un alignement multiple de séquence de facteurs d'élongation mitochondriaux de différents organismes révèle qu'il existe deux classes, l'une définie par GFM1 et l'autre définie par GFM2²⁷⁴.

II.1.2. Rôle de Gfm1

La protéine Gfm1 est un des composants du système de traduction protéique. Différents facteurs catalysent les trois étapes de la traduction que sont, l'initiation, l'élongation et la terminaison. Les eucaryotes contiennent deux systèmes traductionnels, un dans le cytoplasme et un dans la mitochondrie. Dans la mitochondrie, la phase d'élongation nécessite trois facteurs :

- EF-Tu (TUFM), qui amène les ARNt au site A du ribosome afin de lire l'ARNm par interaction codon-anti-codon²⁷⁵⁻²⁷⁶.
- EF-G (GFM), qui catalyse l'étape de translocation durant laquelle les ARNts sur les sites A et P se déplacent sur les sites P et E du ribosome²⁷⁶⁻²⁷⁷.
- EF-T (TSFM), qui convertit EF-Tu :GDP en EF-Tu :GTP actif²⁷⁵.

EF-Tu et EF-G sont toutes deux des protéines capables de lier le GTP (protéines G).

Les ribosomes localisés dans la mitochondrie sont responsables de la traduction des ARNm transcrits à partir de l'ADN mitochondrial. Celui ne code que pour 13 protéines chez les mammifères, qui sont essentiellement des composants de la chaîne respiratoire. Ainsi, la traduction mitochondriale est essentielle pour la maintenance des fonctions mitochondriales et des mutations dans ce système conduisent à une perturbation du système OXPHOS (*Oxydative PHOSphorylation*) et à une mauvaise conservation de l'ADN mitochondrial²⁷⁸⁻²⁷⁹.

Plusieurs maladies humaines ont été spécifiquement corrélées à la traduction mitochondriale²⁷⁸. La plupart des facteurs régulant ce système sont codés par des gènes nucléaires, traduits dans le cytoplasme et transportés dans la mitochondrie.

En 2004, Coenen et al.²⁸⁰ ont identifié une mutation affectant un résidu conservé du domaine de liaison au GTP (521A-G, asn174ser) dans le gène GFM chez 2 jumeaux présentant à la fois une déficience de la phosphorylation oxydative (réduction de 69% du complexe III) avec une réduction du taux de protéines des complexes respiratoires, et une héptoencéphalopathie progressive.

En plus de leur rôle dans la traduction, plusieurs protéines ribosomales et facteurs d'élongation ont été décrits comme impliqués dans d'autres processus biologiques, incluant la réplication, la transcription, la réparation de l'ADN, la régulation de la traduction, la transformation maligne et la régulation du développement²⁸¹. EF-G a également été rapporté comme possédant des propriétés de protéine chaperonne, avec une fonction potentielle dans le repliement protéique et la protection contre le stress²⁷⁷.

II.2. Conséquences sur la fonction de Fbxo47

Les caractéristiques de Gfm1 concordent avec son rôle d'interacteur protéique de Fbxo47.

Tout d'abord, Gfm1 est localisée dans la mitochondrie, ce qui est en accord avec une suspicion forte de localisation dans la mitochondrie pour Fbxo47. Il semble cohérent que deux protéines interagissant se trouvent dans le même compartiment cellulaire, et a fortiori, ici dans la matrice mitochondriale.

Ensuite, il existe un transcrit spécifique dans le tissu testiculaire du gène GFM1. Ce n'est donc pas a priori cette forme d'ARNm que nous avons identifiée puisque que nous ayons utilisé une banque de rein dans la technique du double hybride. Néanmoins, l'existence des trois transcrits de GFM1 serait liée à l'utilisation de signaux de polyadénylation alternatifs²⁷³. Il en a été décrit trois dans la structure génomique, dont deux typiques (AATAAA) situés au niveau des nucléotides 2604-2609 et 3448-3453, et un atypique (ATTAAA) situé aux nucléotides 3013-3018. De plus, Gao *et al.* ont identifié 3 groupes d'ESTs constitués de 39, 19 et 9 représentants différents pour ce gène, respectivement. Cette donnée renforce l'hypothèse de l'emploi des trois signaux, car ils sont concordants avec les taux d'expression des trois transcrits correspondants. Dans le testicule, les trois formes co-existent. Ainsi, l'identification d'un tel partenaire protéique dans le rein pour Fbxo47 subsiste dans le tissu testiculaire. Nous ajouterons également que la séquence nucléotidique séquencée de GFM1 à partir du clone positif s'aligne avec les bases 1894 à 2307, c'est-à-dire en dehors des signaux de polyadénylation alternatifs. Il reste intéressant de savoir si, *in vivo*, Fbxo47 est capable de lier les trois formes de transcrit.

Enfin, le troisième point qui renforce Gfm1 comme un bon interacteur candidat de Fbxo47, est sa capacité à être modifié de façon post-traductionnelle. En effet, comme nous l'avons précédemment évoqué, les protéines F-box reconnaissent dans la très grande majorité des protéines ayant subi des modifications post-traductionnelles de type phosphorylation.

Un certain nombre d'observations montrent que les facteurs d'élongation sont des cibles primaires de la régulation pendant la phase d'élongation. Par exemple, la phosphorylation de résidus thréonine dans la partie N-terminale d'eEF-2, catalysée par la kinase spécifique Ca^{2+} /calmoduline dépendante, réduit fortement son activité ²⁸². Dans le cas de Gfm1, plusieurs sites de phosphorylation ont été prédits pour les protéines kinases cAMP et cGMP dépendantes (T287), pour la protéine kinase C (S57, T63, T70, T389, T479, T518) et pour la kinase caséine II (T61) ²⁷³.

Remarquons que dans le criblage double-hybride, nous avons identifié GFM1 (à 100 % d'identité avec le clone) mais non GFM2 (45 % d'identité avec le clone), contenant également le domaine EFG_C, qui est le domaine codé par le clone positif identifié par la technique de double-hybride. Le rôle de ces deux isoformes dans la mitochondrie reste inconnu à ce jour. Néanmoins, des mutations chez l'homologue de levure de GFM1 résultent en une altération de la synthèse protéique mitochondriale tandis que des mutations dans l'homologue de GFM2 ne présentent pas de phénotype particulier ²⁸³.

Ainsi, Gfm1 pourrait représenter le facteur clé dans l'étape de translocation de la synthèse protéique mitochondriale.

La protéine F-box Fbxo47 pourrait donc appartenir au système de régulation de la synthèse protéique mitochondriale par l'intermédiaire de Gfm1. Fbxo47 pourrait agir en promouvant la dégradation de Gfm1 (bien qu'il n'ait jamais été montré ce type de dégradation dans la mitochondrie), soit agir au niveau de sa transcription. Cette hypothèse nous semble la plus crédible pour trois raisons majeures :

- certaines F-box ont des activités de régulateurs transcriptionnels (cf. introduction).
- Fbxo47 contient un domaine leucine zipper présent dans de nombreuses protéines liant l'ADN.
- Bien que ce clone (C142) ne soit pas en phase avec la protéine de fusion lors de la validation en double hybride, nous avons isolé une interaction avec de l'ADN mitochondrial (Tableau 9 Chapitre 3 : Résultats).

Le chapitre suivant rassemble les conclusions et les hypothèses que nous posons par rapport à l'ensemble des résultats obtenus à partir de l'identification et la caractérisation du gène et de la protéine Fbxo47.

CHAPITRE 5

CONCLUSION ET

PERSPECTIVES

Ce travail a permis de cloner et de caractériser un nouveau gène humain localisé en 17q12, FBXO47. Ce gène est situé à proximité du marqueur microsatellite D17S250 délété au niveau somatique dans différents types de cancers. L'étude de perte d'hétérozygotie sur des patients atteints de cancer papillaire rénal, nous conduit à considérer FBXO47 comme un gène suppresseur de tumeur potentiel. Cette hypothèse a été renforcée par l'étude de son expression comparative entre des tissus tumoraux et les échantillons sains correspondants. Nous n'avons pas pu expérimentalement identifier de double inactivation somatique du gène conformément à la théorie de Knudson.

D'autres travaux seront nécessaires pour valider FBXO47 comme gène suppresseur de tumeur.

Nous avons donc décidé de caractériser la fonction du gène et de la protéine codée dans le contexte cellulaire normal, afin d'identifier le mécanisme cellulaire dans lequel FBXO47 joue un rôle.

Dans l'ensemble des résultats obtenus, au cours de cette recherche, quatre grandes caractéristiques se dégagent et peuvent fournir des indications sur le rôle du produit de FBXO47.

Tout d'abord, l'étude d'expression du transcrite de FBXO47 sur des tissus sains a révélé une expression forte uniquement dans le tissu testiculaire. Ce tissu rassemblant de nombreux types cellulaires, il est nécessaire d'établir spécifiquement le profil d'expression de FBXO47. Ce travail va être très prochainement réalisé en collaboration avec l'équipe « Acteurs moléculaires de la Spermatogenèse » dirigée par Charles Pineau (INSERM U625, Rennes). Cette unité dispose, en effet, de 11 types différents de cellules testiculaires isolées, ainsi que de membranes de Northern et Western blot. Ceci va donc nous permettre d'établir la topographie d'expression testiculaire du gène et de la protéine Fbxo47.

Nous avons également montré que Fbxo47 appartient à la famille des protéines F-box. La plupart des membres de cette famille entre dans la composition d'une ubiquitine ligase E3 sous la forme d'un système SCF, c'est-à-dire qu'elles interagissent à la fois avec la protéine Skp1, la protéine Roc1 et une culline. Nous avons démontré l'interaction de Fbxo47 avec Skp1. Il est nécessaire, à présent, d'identifier si Fbxo47 se lie également aux autres sous-unités d'un complexe SCF pour évaluer son appartenance ou non à la voie de dégradation ubiquitine-protéasome dépendante.

La séquence primaire de Fbxo47 contient dans l'extrémité N-terminale une pré-séquence d'adressage mitochondrial caractéristique. L'étude de visualisation de la protéine dans un système cellulaire montre un résultat cohérent avec cette localisation. Une expérience de transfection stable ou l'identification de la protéine par un anti-corps spécifique à partir de mitochondries isolées permettraient de confirmer définitivement l'appartenance de Fbxo47 aux composants de la matrice mitochondriale.

Enfin, la recherche de protéines partenaires de Fbxo47 a conduit à l'identification de la protéine Gfm1. Il s'agit du facteur d'élongation principal de la traduction mitochondriale²⁸³. Il reste à définir les propriétés structurales et biochimiques de cette interaction, en définissant par exemple quelle région et quels résidus de Fbxo47 sont nécessaires pour la reconnaissance de Gfm1. D'autre part, les deux gènes, FBXO47 et GFM1 possèdent des homologues à la fois chez la souris et le rat, il serait intéressant de savoir si l'interaction est conservée chez les mammifères. La seconde question posée par l'identification de Gfm1 est la conséquence de son interaction avec Fbxo47. En effet, puisque l'on ne sait pas si Fbxo47 appartient à un complexe ubiquitine ligase E3, il est difficile d'avancer que Fbxo47 cible Gfm1 pour sa dégradation. De plus, ces deux protéines appartiennent à la matrice mitochondriale dans laquelle il n'a jamais été décrit ni de protéasome, ni de dégradation protéique ubiquitine-protéasome dépendante.

La technique du double-hybride a été employée afin d'identifier des protéines interagissant avec Fbxo47. Nous avons utilisé une banque d'ADNc de rein, il serait nécessaire de cribler des banques d'autres tissus (notamment de testicule), mais également d'employer d'autres techniques (co-immunoprécipitation, par exemple) permettant d'identifier de nouveaux partenaires protéiques de Fbxo47.

L'ensemble des résultats obtenus offre plusieurs axes de recherche.

En effet, la localisation mitochondriale de la protéine codée par le gène FBXO47, son appartenance à la famille des F-box, et son expression testiculaire combinés aux données sur les phénomènes d'ubiquitination des mitochondries au cours de la spermatogenèse nous permettent de poser la question de l'implication de Fbxo47 dans ces mécanismes.

Par ailleurs, c'est une étape clé de la régulation de l'expression des gènes mitochondriaux, qui est la cible de nombreuses maladies neuro-dégénératives humaines. Les mieux connues affectent des composants ARN de la machinerie de traduction mitochondriale, car ils sont codés par l'ADNmt et donc leurs mutations sont plus faciles à identifier. Au contraire, les protéines nécessaires à la traduction mitochondriale sont principalement

fournies par le génome nucléaire, et les conséquences de leur absence sont méconnues chez l'homme. La régulation de la traduction mitochondriale humaine reste aussi très peu connue à ce jour.

L'existence chez le vers *C.elegans*, d'une protéine F-box nommée FOG-2 capable de former un complexe de répression de la traduction en se liant à la protéine de liaison à l'ARN GLD-1¹⁵⁹, nous apporte une fonction alternative à la voie ubiquitine pour les protéines F-Box. Le complexe FOG-2/GLD-1 se lie à la partie 3' non traduite de l'ARNm tra-2, et inhibe sa traduction provoquant la spermatogenèse chez le vers hermaphrodite¹⁵⁹⁻¹⁶⁰.

La découverte du motif F-box date de seulement 10 ans (1996)⁹⁹. Aujourd'hui, cette famille de protéines, présente chez tous les eucaryotes, rassemble un très grand nombre de membres dans les bases de données. Ce chiffre, en perpétuelle augmentation, correspond en partie à des protéines prédites et non biologiquement identifiées. Un travail important doit être réalisé afin de déterminer la fonction de la plupart de ces protéines.

Le rôle principal des protéines F-box connues, est lié à leur appartenance à des complexes ubiquitine ligases E3 et à la voie de dégradation ubiquitine-protéasome dépendante. Il semble à présent que les exceptions à cette voie soient de plus en plus nombreuses. Les études se développant autour du thème « F-box », révèlent son implication dans des systèmes cellulaires très variés, dont la compréhension et l'analyse restent complexes. Il semble, à présent certain, que les nouveaux rôles attribués aux protéines F-box ne sont pas des exceptions à la voie SCF, et que le nombre de processus dépendants de protéines F-box va continuer son expansion.

Ainsi, l'étude approfondie de chaque nouvelle protéine F-box identifiée est nécessaire afin d'approfondir la connaissance de la régulation F-box-dépendante dans l'ensemble des mécanismes cellulaires.

REFERENCES BIBLIOGRAPHIQUES

1. Jarvinen HJ, Aarnio M. Surveillance on mutation carriers of DNA mismatch repair genes. *Ann Chir Gynaecol* 2000;89(3):207-10.
2. Rodriguez-Bigas MA, Vasen HF, Lynch HT, Watson P, Myrhoj T, Jarvinen HJ, et al. Characteristics of small bowel carcinoma in hereditary nonpolyposis colorectal carcinoma. International Collaborative Group on HNPCC. *Cancer* 1998;83(2):240-4.
3. Delahunt B, Eble JN. Papillary renal cell carcinoma: a clinicopathologic and immunohistochemical study of 105 tumors. *Mod Pathol* 1997;10(6):537-44.
4. Kozak M. Structural features in eukaryotic mRNAs that modulate the initiation of translation. *J Biol Chem* 1991;266(30):19867-70.
5. Kovacs G. Papillary renal cell carcinoma. A morphologic and cytogenetic study of 11 cases. *Am J Pathol* 1989;134(1):27-34.
6. Diakoumis E, Sourvinos G, Kiaris H, Delakas D, Cranidis A, Spandidos DA. Genetic instability in renal cell carcinoma. *Eur Urol* 1998;33(2):227-32.
7. Salama ME, Worsham MJ, DePeralta-Venturina M. Malignant papillary renal tumors with extensive clear cell change: a molecular analysis by microsatellite analysis and fluorescence in situ hybridization. *Arch Pathol Lab Med* 2003;127(9):1176-81.
8. Sanders ME, Mick R, Tomaszewski JE, Barr FG. Unique patterns of allelic imbalance distinguish type 1 from type 2 sporadic papillary renal cell carcinoma. *Am J Pathol* 2002;161(3):997-1005.
9. Slamon DJ, Clark GM, Wong SG, Levin WJ, Ullrich A, McGuire WL. Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. *Science* 1987;235(4785):177-82.
10. Varis A, Zaika A, Puolakkainen P, Nagy B, Madrigal I, Kokkola A, et al. Coamplified and overexpressed genes at ERBB2 locus in gastric cancer. *Int J Cancer* 2004;109(4):548-53.

11. Vernimmen D, Gueders M, Pisvin S, Delvenne P, Winkler R. Different mechanisms are implicated in ERBB2 gene overexpression in breast and in other cancers. *Br J Cancer* 2003;89(5):899-906.
12. Klapper LN, Kirschbaum MH, Sela M, Yarden Y. Biochemical and clinical implications of the ErbB/HER signaling network of growth factor receptors. *Adv Cancer Res* 2000;77:25-79.
13. Jarvinen TA, Liu ET. Topoisomerase IIalpha gene (TOP2A) amplification and deletion in cancer--more common than anticipated. *Cytopathology* 2003;14(6):309-13.
14. Bertucci F, Borie N, Ginestier C, Groulet A, Charafe-Jauffret E, Adelaide J, et al. Identification and validation of an ERBB2 gene expression signature in breast cancers. *Oncogene* 2004;23(14):2564-75.
15. Katoh M. MGC9753 gene, located within PPP1R1B-STARD3-ERBB2-GRB7 amplicon on human chromosome 17q12, encodes the seven-transmembrane receptor with extracellular six-cystein domain. *Int J Oncol* 2003;22(6):1369-74.
16. Godwin AK, Vanderveer L, Schultz DC, Lynch HT, Altomare DA, Buetow KH, et al. A common region of deletion on chromosome 17q in both sporadic and familial epithelial ovarian tumors distal to BRCA1. *Am J Hum Genet* 1994;55(4):666-77.
17. Garcia A, Bussaglia E, Machin P, Matias-Guiu X, Prat J. Loss of heterozygosity on chromosome 17q in epithelial ovarian tumors: association with carcinomas with serous differentiation. *Int J Gynecol Pathol* 2000;19(2):152-7.
18. Kiaris H, Spanakis N, Ergazaki M, Sourvinos G, Spandidos DA. Loss of heterozygosity at 9p and 17q in human laryngeal tumors. *Cancer Lett* 1995;97(1):129-34.
19. Li Y, Chen Z. Molecular cloning and characterization of LCRG1 a novel gene localized to the tumor suppressor locus D17S800-D17S930. *Cancer Lett* 2004;209(1):75-85.

20. Rizos E, Sourvinos G, Spandidos DA. Loss of heterozygosity at 8p, 9p and 17q in laryngeal cytological specimens. *Oral Oncol* 1998;34(6):519-23.
21. Sasiadek M, Stembalska-Kozłowska A, Smigiel R, Krecicki T, Blin N, Mirghomizadeh F. Microsatellite and chromosome instability in squamous cell laryngeal carcinoma. *Int J Oncol* 2001;19(2):401-5.
22. Gao X, Zacharek A, Grignon DJ, Sakr W, Powell IJ, Porter AT, et al. Localization of potential tumor suppressor loci to a < 2 Mb region on chromosome 17q in human prostate cancer. *Oncogene* 1995;11(7):1241-7.
23. Zhao J, Moch H, Scheidweiler AF, Baer A, Schaffer AA, Speel EJ, et al. Genomic imbalances in the progression of endocrine pancreatic tumors. *Genes Chromosomes Cancer* 2001;32(4):364-72.
24. Rahman N, Arbour L, Houlston R, Bonaiti-Pellie C, Abidi F, Tranchemontagne J, et al. Penetrance of mutations in the familial Wilms tumor gene FWT1. *J Natl Cancer Inst* 2000;92(8):650-2.
25. Rahman N, Abidi F, Ford D, Arbour L, Rapley E, Tonin P, et al. Confirmation of FWT1 as a Wilms' tumour susceptibility gene and phenotypic characteristics of Wilms' tumour attributable to FWT1. *Hum Genet* 1998;103(5):547-56.
26. Rahman N, Arbour L, Tonin P, Baruchel S, Pritchard-Jones K, Narod SA, et al. The familial Wilms' tumour susceptibility gene, FWT1, may not be a tumour suppressor gene. *Oncogene* 1997;14(25):3099-102.
27. Paulding CA, Ruvolo M, Haber DA. The Tre2 (USP6) oncogene is a hominoid-specific gene. *Proc Natl Acad Sci U S A* 2003;100(5):2507-11.
28. Strehl S, Borkhardt A, Slany R, Fuchs UE, König M, Haas OA. The human LASP1 gene is fused to MLL in an acute myeloid leukemia with t(11;17)(q23;q21). *Oncogene* 2003;22(1):157-60.

29. Panagopoulos I, Mencinger M, Dietrich CU, Bjerkehagen B, Saeter G, Mertens F, et al. Fusion of the RBP56 and CHN genes in extraskeletal myxoid chondrosarcomas with translocation t(9;17)(q22;q11). *Oncogene* 1999;18(52):7594-8.
30. Shen MH, Harper PS, Upadhyaya M. Molecular genetics of neurofibromatosis type 1 (NF1). *J Med Genet* 1996;33(1):2-17.
31. Puzianowska-Kuznicka M, Krystyniak A, Madej A, Cheng SY, Nauman J. Functionally impaired TR mutants are present in thyroid papillary cancer. *J Clin Endocrinol Metab* 2002;87(3):1120-8.
32. Pandolfi PP. In vivo analysis of the molecular genetics of acute promyelocytic leukemia. *Oncogene* 2001;20(40):5726-35.
33. Yoshida R, Imai T, Hieshima K, Kusuda J, Baba M, Kitaura M, et al. Molecular cloning of a novel human CC chemokine EBI1-ligand chemokine that is a specific functional ligand for EBI1, CCR7. *J Biol Chem* 1997;272(21):13803-9.
34. Neuhausen SL, Marshall CJ. Loss of heterozygosity in familial tumors from three BRCA1-linked kindreds. *Cancer Res* 1994;54(23):6069-72.
35. Mishra S, Murphy LC, Nyomba BL, Murphy LJ. Prohibitin: a potential target for new therapeutics. *Trends Mol Med* 2005;11(4):192-7.
36. Prasad R, Leshkowitz D, Gu Y, Alder H, Nakamura T, Saito H, et al. Leucine-zipper dimerization motif encoded by the AF17 gene fused to ALL-1 (MLL) in acute leukemia. *Proc Natl Acad Sci U S A* 1994;91(17):8107-11.
37. Higashino F, Yoshida K, Fujinaga Y, Kamio K, Fujinaga K. Isolation of a cDNA encoding the adenovirus E1A enhancer binding protein: a new human member of the ets oncogene family. *Nucleic Acids Res* 1993;21(3):547-53.

38. Jacobson KK, Morrison LE, Henderson BT, Blondin BA, Wilber KA, Legator MS, et al. Gene copy mapping of the ERBB2/TOP2A region in breast cancer. *Genes Chromosomes Cancer* 2004;40(1):19-31.
39. Mehta RR, McDermott JH, Hieken TJ, Marler KC, Patel MK, Wild LD, et al. Plasma c-erbB-2 levels in breast cancer patients: prognostic significance in predicting response to chemotherapy. *J Clin Oncol* 1998;16(7):2409-16.
40. Pastorek J, Pastorekova S, Callebaut I, Mornon JP, Zelnik V, Opavsky R, et al. Cloning and characterization of MN, a human tumor-associated protein with a domain homologous to carbonic anhydrase and a putative helix-loop-helix DNA binding segment. *Oncogene* 1994;9(10):2877-88.
41. Backer JM, Mendola CE, Kovesdi I, Fairhurst JL, O'Hara B, Eddy RL, Jr., et al. Chromosomal localization and nucleoside diphosphate kinase activity of human metastasis-suppressor genes NM23-1 and NM23-2. *Oncogene* 1993;8(2):497-502.
42. Bonaventure J, Cohen-Solal L, Lasselin C, Maroteaux P. A dominant mutation in the COL1A1 gene that substitutes glycine for valine causes recurrent lethal osteogenesis imperfecta. *Hum Genet* 1992;89(6):640-6.
43. Uchiumi T, Hinoshita E, Haga S, Nakamura T, Tanaka T, Toh S, et al. Isolation of a novel human canalicular multispecific organic anion transporter, cMOAT2/MRP3, and its expression in cisplatin-resistant cancer cells with decreased ATP-dependent drug transport. *Biochem Biophys Res Commun* 1998;252(1):103-10.
44. Miki T, Kawamata N, Hirosawa S, Aoki N. Gene involved in the 3q27 translocation associated with B-cell lymphoma, BCL5, encodes a Kruppel-like zinc-finger protein. *Blood* 1994;83(1):26-32.

45. Gordon MS, Kato RM, Lansigan F, Thompson AA, Wall R, Rawlings DJ. Aberrant B cell receptor signaling from B29 (Igbeta, CD79b) gene mutations of chronic lymphocytic leukemia B cells. *Proc Natl Acad Sci U S A* 2000;97(10):5504-9.
46. Cantley LC, Auger KR, Carpenter C, Duckworth B, Graziani A, Kapeller R, et al. Oncogenes and signal transduction. *Cell* 1991;64(2):281-302.
47. Aaronson SA. Growth factors and cancer. *Science* 1991;254(5035):1146-53.
48. Knudson AG, Jr. Mutation and cancer: statistical study of retinoblastoma. *Proc Natl Acad Sci U S A* 1971;68(4):820-3.
49. Ishikawa H, Kawano MM, Okada K, Tanaka H, Tanabe O, Sakai A, et al. Expressions of DNA topoisomerase I and II gene and the genes possibly related to drug resistance in human myeloma cells. *Br J Haematol* 1993;83(1):68-74.
50. Hershko A, Ciechanover A. The ubiquitin system. *Annu Rev Biochem* 1998;67:425-79.
51. Hershko A, Ciechanover A. The ubiquitin system. *Annu Rev Biochem*. 1998;67:425-79.
52. Koegl M, Hoppe T, Schlenker S, Ulrich HD, Mayer TU, Jentsch S. A novel ubiquitination factor, E4, is involved in multiubiquitin chain assembly. *Cell* 1999;96(5):635-44.
53. Carthew RW, Xu C. Endocytosis: why not wait to deubiquitinate? *Curr Biol* 2000;10(14):R532-4.
54. Wilkinson KD. Ubiquitination and deubiquitination: targeting of proteins for degradation by the proteasome. *Semin Cell Dev Biol* 2000;11(3):141-8.
55. Verma R, Deshaies RJ. A proteasome howdunit: the case of the missing signal. *Cell* 2000;101(4):341-4.
56. Rodighiero C, Tsai B, Rapoport TA, Lencer WI. Role of ubiquitination in retrotranslocation of cholera toxin and escape of cytosolic degradation. *EMBO Rep* 2002;3(12):1222-7.

57. Spence J, Sadis S, Haas AL, Finley D. A ubiquitin mutant with specific defects in DNA repair and multiubiquitination. *Mol Cell Biol* 1995;15(3):1265-73.
58. Arnason T, Ellison MJ. Stress resistance in *Saccharomyces cerevisiae* is strongly correlated with assembly of a novel type of multiubiquitin chain. *Mol Cell Biol* 1994;14(12):7876-83.
59. Fisk HA, Yaffe MP. A role for ubiquitination in mitochondrial inheritance in *Saccharomyces cerevisiae*. *J Cell Biol* 1999;145(6):1199-208.
60. Garrett WS, Chen LM, Kroschewski R, Ebersold M, Turley S, Trombetta S, et al. Developmental control of endocytosis in dendritic cells by Cdc42. *Cell* 2000;102(3):325-34.
61. Spence J, Gali RR, Dittmar G, Sherman F, Karin M, Finley D. Cell cycle-regulated modification of the ribosome by a variant multiubiquitin chain. *Cell* 2000;102(1):67-76.
62. Pickart CM. Ubiquitin in chains. *Trends Biochem Sci* 2000;25(11):544-8.
63. Aravind L, Koonin EV. The U box is a modified RING finger - a common domain in ubiquitination. *Curr Biol* 2000;10(4):R132-4.
64. Borden KL. RING domains: master builders of molecular scaffolds? *J Mol Biol* 2000;295(5):1103-12.
65. Huibregtse JM, Scheffner M, Beaudenon S, Howley PM. A family of proteins structurally and functionally related to the E6-AP ubiquitin-protein ligase. *Proc Natl Acad Sci U S A* 1995;92(7):2563-7.
66. Laney JD, Hochstrasser M. Substrate targeting in the ubiquitin system. *Cell* 1999;97(4):427-30.
67. Staub O, Gautschi I, Ishikawa T, Breitschopf K, Ciechanover A, Schild L, et al. Regulation of stability and function of the epithelial Na⁺ channel (ENaC) by ubiquitination. *Embo J* 1997;16(21):6325-36.

68. Deshaies RJ. SCF and Cullin/Ring H2-based ubiquitin ligases. *Annu Rev Cell Dev Biol* 1999;15:435-67.
69. Feldman RM, Correll CC, Kaplan KB, Deshaies RJ. A complex of Cdc4p, Skp1p, and Cdc53p/cullin catalyzes ubiquitination of the phosphorylated CDK inhibitor Sic1p. *Cell* 1997;91(2):221-30.
70. Skowyra D, Craig KL, Tyers M, Elledge SJ, Harper JW. F-box proteins are receptors that recruit phosphorylated substrates to the SCF ubiquitin-ligase complex. *Cell* 1997;91(2):209-19.
71. Skowyra D, Koepp DM, Kamura T, Conrad MN, Conaway RC, Conaway JW, et al. Reconstitution of G1 cyclin ubiquitination with complexes containing SCFGrr1 and Rbx1. *Science* 1999;284(5414):662-5.
72. Seol JH, Feldman RM, Zachariae W, Shevchenko A, Correll CC, Lyapina S, et al. Cdc53/cullin and the essential Hrt1 RING-H2 subunit of SCF define a ubiquitin ligase module that activates the E2 enzyme Cdc34. *Genes Dev* 1999;13(12):1614-26.
73. Gagne JM, Downes BP, Shiu SH, Durski AM, Vierstra RD. The F-box subunit of the SCF E3 complex is encoded by a diverse superfamily of genes in Arabidopsis. *Proc Natl Acad Sci U S A* 2002;99(17):11519-24.
74. Kipreos ET, Pagano M. The F-box protein family. *Genome Biol* 2000;1(5):REVIEWS3002.
75. Dias DC, Dolios G, Wang R, Pan ZQ. CUL7: A DOC domain-containing cullin selectively binds Skp1.Fbx29 to form an SCF-like complex. *Proc Natl Acad Sci U S A* 2002;99(26):16601-6.
76. Yu ZK, Gervais JL, Zhang H. Human CUL-1 associates with the SKP1/SKP2 complex and regulates p21(CIP1/WAF1) and cyclin D proteins. *Proc Natl Acad Sci U S A* 1998;95(19):11324-9.

77. Furukawa M, Ohta T, Xiong Y. Activation of UBC5 ubiquitin-conjugating enzyme by the RING finger of ROC1 and assembly of active ubiquitin ligases by all cullins. *J Biol Chem* 2002;277(18):15758-65.
78. Ohta T, Michel JJ, Schottelius AJ, Xiong Y. ROC1, a homolog of APC11, represents a family of cullin partners with an associated ubiquitin ligase activity. *Mol Cell* 1999;3(4):535-41.
79. Clifford SC, Astuti D, Hooper L, Maxwell PH, Ratcliffe PJ, Maher ER. The pVHL-associated SCF ubiquitin ligase complex: molecular genetic analysis of elongin B and C, Rbx1 and HIF-1alpha in renal cell carcinoma. *Oncogene* 2001;20(36):5067-74.
80. Pause A, Lee S, Worrell RA, Chen DY, Burgess WH, Linehan WM, et al. The von Hippel-Lindau tumor-suppressor gene product forms a stable complex with human CUL-2, a member of the Cdc53 family of proteins. *Proc Natl Acad Sci U S A* 1997;94(6):2156-61.
81. Leung SK, Ohh M. Playing Tag with HIF: The VHL Story. *J Biomed Biotechnol* 2002;2(3):131-135.
82. Kamura T, Conaway JW, Conaway RC. Roles of SCF and VHL ubiquitin ligases in regulation of cell growth. *Prog Mol Subcell Biol* 2002;29:1-15.
83. Pintard L, Willems A, Peter M. Cullin-based ubiquitin ligases: Cul3-BTB complexes join the family. *Embo J* 2004;23(8):1681-7.
84. Furukawa M, He YJ, Borchers C, Xiong Y. Targeting of protein ubiquitination by BTB-Cullin 3-Roc1 ubiquitin ligases. *Nat Cell Biol* 2003;5(11):1001-7.
85. Willems AR, Schwab M, Tyers M. A hitchhiker's guide to the cullin ubiquitin ligases: SCF and its kin. *Biochim Biophys Acta* 2004;1695(1-3):133-70.
86. Groisman R, Polanowska J, Kuraoka I, Sawada J, Saijo M, Drapkin R, et al. The ubiquitin ligase activity in the DDB2 and CSA complexes is differentially regulated by the COP9 signalosome in response to DNA damage. *Cell* 2003;113(3):357-67.

87. Shiyarov P, Nag A, Raychaudhuri P. Cullin 4A associates with the UV-damaged DNA-binding protein DDB. *J Biol Chem* 1999;274(50):35309-12.
88. De Sepulveda P, Ilangumaran S, Rottapel R. Suppressor of cytokine signaling-1 inhibits VAV function through protein degradation. *J Biol Chem* 2000;275(19):14005-8.
89. Kamura T, Burian D, Yan Q, Schmidt SL, Lane WS, Querido E, et al. Muf1, a novel Elongin BC-interacting leucine-rich repeat protein that can assemble with Cul5 and Rbx1 to reconstitute a ubiquitin ligase. *J Biol Chem* 2001;276(32):29748-53.
90. Arai T, Kasper JS, Skaar JR, Ali SH, Takahashi C, DeCaprio JA. Targeted disruption of p185/Cul7 gene results in abnormal vascular morphogenesis. *Proc Natl Acad Sci U S A* 2003;100(17):9855-60.
91. Busino L, Chiesa M, Draetta GF, Donzelli M. Cdc25A phosphatase: combinatorial phosphorylation, ubiquitylation and proteolysis. *Oncogene* 2004;23(11):2050-6.
92. Nakayama KI, Nakayama K. Regulation of the cell cycle by SCF-type ubiquitin ligases. *Semin Cell Dev Biol* 2005;16(3):323-33.
93. Liu J, Stevens J, Rote CA, Yost HJ, Hu Y, Neufeld KL, et al. Siah-1 mediates a novel beta-catenin degradation pathway linking p53 to the adenomatous polyposis coli protein. *Mol Cell* 2001;7(5):927-36.
94. Matsuzawa SI, Reed JC. Siah-1, SIP, and Ebi collaborate in a novel pathway for beta-catenin degradation linked to p53 responses. *Mol Cell* 2001;7(5):915-26.
95. Schwob E, Bohm T, Mendenhall MD, Nasmyth K. The B-type cyclin kinase inhibitor p40SIC1 controls the G1 to S transition in *S. cerevisiae*. *Cell* 1994;79(2):233-44.
96. Verma R, Annan RS, Huddleston MJ, Carr SA, Reynard G, Deshaies RJ. Phosphorylation of Sic1p by G1 Cdk required for its degradation and entry into S phase. *Science* 1997;278(5337):455-60.

97. Koepp DM, Harper JW, Elledge SJ. How the cyclin became a cyclin: regulated proteolysis in the cell cycle. *Cell* 1999;97(4):431-4.
98. Zhang H, Kobayashi R, Galaktionov K, Beach D. p19Skp1 and p45Skp2 are essential elements of the cyclin A-CDK2 S phase kinase. *Cell* 1995;82(6):915-25.
99. Bai C, Sen P, Hofmann K, Ma L, Goebel M, Harper JW, et al. SKP1 connects cell cycle regulators to the ubiquitin proteolysis machinery through a novel motif, the F-box; 1996.
100. Willems AR, Lanker S, Patton EE, Craig KL, Nason TF, Mathias N, et al. Cdc53 targets phosphorylated G1 cyclins for degradation by the ubiquitin proteolytic pathway. *Cell* 1996;86(3):453-63.
101. Kipreos ET, Lander LE, Wing JP, He WW, Hedgecock EM. cul-1 is required for cell cycle exit in *C. elegans* and identifies a novel gene family. *Cell* 1996;85(6):829-39.
102. Wu G, Xu G, Schulman BA, Jeffrey PD, Harper JW, Pavletich NP. Structure of a beta-TrCP1-Skp1-beta-catenin complex: destruction motif binding and lysine specificity of the SCF(beta-TrCP1) ubiquitin ligase. *Mol Cell* 2003;11(6):1445-56.
103. Cenciarelli C, Chiaur DS, Guardavaccaro D, Parks W, Vidal M, Pagano M. Identification of a family of human F-box proteins. *Curr Biol* 1999;9(20):1177-9.
104. Winston JT, Koepp DM, Zhu C, Elledge SJ, Harper JW. A family of mammalian F-box proteins. *Curr Biol* 1999;9(20):1180-2.
105. Jin J, Cardozo T, Lovering RC, Elledge SJ, Pagano M, Harper JW. Systematic analysis and nomenclature of mammalian F-box proteins. *Genes Dev* 2004;18(21):2573-80.
106. Yaron A, Hatzubai A, Davis M, Lavon I, Amit S, Manning AM, et al. Identification of the receptor component of the IkappaBalpha-ubiquitin ligase. *Nature* 1998;396(6711):590-4.
107. Carrano AC, Eytan E, Hershko A, Pagano M. SKP2 is required for ubiquitin-mediated degradation of the CDK inhibitor p27. *Nat Cell Biol* 1999;1(4):193-9.

108. Welcker M, Singer J, Loeb KR, Grim J, Bloecher A, Gurien-West M, et al. Multisite phosphorylation by Cdk2 and GSK3 controls cyclin E degradation. *Mol Cell* 2003;12(2):381-92.
109. Nishizawa M, Okazaki K, Furuno N, Watanabe N, Sagata N. The 'second-codon rule' and autophosphorylation govern the stability and activity of Mos during the meiotic cell cycle in *Xenopus* oocytes. *Embo J* 1992;11(7):2433-46.
110. Yoshida Y, Tokunaga F, Chiba T, Iwai K, Tanaka K, Tai T. Fbs2 is a new member of the E3 ubiquitin ligase family that recognizes sugar chains. *J Biol Chem* 2003;278(44):43877-84. Epub 2003 Aug 25.
111. Ellgaard L, Helenius A. Quality control in the endoplasmic reticulum. *Nat Rev Mol Cell Biol* 2003;4(3):181-91.
112. Avots A, Hoffmeyer A, Flory E, Cimanis A, Rapp UR, Serfling E. GABP factors bind to a distal interleukin 2 (IL-2) enhancer and contribute to c-Raf-mediated increase in IL-2 induction. *Mol Cell Biol* 1997;17(8):4381-9.
113. Imaki H, Nakayama K, Delehouzee S, Handa H, Kitagawa M, Kamura T, et al. Cell cycle-dependent regulation of the Skp2 promoter by GA-binding protein. *Cancer Res* 2003;63(15):4607-13.
114. Zachariae W, Nasmyth K. Whose end is destruction: cell division and the anaphase-promoting complex. *Genes Dev* 1999;13(16):2039-58.
115. Schwab M, Neutzner M, Mocker D, Seufert W. Yeast Hct1 recognizes the mitotic cyclin Clb2 and other substrates of the ubiquitin ligase APC. *Embo J* 2001;20(18):5165-75.
116. Wei W, Ayad NG, Wan Y, Zhang GJ, Kirschner MW, Kaelin WG, Jr. Degradation of the SCF component Skp2 in cell-cycle phase G1 by the anaphase-promoting complex. *Nature* 2004;428(6979):194-8.

117. Bashir T, Dorrello NV, Amador V, Guardavaccaro D, Pagano M. Control of the SCF(Skp2-Cks1) ubiquitin ligase by the APC/C(Cdh1) ubiquitin ligase. *Nature* 2004;428(6979):190-3.
118. Sherr CJ. Cancer cell cycles. *Science* 1996;274(5293):1672-7.
119. Murray AW. Recycling the cell cycle: cyclins revisited. *Cell* 2004;116(2):221-34.
120. Pickart CM, Eddins MJ. Ubiquitin: structures, functions, mechanisms. *Biochim Biophys Acta* 2004;1695(1-3):55-72.
121. Weissman AM. Themes and variations on ubiquitylation. *Nat Rev Mol Cell Biol* 2001;2(3):169-78.
122. Nakayama K, Nagahama H, Minamishima YA, Matsumoto M, Nakamichi I, Kitagawa K, et al. Targeted disruption of Skp2 results in accumulation of cyclin E and p27(Kip1), polyploidy and centrosome overduplication. *Embo J* 2000;19(9):2069-81.
123. Bornstein G, Bloom J, Sitry-Shevah D, Nakayama K, Pagano M, Hershko A. Role of the SCFSkp2 ubiquitin ligase in the degradation of p21Cip1 in S phase. *J Biol Chem* 2003;278(28):25752-7. Epub 2003 May 2.
124. Hsu JY, Reimann JD, Sorensen CS, Lukas J, Jackson PK. E2F-dependent accumulation of hEmi1 regulates S phase entry by inhibiting APC(Cdh1). *Nat Cell Biol* 2002;4(5):358-66.
125. Reimann JD, Gardner BE, Margottin-Goguet F, Jackson PK. Emi1 regulates the anaphase-promoting complex by a different mechanism than Mad2 proteins. *Genes Dev* 2001;15(24):3278-85.
126. Busino L, Donzelli M, Chiesa M, Guardavaccaro D, Ganoth D, Dorrello NV, et al. Degradation of Cdc25A by beta-TrCP during S phase and in response to DNA damage. *Nature* 2003;426(6962):87-91.

127. Jin J, Shirogane T, Xu L, Nalepa G, Qin J, Elledge SJ, et al. SCFbeta-TRCP links Chk1 signaling to degradation of the Cdc25A protein phosphatase. *Genes Dev* 2003;17(24):3062-74.
128. Koepf DM, Schaefer LK, Ye X, Keyomarsi K, Chu C, Harper JW, et al. Phosphorylation-dependent ubiquitination of cyclin E by the SCFFbw7 ubiquitin ligase. *Science* 2001;294(5540):173-7.
129. Moberg KH, Bell DW, Wahrer DC, Haber DA, Hariharan IK. Archipelago regulates Cyclin E levels in *Drosophila* and is mutated in human cancer cell lines. *Nature* 2001;413(6853):311-6.
130. Strohmaier H, Spruck CH, Kaiser P, Won KA, Sangfelt O, Reed SI. Human F-box protein hCdc4 targets cyclin E for proteolysis and is mutated in a breast cancer cell line. *Nature* 2001;413(6853):316-22.
131. Watanabe N, Arai H, Nishihara Y, Taniguchi M, Hunter T, Osada H. M-phase kinases induce phospho-dependent ubiquitination of somatic Wee1 by SCFbeta-TrCP. *Proc Natl Acad Sci U S A* 2004;101(13):4419-24.
132. Ayad NG, Rankin S, Murakami M, Jebanathirajah J, Gygi S, Kirschner MW. Tome-1, a trigger of mitotic entry, is degraded during G1 via the APC. *Cell* 2003;113(1):101-13.
133. Margottin-Goguet F, Hsu JY, Loktev A, Hsieh HM, Reimann JD, Jackson PK. Prophase destruction of Emi1 by the SCF(betaTrCP/Slimb) ubiquitin ligase activates the anaphase promoting complex to allow progression beyond prometaphase. *Dev Cell* 2003;4(6):813-26.
134. Guardavaccaro D, Kudo Y, Boulaire J, Barchi M, Busino L, Donzelli M, et al. Control of meiotic and mitotic progression by the F box protein beta-Trcp1 in vivo. *Dev Cell* 2003;4(6):799-812.
135. Peters JM. The anaphase-promoting complex: proteolysis in mitosis and beyond. *Mol Cell* 2002;9(5):931-43.

136. Cardozo T, Pagano M. The SCF ubiquitin ligase: insights into a molecular machine. *Nat Rev Mol Cell Biol* 2004;5(9):739-51.
137. Griparic L, van der Blik AM. The many shapes of mitochondrial membranes. *Traffic* 2001;2(4):235-44.
138. Westermann B. Merging mitochondria matters: cellular role and molecular machinery of mitochondrial fusion. *EMBO Rep* 2002;3(6):527-31.
139. Hales KG, Fuller MT. Developmentally regulated mitochondrial fusion mediated by a conserved, novel, predicted GTPase. *Cell* 1997;90(1):121-9.
140. Berger KH, Yaffe MP. Mitochondrial DNA inheritance in *Saccharomyces cerevisiae*. *Trends Microbiol* 2000;8(11):508-13.
141. Skulachev VP. Mitochondrial filaments and clusters as intracellular power-transmitting cables. *Trends Biochem Sci* 2001;26(1):23-9.
142. Labrousse AM, Shurland DL, van der Blik AM. Contribution of the GTPase domain to the subcellular localization of dynamin in the nematode *Caenorhabditis elegans*. *Mol Biol Cell* 1998;9(11):3227-39.
143. Shaw JM, Nunnari J. Mitochondrial dynamics and division in budding yeast. *Trends Cell Biol* 2002;12(4):178-84.
144. Hermann GJ, Thatcher JW, Mills JP, Hales KG, Fuller MT, Nunnari J, et al. Mitochondrial fusion in yeast requires the transmembrane GTPase Fzo1p. *J Cell Biol* 1998;143(2):359-73.
145. Sesaki H, Jensen RE. UGO1 encodes an outer membrane protein required for mitochondrial fusion. *J Cell Biol* 2001;152(6):1123-34.
146. Wong ED, Wagner JA, Scott SV, Okreglak V, Holewinski TJ, Cassidy-Stone A, et al. The intramitochondrial dynamin-related GTPase, Mgm1p, is a component of a protein complex that mediates mitochondrial fusion. *J Cell Biol* 2003;160(3):303-11.

147. Otsuga D, Keegan BR, Brisch E, Thatcher JW, Hermann GJ, Bleazard W, et al. The dynamin-related GTPase, Dnm1p, controls mitochondrial morphology in yeast. *J Cell Biol* 1998;143(2):333-49.
148. Cipolat S, Martins de Brito O, Dal Zilio B, Scorrano L. OPA1 requires mitofusin 1 to promote mitochondrial fusion. *Proc Natl Acad Sci U S A* 2004;101(45):15927-32. Epub 2004 Oct 27.
149. Fritz S, Weinbach N, Westermann B. Mdm30 is an F-box protein required for maintenance of fusion-competent mitochondria in yeast. *Mol Biol Cell* 2003;14(6):2303-13. Epub 2003 Feb 6.
150. Neutzner A, Youle RJ. Instability of the mitofusin Fzo1 regulates mitochondrial morphology during the mating response of the yeast *Saccharomyces cerevisiae*. *J Biol Chem* 2005;10:10.
151. Lee JC, Peter ME. Regulation of apoptosis by ubiquitination. *Immunol Rev* 2003;193(1):39-47.
152. Maniatis T. A ubiquitin ligase complex essential for the NF-kappaB, Wnt/Wingless, and Hedgehog signaling pathways. *Genes Dev* 1999;13(5):505-10.
153. Suzuki H, Chiba T, Suzuki T, Fujita T, Ikenoue T, Omata M, et al. Homodimer of two F-box proteins betaTrCP1 or betaTrCP2 binds to IkappaBalpha for signal-dependent ubiquitination. *J Biol Chem* 2000;275(4):2877-84.
154. Hays R, Wickline L, Cagan R. Morgue mediates apoptosis in the *Drosophila melanogaster* retina by promoting degradation of DIAP1. *Nat Cell Biol* 2002;4(6):425-31.
155. Wilson R, Goyal L, Ditzel M, Zachariou A, Baker DA, Agapite J, et al. The DIAP1 RING finger mediates ubiquitination of Dronc and is indispensable for regulating apoptosis. *Nat Cell Biol* 2002;4(6):445-50.

156. Russell ID, Grancell AS, Sorger PK. The unstable F-box protein p58-Ctf13 forms the structural core of the CBF3 kinetochore complex. *J Cell Biol* 1999;145(5):933-50.
157. Kaplan KB, Hyman AA, Sorger PK. Regulating the yeast kinetochore by ubiquitin-dependent degradation and Skp1p-mediated phosphorylation. *Cell* 1997;91(4):491-500.
158. Doheny KF, Sorger PK, Hyman AA, Tugendreich S, Spencer F, Hieter P. Identification of essential components of the *S. cerevisiae* kinetochore. *Cell* 1993;73(4):761-74.
159. Clifford R, Lee MH, Nayak S, Ohmachi M, Giorgini F, Schedl T. FOG-2, a novel F-box containing protein, associates with the GLD-1 RNA binding protein and directs male sex determination in the *C. elegans* hermaphrodite germline. *Development* 2000;127(24):5265-76.
160. Jan E, Motzny CK, Graves LE, Goodwin EB. The STAR protein, GLD-1, is a translational regulator of sexual identity in *Caenorhabditis elegans*. *Embo J* 1999;18(1):258-69.
161. Regan-Reimann JD, Duong QV, Jackson PK. Identification of novel F-box proteins in *Xenopus laevis*. *Curr Biol* 1999;9(20):R762-3.
162. Moshe Y, Boulaire J, Pagano M, Hershko A. Role of Polo-like kinase in the degradation of early mitotic inhibitor 1, a regulator of the anaphase promoting complex/cyclosome. *Proc Natl Acad Sci U S A* 2004;101(21):7937-42.
163. Hansen DV, Loktev AV, Ban KH, Jackson PK. Plk1 regulates activation of the anaphase promoting complex by phosphorylating and triggering SCFbetaTrCP-dependent destruction of the APC Inhibitor Emi1. *Mol Biol Cell* 2004;15(12):5623-34.
164. Reimann JD, Jackson PK. Emi1 is required for cytostatic factor arrest in vertebrate eggs. *Nature* 2002;416(6883):850-4.
165. Trimarchi JM, Lees JA. Sibling rivalry in the E2F family. *Nat Rev Mol Cell Biol* 2002;3(1):11-20.

166. Hateboer G, Kerkhoven RM, Shvarts A, Bernards R, Beijersbergen RL. Degradation of E2F by the ubiquitin-proteasome pathway: regulation by retinoblastoma family proteins and adenovirus transforming proteins. *Genes Dev* 1996;10(23):2960-70.
167. Campanero MR, Flemington EK. Regulation of E2F through ubiquitin-proteasome-dependent degradation: stabilization by the pRB tumor suppressor protein. *Proc Natl Acad Sci U S A* 1997;94(6):2221-6.
168. Tsvetkov LM, Yeh KH, Lee SJ, Sun H, Zhang H. p27(Kip1) ubiquitination and degradation is regulated by the SCF(Skp2) complex through phosphorylated Thr187 in p27. *Curr Biol* 1999;9(12):661-4.
169. Ganoth D, Bornstein G, Ko TK, Larsen B, Tyers M, Pagano M, et al. The cell-cycle regulatory protein Cks1 is required for SCF(Skp2)-mediated ubiquitinylation of p27. *Nat Cell Biol* 2001;3(3):321-4.
170. Pelengaris S, Khan M, Evan G. c-MYC: more than just a matter of life and death. *Nat Rev Cancer* 2002;2(10):764-76.
171. von der Lehr N, Johansson S, Larsson LG. Implication of the ubiquitin/proteasome system in Myc-regulated transcription. *Cell Cycle* 2003;2(5):403-7.
172. Bloom J, Pagano M. Deregulated degradation of the cdk inhibitor p27 and malignant transformation. *Semin Cancer Biol* 2003;13(1):41-7.
173. Gstaiger M, Jordan R, Lim M, Catzavelos C, Mestan J, Slingerland J, et al. Skp2 is oncogenic and overexpressed in human cancers. *Proc Natl Acad Sci U S A* 2001;98(9):5043-8.
174. Yada M, Hatakeyama S, Kamura T, Nishiyama M, Tsunematsu R, Imaki H, et al. Phosphorylation-dependent degradation of c-Myc is mediated by the F-box protein Fbw7. *Embo J* 2004;23(10):2116-25.

175. Fortini ME. Notch and presenilin: a proteolytic mechanism emerges. *Curr Opin Cell Biol* 2001;13(5):627-34.
176. Gupta-Rossi N, Le Bail O, Gonen H, Brou C, Logeat F, Six E, et al. Functional interaction between SEL-10, an F-box protein, and the nuclear form of activated Notch1 receptor. *J Biol Chem* 2001;276(37):34371-8.
177. Oberg C, Li J, Pauley A, Wolf E, Gurney M, Lendahl U. The Notch intracellular domain is ubiquitinated and negatively regulated by the mammalian Sel-10 homolog. *J Biol Chem* 2001;276(38):35847-53.
178. Wu G, Lyapina S, Das I, Li J, Gurney M, Pauley A, et al. SEL-10 is an inhibitor of notch signaling that targets notch for ubiquitin-mediated protein degradation. *Mol Cell Biol* 2001;21(21):7403-15.
179. Lai EC. Protein degradation: four E3s for the notch pathway. *Curr Biol* 2002;12(2):R74-8.
180. Keyomarsi K, Tucker SL, Buchholz TA, Callister M, Ding Y, Hortobagyi GN, et al. Cyclin E and survival in patients with breast cancer. *N Engl J Med* 2002;347(20):1566-75.
181. Rajagopalan H, Jallepalli PV, Rago C, Velculescu VE, Kinzler KW, Vogelstein B, et al. Inactivation of hCDC4 can cause chromosomal instability. *Nature* 2004;428(6978):77-81.
182. Spencer E, Jiang J, Chen ZJ. Signal-induced ubiquitination of IkappaBalpha by the F-box protein Slimb/beta-TrCP. *Genes Dev* 1999;13(3):284-94.
183. Fuchs SY, Spiegelman VS, Kumar KG. The many faces of beta-TrCP E3 ubiquitin ligases: reflections in the magic mirror of cancer. *Oncogene* 2004;23(11):2028-36.
184. Spiegelman VS, Tang W, Chan AM, Igarashi M, Aaronson SA, Sassoon DA, et al. Induction of homologue of Slimb ubiquitin ligase receptor by mitogen signaling. *J Biol Chem* 2002;277(39):36624-30.

185. Bhatia N, Herter JR, Slaga TJ, Fuchs SY, Spiegelman VS. Mouse homologue of HOS (mHOS) is overexpressed in skin tumors and implicated in constitutive activation of NF-kappaB. *Oncogene* 2002;21(10):1501-9.
186. Margottin F, Bour SP, Durand H, Selig L, Benichou S, Richard V, et al. A novel human WD protein, h-beta TrCp, that interacts with HIV-1 Vpu connects CD4 to the ER degradation pathway through an F-box motif. *Mol Cell* 1998;1(4):565-74.
187. Besnard-Guerin C, Belaidouni N, Lassot I, Segeral E, Jobart A, Marchal C, et al. HIV-1 Vpu sequesters beta-transducin repeat-containing protein (betaTrCP) in the cytoplasm and provokes the accumulation of beta-catenin and other SCFbetaTrCP substrates. *J Biol Chem* 2004;279(1):788-95.
188. Akari H, Bour S, Kao S, Adachi A, Strebel K. The human immunodeficiency virus type 1 accessory protein Vpu induces apoptosis by suppressing the nuclear factor kappaB-dependent expression of antiapoptotic factors. *J Exp Med* 2001;194(9):1299-311.
189. Bodine SC, Latres E, Baumhueter S, Lai VK, Nunez L, Clarke BA, et al. Identification of ubiquitin ligases required for skeletal muscle atrophy. *Science* 2001;294(5547):1704-8.
190. Gomes MD, Lecker SH, Jagoe RT, Navon A, Goldberg AL. Atrogin-1, a muscle-specific F-box protein highly expressed during muscle atrophy. *Proc Natl Acad Sci U S A* 2001;98(25):14440-5. Epub 2001 Nov 20.
191. Sidow A, Bulotsky MS, Kerrebrock AW, Birren BW, Altshuler D, Jaenisch R, et al. A novel member of the F-box/WD40 gene family, encoding dactylin, is disrupted in the mouse dactylaplasia mutant. *Nat Genet* 1999;23(1):104-7.
192. Li J, Pauley AM, Myers RL, Shuang R, Brashler JR, Yan R, et al. SEL-10 interacts with presenilin 1, facilitates its ubiquitination, and alters A-beta peptide production. *J Neurochem* 2002;82(6):1540-8.

193. Staropoli JF, McDermott C, Martinat C, Schulman B, Demireva E, Abeliovich A. Parkin is a component of an SCF-like ubiquitin ligase complex and protects postmitotic neurons from kainate excitotoxicity. *Neuron* 2003;37(5):735-49.
194. Pagano M. Cell cycle regulation by the ubiquitin pathway. *Faseb J* 1997;11(13):1067-75.
195. Orlicky S, Tang X, Willems A, Tyers M, Sicheri F. Structural basis for phosphodependent substrate selection and orientation by the SCFCdc4 ubiquitin ligase. *Cell* 2003;112(2):243-56.
196. Zhou P, Bogacki R, McReynolds L, Howley PM. Harnessing the ubiquitination machinery to target the degradation of specific cellular proteins. *Mol Cell* 2000;6(3):751-6.
197. Zhang J, Zheng N, Zhou P. Exploring the functional complexity of cellular proteins by protein knockout. *Proc Natl Acad Sci U S A* 2003;100(24):14127-32. Epub 2003 Oct 30.
198. Cong F, Zhang J, Pao W, Zhou P, Varmus H. A protein knockdown strategy to study the function of beta-catenin in tumorigenesis. *BMC Mol Biol* 2003;4(1):10.
199. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol* 1990;215(3):403-10.
200. Boguski MS, Schuler GD. ESTablishing a human transcript map. *Nat Genet* 1995;10(4):369-71.
201. Frohman MA, Dush MK, Martin GR. Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. *Proc Natl Acad Sci U S A* 1988;85(23):8998-9002.
202. Borson ND, Salo WL, Drewes LR. A lock-docking oligo(dT) primer for 5' and 3' RACE PCR. *PCR Methods Appl* 1992;2(2):144-8.
203. Feinberg AP, Vogelstein B. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal Biochem* 1983;132(1):6-13.

204. Herrler M. Use of SMART-generated cDNA for differential gene expression studies. *J Mol Med* 2000;78(7):B23.
205. Zhumabayeva B, Chenchik A, Siebert PD, Herrler M. Disease profiling arrays: reverse format cDNA arrays complimentary to microarrays. *Adv Biochem Eng Biotechnol* 2004;86:191-213.
206. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 1994;22(22):4673-80.
207. Dove SL, Hochschild A. Conversion of the omega subunit of Escherichia coli RNA polymerase into a transcriptional activator or an activation target. *Genes Dev* 1998;12(5):745-54.
208. Dove SL, Joung JK, Hochschild A. Activation of prokaryotic transcription through arbitrary protein-protein contacts. *Nature* 1997;386(6625):627-30.
209. Joung JK, Ramm EI, Pabo CO. A bacterial two-hybrid selection system for studying protein-DNA and protein-protein interactions. *Proc Natl Acad Sci U S A* 2000;97(13):7382-7.
210. Cavenee WK, Dryja TP, Phillips RA, Benedict WF, Godbout R, Gallie BL, et al. Expression of recessive alleles by chromosomal mechanisms in retinoblastoma. *Nature* 1983;305(5937):779-84.
211. Batzer MA, Deininger PL, Hellmann-Blumberg U, Jurka J, Labuda D, Rubin CM, et al. Standardized nomenclature for Alu repeats. *J Mol Evol* 1996;42(1):3-6.
212. Schmid CW. Alu: structure, origin, evolution, significance and function of one-tenth of human DNA. *Prog Nucleic Acid Res Mol Biol* 1996;53:283-319.
213. Smit AF, Riggs AD. MIRs are classic, tRNA-derived SINEs that amplified before the mammalian radiation. *Nucleic Acids Res* 1995;23(1):98-102.

214. Smit AF, Toth G, Riggs AD, Jurka J. Ancestral, mammalian-wide subfamilies of LINE-1 repetitive sequences. *J Mol Biol* 1995;246(3):401-417.
215. Smit AF. Identification of a new, abundant superfamily of mammalian LTR-transposons. *Nucleic Acids Res* 1993;21(8):1863-72.
216. Bernardi G. The human genome: organization and evolutionary history. *Annu Rev Genet* 1995;29:445-76.
217. Takai D, Jones PA. Comprehensive analysis of CpG islands in human chromosomes 21 and 22. *Proc Natl Acad Sci U S A* 2002;99(6):3740-5.
218. Guruprasad K, Reddy BV, Pandit MW. Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence. *Protein Eng* 1990;4(2):155-61.
219. Schulman BA, Carrano AC, Jeffrey PD, Bowen Z, Kinnucan ER, Finnin MS, et al. Insights into SCF ubiquitin ligases from the structure of the Skp1-Skp2 complex. *Nature* 2000;408(6810):381-6.
220. Gavel Y, Nilsson L, von Heijne G. Mitochondrial targeting sequences. Why 'non-amphiphilic' peptides may still be amphiphilic. *FEBS Lett* 1988;235(1-2):173-7.
221. Simon-Kayser B, Scoul C, Renaudin K, Jezequel P, Bouchot O, Rigaud J, et al. Molecular cloning and characterization of FBXO47, a novel gene containing an F-box domain, located in the 17q12 band deleted in papillary renal cell carcinoma. *Genes Chromosomes Cancer* 2005;43(1):83-94.
222. Bhattacharya S, Garriga J, Calbo J, Yong T, Haines DS, Grana X. SKP2 associates with p130 and accelerates p130 ubiquitylation and degradation in human cells. *Oncogene* 2003;22(16):2443-51.
223. Craig KL, Tyers M. The F-box: a new motif for ubiquitin dependent proteolysis in cell cycle regulation and signal transduction. *Prog Biophys Mol Biol* 1999;72(3):299-328.

224. Marshall CJ. Tumor suppressor genes. *Cell* 1991;64(2):313-26.
225. Weinberg RA. Tumor suppressor genes. *Science* 1991;254(5035):1138-46.
226. Crawley JJ, Furge KA. Identification of frequent cytogenetic aberrations in hepatocellular carcinoma using gene-expression microarray data. *Genome Biol* 2002;3(12):RESEARCH0075.
227. Gomez L, Rubio MP, Martin MT, Vazquez JJ, Idoate M, Pastorfide G, et al. Chromosome 17 allelic loss and NF1-GRD mutations do not play a significant role as molecular mechanisms leading to melanoma tumorigenesis. *J Invest Dermatol* 1996;106(3):432-6.
228. Iacobuzio-Donahue CA, van der Heijden MS, Baumgartner MR, Troup WJ, Romm JM, Doheny K, et al. Large-scale allelotype of pancreaticobiliary carcinoma provides quantitative estimates of genome-wide allelic loss. *Cancer Res* 2004;64(3):871-5.
229. Presti JC, Jr., Reuter VE, Cordon-Cardo C, Mazumdar M, Fair WR, Jhanwar SC. Allelic deletions in renal tumors: histopathological correlations. *Cancer Res* 1993;53(23):5780-3.
230. Renshaw AA, Zhang H, Corless CL, Fletcher JA, Pins MR. Solid variants of papillary (chromophil) renal cell carcinoma: clinicopathologic and genetic features. *Am J Surg Pathol* 1997;21(10):1203-9.
231. Kovacs G. Molecular genetics of human renal cell tumours. *Nephrol Dial Transplant* 1996;11 Suppl 6:62-5.
232. Bugert P, Kovacs G. Molecular differential diagnosis of renal cell carcinomas by microsatellite analysis. *Am J Pathol* 1996;149(6):2081-8.
233. Thrash-Bingham CA, Greenberg RE, Howard S, Bruzel A, Bremer M, Goll A, et al. Comprehensive allelotyping of human renal cell carcinomas using microsatellite DNA probes. *Proc Natl Acad Sci U S A* 1995;92(7):2854-8.

234. Thrash-Bingham CA, Salazar H, Freed JJ, Greenberg RE, Tartof KD. Genomic alterations and instabilities in renal cell carcinomas and their relationship to tumor pathology. *Cancer Res* 1995;55(24):6189-95.
235. St Croix B, Rago C, Velculescu V, Traverso G, Romans KE, Montgomery E, et al. Genes expressed in human tumor endothelium. *Science* 2000;289(5482):1197-202.
236. Shen CY, Yu JC, Lo YL, Kuo CH, Yue CT, Jou YS, et al. Genome-wide search for loss of heterozygosity using laser capture microdissected tissue of breast carcinoma: an implication for mutator phenotype and breast cancer pathogenesis. *Cancer Res* 2000;60(14):3884-92.
237. Esteller M, Silva JM, Dominguez G, Bonilla F, Matias-Guiu X, Lerma E, et al. Promoter hypermethylation and BRCA1 inactivation in sporadic breast and ovarian tumors. *J Natl Cancer Inst* 2000;92(7):564-9.
238. Macleod K. Tumor suppressor genes. *Curr Opin Genet Dev* 2000;10(1):81-93.
239. Fero ML, Randel E, Gurley KE, Roberts JM, Kemp CJ. The murine gene p27Kip1 is haplo-insufficient for tumour suppression. *Nature* 1998;396(6707):177-80.
240. Prowse AH, Webster AR, Richards FM, Richard S, Olschwang S, Resche F, et al. Somatic inactivation of the VHL gene in Von Hippel-Lindau disease tumors. *Am J Hum Genet* 1997;60(4):765-71.
241. Herman JG, Umar A, Polyak K, Graff JR, Ahuja N, Issa JP, et al. Incidence and functional consequences of hMLH1 promoter hypermethylation in colorectal carcinoma. *Proc Natl Acad Sci U S A* 1998;95(12):6870-5.
242. Ben-Ezra J, Johnson DA, Rossi J, Cook N, Wu A. Effect of fixation on the amplification of nucleic acids from paraffin-embedded material by the polymerase chain reaction. *J Histochem Cytochem* 1991;39(3):351-4.

243. Giannella C, Zito FA, Colonna F, Paradiso A, Marzullo F, Alaibac M, et al. Comparison of formalin, ethanol, and Histochoice fixation on the PCR amplification from paraffin-embedded breast cancer tissue. *Eur J Clin Chem Clin Biochem* 1997;35(8):633-5.
244. Kalachikov S, Migliazza A, Cayanis E, Fracchiolla NS, Bonaldo MF, Lawton L, et al. Cloning and gene mapping of the chromosome 13q14 region deleted in chronic lymphocytic leukemia. *Genomics* 1997;42(3):369-77.
245. Byrne BC, Li JJ, Sninsky J, Poiesz BJ. Detection of HIV-1 RNA sequences by in vitro DNA amplification. *Nucleic Acids Res* 1988;16(9):4165.
246. Mocharla H, Mocharla R, Hodes ME. Coupled reverse transcription-polymerase chain reaction (RT-PCR) as a sensitive and rapid method for isozyme genotyping. *Gene* 1990;93(2):271-5.
247. Bellve AR, Millette CF, Bhatnagar YM, O'Brien DA. Dissociation of the mouse testis and characterization of isolated spermatogenic cells. *J Histochem Cytochem* 1977;25(7):480-94.
248. Bellve AR, Cavicchia JC, Millette CF, O'Brien DA, Bhatnagar YM, Dym M. Spermatogenic cells of the prepuberal mouse. Isolation and morphological characterization. *J Cell Biol* 1977;74(1):68-85.
249. Yamanaka A, Yada M, Imaki H, Koga M, Ohshima Y, Nakayama K. Multiple Skp1-related proteins in *Caenorhabditis elegans*: diverse patterns of interaction with Cullins and F-box proteins. *Curr Biol* 2002;12(4):267-75.
250. Creighton TE, Freedman RB. Protein folding. An unfolding story
A model catalyst of protein disulphide bond formation. *Curr Biol* 1995;5(4):353-6.
251. Rehling P, Brandner K, Pfanner N. Mitochondrial import and the twin-pore translocase. *Nat Rev Mol Cell Biol* 2004;5(7):519-30.

252. Wiedemann N, Frazier AE, Pfanner N. The protein import machinery of mitochondria. *J Biol Chem* 2004;279(15):14473-6.
253. Donzeau M, Kaldi K, Adam A, Paschen S, Wanner G, Guiard B, et al. Tim23 links the inner and outer mitochondrial membranes. *Cell* 2000;101(4):401-12.
254. Martin J, Mahlke K, Pfanner N. Role of an energized inner membrane in mitochondrial protein import. Delta psi drives the movement of presequences. *J Biol Chem* 1991;266(27):18051-7.
255. Herrmann JM, Neupert W, Stuart RA. Insertion into the mitochondrial inner membrane of a polytopic protein, the nuclear-encoded Oxa1p. *Embo J* 1997;16(9):2217-26.
256. Glick B, Schatz G. Import of proteins into mitochondria. *Annu Rev Genet* 1991;25:21-44.
257. Orłowski M, Wilk S. Purification and specificity of a membrane-bound metalloendopeptidase from bovine pituitaries. *Biochemistry* 1981;20(17):4942-50.
258. Goldberg AL, Akopian TN, Kisselev AF, Lee DH, Rohrwild M. New insights into the mechanisms and importance of the proteasome in intracellular protein degradation. *Biol Chem* 1997;378(3-4):131-40.
259. Takeshige K, Baba M, Tsuboi S, Noda T, Ohsumi Y. Autophagy in yeast demonstrated with proteinase-deficient mutants and conditions for its induction. *J Cell Biol* 1992;119(2):301-11.
260. Rep M, Grivell LA. The role of protein degradation in mitochondrial function and biogenesis. *Curr Genet* 1996;30(5):367-80.
261. Langer T, Neupert W. Regulated protein degradation in mitochondria. *Experientia* 1996;52(12):1069-76.
262. Luciano P, Geli V. The mitochondrial processing peptidase: function and specificity. *Experientia* 1996;52(12):1077-82.

263. Wickner S, Maurizi MR, Gottesman S. Posttranslational quality control: folding, refolding, and degrading proteins. *Science* 1999;286(5446):1888-93.
264. Gottesman S, Maurizi MR, Wickner S. Regulatory subunits of energy-dependent proteases. *Cell* 1997;91(4):435-8.
265. Lotan R, Rotem A, Gonen H, Finberg JP, Kemeny S, Steller H, et al. Regulation of the proapoptotic ARTS protein by ubiquitin-mediated degradation. *J Biol Chem* 2005;280(27):25802-10.
266. Flood WD, Moyer RW, Tsykin A, Sutherland GR, Koblar SA. Nxf and Fbxo33: novel seizure-responsive genes in mice. *Eur J Neurosci* 2004;20(7):1819-26.
267. Sutovsky P, Moreno RD, Ramalho-Santos J, Dominko T, Simerly C, Schatten G. Ubiquitin tag for sperm mitochondria. *Nature* 1999;402(6760):371-2.
268. Sutovsky P, Moreno RD, Ramalho-Santos J, Dominko T, Simerly C, Schatten G. Ubiquitinated sperm mitochondria, selective proteolysis, and the regulation of mitochondrial inheritance in mammalian embryos
Ubiquitin tag for sperm mitochondria. *Biol Reprod* 2000;63(2):582-90.
269. Thompson WE, Ramalho-Santos J, Sutovsky P. Ubiquitination of prohibitin in mammalian sperm mitochondria: possible roles in the regulation of mitochondrial inheritance and sperm quality control. *Biol Reprod* 2003;69(1):254-60.
270. Nijtmans LG, Artal SM, Grivell LA, Coates PJ. The mitochondrial PHB complex: roles in mitochondrial respiratory complex assembly, ageing and degenerative disease. *Cell Mol Life Sci* 2002;59(1):143-55.
271. Artal-Sanz M, Tsang WY, Willems EM, Grivell LA, Lemire BD, van der Spek H, et al. The mitochondrial prohibitin complex is essential for embryonic viability and germline function in *Caenorhabditis elegans*. *J Biol Chem* 2003;278(34):32091-9.

272. Vidal M, Legrain P. Yeast forward and reverse 'n'-hybrid systems. *Nucleic Acids Res* 1999;27(4):919-29.
273. Gao J, Yu L, Zhang P, Jiang J, Chen J, Peng J, et al. Cloning and characterization of human and mouse mitochondrial elongation factor G, GFM and Gfm, and mapping of GFM to human chromosome 3q25.1-q26.2. *Genomics* 2001;74(1):109-14.
274. Hammarsund M, Wilson W, Corcoran M, Merup M, Einhorn S, Grander D, et al. Identification and characterization of two novel human mitochondrial elongation factor genes, hEFG2 and hEFG1, phylogenetically conserved through evolution. *Hum Genet* 2001;109(5):542-50.
275. Caraglia M, Budillon A, Vitale G, Lupoli G, Tagliaferri P, Abbruzzese A. Modulation of molecular mechanisms involved in protein synthesis machinery as a new tool for the control of cell proliferation. *Eur J Biochem* 2000;267(13):3919-36.
276. Rodnina MV, Stark H, Savelsbergh A, Wieden HJ, Mohr D, Matassova NB, et al. GTPases mechanisms and functions of translation factors on the ribosome. *Biol Chem* 2000;381(5-6):377-87.
277. Caldas T, Laalami S, Richarme G. Chaperone properties of bacterial elongation factor EF-G and initiation factor IF2. *J Biol Chem* 2000;275(2):855-60.
278. Schon EA. Mitochondrial genetics and disease. *Trends Biochem Sci* 2000;25(11):555-60.
279. Vambutas A, Ackerman SH, Tzagoloff A. Mitochondrial translational-initiation and elongation factors in *Saccharomyces cerevisiae*. *Eur J Biochem* 1991;201(3):643-52.
280. Coenen MJ, Antonicka H, Ugalde C, Sasarman F, Rossi R, Heister JG, et al. Mutant mitochondrial elongation factor G1 and combined oxidative phosphorylation deficiency. *N Engl J Med* 2004;351(20):2080-6.

281. Wool IG. Extraribosomal functions of ribosomal proteins. Trends Biochem Sci 1996;21(5):164-5.
282. Ryazanov AG, Shestakova EA, Natapov PG. Phosphorylation of elongation factor 2 by EF-2 kinase affects rate of translation. Nature 1988;334(6178):170-3.
283. Winzeler EA, Shoemaker DD, Astromoff A, Liang H, Anderson K, Andre B, et al. Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. Science 1999;285(5429):901-6.

ANNEXES

Annexe A :
Diagramme *Cancer Profiling Array II* (Clontech)

PRODUCT: Cancer Profiling Array II

CATALOG #: 7847-1 or 631777

LOT #: Specified on product label.

STORAGE CONDITIONS

- Store Array at -20°C sealed in a plastic bag away from light.
- Store Control Probe at -20°C .
- Store BD ExpressHyb Solution at room temperature. Dissolve any precipitate with gentle warming.

SHELF LIFE

1 year from date of receipt under proper storage conditions.

SHIPPING CONDITIONS

Box 1: Room temperature

Box 2: -20°C

DESCRIPTION

The Cancer Profiling Array II includes normalized cDNA from 154 tumor and corresponding normal tissues from individual patients, amplified using BD SMART™ technology. Hybridization controls, positive controls, and cDNA from nine cancer cell lines are also included. See the attached list for complete tissue representation and cDNA location.

PACKAGE CONTENTS

Box 1:

- 1 Cancer Profiling Array II
- 25 ml BD ExpressHyb™ Hybridization Solution
- Orientation Grid (PT3711-3)

Box 2:

- 100 ng Human Ubiquitin cDNA Control Probe (5 ng/ μl)

Other:

- User Manual (PT3578-1)

FOR RESEARCH USE ONLY

QUALITY CONTROL DATA

The array was tested with [$-^{32}\text{P}$]dATP labelled ubiquitin according to the User Manual and the expected pattern was observed.

A probe was made from this lot of ubiquitin. It was hybridized to a sample array and the expected pattern was observed.

Gelsolin was labeled with [$-^{32}\text{P}$]dATP and hybridized to a sample array. The expected hybridization pattern was observed.

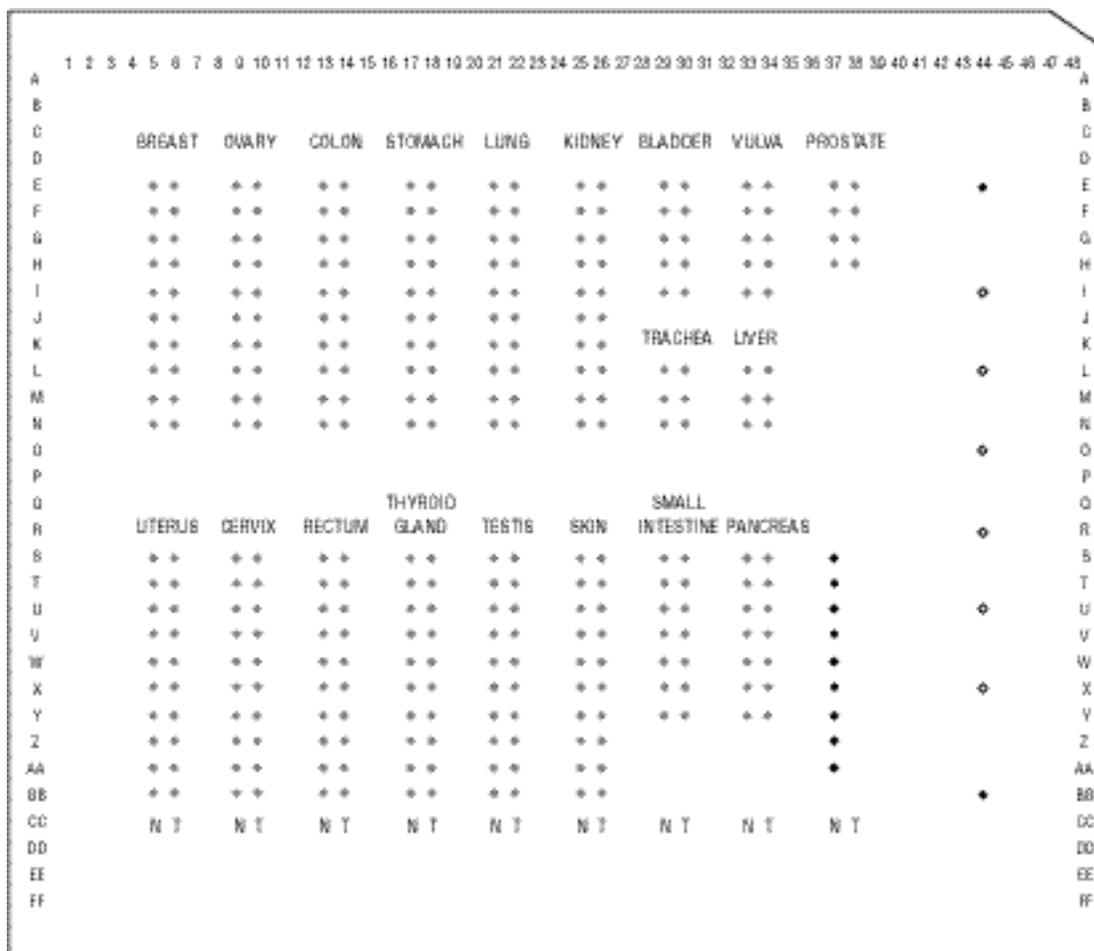
APPROVED BY: _____

Notice to Purchaser

This product is intended to be used for research purposes only. It is not to be used for drug or diagnostic purposes nor is it intended for human use. BD Biosciences Clontech products may not be resold, modified for resale, or used to manufacture commercial products without written approval of BD Biosciences Clontech.

BD SMART™ technology is covered by U.S. Patents #5,962,271 & #5,962,272.

This product was manufactured using the PCR process under license from Roche Molecular Systems and F. Hoffmann-La Roche. No license to use the PCR process is conveyed expressly or by implication to the purchaser by the purchase of this product.



cDNAs and controls immobilized on the Cancer Profiling Array II.

The locations of individual samples on the Cancer Profiling Array II are diagramed above. Tissue sources are listed above each sample group. Light gray spots represent tumor/normal cDNAs; black spots represent cell line cDNAs and positive controls. Open circles represent hybridization controls. N = normal; T = tumor.

Tissue Source/Additional Information	Coordinates Tumor/Normal	Matched Pair Cat. #
BREAST Infiltrating ductal carcinoma Gender: Female Age: 59 Race: Caucasian Tumor Tissue ID #: 51T Stage: I;Unknown ICD Neoplasm Classification: N/A ICD Morphology of Neoplasms: N/A	6E/5E	
BREAST Infiltrating duct carcinoma Gender: Female Age: 60 Race: Caucasian Tumor Tissue ID #: P5225 Stage: III A;T1N2M0 ICD Neoplasm Classification: 174.9 ICD Morphology of Neoplasms: M8500/3	6F/5F	

Tissue Source/Additional Information		Coordinates Tumor/Normal	Matched Pair Cat. #
BREAST	Mucinous adenocarcinoma Gender: Female Age: 53 Race: Caucasian Tumor Tissue ID #: P3925	Stage: I;T1N0M0 ICD Neoplasm Classification: 174.9 ICD Morphology of Neoplasms: M8480/3	6G/5G
BREAST	Infiltrating duct carcinoma Gender: Female Age: 42 Race: Caucasian Tumor Tissue ID #: P1885	Stage: III A;T3N1M0 ICD Neoplasm Classification: 174.9 ICD Morphology of Neoplasms: M8500/3	6H/5H
BREAST	Paget's disease and infiltrating duct carcinoma of breast Gender: Female Age: 56 Race: Caucasian Tumor Tissue ID #: T 13891/00	Stage: II;T2N0M0 ICD Neoplasm Classification: 174.9 ICD Morphology of Neoplasms: M8541/3	6I/5I
BREAST	Lobular carcinoma NOS Gender: Female Age: 52 Race: Caucasian Tumor Tissue ID #: T 16182/00	Stage: III B;T4N1M0 ICD Neoplasm Classification: 174.9 ICD Morphology of Neoplasms: M8520/3	6J/5J
BREAST	Infiltrating duct carcinoma Gender: Female Age: 71 Race: Caucasian Tumor Tissue ID #: T 16187/00	Stage: II B;T2N1M0 ICD Neoplasm Classification: 174.4 ICD Morphology of Neoplasms: M8500/3	6K/5K
BREAST	Infiltrating duct carcinoma Gender: Female Age: 48 Race: Caucasian Tumor Tissue ID #: T 13768/00	Stage: II B;T2N1M0 ICD Neoplasm Classification: 174.9 ICD Morphology of Neoplasms: M8500/3	6L/5L
BREAST	Infiltrating duct carcinoma Gender: Female Age: 46 Race: Caucasian Tumor Tissue ID #: T 15536/00	Stage: I;T1N0M0 ICD Neoplasm Classification: 174.5 ICD Morphology of Neoplasms: M8500/3	6M/5M
BREAST	Infiltrating duct carcinoma Gender: Female Age: 45 Race: Caucasian Tumor Tissue ID #: T 14338/00	Stage: II B;T2N1M0 ICD Neoplasm Classification: 174.9 ICD Morphology of Neoplasms: M8500/3	6N/5N
UTERUS	Adenocarcinoma NOS Gender: Female Age: 71 Race: Caucasian Tumor Tissue ID #: T 14221/00	Stage: II B;T3N0M0 ICD Neoplasm Classification: 182 ICD Morphology of Neoplasms: M8140/3	6AA/5AA

Tissue Source/Additional Information	Coordinates Tumor/Normal	Matched Pair Cat. #
UTERUS Adenocarcinoma NOS Gender: Female Age: 69 Race: Caucasian Tumor Tissue ID #: T 14727/00	Stage: I;T1cN0M0 ICD Neoplasm Classification: 182 ICD Morphology of Neoplasms: M8140/3	6BB/5BB
UTERUS Adenocarcinoma NOS Gender: Female Age: 50 Race: Caucasian Tumor Tissue ID #: P15236	Stage: I;T1N0M0 ICD Neoplasm Classification: 182 ICD Morphology of Neoplasms: M8140/3	6S/5S
UTERUS Adenocarcinoma NOS Gender: Female Age: 75 Race: Caucasian Tumor Tissue ID #: P15698	Stage: I;T1b1N0M0 ICD Neoplasm Classification: 182 ICD Morphology of Neoplasms: M8140/3	6T/5T
UTERUS Malignant mixed Mullerian tumor (MMMT) Gender: Female Age: 77 Race: Unknown Tumor Tissue ID #: 12140	Stage: Unknown;Unknown ICD Neoplasm Classification: N/A ICD Morphology of Neoplasms: N/A	6U/5U HP105U
UTERUS Squamous cell carcinoma, keratinizing type NOS Gender: Female Age: 46 Race: Caucasian Tumor Tissue ID #: P12367	Stage: Unknown;Unknown ICD Neoplasm Classification: 180.1 ICD Morphology of Neoplasms: M8071/3	6V/5V
UTERUS Uterus benign tumor Gender: Female Age: 55 Race: Unknown Tumor Tissue ID #: 14730	Stage: Unknown;Unknown ICD Neoplasm Classification: N/A ICD Morphology of Neoplasms: N/A	6W/5W
UTERUS Squamous cell carcinoma, keratinizing type NOS Gender: Female Age: 46 Race: Caucasian Tumor Tissue ID #: T 8897/00	Stage: II B;T2N0M0 ICD Neoplasm Classification: 180 ICD Morphology of Neoplasms: M8071/3	6X/5X
UTERUS Adenocarcinoma NOS Gender: Female Age: 43 Race: Caucasian Tumor Tissue ID #: T 8184/00	Stage: I A;T1aN0M0 ICD Neoplasm Classification: 182 ICD Morphology of Neoplasms: M8140/3	6Y/5Y
UTERUS Squamous cell carcinoma NOS Gender: Female Age: 40 Race: Caucasian Tumor Tissue ID #: T 7707	Stage: I B;T1b1N0M0 ICD Neoplasm Classification: 180.9 ICD Morphology of Neoplasms: M8070/3	6Z/5Z

Tissue Source/Additional Information	Coordinates Tumor/Normal	Matched Pair Cat. #
OVARY Papillary serous cystadenoma, borderline malignancy Gender: Female Stage: I A;T1a2N0M0 Age: 38 ICD Neoplasm Classification: 183 Race: Caucasian ICD Morphology of Neoplasms: M8460/1 Tumor Tissue ID #: P 3060/01	10E/9E	
OVARY Leiomyoma NOS Gender: Female Stage: Unknown;Unknown Age: 53 ICD Neoplasm Classification: 182 Race: Caucasian ICD Morphology of Neoplasms: M8890/0 Tumor Tissue ID #: P 2844/01	10F/9F	
OVARY Adenocarcinoma NOS Gender: Female Stage: III;T3NxM0 Age: 61 ICD Neoplasm Classification: 183 Race: Caucasian ICD Morphology of Neoplasms: M8140/3 Tumor Tissue ID #: P 4306/01	10G/9G	
OVARY Clear cell adenocarcinoma NOS Gender: Female Stage: III C;T3cN1M0 Age: 38 ICD Neoplasm Classification: 183 Race: Asian ICD Morphology of Neoplasms: M8310/3 Tumor Tissue ID #: P 3317/01	10H/9H	
OVARY Adenocarcinoma NOS Gender: Female Stage: Unknown;T2NxM0 Age: 55 ICD Neoplasm Classification: 179 Race: Caucasian ICD Morphology of Neoplasms: M8140/3 Tumor Tissue ID #: P 2412/01	10I/9I	
OVARY Leiomyoma NOS Gender: Female Stage: II C;T2cN0M0 Age: 54 ICD Neoplasm Classification: 182 Race: Caucasian ICD Morphology of Neoplasms: M8890/0 Tumor Tissue ID #: P 2308/01	10J/9J	
OVARY Mucinous cystadenocarcinoma NOS Gender: Female Stage: I A;Unknown Age: 63 ICD Neoplasm Classification: 183 Race: Caucasian ICD Morphology of Neoplasms: M8470/3 Tumor Tissue ID #: P 16485/00	10K/9K	
OVARY Adenocarcinoma NOS Gender: Female Stage: IV B;Unknown Age: 72 ICD Neoplasm Classification: 180.9 Race: Caucasian ICD Morphology of Neoplasms: M8140/3 Tumor Tissue ID #: P 17162/00	10L/9L	
OVARY Serous surface papillary carcinoma Gender: Female Stage: III C;T3N0M0 Age: 36 ICD Neoplasm Classification: 183 Race: Caucasian ICD Morphology of Neoplasms: M8461/3 Tumor Tissue ID #: P 14803/00	10M/9M	

Tissue Source/Additional Information	Coordinates Tumor/Normal	Matched Pair Cat. #
OVARY Papillary serous cystadenoma, borderline malignancy Gender: Female Stage: III A;T3NxM0 Age: 53 ICD Neoplasm Classification: 183 Race: Caucasian ICD Morphology of Neoplasms: M8460/1 Tumor Tissue ID #: P5998/01	10N/9N	
CERVIX Squamous cell carcinoma NOS Gender: Female Stage: I B;T1b2N0M0 Age: 26 ICD Neoplasm Classification: 180.9 Race: Caucasian ICD Morphology of Neoplasms: M8070/3 Tumor Tissue ID #: P1601/01	10AA/9AA	
CERVIX Squamous cell carcinoma NOS Gender: Female Stage: I B;T1b1N0M0 Age: 44 ICD Neoplasm Classification: 180.9 Race: Caucasian ICD Morphology of Neoplasms: M8070/3 Tumor Tissue ID #: P8151/01	10BB/9BB	
CERVIX Squamous cell carcinoma, keratinizing type NOS Gender: Female Stage: III B;T1b1N1M0 Age: 52 ICD Neoplasm Classification: 180.9 Race: Caucasian ICD Morphology of Neoplasms: M8071/3 Tumor Tissue ID #: P 1192/01	10S/9S	
CERVIX Adenocarcinoma NOS Gender: Female Stage: II A;T2N0M0 Age: 69 ICD Neoplasm Classification: 180.9 Race: Caucasian ICD Morphology of Neoplasms: M8140/3 Tumor Tissue ID #: T 10579/00	10T/9T	
CERVIX Squamous cell carcinoma, keratinizing type NOS Gender: Female Stage: III B;T1b2N1M0 Age: 36 ICD Neoplasm Classification: 180.9 Race: Caucasian ICD Morphology of Neoplasms: M8071/3 Tumor Tissue ID #: T 10572/00	10U/9U	
CERVIX Squamous cell carcinoma, keratinizing type NOS Gender: Female Stage: II A;T2aN0M0 Age: 52 ICD Neoplasm Classification: 180 Race: Caucasian ICD Morphology of Neoplasms: M8071/3 Tumor Tissue ID #: T 12505/00	10V/9V	
CERVIX Adenocarcinoma NOS Gender: Female Stage: II;T2aN0M0 Age: 63 ICD Neoplasm Classification: 180.9 Race: Caucasian ICD Morphology of Neoplasms: M8140/3 Tumor Tissue ID #: P 3161/01	10W/9W	
CERVIX Adenocarcinoma NOS Gender: Female Stage: I-B2;T1b2N0M0 Age: 50 ICD Neoplasm Classification: 180.9 Race: Caucasian ICD Morphology of Neoplasms: M8140/3 Tumor Tissue ID #: P 15665/00	10X/9X	

Tissue Source/Additional Information		Coordinates Tumor/Normal	Matched Pair Cat. #
CERVIX	Squamous cell carcinoma NOS Gender: Female Age: 50 Race: Caucasian Tumor Tissue ID #: P 6263/01	Stage: I A1;T1a1N0M0 ICD Neoplasm Classification: 180.1 ICD Morphology of Neoplasms: M8070/3	10Y/9Y
CERVIX	Adenocarcinoma NOS Gender: Female Age: 56 Race: Caucasian Tumor Tissue ID #: P 6847/01	Stage: I B1;T1b1N0M0 ICD Neoplasm Classification: 180 ICD Morphology of Neoplasms: M8140/3	10Z/9Z
COLON	Adenocarcinoma NOS Gender: Female Age: 67 Race: Caucasian Tumor Tissue ID #: P14251	Stage: II;T4N0M0 ICD Neoplasm Classification: 154.1 ICD Morphology of Neoplasms: M8140/3	14E/13E
COLON	Tubulovillous adenoma Gender: Female Age: 58 Race: Caucasian Tumor Tissue ID #: P12152	Stage: I;T1N0M0 ICD Neoplasm Classification: 153.9 ICD Morphology of Neoplasms: M8263/0	14F/13F
COLON	Adenocarcinoma NOS Gender: Female Age: 43 Race: Caucasian Tumor Tissue ID #: P13978	Stage: IV;T4N1M1 ICD Neoplasm Classification: 153.9 ICD Morphology of Neoplasms: M8140/3	14G/13G
COLON	Adenocarcinoma NOS Gender: Female Age: 69 Race: Caucasian Tumor Tissue ID #: P1520	Stage: III B;T3N1M0 ICD Neoplasm Classification: 153.9 ICD Morphology of Neoplasms: M8140/3	14H/13H
COLON	Adenocarcinoma NOS Gender: Female Age: 35 Race: Caucasian Tumor Tissue ID #: P1288	Stage: III B;T3N1M0 ICD Neoplasm Classification: 153.3 ICD Morphology of Neoplasms: M8140/3	14I/13I HP108C
COLON	Adenocarcinoma NOS Gender: Male Age: 58 Race: Caucasian Tumor Tissue ID #: P3837	Stage: III A;T4N0M0 ICD Neoplasm Classification: 153.9 ICD Morphology of Neoplasms: M8140/3	14J/13J HP105C
COLON	Adenocarcinoma NOS Gender: Male Age: 63 Race: Caucasian Tumor Tissue ID #: T 12558/00	Stage: III B;T4N1M0 ICD Neoplasm Classification: 153.9 ICD Morphology of Neoplasms: M8140/3	14K/13K

Tissue Source/Additional Information	Coordinates Tumor/Normal	Matched Pair Cat. #
COLON Adenocarcinoma NOS Gender: Female Age: 73 Race: Caucasian Tumor Tissue ID #: T 13689/00	Stage: II A;T2N0M0 ICD Neoplasm Classification: 153.3 ICD Morphology of Neoplasms: M8140/3	14L/13L
COLON Adenocarcinoma NOS Gender: Female Age: 65 Race: Caucasian Tumor Tissue ID #: T 14509/00	Stage: III B;T4N0Mx ICD Neoplasm Classification: 153.9 ICD Morphology of Neoplasms: M8140/3	14M/13M
COLON Adenocarcinoma NOS Gender: Female Age: 65 Race: Caucasian Tumor Tissue ID #: T 13992/00	Stage: III A;T3N1M0 ICD Neoplasm Classification: 153.3 ICD Morphology of Neoplasms: M8140/3	14N/13N
RECTUM Mucinous adenocarcinoma Gender: Female Age: 44 Race: Asian Tumor Tissue ID #: P 725/01	Stage: III B;T4N1M0 ICD Neoplasm Classification: 154.1 ICD Morphology of Neoplasms: M8480/3	14AA/13AA
RECTUM Adenocarcinoma NOS Gender: Male Age: 68 Race: Caucasian Tumor Tissue ID #: C1138T	Stage: III;T3N1M0 ICD Neoplasm Classification: 154.1 ICD Morphology of Neoplasms: M8140/3	14BB/13BB
RECTUM Adenocarcinoma NOS Gender: Female Age: 49 Race: Caucasian Tumor Tissue ID #: T 13149/00	Stage: IV B;T4N2M1 ICD Neoplasm Classification: 154 ICD Morphology of Neoplasms: M8140/3	14S/13S
RECTUM Adenocarcinoma NOS Gender: Female Age: 59 Race: Caucasian Tumor Tissue ID #: T 15134/00	Stage: III A;T4N0M0 ICD Neoplasm Classification: 154.2 ICD Morphology of Neoplasms: M8140/3	14T/13T
RECTUM Mucinous adenocarcinoma Gender: Female Age: 53 Race: Caucasian Tumor Tissue ID #: P 2615/01	Stage: IV;T4N2M1 ICD Neoplasm Classification: 154 ICD Morphology of Neoplasms: M8480/3	14U/13U
RECTUM Adenocarcinoma NOS Gender: Female Age: 70 Race: Caucasian Tumor Tissue ID #: P 5778/01	Stage: III B;T3N2M0 ICD Neoplasm Classification: 154.1 ICD Morphology of Neoplasms: M8140/3	14V/13V

Tissue Source/Additional Information	Coordinates Tumor/Normal	Matched Pair Cat. #
RECTUM Adenocarcinoma NOS Gender: Male Age: 42 Race: Caucasian Tumor Tissue ID #: P 3316/01	Stage: III;T3N1M0 ICD Neoplasm Classification: 154 ICD Morphology of Neoplasms: M8140/3	14W/13W
RECTUM Adenocarcinoma NOS Gender: Female Age: 71 Race: Caucasian Tumor Tissue ID #: P 7240/01	Stage: II;T3N0M0 ICD Neoplasm Classification: 154.1 ICD Morphology of Neoplasms: M8140/3	14X/13X
RECTUM Adenocarcinoma NOS Gender: Female Age: 59 Race: Caucasian Tumor Tissue ID #: P 6670/01	Stage: II;T3N0M0 ICD Neoplasm Classification: 154.1 ICD Morphology of Neoplasms: M8140/3	14Y/13Y
RECTUM Adenocarcinoma NOS Gender: Male Age: 43 Race: Caucasian Tumor Tissue ID #: P 8231/01	Stage: II;T3N0M0 ICD Neoplasm Classification: 154.1 ICD Morphology of Neoplasms: M8140/3	14Z/13Z
STOMACH Adenocarcinoma NOS Gender: Male Age: 63 Race: Caucasian Tumor Tissue ID #: T 16683/00	Stage: II B;T3N0M0 ICD Neoplasm Classification: 151.9 ICD Morphology of Neoplasms: M8140/3	18E/17E
STOMACH Gastric cancer Gender: Male Age: 57 Race: Caucasian Tumor Tissue ID #: 522T	Stage: Unknown;T3N1Mx ICD Neoplasm Classification: 151 ICD Morphology of Neoplasms: N/A	18F/17F
STOMACH Gastric cancer Gender: Male Age: 47 Race: Caucasian Tumor Tissue ID #: 526T	Stage: Unknown;T3N0M0 ICD Neoplasm Classification: 151 ICD Morphology of Neoplasms: N/A	18G/17G
STOMACH Adenocarcinoma NOS Gender: Male Age: 65 Race: Caucasian Tumor Tissue ID #: P 3309/01	Stage: I B;T2N0M0 ICD Neoplasm Classification: 151.9 ICD Morphology of Neoplasms: M8140/3	18H/17H
STOMACH Adenocarcinoma NOS Gender: Male Age: 61 Race: Caucasian Tumor Tissue ID #: P 15379/00	Stage: I B;T2N0M0 ICD Neoplasm Classification: 151.9 ICD Morphology of Neoplasms: M8140/3	18I/17I

Tissue Source/Additional Information	Coordinates Tumor/Normal	Matched Pair Cat. #
STOMACH Mucinous adenocarcinoma Gender: Male Age: 50 Race: Caucasian Tumor Tissue ID #: P 17293/00	18J/17J	
Stage: IV;T4N3M0 ICD Neoplasm Classification: 151.9 ICD Morphology of Neoplasms: M8480/3		
STOMACH Adenocarcinoma NOS Gender: Female Age: 72 Race: Caucasian Tumor Tissue ID #: P6505/01	18K/17K	
Stage: IV;T4N2M0 ICD Neoplasm Classification: 151.9 ICD Morphology of Neoplasms: M8140/3		
STOMACH Adenocarcinoma NOS Gender: Female Age: 67 Race: Caucasian Tumor Tissue ID #: P4842/01	18L/17L	
Stage: III A;T3N1M0 ICD Neoplasm Classification: 151.8 ICD Morphology of Neoplasms: M8140/3		
STOMACH Adenocarcinoma NOS Gender: Male Age: 75 Race: Caucasian Tumor Tissue ID #: P5558/01	18M/17M	
Stage: III A;T3N1M0 ICD Neoplasm Classification: 151.9 ICD Morphology of Neoplasms: M8140/3		
STOMACH Adenocarcinoma NOS Gender: Female Age: 68 Race: Caucasian Tumor Tissue ID #: P8460/01	18N/17N	
Stage: III;T4N0M0 ICD Neoplasm Classification: 151.9 ICD Morphology of Neoplasms: M8140/3		
THYROID GLAND Papillary adenocarcinoma NOS Gender: Female Age: 49 Race: Caucasian Tumor Tissue ID #: C1115T	18AA/17AA	
Stage: II;T2N0M0 ICD Neoplasm Classification: 193 ICD Morphology of Neoplasms: M8260/3		
THYROID GLAND Follicular adenocarcinoma NOS Gender: Female Age: 63 Race: Caucasian Tumor Tissue ID #: C1146T	18BB/17BB	
Stage: II;T2N0M0 ICD Neoplasm Classification: 193 ICD Morphology of Neoplasms: M8330/3		
THYROID GLAND Papillary adenocarcinoma NOS Gender: Female Age: 29 Race: Caucasian Tumor Tissue ID #: C614	18S/17S	
Stage: I;T3N1M0 ICD Neoplasm Classification: 193 ICD Morphology of Neoplasms: M8260/3		
THYROID GLAND Medullary carcinoma NOS Gender: Male Age: 66 Race: Caucasian Tumor Tissue ID #: C804	18T/17T	
Stage: IV;T4N1M0 ICD Neoplasm Classification: 193 ICD Morphology of Neoplasms: M8510/3		

Tissue Source/Additional Information	Coordinates Tumor/Normal	Matched Pair Cat. #
THYROID GLAND Papillary adenocarcinoma NOS Gender: Female Age: 49 Race: Asian Tumor Tissue ID #: C777	Stage: Unknown;T3N0M0 ICD Neoplasm Classification: 193 ICD Morphology of Neoplasms: M8260/3	18U/17U
THYROID GLAND Papillary adenocarcinoma NOS Gender: Female Age: 62 Race: Caucasian Tumor Tissue ID #: C805	Stage: II;T3N0M0 ICD Neoplasm Classification: 193 ICD Morphology of Neoplasms: M8260/3	18V/17V
THYROID GLAND Papillary adenocarcinoma NOS Gender: Female Age: 69 Race: Caucasian Tumor Tissue ID #: C1122	Stage: III;T4N1aM0 ICD Neoplasm Classification: 193 ICD Morphology of Neoplasms: M8260/3	18W/17W
THYROID GLAND Papillary adenocarcinoma NOS Gender: Female Age: 70 Race: Caucasian Tumor Tissue ID #: C1128	Stage: II;T3N0M0 ICD Neoplasm Classification: 193 ICD Morphology of Neoplasms: M8260/3	18X/17X
THYROID GLAND Papillary adenocarcinoma NOS Gender: Female Age: 54 Race: Caucasian Tumor Tissue ID #: C1118	Stage: II;T2N0M0 ICD Neoplasm Classification: 193 ICD Morphology of Neoplasms: M8260/3	18Y/17Y
THYROID GLAND Papillary adenocarcinoma NOS Gender: Female Age: 22 Race: Asian Tumor Tissue ID #: C1129	Stage: I;T4N1bM0 ICD Neoplasm Classification: 193 ICD Morphology of Neoplasms: M8260/3	18Z/17Z
LUNG Bronchiolo-alveolar adenocarcinoma Gender: Female Age: 69 Race: Caucasian Tumor Tissue ID #: P14476	Stage: II;T2N1M0 ICD Neoplasm Classification: 162.9 ICD Morphology of Neoplasms: M8250/3	22E/21E HP106L
LUNG Central cancer of the right lung Gender: Male Age: 72 Race: Caucasian Tumor Tissue ID #: 160 T	Stage: Unknown;T3N2M0 ICD Neoplasm Classification: N/A ICD Morphology of Neoplasms: N/A	22F/21F
LUNG Central cancer of the middle lobe of right lung Gender: Male Age: 66 Race: Caucasian Tumor Tissue ID #: 177T	Stage: Unknown;T3N1M0 ICD Neoplasm Classification: N/A ICD Morphology of Neoplasms: N/A	22G/21G

Tissue Source/Additional Information	Coordinates Tumor/Normal	Matched Pair Cat. #
LUNG Peripheral cancer of the upper lobe of the right lung Gender: Male Age: 51 Race: Caucasian Tumor Tissue ID #: 509T	Stage: Unknown;T3N0M0 ICD Neoplasm Classification: N/A ICD Morphology of Neoplasms: N/A	22H/21H
LUNG Squamous cell carcinoma, keratinizing type NOS Gender: Male Age: 63 Race: Caucasian Tumor Tissue ID #: P 18175/00	Stage: I B;T2N0M0 ICD Neoplasm Classification: 162.3 ICD Morphology of Neoplasms: M8071/3	22I/21I
LUNG Squamous cell carcinoma, keratinizing type NOS Gender: Male Age: 62 Race: Caucasian Tumor Tissue ID #: P2022/01	Stage: I;T2N0M0 ICD Neoplasm Classification: 162.9 ICD Morphology of Neoplasms: M8071/3	22J/21J
LUNG Squamous cell carcinoma NOS Gender: Male Age: 67 Race: Caucasian Tumor Tissue ID #: P14980/00	Stage: III A;T3N1M0 ICD Neoplasm Classification: 162.9 ICD Morphology of Neoplasms: M8070/3	22K/21K
LUNG Squamous cell carcinoma, keratinizing type NOS Gender: Male Age: 71 Race: Caucasian Tumor Tissue ID #: P18281/00	Stage: II B;T2N1M0 ICD Neoplasm Classification: 162.9 ICD Morphology of Neoplasms: M8071/3	22L/21L
LUNG Adenocarcinoma NOS Gender: Male Age: 65 Race: Caucasian Tumor Tissue ID #: C1109	Stage: Unknown;Unknown ICD Neoplasm Classification: 162.9 ICD Morphology of Neoplasms: M8140/3	22M/21M
LUNG Squamous cell carcinoma NOS Gender: Male Age: 52 Race: Asian Tumor Tissue ID #: C921	Stage: I B;T2N0M0 ICD Neoplasm Classification: 162.9 ICD Morphology of Neoplasms: M8070/3	22N/21N
TESTIS Seminoma Gender: Male Age: 38 Race: Caucasian Tumor Tissue ID #: C1148T	Stage: IV;T1N3M1 ICD Neoplasm Classification: 162.9 ICD Morphology of Neoplasms: M9061/6	22AA/21AA
TESTIS Seminoma NOS Gender: Male Age: 28 Race: Caucasian Tumor Tissue ID #: C1149T	Stage: III;T3N2M0 ICD Neoplasm Classification: 186.9 ICD Morphology of Neoplasms: M9061/3	22BB/21BB

Tissue Source/Additional Information	Coordinates Tumor/Normal	Matched Pair Cat. #
TESTIS Seminoma NOS Gender: Male Age: 37 Race: Caucasian Tumor Tissue ID #: P 16034/00	Stage: I;T2N0M0 ICD Neoplasm Classification: 186 ICD Morphology of Neoplasms: M9061/3	22S/21S
TESTIS Seminoma NOS Gender: Male Age: 45 Race: Caucasian Tumor Tissue ID #: C1125	Stage: II;T2N0M0 ICD Neoplasm Classification: 186.9 ICD Morphology of Neoplasms: M9061/3	22T/21T
TESTIS Embryonal carcinoma NOS Gender: Male Age: 21 Race: Caucasian Tumor Tissue ID #: C1123	Stage: I;T1N0M0 ICD Neoplasm Classification: 186.9 ICD Morphology of Neoplasms: M9070/3	22U/21U
TESTIS Seminoma NOS Gender: Male Age: 32 Race: Caucasian Tumor Tissue ID #: C1113	Stage: I;T1N0M0 ICD Neoplasm Classification: 186.9 ICD Morphology of Neoplasms: M9061/3	22V/21V
TESTIS Seminoma NOS Gender: Male Age: 26 Race: Caucasian Tumor Tissue ID #: C1111	Stage: I;T1N0M0 ICD Neoplasm Classification: 186.9 ICD Morphology of Neoplasms: M9061/3	22W/21W
TESTIS Seminoma NOS Gender: Male Age: 18 Race: Caucasian Tumor Tissue ID #: C1112	Stage: I;T1N0M0 ICD Neoplasm Classification: 186.9 ICD Morphology of Neoplasms: M9061/3	22X/21X
TESTIS Endodermal sinus tumor Gender: Male Age: 22 Race: Caucasian Tumor Tissue ID #: C1133	Stage: IV;T2NxM1 ICD Neoplasm Classification: 186.9 ICD Morphology of Neoplasms: M9071/3	22Y/21Y
TESTIS Seminoma NOS Gender: Male Age: 32 Race: Caucasian Tumor Tissue ID #: C1142T	Stage: I;T1N0M0 ICD Neoplasm Classification: 186.9 ICD Morphology of Neoplasms: M9061/3	22Z/21Z
KIDNEY Renal cell carcinoma Gender: Male Age: 59 Race: Caucasian Tumor Tissue ID #: T 14403/00	Stage: II;T2N0M0 ICD Neoplasm Classification: 189 ICD Morphology of Neoplasms: M8312/3	26E/25E

Tissue Source/Additional Information	Coordinates Tumor/Normal	Matched Pair Cat. #
KIDNEY Oxyphilic adenoma Gender: Male Age: 60 Race: Caucasian Tumor Tissue ID #: T 9260/00	Stage: Unknown;Unknown ICD Neoplasm Classification: 189 ICD Morphology of Neoplasms: M8290/0	26F/25F
KIDNEY Renal cell carcinoma Gender: Female Age: 69 Race: Caucasian Tumor Tissue ID #: T 9480/00	Stage: IV;T3aN0M1 ICD Neoplasm Classification: 170.3 ICD Morphology of Neoplasms: M8312/6	26G/25G
KIDNEY Renal cell carcinoma Gender: Male Age: 55 Race: Asian Tumor Tissue ID #: T 16877/00	Stage: III;T3aN0M0 ICD Neoplasm Classification: 189 ICD Morphology of Neoplasms: M8312/3	26H/25H
KIDNEY Transitional cell carcinoma NOS Gender: Male Age: 58 Race: Caucasian Tumor Tissue ID #: T 7596/00	Stage: III;T3bN0M0 ICD Neoplasm Classification: 189 ICD Morphology of Neoplasms: M8120/3	26I/25I
KIDNEY Clear cell adenocarcinoma NOS Gender: Male Age: 41 Race: Caucasian Tumor Tissue ID #: T 15938/00	Stage: III;T3aN0M0 ICD Neoplasm Classification: 189 ICD Morphology of Neoplasms: M8310/3	26J/25J
KIDNEY Transitional cell carcinoma NOS Gender: Male Age: 79 Race: Caucasian Tumor Tissue ID #: T 14412	Stage: III;T2N1M0 ICD Neoplasm Classification: 189 ICD Morphology of Neoplasms: M8120/3	26K/25K
KIDNEY Renal cell carcinoma Gender: Female Age: 69 Race: Caucasian Tumor Tissue ID #: T 14250	Stage: IV;T3bN1M0 ICD Neoplasm Classification: 189 ICD Morphology of Neoplasms: M8312/3	26L/25L
KIDNEY Epithelial nephroblastoma Gender: Female Age: 22 Race: Caucasian Tumor Tissue ID #: T 14327	Stage: II B;T3N0M0 ICD Neoplasm Classification: 189 ICD Morphology of Neoplasms: M8961/3	26M/25M
KIDNEY Cancer of the left kidney Gender: Male Age: 58 Race: Caucasian Tumor Tissue ID #: 524T	Stage: Unknown;T3aNxM0 ICD Neoplasm Classification: 189 ICD Morphology of Neoplasms: N/A	26N/25N

Tissue Source/Additional Information		Coordinates Tumor/Normal	Matched Pair Cat. #
SKIN	Malignant melanoma NOS Gender: Female Age: 40 Race: Caucasian Tumor Tissue ID #: C620T	Stage: III;T4N0M0 ICD Neoplasm Classification: 173.7 ICD Morphology of Neoplasms: M8720/3	26AA/25AA
SKIN	Malignant melanoma NOS Gender: Female Age: 43 Race: Caucasian Tumor Tissue ID #: C 791T	Stage: I;T2N0M0 ICD Neoplasm Classification: 173.7 ICD Morphology of Neoplasms: M8720/3	26BB/25BB
SKIN	Malignant melanoma NOS Gender: Female Age: 49 Race: Caucasian Tumor Tissue ID #: P 12549/00	Stage: I;T2N0M0 ICD Neoplasm Classification: 173.6 ICD Morphology of Neoplasms: M8720/3	26S/25S
SKIN	Malignant melanoma NOS Gender: Female Age: 45 Race: Caucasian Tumor Tissue ID #: P 7619/00	Stage: III;T4N0M0 ICD Neoplasm Classification: 173.7 ICD Morphology of Neoplasms: M8720/3	26T/25T
SKIN	Malignant melanoma NOS Gender: Male Age: 73 Race: Caucasian Tumor Tissue ID #: P 7683/00	Stage: III;T4N0M0 ICD Neoplasm Classification: 173.7 ICD Morphology of Neoplasms: M8720/3	26U/25U
SKIN	Malignant melanoma NOS Gender: Male Age: 58 Race: Caucasian Tumor Tissue ID #: C134	Stage: IV;Unknown ICD Neoplasm Classification: 171.4 ICD Morphology of Neoplasms: M8720/3	26V/25V
SKIN	Malignant melanoma NOS Gender: Female Age: 74 Race: Caucasian Tumor Tissue ID #: C161	Stage: III;T4N0M0 ICD Neoplasm Classification: 173.7 ICD Morphology of Neoplasms: M8720/3	26W/25W
SKIN	Unknown Gender: Male Age: 64 Race: Caucasian Tumor Tissue ID #: C915T	Stage: Unknown;Unknown ICD Neoplasm Classification: 142.9 ICD Morphology of Neoplasms: M8982/3	26X/25X
SKIN	Squamous cell carcinoma NOS Gender: Female Age: 70 Race: Caucasian Tumor Tissue ID #: C689T	Stage: II;T2N0M0 ICD Neoplasm Classification: 173.5 ICD Morphology of Neoplasms: M8070/3	26Y/25Y

Tissue Source/Additional Information	Coordinates Tumor/Normal	Matched Pair Cat. #
SKIN Squamous cell carcinoma, keratinizing type NOS Gender: Male Age: 68 Race: Caucasian Tumor Tissue ID #: C922T	Stage: II;T2N0M0 ICD Neoplasm Classification: 173.9 ICD Morphology of Neoplasms: M8071/3	26Z/25Z
BLADDER Transitional cell carcinoma NOS Gender: Male Age: 60 Race: Caucasian Tumor Tissue ID #: C1200	Stage: II;T2N0M0 ICD Neoplasm Classification: 188.9 ICD Morphology of Neoplasms: M8120/3	30E/29E
BLADDER Transitional cell carcinoma NOS Gender: Male Age: 54 Race: Caucasian Tumor Tissue ID #: C1174	Stage: III;T3N0M0 ICD Neoplasm Classification: 188.3 ICD Morphology of Neoplasms: M8120/3	30F/29F
BLADDER Transitional cell carcinoma NOS Gender: Male Age: 62 Race: Asian Tumor Tissue ID #: C1175	Stage: IV;T4bN0M0 ICD Neoplasm Classification: 188.9 ICD Morphology of Neoplasms: M8120/3	30G/29G
BLADDER Transitional cell carcinoma NOS Gender: Female Age: 59 Race: Caucasian Tumor Tissue ID #: C1177	Stage: II;T2N0M0 ICD Neoplasm Classification: 188.9 ICD Morphology of Neoplasms: M8120/3	30H/29H
BLADDER Transitional cell carcinoma NOS Gender: Male Age: 56 Race: Caucasian Tumor Tissue ID #: C1180	Stage: III;T3aN0M0 ICD Neoplasm Classification: 188.9 ICD Morphology of Neoplasms: M8120/3	30I/29I
TRACHEA Cancer of the trachea Gender: Female Age: 44 Race: Caucasian Tumor Tissue ID #: 500T	Stage: Unknown;T4N1M0 ICD Neoplasm Classification: N/A ICD Morphology of Neoplasms: N/A	30L/29L
TRACHEA Cancer of the trachea Gender: Female Age: 54 Race: Caucasian Tumor Tissue ID #: 503T	Stage: Unknown;T4N0M0 ICD Neoplasm Classification: N/A ICD Morphology of Neoplasms: N/A	30M/29M
TRACHEA Cancer of the trachea Gender: Male Age: 54 Race: Caucasian Tumor Tissue ID #: 506T	Stage: Unknown;T3N1M0 ICD Neoplasm Classification: N/A ICD Morphology of Neoplasms: N/A	30N/29N

Tissue Source/Additional Information	Coordinates Tumor/Normal	Matched Pair Cat. #
SMALL INTESTINE Adenocarcinoma NOS Gender: Female Age: 49 Race: Caucasian Tumor Tissue ID #: P13016	Stage: III;T1N1M0 ICD Neoplasm Classification: 156.2 ICD Morphology of Neoplasms: M8140/3	30S/29S
SMALL INTESTINE Adenocarcinoma NOS Gender: Male Age: 46 Race: Caucasian Tumor Tissue ID #: P1063	Stage: II;T3N0M0 ICD Neoplasm Classification: 152.9 ICD Morphology of Neoplasms: M8140/3	30T/29T HP1011
SMALL INTESTINE Adenocarcinoma NOS Gender: Female Age: 65 Race: Caucasian Tumor Tissue ID #: P 4671/01	Stage: II;T3N0M0 ICD Neoplasm Classification: 152 ICD Morphology of Neoplasms: M8140/3	30U/29U
SMALL INTESTINE Adenocarcinoma NOS Gender: Female Age: 63 Race: Caucasian Tumor Tissue ID #: P 4417/01	Stage: II;T2N0M0 ICD Neoplasm Classification: 156.2 ICD Morphology of Neoplasms: M8140/3	30V/29V
SMALL INTESTINE Carcinoid tumor, malignant Gender: Male Age: 59 Race: Caucasian Tumor Tissue ID #: P 14804/00	Stage: Unknown;Unknown ICD Neoplasm Classification: 152.9 ICD Morphology of Neoplasms: M8240/3	30W/29W
SMALL INTESTINE Adenocarcinoma NOS Gender: Female Age: 70 Race: Caucasian Tumor Tissue ID #: P 3013/01	Stage: III;T3N1M0 ICD Neoplasm Classification: 156.2 ICD Morphology of Neoplasms: M8140/3	30X/29X
SMALL INTESTINE Adenocarcinoma NOS Gender: Male Age: 52 Race: Caucasian Tumor Tissue ID #: C1144T	Stage: IV;T4N1M1 ICD Neoplasm Classification: 156.2 ICD Morphology of Neoplasms: M8140/3	30Y/29Y
VULVA Squamous cell carcinoma, keratinizing type NOS Gender: Female Age: 80 Race: Caucasian Tumor Tissue ID #: P 5305/01	Stage: II;T2N0M0 ICD Neoplasm Classification: 184.4 ICD Morphology of Neoplasms: M8071/3	34E/33E
VULVA Squamous cell carcinoma NOS Gender: Female Age: 78 Race: Caucasian Tumor Tissue ID #: C1181	Stage: II;T2N0M0 ICD Neoplasm Classification: 184.4 ICD Morphology of Neoplasms: M8070/3	34F/33F

Tissue Source/Additional Information		Coordinates Tumor/Normal	Matched Pair Cat. #
VULVA	Squamous cell carcinoma NOS Gender: Female Age: 78 Race: Caucasian Tumor Tissue ID #: C1182	Stage: II;T2N0M0 ICD Neoplasm Classification: 184.4 ICD Morphology of Neoplasms: M8070/3	34G/33G
VULVA	Squamous cell carcinoma NOS Gender: Female Age: 74 Race: Caucasian Tumor Tissue ID #: C1183	Stage: II;T2N0M0 ICD Neoplasm Classification: 184.4 ICD Morphology of Neoplasms: M8070/3	34H/33H
VULVA	Squamous cell carcinoma NOS Gender: Female Age: 60 Race: Caucasian Tumor Tissue ID #: C1184	Stage: Unknown;T2aN1M0 ICD Neoplasm Classification: 184.4 ICD Morphology of Neoplasms: M8070/3	34I/33I
LIVER	Primary cancer of the liver Gender: Male Age: 59 Race: Caucasian Tumor Tissue ID #: 570T	Stage: Unknown;T3N0M0 ICD Neoplasm Classification: N/A ICD Morphology of Neoplasms: N/A	34L/33L
LIVER	Primary cancer of the liver Gender: Male Age: 36 Race: Caucasian Tumor Tissue ID #: 585T	Stage: Unknown;T2N0M0 ICD Neoplasm Classification: N/A ICD Morphology of Neoplasms: N/A	34M/33M
LIVER	Primary cancer of the liver Gender: Male Age: 58 Race: Caucasian Tumor Tissue ID #: 589T	Stage: Unknown;T2N0M0 ICD Neoplasm Classification: N/A ICD Morphology of Neoplasms: N/A	34N/33N
PANCREAS	Adenocarcinoma NOS Gender: Female Age: 56 Race: Caucasian Tumor Tissue ID #: 45T	Stage: I;Unknown ICD Neoplasm Classification: N/A ICD Morphology of Neoplasms: N/A	34S/33S
PANCREAS	Adenocarcinoma NOS Gender: Female Age: 63 Race: Caucasian Tumor Tissue ID #: P17286/00	Stage: I;T2N0M0 ICD Neoplasm Classification: 157 ICD Morphology of Neoplasms: M8140/3	34T/33T
PANCREAS	Adenocarcinoma NOS Gender: Male Age: 40 Race: Caucasian Tumor Tissue ID #: P639/01	Stage: III;T2N1M0 ICD Neoplasm Classification: 157.9 ICD Morphology of Neoplasms: M8140/3	34U/33U

Tissue Source/Additional Information		Coordinates Tumor/Normal	Matched Pair Cat. #
PANCREAS	Squamous cell carcinoma NOS Gender: Male Age: 65 Race: Caucasian Tumor Tissue ID #: P553/01	Stage: II;T3N0M0 ICD Neoplasm Classification: 157.3 ICD Morphology of Neoplasms: M8070/3	34V/33V
PANCREAS	Adenocarcinoma NOS Gender: Female Age: 51 Race: Caucasian Tumor Tissue ID #: P17958/00	Stage: IV A;T3N1M1 ICD Neoplasm Classification: 157.9 ICD Morphology of Neoplasms: M8140/3	34W/33W
PANCREAS	Carcinoid tumor, malignant Gender: Female Age: 53 Race: Caucasian Tumor Tissue ID #: C1153T	Stage: Unknown;Unknown ICD Neoplasm Classification: 157 ICD Morphology of Neoplasms: M8240/3	34X/33X
PANCREAS	Cancer of caput of pancreas Gender: Male Age: 40 Race: Caucasian Tumor Tissue ID #: 543T	Stage: Unknown;T3NxM0 ICD Neoplasm Classification: 157 ICD Morphology of Neoplasms: N/A	34Y/33Y
PROSTATE	Adenocarcinoma NOS Gender: Male Age: 62 Race: Caucasian Tumor Tissue ID #: P7331/00	Stage: I;T1N0M0 ICD Neoplasm Classification: 185 ICD Morphology of Neoplasms: M8140/3	38E/37E
PROSTATE	Adenocarcinoma Gender: Male Age: 65 Race: Unknown Tumor Tissue ID #: 750B	Stage: Unknown;Unknown ICD Neoplasm Classification: N/A ICD Morphology of Neoplasms: N/A	38F/37F HP101P
PROSTATE	Adenocarcinoma Gender: Male Age: 67 Race: Unknown Tumor Tissue ID #: 842B	Stage: Unknown;Unknown ICD Neoplasm Classification: N/A ICD Morphology of Neoplasms: N/A	38G/37G
PROSTATE	Adenocarcinoma Gender: Male Age: 68 Race: Caucasian Tumor Tissue ID #: 1867	Stage: Unknown;Unknown ICD Neoplasm Classification: N/A ICD Morphology of Neoplasms: N/A	38H/37H HP103P

Tissue Source/Additional Information	Coordinates Tumor/Normal	Matched Pair Cat. #
HUMAN CANCER CELL LINES		
HeLa	37S	
Daudi	37T	
K562	37U	
HL60	37V	
G361	37W	
A549	37X	
MOLT4	37Y	
SW480	37Z	
Raji	37AA	
CONTROLS		
yeast total RNA	44I	
yeast tRNA	44L	
<i>E. coli</i> DNA	44O	
Poly(A)	44R	
human C ₀ t-1 DNA	44U	
human genomic DNA	44X	
ubiquitin cDNA	44E 44BB	

Annexe B :

**Milieus utilisés dans la technique du double –hybride
BacterioMatch II System (Stratagene)**

BacterioMatch® II Two-Hybrid System Vector Kit

INSTRUCTION MANUAL

Catalog #240065

Revision #016009

For In Vitro Use Only



LIMITED PRODUCT WARRANTY

This warranty limits our liability to replacement of this product. No other warranties of any kind, express or implied, including without limitation, implied warranties of merchantability or fitness for a particular purpose, are provided by Stratagene. Stratagene shall have no liability for any direct, indirect, consequential, or incidental damages arising out of the use, the results of use, or the inability to use this product.

ORDERING INFORMATION AND TECHNICAL SERVICES

United States and Canada

Stratagene

11011 North Torrey Pines Road

La Jolla, CA 92037

Telephone (858) 373-6300**Order Toll Free** (800) 424-5444**Technical Services** (800) 894-1304**Internet** tech_services@stratagene.com**World Wide Web** www.stratagene.com

Stratagene European Contacts

Location	Telephone	Fax	Technical Services
Austria	0800 292 499	0800 292 496	0800 292 498
Belgium	00800 7000 7000	00800 7001 7001	00800 7400 7400
	0800 15775	0800 15740	0800 15720
France	00800 7000 7000	00800 7001 7001	00800 7400 7400
	0800 919 288	0800 919 287	0800 919 289
Germany	00800 7000 7000	00800 7001 7001	00800 7400 7400
	0800 182 8232	0800 182 8231	0800 182 8234
Netherlands	00800 7000 7000	00800 7001 7001	00800 7400 7400
	0800 023 0446	+31 (0)20 312 5700	0800 023 0448
Switzerland	00800 7000 7000	00800 7001 7001	00800 7400 7400
	0800 563 080	0800 563 082	0800 563 081
United Kingdom	00800 7000 7000	00800 7001 7001	00800 7400 7400
	0800 917 3282	0800 917 3283	0800 917 3281

All Other Countries

Please contact your local distributor. A complete list of distributors is available at www.stratagene.com.

BacterioMatch® II Two-Hybrid System Vector Kit

CONTENTS

Critical Success Factors	1
Materials Provided	2
Storage Conditions	2
Additional Materials Required	2
Notice to Purchaser	2
Introduction	3
Advantages of the Bacterial Two-Hybrid System.....	3
Overview of the Interaction	4
Positive Selection Using <i>HIS3</i> Transcriptional Activation.....	5
Reporter Strain.....	6
Plasmids	7
pBT Vector Map.....	8
pTRG Vector Map.....	9
Control Plasmids.....	10
Recombinant pBT and pTRG Plasmid Construction and Validation	11
Cloning into the pBT and pTRG Plasmids.....	11
Verifying Expression.....	14
Preparation of Recombinant pBT and pTRG DNA.....	16
Testing Self-Activation by Recombinant pBT or pTRG	16
Cotransformation Protocol.....	18
Performing Protein-Protein Interaction Assays	21
Protein-Protein Interaction Assay Plating and Analysis.....	22
Preserving Putative Positive Colonies.....	23
Verification of Positives using the Streptomycin Resistance Reporter.....	23
Troubleshooting	24
Preparation of Media and Reagents	26
Minimal Media for BacterioMatch® II Two-Hybrid Assays.....	26
Rich Media and Other Reagents.....	28
References	29
Endnotes	29
MSDS Information	29

CRITICAL SUCCESS FACTORS

Before beginning experiments using the BacterioMatch® II two-hybrid system vector kit, please read the following general guidelines. Following these recommendations is critical to the success of your experiments.

1. It is important that the strain used to propagate pBT and pTRG plasmids contains the *lacI^a* gene. The XL1-Blue MRF['] Kan strain supplied in this kit as a glycerol stock is an appropriate host for propagation of pBT and pTRG. Do not substitute another strain when growing these plasmids.
2. pBT and pTRG are both low-copy-number plasmids; pBT is present at 5–10 copies per cell and pTRG is present at about 20–30 copies per cell. Depending on the yield of plasmid desired, it may be necessary to purify plasmid DNA from a larger volume of culture. It is also suggested that, during purification of plasmid DNA, protocol modifications recommended by the purification kit manufacturer for low-copy-number plasmids are employed.
3. The BacterioMatch II system uses a new *HIS3-aadA* reporter cassette. Detection of protein-protein interactions is based on transcriptional activation of the *HIS3* reporter gene, which allows growth in the presence of 3-amino-1,2,4-triazole (3-AT), a competitive inhibitor of His3 enzyme. Prepare the media exactly as described in *Preparation of Media and Reagents*. In particular, ensure that the Selective Screening Medium plates contain the correct amount of 3-amino-1,2,4-triazole (3-AT). Small deviations from the media recipes can result in significant variation in experimental results.

BacterioMatch® II Two-Hybrid System Vector Kit

MATERIALS PROVIDED

Materials provided	Amount
BacterioMatch® II two-hybrid system plasmids ^a	
pTRG target plasmid, supercoiled, 10 µg [1 µg/µl in TE buffer (Catalog #240066)]	10 µl
pBT bait plasmid, supercoiled 10 µg [1 µg/µl in TE buffer (Catalog #240067)]	10 µl
BacterioMatch® II two-hybrid system control plasmids ^a	
pTRG-Gal11 ^P control plasmid, supercoiled, 150 ng [50 ng/µl in TE buffer]	3 µl
pBT-LGF2 control plasmid, supercoiled, 150 ng [50 ng/µl in TE buffer]	3 µl
XL1-Blue MRF' Kan, glycerol stock ^b (host strain for propagating pBT and pTRG recombinants)	0.5 ml

^a Store at -20°C.

^b Store at -80°C. Genotype: $\Delta(mcrA)183 \Delta(mcrCB-hsdSMR-mrr)173 endA1 supE44 thi-1 recA1 gyrA96 relA1 lac [F' proAB lac^qZ\Delta M15 Tn5 (Kan^r)]$.

STORAGE CONDITIONS

All Plasmids: -20°C

XL1-Blue MRF' Kan glycerol stock: -80°C

ADDITIONAL MATERIALS REQUIRED

BacterioMatch® II Validation Reporter Competent Cells (Catalog #200192)
14-ml BD Falcon polypropylene round-bottom tubes (BD Biosciences Catalog #352059)
components for minimal, his-dropout bacteriological media,[§] including:

 M9 salts[§] (Qbiogene, Cat. # 3037-032)

 adenine HCl (Sigma, Cat. #A-8751)

 -His Dropout Supplement (BD/Clontech, Cat. #630415)

chloramphenicol

tetracycline

streptomycin

isopropyl-1-thio-β-D-galactopyranoside (IPTG)

3-amino-1,2,4-triazole (3-AT, Sigma, Catalog #A-8056)

dimethyl sulfoxide (DMSO)

water baths (12°C and 42°C)

incubators (30°C and 37°C)

NOTICE TO PURCHASER

U.S. Patent No. 5,925,523, covering the BacterioMatch® two-hybrid system, is licensed exclusively by Stratagene. Research use of the BacterioMatch two-hybrid system by commercial entities requires a license from Stratagene. For license information, please contact: Director of Business Development at (858) 535-5400.

[§] See *Preparation of Media and Reagents*.

INTRODUCTION

Stratagene's exclusive BacterioMatch® II two-hybrid system is an efficient method for detecting protein-protein interactions *in vivo*. The BacterioMatch II two-hybrid system is based on a methodology developed by Dove, Joung, and Hochschild of Harvard Medical School and further refined by Joung and Pabo of the Massachusetts Institute of Technology.^{1, 2, 3} This *E. coli*-based system offers advantages over Stratagene's original BacterioMatch two-hybrid system. Featuring a new *HIS3-aadA* reporter cassette, the BacterioMatch II system offers the ability to screen libraries for harder-to-find binding partners with reduced background. Detection of protein-protein interactions is based on transcriptional activation of the *HIS3* reporter gene, which allows growth in the presence of 3-amino-1,2,4-triazole (3-AT), a competitive inhibitor of His3 enzyme. Positives are verified by using the *aadA* gene, which confers streptomycin resistance, as a secondary reporter.

The BacterioMatch II Two-Hybrid System Vector Kit (Catalog #240065) is designed for the characterization of protein-protein interactions between a pair of proteins cloned in the pBT (bait) and the pTRG (target) vectors. Once an interaction pair has been identified, the system may be used to map the interaction domain of either the bait or target protein after site-directed mutagenesis.

For two-hybrid screening using the pBT-bait plasmid to identify protein-protein interaction partners from a pTRG cDNA plasmid library, Stratagene offers the BacterioMatch II Two-Hybrid System Library Construction Kit (Catalog #200412) and the BacterioMatch II Two-Hybrid System Library Construction Kit with Electrocompetent Cells (Catalog #200414). In addition, Stratagene offers a variety of BacterioMatch II premade cDNA library kits. See <http://www.stratagene.com/index.asp?catID=75> for a complete list of premade pTRG plasmid libraries.

Advantages of the Bacterial Two-Hybrid System

While the yeast two-hybrid system has been widely and successfully exploited, a method that utilizes *E. coli* is valuable for many reasons: *E. coli* grows much faster than yeast, it is transformed with higher efficiency so larger numbers of interactions can be more rapidly and easily screened, and isolating plasmid DNA from *E. coli* is easier than DNA isolation from yeast. Furthermore, using *E. coli* for two-hybrid screening reduces the chance that the host harbors a eukaryotic homologue of one of the interacting protein partners. Some eukaryotic regulatory proteins, such as cell cycle checkpoint proteins and signal transduction pathway proteins, may be toxic in yeast because they interfere with the function of yeast homologues; presumably they would not be as harmful in *E. coli*. For the same reason, use of a bacterial system could also reduce the number of false positives observed. Although some heterologous proteins could be toxic to an *E. coli* host, and bacteria lack the ability to perform some posttranslational modifications, an *E. coli* two-hybrid system makes for an important experimental alternative.

OVERVIEW OF THE INTERACTION

The BacterioMatch II two-hybrid system detects protein-protein interactions based on transcriptional activation. A protein of interest (the bait) is fused to the full-length bacteriophage λ repressor protein (λ cI, 237 amino acids), containing the amino-terminal DNA-binding domain and the carboxyl-terminal dimerization domain. The corresponding target protein is fused to the N-terminal domain of the α -subunit of RNA polymerase (248 amino acids). The bait is tethered to the λ operator sequence upstream of the reporter promoter through the DNA-binding domain of λ cI. When the bait and target interact, they recruit and stabilize the binding of RNA polymerase at the promoter and activate the transcription of the *HIS3* reporter gene. A second reporter gene, *aadA*, encoding a protein that confers streptomycin resistance, provides an additional mechanism to validate the bait and target interaction (Figure 1).

Due to the tendency of both the λ repressor protein and the N-terminal domain of the α -subunit of RNA polymerase to dimerize, this system might not be optimal for the analysis of proteins that self-associate unless their interaction with other protein partners depends on the oligomerization.

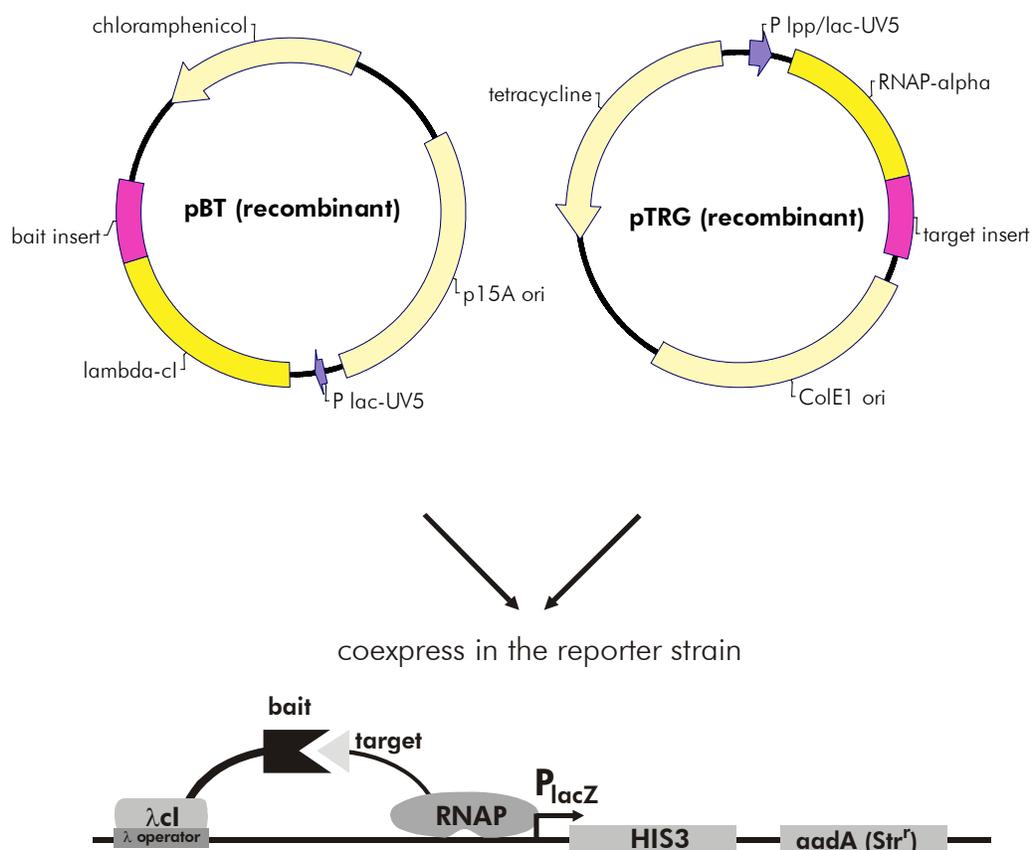


Figure 1 Schematic of the BacterioMatch II two-hybrid system dual reporter construct

Positive Selection Using *HIS3* Transcriptional Activation

Transcriptional activation of the *HIS3* gene is used as the initial test for interaction of the bait and target hybrid proteins. *HIS3* encodes a component of the histidine biosynthetic pathway that complements a *hisB* mutation in the reporter strain. The *HIS3* gene product is produced from the reporter gene cassette at low levels in the absence of transcriptional activation, allowing the reporter strain to grow on minimal medium lacking histidine.

The compound 3-amino-1,2,4-triazole (3-AT) acts as a competitive inhibitor of the *HIS3* gene product. In the presence of 5 mM 3-AT, the reporter strain is unable to grow on media lacking histidine.

When the reporter strain is cotransformed with hybrid bait and target proteins that interact, the RNA polymerase is recruited to the promoter, activating the transcription of *HIS3*. Growth of the reporter strain on media lacking histidine and containing 5 mM 3-AT occurs when transcriptional activation increases expression of the *HIS3* gene product to levels that are sufficient to overcome the competitive inhibition by 3-AT. This allows for positive selection for plasmids encoding interacting proteins on media containing 5 mM 3-AT.

The reporter gene cassette (Figure 1) is present in the BacterioMatch II reporter strain on an F' episome. The cassette contains the *HIS3* gene, conferring the ability to grow on media containing 3-AT, and the *aadA* gene, conferring resistance to streptomycin. The activatable promoter in the reporter gene cassette is a modified *lac* promoter that contains a single λ operator (O_{R2}) centered at position -62, replacing the CRP-binding site originally associated with the *lac* promoter. This modified *lac* promoter is not inducible by IPTG. The Shine-Dalgarno translational signal is also included upstream of the reporter genes.

Reporter Strain

The BacterioMatch II two-hybrid system uses a reporter strain derived from Stratagene's XL1-Blue MR, providing a high transformation efficiency in a restriction minus host. The reporter strain harbors *lacI^q* on the F' episome to repress synthesis of the bait and target in the absence of IPTG, but a basal level of transcription of these proteins is always occurring.

Note *The BacterioMatch II reporter strain contains the supE44 mutation, an amber suppressor mutation that allows some readthrough of TAG stop codons with insertion of glutamine.*

Stratagene offers three varieties of the BacterioMatch II reporter strain competent cells that are optimized for specific applications.

Catalog #	Description	Applications	Efficiency	Format
200190	BacterioMatch® II Screening Reporter Competent Cells (chemically-competent)	Library screening (cotransformation of pBT-bait and the pTRG library)	$\geq 7.5 \times 10^7$ cfu/ μ g pUC18	6 \times 500 μ l aliquots
200192	BacterioMatch® II Validation Reporter Competent Cells (chemically-competent)	Interaction assays, expression testing, validation of screen isolates	$\geq 1 \times 10^7$ cfu/ μ g pUC18	5 \times 200 μ l aliquots
200195	BacterioMatch® II Electrocompetent Reporter Cells (electroporation-competent)	Library screening (cotransformation of pBT-bait and the pTRG library) with increased efficiency compared to chemically-competent reporter cells	$\geq 1 \times 10^9$ cfu/ μ g pUC18	5 \times 100 μ l aliquots

PLASMIDS

The pBT bait plasmid and the pTRG target plasmid are designed to allow detection of protein-protein interactions when used to cotransform a host strain containing the appropriate reporter gene cassette. The 3.2 kb pBT bait plasmid (Figure 2) carries a low-copy p15A replication origin and confers chloramphenicol resistance. The plasmid encodes the full-length bacterial phage λ cI protein under the control of the IPTG-inducible *lac-UV5* promoter. The multiple cloning site* contains several restriction sites to facilitate fusion gene construction. From a 5 ml LB-chloramphenicol culture (34 μ g/ml chloramphenicol), yields of approximately 200 ng are obtained.

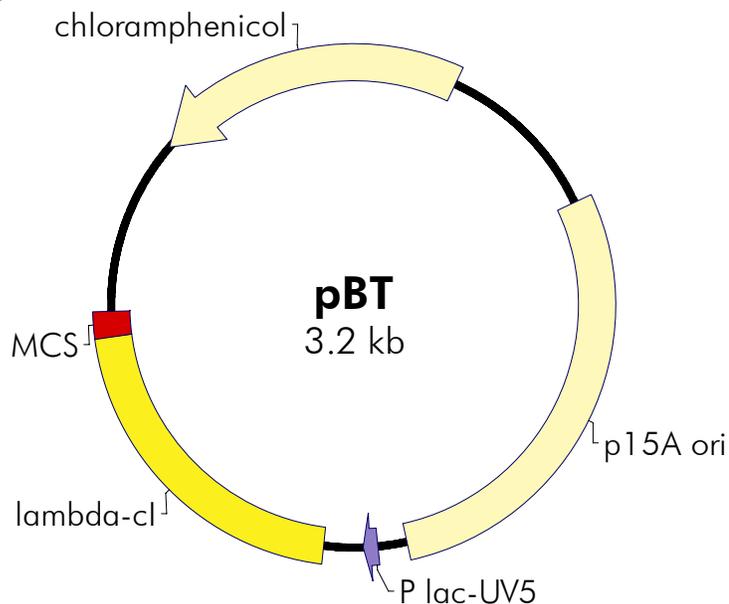
The 4.4 kb target plasmid, pTRG (Figure 3), carries the low-copy ColE1 replication origin and confers tetracycline resistance. The plasmid directs transcription of the amino-terminal domain of RNA polymerase α subunit through a multiple cloning site at the 3' end of the α subunit gene and is under the control of the IPTG-inducible, tandem promoter *lpp/lac-UV5*. The expression level from the *lpp/lac-UV5* promoter operator is higher than the levels of *lac-UV5*-driven expression. The arrangement of *EcoR* I and *Xho* I restriction sites in the pTRG MCS makes the plasmid compatible with inserts produced using Stratagene's cDNA Synthesis Kit (Catalog #200401). From a 5 ml LB-tetracycline culture (12.5 μ g/ml tetracycline), yields of approximately 1 μ g are obtained.

Both bait and target plasmids contain the *Not* I restriction site in the MCS which encodes a short alanine linker to facilitate the orientation and folding of the fused bait and target proteins.

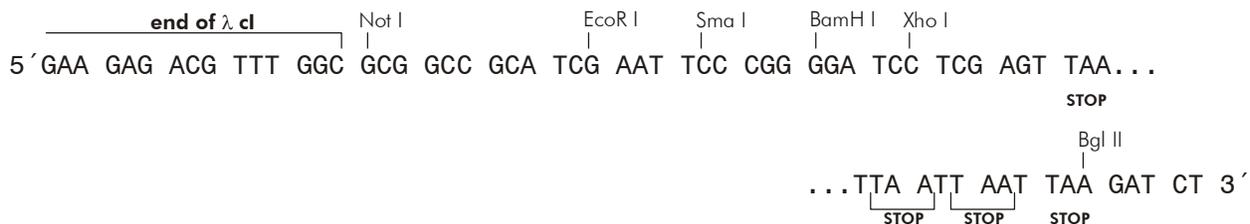
Note *The pTRG vector (with or without insert present), as well as recombinant pBT plasmids, can be toxic to the bacterial host. To minimize host toxicity, propagate all forms of the pBT and pTRG plasmids using a host strain containing the lacI^q gene (such as the XLI-Blue MRF' strain provided). When growing cells harboring either plasmid on rich medium, incubate the cells at 30°C, in order to further reduce any potential toxicity. For growth on minimal medium, incubate the cells at 37°C.*

* Sites in the MCS of the pBT and pTRG are not directly compatible with each other. If swapping the insert between the two plasmids is desired, indirect cloning methods will have to be employed.

pBT Vector Map



pBT Multiple Cloning Site Region (sequence shown 2327–2394)

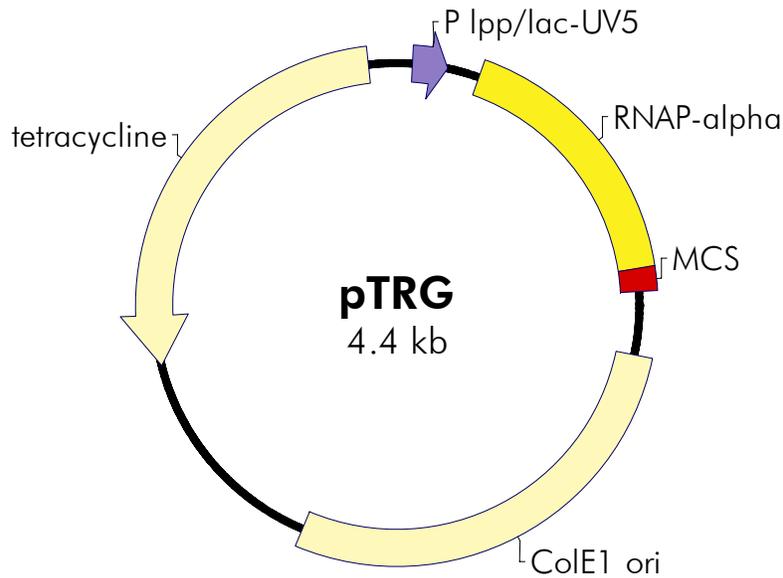


Feature	Position
chloramphenicol resistance ORF	2770–219
p15A origin of replication	581–1493
<i>lac-UV5</i> promoter	1556–1586
λ cl ORF	1631–2341
pBT forward primer [5' TCCGTTGTGGGGAAAGTTATC 3']	2291–2311
multiple cloning site	2342–2394
pBT reverse primer [5' GGGTAGCCAGCAGCATCC 3']	2419–2436

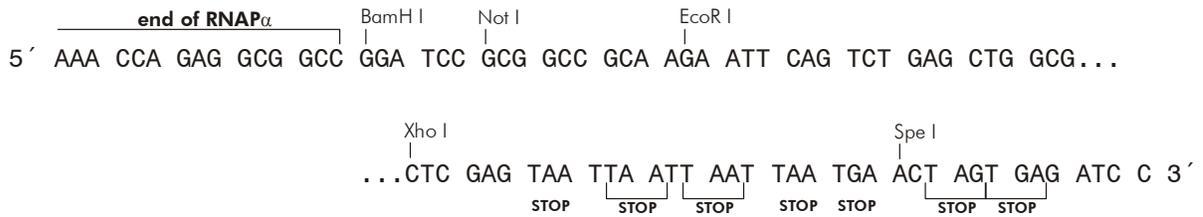
Note The nucleotide sequence and list of restriction sites for the pBT bait plasmid can be found at www.stratagene.com.

FIGURE 2 The pBT bait plasmid

pTRG Vector Map



pTRG Multiple Cloning Site Region (sequence shown 978–1065)



Feature	Position
<i>lpp</i> promoter	47–76
<i>lac-UV5</i> promoter	119–148
RNAP α ORF	243–992
pTRG forward primer [5' TGGCTGAACAACCTGGAAGCT 3']	913–932
multiple cloning site	993–1058
pTRG reverse primer [5' ATTCGTCGCCCGCCATAA 3']	1102–1119
ColE1 origin of replication	1243–2475
tetracycline resistance ORF	3120–4310

Note The nucleotide sequence and list of restriction sites for the pTRG target plasmid can be found at www.stratagene.com.

FIGURE 3 The pTRG target plasmid

Control Plasmids

Description

The BacterioMatch II two-hybrid system includes the pTRG-Gal11^P and the pBT-LGF2 positive control plasmids (Table I and Figure 4).

TABLE I
Description Of Control Plasmids

Control plasmid	Description
pBT-LGF2	interaction control plasmid encoding the dimerization domain (40 amino acids) of the Gal4 transcriptional activator protein
pTRG-Gal11 ^P	interaction control plasmid encoding a domain (90 amino acids) of the mutant form of the Gal11 protein

Control Plasmid Maps

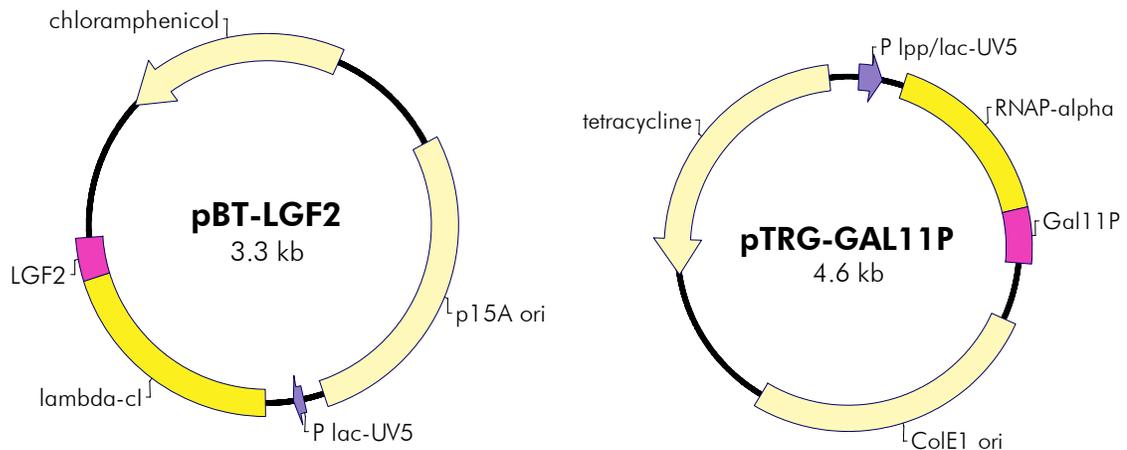


FIGURE 4 Circular maps of the interaction control plasmids

Applications

The dimerization domain of the yeast transcriptional activator Gal4 and a domain derived from a mutant form of Gal11 protein, called Gal11^P, have been shown to interact in *E. coli* cells.¹ These polypeptides are expressed by the pBT-LGF2 and the pTRG-Gal11^P control plasmids respectively. When these two plasmids are used to cotransform the BacterioMatch II two-hybrid system reporter strain competent cells, a positive interaction is indicated by the growth of colonies on M9⁺ His-dropout medium containing 5 mM 3-AT.

RECOMBINANT pBT AND pTRG PLASMID CONSTRUCTION AND VALIDATION

Cloning into the pBT and pTRG Plasmids

Bait and Target Insert Preparation

DNA is prepared for insertion into the pBT or pTRG plasmids either by restriction digestion or PCR amplification. DNA encoding the bait protein must be inserted so that the bait protein is expressed in the same reading frame as the λ cI protein. DNA encoding the target protein must be inserted in frame with the RNAP α protein. In the MCS of the pBT plasmid, the *Not* I, *Eco*R I, *Sma* I, *Bam*H I, *Xho* I, and *Bgl* II, sites are unique (see Figure 2). In the MCS of the pTRG plasmid, the *Bam*H I, *Not* I, *Eco*R I, *Xho* I, and *Spe* I sites are unique (see Figure 3).

If the insert DNA is phosphorylated, Stratagene suggests dephosphorylating and agarose-gel isolating the digested pBT and pTRG plasmid DNA prior to setting up the ligation reaction. If more than one restriction enzyme is used, the background can be reduced further by selective precipitation using ammonium acetate, eliminating the small fragment from the MCS that results from digestion.

Constructing cDNA Libraries in the pTRG Plasmid

For two-hybrid screening of a pTRG plasmid cDNA library, Stratagene offers BacterioMatch II two-hybrid system library construction kits that include either chemically-competent reporter cells (Catalog #200412) or electrocompetent reporter cells (Catalog #200414). In addition to a complete set of BacterioMatch II vectors, control plasmids and reporter cells, these kits provide the reagents required to prepare a unidirectional pTRG cDNA plasmid library. The kit features Stratagene's cDNA Synthesis Kit (also available separately, Catalog # 200401), with optimized reagents for converting mRNA to cDNA inserts that are suitable for unidirectional insertion into the pTRG plasmid.

A variety of BacterioMatch II premade libraries, prepared from mRNA originating from a number of different organisms and cell types, are also available from Stratagene. For a list of premade pTRG plasmid cDNA libraries, see <http://www.stratagene.com/index.asp?catID=75>.

For complete library construction and library screening protocols, see the appropriate BacterioMatch II library construction kit manual, available online at <http://www.stratagene.com/manuals/index.asp>.

Ligation of DNA Inserts into the pBT and pTRG Plasmids

1. Digest 2 µg of both the pBT plasmid and the pTRG plasmid separately in 50-µl reaction volumes.

Optional *If appropriate, treat the digested plasmid DNA with alkaline phosphatase (following the manufacturer's instructions) and isolate using agarose gel electrophoresis.*

2. Extract each digestion reaction with an equal volume of phenol-chloroform until a clear interface is obtained.
3. Repeat the extraction with an equal volume of chloroform for each DNA sample.
4. Add an equal volume of 4 M ammonium acetate to the aqueous phase in each tube.
5. To each of the tubes, add 2 volumes of 100% (v/v) ethanol at room temperature. Immediately spin in a microcentrifuge at room temperature to precipitate the plasmid DNA.
6. Wash the pellets twice with 70% (v/v) ethanol and then dry the pellets.
7. Resuspend the pellet in each tube in the appropriate volume of TE buffer (see *Preparation of Media and Reagents*) such that the concentration of the plasmid DNA is the same as the concentration of the insert DNA (~0.1 µg/µl).
8. Verify that the plasmids have been digested completely using agarose gel electrophoresis.
9. Calculate the amount of insert required (see equation below), and prepare ligation reactions according to the table below.
10. Ligate overnight at 12°C.

For ligation, the ideal ratio of insert-to-vector DNA is variable; however, 2:1 or 3:1 ratios are recommended (insert-to-vector molar ratio). The amount of insert DNA required for a 1:1 ratio is calculated as follows:

$$X \text{ ng of insert} = \frac{(\text{number of base pairs of insert}) (100 \text{ ng of vector})}{Y \text{ bp of vector (either pBT or pTRG)}}$$

Stratagene suggests the following ligation protocol which includes three control ligations. **Set up this ligation panel for each pBT and pTRG construct.**

Ligation Reaction Components	Experimental	Control		
	1	2 ^a	3 ^b	4 ^c
Prepared plasmid DNA (0.1 µg/µl)	1.0 µl	1.0 µl	1.0 µl	0 µl
Prepared insert (0.1 µg/µl)	Y µl to make a 2:1 or a 3:1 ratio	0 µl	0 µl	1.0 µl
10 mM rATP (pH 7.0)	1.0 µl	1.0 µl	1.0 µl	1.0 µl
10× ligase buffer [§]	1.0 µl	1.0 µl	1.0 µl	1.0 µl
T4 DNA ligase (4 U/µl)	0.5 µl	0.5 µl	0 µl	0.5 µl
Double-distilled water (to 10 µl)	X µl	X µl	X µl	X µl

^a Control 2 is used to test the effectiveness of the digestion and alkaline phosphatase treatment (if performed) of the plasmid.

^b Control 3 is used to test whether the plasmid was digested completely and whether residual uncut plasmid remains.

^c Control 4 is used to test whether the insert alone is contaminated with the plasmid DNA.

Transformation of the Ligation Reactions

Note *In this section, bacteria are grown on rich medium, using a 30°C incubation temperature.*

1. Use the ligation reactions to transform the XL1-Blue MRF⁺ Kan strain (provided with the kit as a glycerol stock). Protocols for preparing and transforming *E. coli* competent cells can be found in reference 4. **It is important to use a host strain (such as XL1-Blue MRF⁺ Kan) that contains the *lacI^q* gene in order to reduce the expression of potentially toxic bait and target proteins.** This limits the selection against individual bait or target proteins. Also, the strain must be chloramphenicol- and tetracycline sensitive to allow selection for pBT and pTRG transformants, respectively.
2. Plate the recombinant pBT transformants on LB-chloramphenicol agar[§] plates.
3. Plate the recombinant pTRG transformants on LB-tetracycline agar[§] plates.
4. Incubate the plates at 30°C for approximately 24 hours. Colonies will be tiny and may require additional growth at 30°C. Incubation at 30°C is necessary to minimize potential toxicity.

[§] See *Preparation of Media and Reagents*.

Verifying the Insert in the pBT and the pTRG Plasmids

To verify insert cloning, prepare miniprep DNA from isolated colonies. Use this DNA to perform restriction analysis, DNA sequence analysis, or PCR analysis (using the appropriate primers shown in the table below). Identify recombinants containing the DNA insert in frame with the λ cI protein (in the case of recombinant pBT) or with the RNAP α protein (in the case of recombinant pTRG).

Primers for use with pBT and pTRG

Primer	Sequence
pBT forward primer	5'- TCC GTT GTG GGG AAA GTT ATC - 3'
pBT reverse primer	5'- GGG TAG CCA GCA GCA TCC - 3'
pTRG forward primer	5'- TGG CTG AAC AAC TGG AAG CT- 3'
pTRG reverse primer	5'- ATT CGT CGC CCG CCA TAA - 3'

Verifying Expression

Note *In this section, bacterial cell growth is performed either on rich medium using a 30°C incubation temperature or in minimal medium using a 37°C incubation temperature.*

It is often possible to detect insert expression using Western blot analysis, although protein expression from both recombinant pBT and pTRG is expected to be low due to the low copy number of both of these plasmids.

Stratagene offers the lambda cI antibody (Catalog #240110) as a tool for verifying the expression of the recombinant bait protein.

Note *Perform expression analysis in the BacterioMatch II two-hybrid reporter strain, as this is the strain that will ultimately be used to detect the interaction of target and bait. Do not substitute another strain for protein expression experiments.*

For expression analysis, transform BacterioMatch II validation reporter competent cells (Catalog #200192) side-by-side with (1) non-recombinant pBT and pTRG in separate transformation reactions and (2) plasmid DNA purified from sequence-verified pBT and pTRG recombinants in separate transformation reactions. Follow the transformation protocol supplied with the reporter competent cells, with the plating modifications given below. General guidelines for expression analysis follow.

Note *When preparing pBT or pTRG plasmid DNA for transformation into the reporter strain, use protocol modifications for low-copy-number plasmids (i.e. grow a larger culture volume and implement any specific protocol modifications for low-copy number plasmids suggested by the purification kit manufacturer). The pBT plasmid is present at about 5–10 copies per cell and the pTRG plasmid is present at about 20–30 copies per cell.*

- ◆ Plate the reporter strain pBT transformants on LB-chloramphenicol agar^s plates and the pTRG transformants on LB-tetracycline agar^s plates. Incubate the plates at 30°C for approximately 24 hours.

- ◆ Grow individual pBT-containing colonies in 2 ml of M9⁺ His-dropout broth[§] containing 25 µg/ml chloramphenicol at 37°C overnight. Use these cultures to inoculate fresh 2-ml cultures of M9⁺ His-dropout broth containing 25 µg/ml chloramphenicol and 10 µM IPTG, to induce protein expression. (Optimizing the IPTG concentration in the 10–100 µM range may be beneficial for some fusion proteins.) Grow the cells at 37°C until the OD₆₀₀ is 0.5–0.6.
- ◆ Grow individual pTRG-containing colonies in 2 ml of M9⁺ His-dropout broth containing 12.5 µg/ml tetracycline at 37°C overnight. Use these cultures to inoculate fresh 2-ml cultures of M9⁺ His-dropout broth containing 12.5 µg/ml tetracycline and 10 µM IPTG, to induce protein expression. (Optimizing the IPTG concentration in the 10–100 µM range may be beneficial for some fusion proteins.) Grow the cells at 37°C until the OD₆₀₀ is 0.5–0.6.
- ◆ Mix 20 µl of each induced culture with 20 µl of 2× SDS gel sample buffer[§]. Boil the samples for 5 minutes, and resolve using SDS-PAGE. Expression of the recombinant proteins may be verified by Western blot analysis with an antibody that immunoreacts with either the bait or target protein.

Notes *When comparing protein expression in different strains (different reporter strain transformants) by Western analysis, it is important to grow all of the bacterial cultures to a similar density before preparing the cell lysates.*

Failure to detect expression of the bait or target fusion protein with a given antibody need not preclude use of the fusion protein-encoding vector. Some proteins may be expressed at levels that are too low for detection by Western blotting with a particular antibody but be expressed at levels sufficient for transcriptional activation of the reporter genes. Verify that the bait or target protein is in frame with the fusion protein before proceeding. See Troubleshooting for additional suggestions.

[§] See *Preparation of Media and Reagents*.

Preparation of Recombinant pBT and pTRG DNA

Note *In this section, bacteria should be grown in **rich medium**, using a **30°C incubation temperature**.*

Following sequence validation and, optionally, confirmation of the expression of the protein of interest, purify the recombinant pBT and pTRG plasmids in sufficient quantity for the subsequent experiments. Grow cultures of XL1-Blue MRF⁺ Kan cells harboring the verified plasmids, and then use standard alkaline lysis protocols⁴ or commercially available plasmid DNA purification columns to prepare the DNA. Both methods usually produce plasmid DNA of sufficient quality for subsequent experiments.

Note *The pBT plasmid and the pTRG plasmid are maintained at about 5–10 and 20–30 copies per cell, respectively. When preparing pBT or pTRG plasmid DNA, use protocol modifications for low-copy-number plasmids (i.e. grow a larger culture volume and implement any protocol modifications for low-copy number plasmids suggested by the purification kit manufacturer).*

TESTING SELF-ACTIVATION BY RECOMBINANT pBT OR pTRG

Prior to performing two-hybrid interaction assays using the recombinant bait and target plasmids (pBT and pTRG, each containing their respective genes of interest) it is prudent to determine whether a specific recombinant is suitable for detecting protein-protein interactions in the BacterioMatch II two-hybrid system. Cotransformation of the BacterioMatch II reporter strain using recombinant pBT plasmid (containing the bait gene of interest) and the empty pTRG vector should not produce a significant number of colonies on Selective Screening Medium (5 mM 3-AT). Likewise, cotransformation of the reporter strain using recombinant pTRG plasmid (containing the target gene of interest) and the empty pBT vector should not produce a significant number of colonies on this medium. If significant growth is observed in either of the cotransformation experiments, the recombinant is not suitable for use in the BacterioMatch II two-hybrid system without selective alteration of certain residues or regions.

Notes *Use BacterioMatch II validation reporter competent cells (Stratagene Catalog #200192) to complete this set of control experiments. Follow the Cotransformation Protocol on the following pages. Do not substitute another host strain for this control experiment.*

*In this section, bacteria are grown on **minimal medium** using a **37°C incubation temperature**.*

TABLE II

Cotransformations Carried Out in this Experiment

Plasmids	Plates	Purpose	Expected Results and Guidelines for Interpretation
recombinant pBT + pTRG (empty vector) (50 ng each)	Nonselective Screening Medium (no 3-AT) [§]	Measures number of cotransformants	10 ² –10 ³ cfu per 200 µl of 1:100 dilution plated (efficiency varies for different bait plasmids)
	Selective Screening Medium (5 mM 3-AT) [§]	Tests whether the bait protein is able, on its own, to activate the reporter cassette	<p>≤0.1% of total cotransformants verifies absence of bait self-activation and suitability of bait both for assays with individual pTRG target plasmids and for library screens</p> <p>0.1–1% of total cotransformants verifies suitability of bait for assays with individual pTRG target plasmids; troubleshooting required prior to use in library screens</p> <p>>1% of total cotransformants indicates bait self-activation or media problem; check pBT +pTRG-Gal11^P negative control cotransformation results to assess media quality</p>
pBT (empty vector) + recombinant pTRG (50 ng each)	Nonselective Screening Medium (no 3-AT)	Measures number of cotransformants	10 ² –10 ³ cfu per 200 µl of 1:100 dilution plated (efficiency varies for different target plasmids)
	Selective Screening Medium (5 mM 3-AT)	Tests whether a given target protein is able, on its own, to activate the reporter cassette	<p>≤1% of total cotransformants verifies suitability of target for assays with individual pBT bait plasmids</p> <p>>1% of total cotransformants indicates self-activation or media problem. Check negative control cotransformation results to assess media quality</p>
pBT (empty vector) + pTRG-Gal11 ^P (50 ng each)	Nonselective Screening Medium (no 3-AT)	Measures number of cotransformants	~10 ³ cfu per 200 µl of 1:100 dilution plated
	Selective Screening Medium (5 mM 3-AT)	Negative control (non-interacting pair) used to verify media quality	<p>≤0.1% of total cotransformants obtained is expected result; verifies that 3-AT selection is working properly</p> <p>>0.1% of total cotransformants indicates failure of 3-AT selection. Troubleshoot media preparation and repeat experiment</p>

[§] See *Preparation of Media and Reagents*.

Cotransformation Protocol

Note *It is important that the agar plates used in this transformation do not have excessive surface moisture, which can lead to an increase in the appearance of false positives. To achieve proper moisture levels, place the plates in a 37°C incubator with lids ajar for 30 minutes to 1 hour, until the moisture on the lid and agar surface has just evaporated. Avoid overdrying the plates, which results in delayed colony formation.*

If desired, test the transformation efficiency of the reporter strain competent cells using the pUC18 plasmid DNA provided with the competent cells in parallel with the cotransformations performed in this experiment. Follow the transformation protocol provided with the reporter strain competent cells for the pUC18 efficiency verification. Expect an efficiency of $\geq 1 \times 10^7$ cfu/ μ g pUC18 DNA.

1. Pre-warm SOC medium[§] to 42°C.
2. Thaw the BacterioMatch II validation reporter competent cells on ice. Gently mix the cells by tapping the tube. Once thawed, aliquot 100 μ l of cells into the appropriate number of prechilled 14-ml BD Falcon tubes. Prepare one aliquot for each pBT/pTRG test pair. **It is critical to use 14-ml BD Falcon polypropylene round-bottom tubes for this procedure.**
3. Add 1.7 μ l of the β -ME provided with the reporter strain competent cells to each 100- μ l aliquot of cells. Swirl the tube contents gently. **Do not substitute with another preparation of β -ME.**
4. Incubate the tubes on ice for 10 minutes, swirling every 2 minutes.
5. To each tube, add 50 ng each of the appropriate pBT bait vector plus pTRG target vector. Swirl the tubes gently.
6. Incubate the tubes on ice for 30 minutes, swirling gently after the first 15 minutes of incubation.
7. Swirl each tube gently, and then heat-pulse the tubes in a 42°C water bath for 35 seconds. **The duration and temperature of the heat pulse is critical for obtaining the highest efficiencies.**
8. Incubate the tubes on ice for 2 minutes.
9. Add 0.9 ml of pre-heated SOC medium to each tube.

[§] See *Preparation of Media and Reagents*.

10. Incubate the tubes at 37 °C with shaking at 225 rpm for 90 minutes.

Note *Ensure that the shaker speed does not exceed 250 rpm.*

11. Spin down the cells in a tabletop centrifuge at 2000 × g for 10 minutes. This step may be completed either at room temperature or at 4°C.
12. Aspirate the supernatant, taking care to avoid disturbing the pellet. To remove the rich medium, wash the cells once by resuspending the cells in 1 ml of room temperature M9⁺ His-dropout broth.
13. Collect the cells as described in step 11, gently aspirate the supernatant, and then resuspend the cells in a fresh 1 ml of M9⁺ His-dropout broth.
14. Incubate cells at 37°C with shaking at 225 rpm for 2 hours. This allows the cells to adapt to growth in minimal medium prior to plating.

Note *Ensure that the shaker speed does not exceed 250 rpm.*

15. Plate each of the cotransformation mixtures on the appropriate media. For the cotransformation experiments listed in Table II, proceed to step 16. For other experiments that employ the *Cotransformation Protocol*, proceed to the relevant section of the manual for plating instructions.

Note *It is important to plate the recommended volume and to spread the mixture evenly on the plates as this promotes even colony distribution. Using sterile glass beads to spread the cells on the plates may facilitate even distribution.*

Plating and Analysis of the Recombinant pBT or pTRG Self-Activation Tests

16. For each of the cotransformations listed in Table II, plate 200 µl of cells from the adapted outgrowth culture on Selective Screening Medium (5 mM 3-AT) plates.
17. For plating on Nonselective Screening Medium (no 3-AT), remove 100 µl of cells from the adapted outgrowth culture. Use this aliquot to prepare a 1:100 dilution of the culture in M9⁺ His-dropout broth. Plate 20-µl and 200-µl aliquots of the diluted cell suspension on Nonselective Screening Medium (no 3-AT) plates.
18. Incubate the plates at 37°C for 24 hours. If colonies are not apparent, transfer the plates to room temperature and continue to incubate the plates in a dark location (to preserve the tetracycline) for an additional 16 hours. This secondary incubation may allow the growth of cells containing toxic proteins or weak interactors.

19. Count the cfu obtained on the Nonselective Screening Medium (no 3-AT) plate that contains the appropriate number of colonies for counting (optimally 10–1000 cfu). This value is used to determine the number of cotransformants obtained.

Significant numbers of colonies should be observed, indicating that the reporter strain competent cells were successfully cotransformed with the plasmid pair. The number of colonies obtained is variable for different plasmid pairs.

20. Count the cfu obtained on the Selective Screening Medium (5 mM 3-AT) plate.

Note *It is normal for reporter strain transformants of various sizes to appear on both Nonselective- and Selective Screening Medium plates.*

21. Calculate the percentage of cotransformants that are able to grow on Selective Screening Medium (5 mM 3-AT) for each plasmid pair. When performing this calculation, take into account the 100-fold dilution, as well as the different amounts of the dilution plated. It is generally useful to standardize the calculations to the cfu expected from 200 μ l of undiluted cell suspension, since this single amount was used for plating on selective medium. See below for two sample calculations from two theoretical experiments.

Calculation 1: Number of cotransformants in 200 μ l of culture:

cfu obtained on Nonselective Screening Medium (no 3-AT) plate	Dilution factor	Plating volume factor (relative to 200 μ l)	Adjusted cotransformants from 200 μ l culture
50 cfu (20 μ l plated)	$\times 100$	$\times 10$	50,000 cfu
100 cfu (200 μ l plated)	$\times 100$	$\times 1$	10,000 cfu

Calculation 2: Percent of cotransformants able to grow on 5 mM 3-AT:

Adjusted cotransformants (200 μ l culture)	cfu obtained on Selective Screening Medium (5 mM 3-AT) (200 μ l culture)	Percent of cotransformants able to grow on 5 mM 3-AT
50,000 cfu	20 cfu	0.04%
10,000 cfu	5 cfu	0.05%

22. Compare the results of these calculations to the guidelines presented in Table II, to assess the suitability of the bait and target vectors for BacterioMatch II two-hybrid assay experiments and library screens.

If either the bait or target protein is found to be unsuitable for BacterioMatch II two-hybrid system experiments, it may be possible to identify the regions/residues of the protein responsible for the non-specific interaction. Alteration of these residues may be used to create a suitable hybrid bait or target protein.

PERFORMING PROTEIN-PROTEIN INTERACTION ASSAYS

To assay for a two-hybrid interaction between a specific bait and target plasmid pair, cotransform the BacterioMatch II validation reporter competent cells with the plasmid pairs listed in Table III. For these experiments, complete steps 1–14 of the protocol in *Cotransformation Protocol*, and then proceed to the *Protein-Protein Interaction Assay Plating and Analysis* section, below.

Notes *The protein-protein interaction assays must be performed in the BacterioMatch II two-hybrid reporter strain, containing the reporter gene cassette. Do not substitute another strain.*

For two-hybrid assays, bacterial cell growth is performed on minimal medium using a 37°C incubation temperature. Any follow-up steps that require growth on rich medium should employ a 30°C incubation temperature.

TABLE III

Cotransformations for Detecting Protein-Protein Interactions

Plasmids	Purpose	Expected Results
pBT-LGF2 + pTRG-Gal11 ^P (50 ng each)	Positive control demonstrating the behavior of a robust interaction pair	Significant growth (10 ⁴ –10 ⁵ cfu) should be observed on Selective Screening Medium (5 mM 3-AT)
recombinant pBT + pTRG (empty vector) (50 ng each)	Negative control indicative of background growth due to activation by the recombinant bait alone or to reporter strain background growth	Few or no colonies should be observed on Selective Screening Medium (5 mM 3-AT)
pBT (empty vector) + recombinant pTRG (50 ng each)	Negative control indicative of background growth due to activation by the recombinant target alone or to reporter strain background growth	Few or no colonies should be observed on Selective Screening Medium (5 mM 3-AT)
recombinant pTRG + recombinant pBT (50 ng each)	Tests for interaction between the bait and target proteins	Growth on Selective Screening Medium (5 mM 3-AT) indicates a putative positive interaction pair

Protein-Protein Interaction Assay Plating and Analysis

1. After completing steps 1–14 of the *Cotransformation Protocol*, plate 100- μ l aliquots of each of the cotransformation mixtures on both Nonselective Screening Medium (no 3-AT) and on Selective Screening Medium (5 mM 3-AT).

Note *It is important to plate the recommended volume and to spread the mixture evenly on the plates as this promotes even colony distribution. Using sterile glass beads to spread the cells on the plates may facilitate even distribution.*

2. Incubate the plates at 37°C for 24 hours. If colonies are not apparent, transfer the plates to room temperature and continue to incubate the plates in a dark location (to preserve the tetracycline) for an additional 16 hours. This secondary incubation may allow the growth of cells containing toxic proteins or weak interactors.
3. Score the plates. Interactions between a pair of hybrid proteins, including the Gal4 + Gal11^P hybrid protein pair, are indicated by growth of colonies on Selective Screening Medium (5 mM 3-AT). The strength of the interaction typically correlates with the ratio of colonies obtained on the selective plates compared to on the nonselective plates.

Notes *Reporter strain cotransformants of various sizes may appear on the plates.*

*The interaction of the Gal4-Gal11^P hybrid pair is robust and is readily detectable in the BacterioMatch II two-hybrid system. **It is important to note that cells harboring weaker interactors are expected to grow more slowly, requiring longer incubation times for colony development.***

Certain bait-target interactions may require cofactors, such as metal ions, that are not present in the minimal media at sufficient levels to allow detection of the interaction. It may be possible to supplement the media with the relevant cofactor(s) in order to increase transcriptional activation to detectable levels.

For a given plasmid pair, if colonies are observed on Nonselective Screening Medium (no 3-AT) but not on Selective Screening Medium (5 mM 3-AT), cotransformation was achieved but the protein pair does not demonstrate an interaction that is detectable by this two-hybrid assay.

If colonies are not observed on Nonselective Screening Medium (no 3-AT) or on Selective Screening Medium (5 mM 3-AT), then cotransformation was not successful. Perform troubleshooting to determine whether both plasmids are of high quality and that the correct quantity was added. Perform transformation of the reporter strain with the pUC18 control plasmid to verify the transformation efficiency of the reporter strain, using the transformation protocol provided with the reporter strain competent cells. See *Troubleshooting* for additional suggestions.

Preserving Putative Positive Colonies

After identifying putative positive colonies on the Selective Screening Medium (5 mM 3-AT) plates, patch several colonies for each interaction pair onto LB-Tetracycline/Chloramphenicol agar plates.[§] These patched plates serve as the primary source of plasmid DNA for the putative positive colonies in subsequent analyses.

Verification of Positives using the Streptomycin Resistance Reporter

Activation of the second reporter gene, *aadA*, encoding streptomycin resistance, may be used as a verification of the interaction between the bait and target proteins. To assay for transcriptional activation of the *aadA* reporter, patch cells from a putative positive colony from a Selective Screening Medium (5mM 3-AT) plate onto a Dual Selective Screening Medium (5 mM 3-AT + Strep) plate.[§] As indicators of growth expected on the Dual Selective Screening Medium, include patches of positive control cotransformants (pBT-LGF2 + pTRG-Gal11^P, taken from a Selective Screening Medium plate) and negative control cotransformants (e.g. recombinant pBT + pTRG empty vector, taken from a Nonselective Screening Medium plate).

Notes *Patching should be performed by lightly brushing cells from the toothpick onto the surface of the agar plate. Avoid piercing the surface of the agar with the toothpick during patching.*

Use fresh colonies from the Selective Screening Medium (5mM 3-AT) plates for patching on Dual Selective Screening Medium (5 mM 3-AT + Strep) plates. Colonies should be maintained on the selective medium for less than three days prior to patching on the dual selective medium.

*Do not use cells maintained on rich medium for the *aadA* reporter gene activation assay. If it is necessary to use cells kept on rich medium, adapt the cells first by growth on Nonselective Screening Medium and then by growth on Selective Screening Medium prior to patching on the Dual Selective Screening Medium plates. Switching cells from rich medium to minimal medium results in a growth delay, and the cells must be adapted to growth on minimal medium prior to testing the activation of the *aadA* reporter gene.*

[§] See *Preparation of Media and Reagents*.

TROUBLESHOOTING

Observation	Suggestion
The yield of bait plasmid is low	The copy number of the bait plasmid is low in the host strain. Increase the volume of culture or apply a standard spectinomycin treatment to the culture (see reference 4). Ensure that the host strain contains the <i>lacI^a</i> gene and that growth is performed at 30°C.
The yield of target plasmid is low	The copy number of the target plasmid is low in the host strain. Increase the volume of culture or apply a standard chloramphenicol treatment to the culture (see reference 4). Ensure that the host strain contains the <i>lacI^a</i> gene and that growth is performed at 30°C.
During construction of bait or target plasmids, transformation results in few colonies or in small, slow-growing colonies	Use the XL1-Blue MRF ⁺ Kan strain provided with the kit for bait and target vector propagation to ensure the presence of the <i>lacI^a</i> gene. Ensure that growth is performed at 30°C during plasmid production. Both of these measures decrease any potential toxicity of individual bait or target proteins to the host strain.
	Ensure that an appropriate protocol was used to transform the XL1-Blue MRF ⁺ Kan strain. The several transformation protocols provided in this manual differ and were optimized for a specific host strain and a specific cell volume.
	In some cases, diluting the ligation reaction 1:5 prior to the transformation results in higher transformation efficiencies.
Efficiency of bait and target plasmid cotransformation into the reporter strain is low	Evaluate whether the low efficiency may be attributed to one of the two plasmids. When repeating the cotransformation, plate aliquots of the cotransformation reaction on both LB-chloramphenicol (to determine the efficiency for the bait plasmid) and LB-tetracycline (to determine the efficiency for the target plasmid). If one of the plasmids demonstrates a low efficiency, reassess the concentration and purity of the corresponding DNA preparation. In addition, plate the cotransformation reaction on LB-Tetracycline/Chloramphenicol agar plates to assess the potential of the reporter strain to be cotransformed with the specific bait and target plasmid pair.
	Check the concentration and purity of each plasmid preparation by spectrophotometry and agarose gel electrophoresis. If either plasmid is damaged or impure, perform additional rounds of plasmid purification and concentration or prepare new plasmid DNA stocks.
	Ensure that the reporter strain competent cells are stored and handled properly. Competent cells must be stored at the bottom of a -80°C freezer and should be handled as described in the cotransformation protocol.
The bait and/or target protein is not detected in Western blot analysis	Verify that the insert DNA is in the same reading frame as the fusion protein.
	If the antibody does not have a sufficiently high affinity for the protein, the protein may be expressed but may not be detectable.
	Try growing the reporter strain transformed with the bait or target plasmid in rich medium prior to Western blot analysis. Growth in rich medium may increase the abundance of the protein to detectable levels.
	Optimize the induction conditions including IPTG concentration (10–100 μM) and induction time.
	Increase the amount of cell extract used in the SDS-PAGE/Western blot procedure.

The positive control plasmids do not give the expected results in the reporter strain	Verify that the correct control plasmid pairs are used.
	Ensure that 14-ml BD Falcon polypropylene round-bottom tubes are used for the transformation procedure. The protocol is optimized using these tubes.
	Use sterile technique when preparing and transforming the BacterioMatch II reporter strain competent cells to avoid contamination of the culture with different <i>E. coli</i> strains.
Unexpectedly large number of colonies on Selective Screening Medium (5 mM 3-AT) plates containing negative control	3-AT concentration on the plate is too low. Ensure that the correct amount of 3-AT was added to the plates (see <i>Preparation of Media and Reagents</i>).
	Ensure that the plates were prepared exactly according to the recipe in <i>Preparation of Media and Reagents</i> . In particular, ensure that the medium includes all of the selective agents (chloramphenicol, tetracycline, and 3-AT), that the medium is allowed to cool to 50°C prior to 3-AT addition, and that the plates were stored properly (at 4°C, covered in foil, for up to one month).
	Ensure that the plates are incubated for approximately 24 hours at 37°C.
Colonies on Selective Screening Medium (5 mM 3-AT) plates grow slowly	Weak interaction pairs generate slow-growing colonies.
	Ensure that the plates are not too dry.
No colonies on Selective Screening Medium (5 mM 3-AT) plates following cotransformation with a known interacting protein pair	Protein is not expressed; ensure the gene of interest is ligated in frame with the fusion protein; ensure that the DNA used in the cotransformation is of high quality.
	The bait or target fusion protein may be toxic to the reporter strain host cells. Examine the growth properties of the cotransformants on Nonselective Screening Medium (no 3-AT). If growth is weak on the nonselective medium, try reducing the amount of IPTG inducer added to the medium in order to reduce the fusion gene expression levels. Test whether fusion proteins containing truncated versions of the proteins of interest are less toxic to the reporter strain host.
	The bait or target fusion protein may be unstable or improperly folded. Construct fusion proteins containing different domains of the proteins of interest.
	The interaction may not be robust enough to be detected on 5mM 3-AT. Reduce the amount of 3-AT and re-test the plasmid pair.
	The interaction may require certain protein modifications that are not made in <i>E. coli</i> cells.
	The interaction may require a cofactor, such as a metal ion, that is not present in the media in sufficient quantity. Test supplementation of the screening media recipes with any suspected small-molecule cofactors.
Colonies obtained from Selective Screening Medium (5 mM 3-AT) plates do not grow on Dual Selective Screening Medium (5 mM 3-AT + Strep) plates	The colony may be a false positive. Resistance to 5 mM 3-AT may not be dependent on two-hybrid transcriptional activation of the <i>HIS3</i> gene.
	The interaction may not be robust enough to be detected on 12.5 µg/ml streptomycin. Reduce the amount of streptomycin to 10 µg/ml and re-test the plasmid pair.

PREPARATION OF MEDIA AND REAGENTS

Minimal Media for BacterioMatch® II Two-Hybrid Assays

Notes 3-AT is heat-labile and will be inactivated if added to medium at a temperature >55 °C.

All media components should be added in the order listed.

M9 salts are commercially available from Qbiogene (Catalog #3037-032) or may be prepared using standard bacteriological media recipes. The components in 1 liter water should be: Na₂HPO₄: 67.8 g, KH₂PO₄: 30 g, NaCl: 5 g, NH₄Cl: 10 g. Qbiogene M9 salts prepared as directed will yield a 5× solution, and should be doubled for use with this kit.

M9 Media Additives (sufficient for 500 ml of M9+ medium)

Prepare Solution I and Solution II separately, by mixing the components listed in order. Add Solution II to Solution I and then mix well. This mixture may then be added to the different variants of M9+ media.

Solution I:

10 ml of 20% glucose (filter-sterilized)
5 ml of 20 mM adenine HCl (filter-sterilized)
50 ml of 10× His dropout amino acid supplement (BD/Clontech, Cat. #630415)

Note: Sterilize the 10× His dropout supplement by autoclaving at 121 °C for 15 minutes prior to addition to Solution I. Do not exceed 15 minutes.

Solution II:

0.5 ml of 1 M MgSO₄
0.5 ml of 1 M Thiamine HCl
0.5 ml of 10 mM ZnSO₄
0.5 ml of 100 mM CaCl₂
0.5 ml of 50 mM IPTG

Note: Filter-sterilize the thiamine-HCl and the IPTG prior to use. Sterilize the remaining components of Solution II using the method of choice.

M9+ His-dropout Broth (500 ml)

In a 500-ml flask, combine and mix well:

380 ml of sterile, deionized H₂O

50 ml of 10× M9 salts

To this mixture, add one preparation of M9 Media Additives (67.5 ml, see separate recipe)

Store at 4 °C for up to one month

Bring to room temperature prior to use

Nonselective Screening Medium (500 ml)

In a 500-ml flask, combine:

380 ml deionized H₂O

7.5g Bacto agar

Autoclave at 121 °C for 20 minutes

Cool the agar to 70 °C, and then add 50 ml of 10× M9 salts

Combine one preparation of M9 Media Additives (67.5 ml, see separate recipe) with the following agents:

0.5 ml of 25 mg/ml chloramphenicol

0.5 ml of 12.5 mg/ml tetracycline

When the agar mixture has cooled to 50 °C, **immediately** add the M9 Media Additives/agent mixture to the agar

Pour into petri dishes (~25 ml/100-mm plate)

After the agar solidifies, wrap the plates in aluminum foil

Store plates at 4 °C for up to 1 month

Note: It is important to add the M9 Media Additives/agent mixture to the agar at 50 °C. Cooling the agar to lower temperatures prior to addition may interfere with the even distribution of the selective agents in the agar mixture.

Minimal Media for BacterioMatch® II Two-Hybrid Assays (Continued)

<p>3-AT Stock Solution, 1 M (10 ml) Dissolve 840.8 mg of 3-AT (Sigma, Cat. #A-8056) in 10 ml of DMSO Aliquot into 4 × 2.5 ml tubes Store at –20°C for up to one month</p>	<p>Dual Selective Screening Medium (5 mM 3-AT + Strep)(500 ml) In a 500-ml flask, combine: 7.5g Bacto agar 380 ml H₂O Autoclave at 121°C for 20 minutes Cool the agar to 70 °C, and then add 50 ml of 10× M9 salts Combine one preparation of M9 Media Additives (67.5 ml, see separate recipe) with the following agents: 0.5 ml of 25 mg/ml chloramphenicol 0.5 ml of 12.5 mg/ml tetracycline 2.5 ml of 1 M 3-AT (dissolved in DMSO) 0.5 ml of 12.5 mg/ml streptomycin When the agar mixture has cooled to 50°C, immediately add the M9 Media Additives/agent mixture to the agar Pour into petri dishes (~25 ml/100-mm plate) After the agar solidifies, wrap the plates in aluminum foil Store plates at 4°C for up to 1 month Note: <i>It is important to add the M9 Media Additives/agent mixture to the agar at 50°C. Cooling the agar to lower temperatures prior to addition may interfere with the even distribution of the selective agents in the agar mixture.</i></p>
<p>Selective Screening Medium (5 mM 3-AT) (500 ml) In a 500-ml flask, combine: 7.5g Bacto agar 380 ml H₂O Autoclave at 121°C for 20 minutes Cool the agar to 70°C, and then add 50 ml of 10× M9 salts Combine one preparation of M9 Media Additives (67.5 ml, see separate recipe) with the following agents: 0.5 ml of 25 mg/ml chloramphenicol 0.5 ml of 12.5 mg/ml tetracycline 2.5 ml of 1 M 3-AT (dissolved in DMSO) When the agar mixture has cooled to 50°C, immediately add the M9 Media Additives/agent mixture to the agar Pour into petri dishes (~25 ml/100-mm plate; ~60 ml for 150-mm plate) After the agar solidifies, wrap the plates in aluminum foil Store plates at 4°C for up to 1 month Notes: <i>Stratagene recommends using freshly prepared plates for library screening experiments.</i> <i>It is important to add the M9 Media Additives/agent mixture to the agar at 50°C. Cooling the agar to lower temperatures prior to addition may interfere with the even distribution of the selective agents in the agar mixture.</i></p>	

Rich Media and Other Reagents

<p>LB Agar (per Liter) 10 g of NaCl 10 g of tryptone 5 g of yeast extract 20 g of agar Add deionized H₂O to a final volume of 1 liter Adjust pH to 7.0 with 5 N NaOH Autoclave Pour into petri dishes (~25 ml/100-mm plate; 60 ml for 150-mm plate)</p>	<p>LB–Tetracycline Agar (per Liter) Prepare 1 liter of LB agar Autoclave Cool to 45°C Add 1.25 ml of 10-mg/ml tetracycline (prepared in 50% EtOH) Pour into petri dishes (~25 ml/100-mm plate; 60 ml for 150-mm plate) Store plates in a dark, cool place or cover plates with foil if left out at room temperature for extended time periods as tetracycline is light-sensitive</p>
<p>LB–Chloramphenicol Agar (per Liter) Prepare 1 liter of LB agar Autoclave Cool to 45°C Add 2.5 ml of 10-mg/ml chloramphenicol (prepared in 100% ethanol) Pour into petri dishes (~25 ml/100-mm plate; 60 ml for 150-mm plate)</p>	<p>LB–Chloramphenicol/Tetracycline Agar (per Liter) Prepare 1 liter of LB agar Autoclave Cool to 45°C, and then add: 0.5 ml of 25 mg/ml chloramphenicol 0.5 ml of 12.5 mg/ml tetracycline Pour into petri dishes (~25 ml/100-mm plate; 60 ml for 150-mm plate) Store plates in a dark, cool place or cover with foil if left out at room temperature as tetracycline is light-sensitive</p>
<p>SOB Medium (per Liter) 20.0 g of tryptone 5.0 g of yeast extract 0.5 g of NaCl Add deionized H₂O to a final volume of 1 liter Autoclave Add 10 ml of filter-sterilized 1 M MgCl₂ and 10 ml of filter-sterilized 1 M MgSO₄ prior to use</p>	<p>SOC Medium (per 100 ml) Note <i>This medium should be prepared immediately before use.</i> 2 ml of filter-sterilized 20% (w/v) glucose or 1 ml of filter-sterilized 2 M glucose SOB medium (autoclaved) to a final volume of 100 ml</p>
<p>LB Broth (per Liter) 10 g of NaCl 10 g of tryptone 5 g of yeast extract Add deionized H₂O to a final volume of 1 liter Adjust to pH 7.0 with 5 N NaOH Autoclave</p>	<p>LB–Ampicillin Agar (per Liter) Prepare 1 liter of LB agar Autoclave Cool to 55°C Add 10 ml of 10-mg/ml-filter-sterilized ampicillin Pour into petri dishes (~25 ml/100-mm plate)</p>

Rich Media and Other Reagents (Continued)

10× Ligase Buffer 500 mM Tris-HCl (pH 7.5) 70 mM MgCl ₂ 10 mM dithiothreitol (DTT) Note <i>rATP is added separately in the ligation reaction</i>	TE Buffer 10 mM Tris-HCl (pH 7.5) 1 mM EDTA
2× SDS gel sample buffer 100 mM Tris-HCl (pH 6.5) 4% SDS (electrophoresis grade) 0.2% bromophenol blue 20% glycerol Note <i>Add dithiothreitol to a final concentration in the 2× buffer of 200 mM prior to use. This sample buffer is useful for denaturing, discontinuous acrylamide gel systems only.</i>	

REFERENCES

1. Dove, S. L. and Hochschild, A. (1998) *Genes Dev* 12(5):745-54.
2. Dove, S. L., Joung, J. K. and Hochschild, A. (1997) *Nature* 386(6625):627-30.
3. Joung, J. K., Ramm, E. I. and Pabo, C. O. (2000) *Proc Natl Acad Sci U S A* 97(13):7382-7.
4. Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

ENDNOTES

BacterioMatch® is a registered trademark of Stratagene in the United States.

MSDS INFORMATION

The Material Safety Data Sheet (MSDS) information for Stratagene products is provided on Stratagene's website at <http://www.stratagene.com/MSDS>. Simply enter the catalog number to retrieve any associated MSDS's in a print-ready format. MSDS documents are not included with product shipments.

Annexe C : Article
Genes Chromosomes & Cancer
B.SIMON-KAYSER mai 2005

Molecular Cloning and Characterization of *FBXO47*, a Novel Gene Containing an F-Box Domain, Located in the 17q12 Band Deleted in Papillary Renal Cell Carcinoma

Barbara Simon-Kayser,^{1*} Catherine Scoul,¹ Karine Renaudin,² Pascal Jezequel,³ Olivier Bouchot,⁴ Jérôme Rigaud,⁴ and Stéphane Bezieau¹

¹Laboratoire d'Etude du Polymorphisme de l'ADN, Faculté de Médecine, Nantes, France

²Service d'anatomo-pathologie, Hôtel-Dieu, Centre Hospitalier Universitaire, Nantes, France

³Centre de Recherche de Lutte Contre le Cancer René Gauducheau, Saint Herblain, France

⁴Service d'urologie, Hôtel-Dieu, Centre Hospitalier Universitaire, Nantes, France

Genetic alterations of chromosome arm 17q occur in numerous tumor types, including breast and ovarian tumors, suggesting the presence of a tumor-suppressor gene on the long arm of chromosome 17 that is critical for carcinogenesis. Previous studies have shown an allelic imbalance (70% gain or loss) of 17q in papillary renal cell carcinoma (pRCC). In this study, we analyzed 15 cases of pRCC for loss of heterozygosity with the use of 7 microsatellite markers between 17q11 and 17q23. We identified a minimal deleted region in which the *D17S250* marker (17q12) was deleted in 50% (7 of 14) of informative cases. We isolated the cDNA of a novel gene named *FBXO47*, which is near *D17S250*. Human *FBXO47* is composed of 11 exons and spans approximately 30 kb of genomic DNA. *FBXO47* cDNA consists of 2,269 bp with a 1,359-bp open-reading frame. Of note is that *FBXO47* is preferentially expressed in normal tissue relative to the corresponding tumor tissue, particularly in the kidney, liver, and pancreas and to a lesser extent in the thyroid gland, stomach, and small intestine. The putative protein encoded by this gene is made up of 453 amino acids and belongs to the F-box family, most of whose members, such as SKP2 and FBW7, have been implicated in carcinogenesis. Together, these results indicate that *FBXO47* has a potential role as a tumor-suppressor gene. © 2005 Wiley-Liss, Inc.

INTRODUCTION

Abnormalities in chromosome region 17q12–q21 are a common feature of many cancers. The detection of loss or gain of genomic material may lead to the identification of oncogenes or tumor-suppressor genes. The *HER2* gene (also known as *ERBB2/cERB2/HER2/NEU*), which is in the 17q12 band, is the most frequently amplified (15%–30%) oncogene in breast cancer (Slamon et al., 1987) and is commonly amplified in other forms of cancer (Klapper et al., 2000; Vernimmen et al., 2003; Varis et al., 2004) as well. In breast cancer, the *HER2* amplicon contains other genes in 17q12–q21 that have altered copy numbers (Jarvinen and Liu, 2003). A recent study identified five genes in the immediate vicinity of *HER2*: *PSMB3*, *RPL19*, *PPARBP*, *PPP1R1B*, and *GRB7* (Bertucci et al., 2004). Another gene, *TOP2A*, which is adjacent to the *HER2* oncogene, has been found to be amplified or deleted with equal frequency in breast carcinoma (Jarvinen and Liu, 2003). Similar data were obtained for gastric cancer, with coamplification of several putative oncogenes including *HER2*, *PPP1R1B*, *GRB7*, and *SARD3* (Kato, 2003). On

the other hand, some studies have provided evidence of tumor-suppressor genes in the 17q12–q21 region that are implicated in sporadic cancers. Loss of heterozygosity (LOH) has been documented in epithelial ovarian tumors (Godwin et al., 1994; Garcia et al., 2000), laryngeal carcinoma (Kiaris et al., 1995; Rizo et al., 1998; Sasiadek et al., 2001; Li and Chen, 2004), prostate cancer (Gao et al., 1995), and endocrine pancreatic tumors (Zhao et al., 2001). In two families with Wilms tumor, a locus named *FWT1* was localized within the 17q12–q21 region by linkage analysis (Rahman et al., 2000). Despite these observations, few

Sequence data from this article have been deposited with the GenBank Data Library under Accession No. AY700575.

Supported by: Association Anne de Bretagne Génétique; University Hospital of Nantes (Service Génétique Médicale); Association Ligue contre le Cancer of Loire-Atlantique (to B.S.-K.).

*Correspondence to: Barbara Simon-Kayser, Laboratoire d'Etude du Polymorphisme de l'ADN, Faculté de Médecine, 1 rue Gaston Veil, 44035 Nantes cedex, France.

E-mail: bsimon@sante.univ-nantes.fr

Received 6 August 2004; Accepted 5 December 2004

DOI 10.1002/gcc.20170

Published online 18 February 2005 in

Wiley InterScience (www.interscience.wiley.com).

TABLE 1. Characteristics of Repeat Markers on Chromosome 17 Selected for the Study of LOH

Locus	Map position ^a	Maximum heterozygosity	Type of sequence	GDB database accession number
<i>D17S969</i>	17p12	ND	Tetranucleotide	228553
<i>D17S1532</i>	17q11.2	ND	Trinucleotide	455733
<i>D17S933</i>	17q11.2	0.8240	Dinucleotide	199552
<i>D17S250</i>	17q12	0.8270	Dinucleotide	177030
<i>D17S800</i>	17q21.2	0.7433	Dinucleotide	188300
<i>D17S579</i>	17q21.31	0.87	Dinucleotide	191691
<i>D17S588</i>	17q21.33	0.8510	Dinucleotide	185206
<i>D17S1290</i>	17q23.2	ND	Tetranucleotide	364316

^aPrecise location of the markers on respective chromosome arms. ND, not determined.

tumor-suppressor genes have been discovered in this subregion. Li and Chen (2004) described a putative tumor-suppressor gene, named *LCRG1* (laryngeal carcinoma related gene 1), at 17q21. In addition, previous studies of papillary renal cell carcinoma (pRCC) have shown 70% allelic imbalance (AI) of 17q (Sanders et al., 2002). AI concerns both loss of heterozygosity and selective allelic gain/duplication. Two genes have been identified and implicated in the familial form of renal carcinoma: the *MET* oncogene in HPRC (hereditary papillary renal cell of type 1; Zbar et al., 1995) and the furamate hydratase gene in HLRC (solitary type 2 papillary renal carcinoma; Toro et al., 2003). Nevertheless, so far, no genes implicated in sporadic cases of pRCC in 17q have been identified.

In this study, we investigated LOH in 15 cases of pRCC with the use of 7 microsatellite markers in the 17q11–q23 region. Fifty percent of the informative tumors analyzed showed LOH at the *D17S250* (17q12) locus. We have thus identified and characterized a novel gene that is 30 kb proximal of the *D17S250* marker. The protein encoded by this gene (*FBXO47*) is a putative member of the F-box family, most of whose members function as oncogenes or tumor-suppressor genes (Guardavacaro and Pagano, 2004).

MATERIALS AND METHODS

DNA restriction enzymes were from Takara (Taufkirchen, Germany), PCR was performed using Taq platinum (Invitrogen Life Technologies, Cergy Pontoise, France), and all reagents were from Sigma (Saint Quentin, France), unless otherwise indicated. DNA probes were radiolabeled using a Rediprime DNA labeling system (Amersham-Pharmacia Biotech, Saclay, France). RNA was retrotranscribed using the SuperScript RT-PCR System (Invitrogen Life Technologies, Cergy Pontoise, France). Automated DNA sequencing was carried out using an Applied Biosystems (Les Ulis, France) ABI3100 fluorescence sequencer. All

techniques were carried out according to the relevant manufacturer's recommendations.

Tissue Samples and Genomic DNA Extraction

Ten-micrometer serial sections were prepared from formalin-fixed, paraffin-embedded tumor or the corresponding normal tissue. Areas of representative tumor and normal kidney tissue were outlined on the cover glass with a permanent marker. All sections were reviewed and their type and grade characterized by a single pathologist. The scraped tissue was incubated at 60°C for a minimum of 5 hr in 500 µL of buffer solution (1 M Tris HCl, pH 8; 5 M NaCl; 0.5 M EDTA, pH 8; 10% SDS; and 10 ng/mL proteinase K in water). Twenty microliters of proteinase K was subsequently added, and the tubes were further incubated at 60°C overnight. Finally, DNA was extracted by the phenol-chloroform method, concentrated in microcon tubes, and resuspended in 10:1 TE to a concentration of 20 ng/µL, estimated using Nanodrop technology (Nyxor Biotech, Paris, France).

PCR-LOH Assay

A total of 8 microsatellite markers distributed along chromosome 17 (*D17S969*, *D17S1532*, *D17S588*, *D17S250*, *D17S1290*, *D17S933*, *D17S800*, and *D17S579*) were examined. The primers were synthesized by Eurogentec (Angers, France), and all forward primers were labeled with a fluorochrome (Tables 1 and 2).

Amplifications were performed on 50 ng of DNA with 5 pmol of each primer, 10 µM dNTPs, 0.125 U Platinum Taq polymerase, 1× PCR buffer minus Mg, and 2 mM MgCl₂ (Invitrogen Life Technologies, Cergy Pontoise, France). For *D17S969*, *D17S1532*, *D17S588*, and *D17S1290*, the PCR program consisted of an initial denaturation step of 95°C for 5 min, followed by 30 cycles of 94°C for 30 sec, 55°C for 50 sec, and 72°C for 1 min 30 sec, with a final extension cycle at 72°C for 10 min. For *D17S800*, *D17S933*, *D17S579*, and

TABLE 2. Sense and Antisense Primers of Microsatellite Markers Used in the Study of LOH of Chromosome 17

Locus	Primer sequence ^a	Color	Size of PCR product (pb)
<i>D17S969</i>	ATCTAATCTGTCATTCATTCATCTATCCA-6-Fam/AACTGCAGTGCTGCATCATA	Blue	116-140
<i>D17S1532</i>	CAACGGTAGGCAAGTAAGCT-6-Hex/TGCTGTGTGCAGTTTGATCT	Green	165-185
<i>D17S933</i>	ACTCACTGGGGTCTGG-6-Fam/TGTGGTTTCCTTATAGACTGTAGA	Blue	187-206
<i>D17S250</i>	GGAAGAATCAAATAGACAAT-6-Ned/GCTGGCCATATATATATTTAAACC	Yellow	135-165
<i>D17S800</i>	GGTCTCATCCATCAGGTTTT-6-Hex/ATAGACTGTGTACTGGGCATTGA	Green	168-178
<i>D17S579</i>	AGTCCTGTAGACAAAACCTG-6-Fam/CAGTTTCATACCAAGTTCCT	Blue	111-125
<i>D17S588</i>	CCTGGTCTAGGAAGAGTGTCA-6-Hex/GTGTAAGCATCTGTGTATACTAC	Green	140-160
<i>D17S1290</i>	GCCAACAGAGCAAGACTGTC-6-Fam/GGAAACAGTTAAATGGCCAA	Blue	170-210

^aPCR primers used to amplify microsatellite sequences are listed forward/reverse in a 5' → 3' orientation.

D17S250, the PCR program consisted of an initial denaturation step of 95°C for 5 min, followed by 30 cycles of 94°C for 30 sec, 52°C or 56°C for *D17S250* for 1 min, and 72°C for 1 min 30 sec, with a final extension cycle at 72°C for 20 min.

All PCR amplifications were performed using an ICycler thermocycler (Biorad, Marnes, France). After denaturation at 90°C for 3 min, the PCR products were evaluated with a 3100 ABI Prism Genetic Analyzer (Applied Biosystems), on the basis of automated capillary electrophoresis and automated allele sizing by GeneScan 2.0.2 and Genotyper 2.1 software (Applied Biosystems).

Data Analysis

DNA samples heterozygous for any of the eight loci were considered informative, whereas homozygous samples were deemed noninformative. Positive LOH was defined as a visible absence or a reduction in the intensity of one allele in the tumor compared to the corresponding normal tissue. In cases of reduced LOH, an LOH index value (LIV) was calculated by: $LIV = (\text{height of peak 2 in normal tissue} / \text{height of peak 1 in normal tissue}) / (\text{height of peak 2 in tumor tissue} / \text{height of peak 1 in tumor tissue})$. A value above 2 was defined as loss of the larger allele in tumor tissue, whereas a value below 0.67 meant a loss of the smaller allele in tumor tissue.

In Silico Cloning

The search for putative genes was carried out using four prediction programs available on the Web: Genscan (<http://genes.mit.edu/GENSCAN.html>), GrailExp (<http://grail.Isd.ornl.gov/grailexp>), Fgenes (<http://genomic.sanger.ac.uk/gf/gf.shtml>), and Genmark (<http://www2.ebi.ac.uk/genemark>). The retained genomic sequence ACAC006441 was masked using Repeat Masker (<http://ftp.genome.washington.edu/cgi-bin/RepeatMasker>). The deduced nucleotide sequence was compared to nonredundant

animal sequence databases at the Web sites of the NCBI (National Center for Biotechnology Information) and Human Genome Browser at UCSC by use of several BLAST programs.

The promoter region was analyzed by searching the 5' region flanking exon 1 for putative binding sites of known transcription factors and the TATA box, using the following computational analysis and recognition programs via the Infobiogen servers (<http://www.infobiogen.fr/services/deambulumb/fr/prog2.html>): Fgenes, McPromoter, CorePromoter, Promoter Inspector, TFSEARCH, and PROSCAN.

Potential polyadenylation sites were deduced in masked sequences by combining the results obtained by methods including the GenScan, GraiExp, and Fgenes programs and by running the recognition program PolyAH (<http://genomic.sanger.ac.uk/gf/gf.shtml>).

A search for CpG islands was carried out using the Grail program (http://compbio.ornl.gov/GAT_tmp).

Biocomputing Analysis

Sequence assembly and editing were performed using both the AutoAssembler (Applied Biosystems) and the DNA Strider 1.2 programs. The search for a motif in the deduced amino acid sequence was carried out using a phi BLAST (NCBI), PFSCAN algorithm (<http://www.isrec.isb-sib.ch>; Altschul et al., 1997) and ScanProsite (<http://www.expasy.org/cgi-bin/scanprosite>). Multiple sequence alignments of *FBXO47* and other F-box family members were performed using the Clustal W and SeqVu programs (Thompson et al., 1994). Cellular localization was predicted by the PSORTII program (<http://psort.nibb.ac.jp/form2.html>).

3' and 5' Race

First-strand cDNAs were synthesized from 1 µg of human testis total RNA (Clontech, BDBiosciences, Erembodegem, Belgium) and used as

TABLE 3. PCR Primers Used to Amplify and Sequence *FBX047* Exons, Listed in a 5' → 3' Orientation

Primer	Sequence	PCR product	Annealing temperature (°C)	MgCl ₂ concentration (mM)
1	GAGCGGGCACTACGAGAGC	Exon 1	67	1.6
2	TGTTTCGCTCCCTTCAGTCCAT			
3	TGTACTTGTGTTTTCTGACC	Exon 2	52	2
4	GTGTCTTGTGTAGGCTATTC			
5	GATGATGGTGATAATGATG	Exon 3	54	1.5
6	AAAAACATCATAGAAACAGG			
7	ATTCTGTGCTCTATTGATGT	Exon 4	52	2
8	AGAGTAAAATGAGCATAAGAT			
9	CCATTGTTCCACCATTGTC	Exon 5	55	1.5
10	GGAGTAACCTTCATTTCAGC			
11	TGGTGTGACTTATGTTTTGA	Exon 6	55	1.5
12	AAGTGCTGGGATTTTAGG			
13	TCAGGAAGGTAAGTAAGT	Exon 7	51	2
14	AGTAAAATAATGAACCAGT			
15	AAAGCCAACACTATCTGTGCAT	Exon 8	54	2
16	TCCAACAGGCACATCATT			
17	CAATAACCTGTCTCTGACTG	Exon 9	54	2
18	TCTTTTGTCCACATTTTCAGT			
19	TCAGTATGGTGTCCAGGTATC	Exon 10	55	1.5
20	CCGCTTCCTCTACTTATTGT			
21	AACTGTCATTTTCTTTTCCCC	Exon 11	54	2
22	AGCACTCCTCTAATCATTCCG			

templates for 5' and 3' rapid amplification of cDNA ends (RACE) by use of SMART technology (SMART RACE cDNA Amplification kit, Clontech, BDBiosciences, Erembodegem, Belgium). RACE PCR reactions were primed with gene-specific primers (GSPs) located within the cDNA coding sequence (GSP1: 5'-ACCAAAGCCAAAAAG-ACTACGAGCC-3', designed for 5'RACE; GSP2: 5'-CAGTGGTTCCTCCGTGAGTGGCTTCTA-3', designed for 3'RACE). The reaction consisted of 1 cycle at 95°C for 1 min; 5 cycles at 94°C for 5 sec and 72°C for 3 min; 5 cycles at 94°C for 5 sec, 70°C for 10 sec, and 72°C for 3 min; 30 cycles at 94°C for 5 sec, 68°C for 10 sec, 72°C for 3 min; and 1 cycle at 72°C for 10 min. For both reactions, a nested PCR reaction was performed using a nested GSP (5'-TAGAAGCCACTCACGGGGAACCACTG-3', designed for nested 5'RACE, and 5'-CAGTGGTTCCTCCGTGAGTGGCTTCTA-3', designed for nested 3'RACE). The reaction consisted of 1 cycle at 95°C for 1 min; 20 cycles at 94°C for 5 sec, 68°C for 10 sec, and 72°C for 3 min; and 1 cycle at 70°C for 10 min. PCR products of nested 5' and 3' RACE were subcloned via a pGEMT easy vector (Promega, Carbondières, France) and sequenced automatically by an ABI 3100 (Applied Biosystems).

cDNA Expression Cancer Analysis

FBX047 gene expression was analyzed by a cDNA blot assay using BDTM Cancer Profiling

Array II (Clontech, BDBiosciences). This array consists of a nylon membrane spotted with 154 paired cDNA samples generated from the total RNA of 19 tissue types (amplified using BD SMARTTM Technology). Each pair consists of a tumor sample and adjacent corresponding normal tissue from the same patient. The panels are composed of 19 tissue types: breast, ovary, colon, stomach, lung, kidney, bladder, vulva, prostate, trachea, liver, uterus, cervix, rectum, thyroid gland, testis, skin, small intestine, and pancreas. This array also contains cancer cell line samples and positive and negative controls. The normalization procedure enabled an accurate estimation of the relative abundance of target gene transcripts in different tissues. Hybridization was carried out with ExpressHyb hybridization solution (Clontech, BDBiosciences) in the presence of a ³²P-labeled cDNA probe according to the manufacturer's instructions.

This probe was obtained by PCR amplification of testis cDNA with primers specific for the 3'UTR and encompassing the three last exons of the *FBX047* gene. This array was scanned by PhosphorImager and the intensity signal of each spot was analyzed by IPLab Gel software (BioSystematica, Plymouth, UK). Each spot was quantified by a corrected-background area value.

Mutation Analysis of the Candidate Gene *FBX047*

Fresh tumor samples were available from four individuals with LOH. DNA was extracted by the

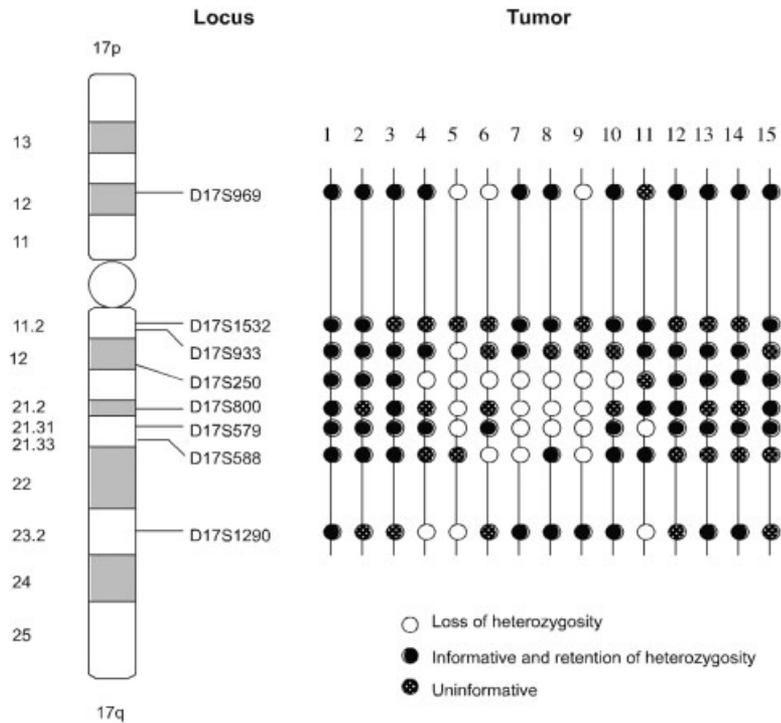


Figure 1. LOH analysis of 17q in papillary renal cell carcinoma. Cases numbers are indicated at the top, microsatellite markers used and their chromosomal locations on the left. Markers are placed in the predicted order from 17cen to 17qter and are listed in the same chromosomal order as in Tables 1 and 2.

same method used for extraction from blocks of formalin-fixed, paraffin-embedded tissue. The 10 coding exons of human *FBXO47* were amplified using 20 ng of tumoral genomic DNA per reaction (Table 3). PCR products were sequenced automatically (ABI3100).

RESULTS

LOH Analysis

LOH analyses were performed in 15 cases of pRCC with a panel of microsatellites spanning the long arm of chromosome 17, and one at 17p (Fig. 1). Eight samples showed allelic alterations (LOH) in at least one chromosomal locus of the eight studied. The frequency of informative cases showing LOH is summarized in Table 4. Allelic alterations were identified at seven of the eight loci examined, with frequency ranging from 8.3%, at 17q11.2 (*D17S933*), to 50%, at 17q12 (*D17S250*) (Table 4). For informative cases, 50% (7 of 14) LOH was detected at *D17S250* and 44% (4 of 9) LOH at *D17S800*. Only 8.3%, 33%, 37.5% and 30% of the informative cases were found to be LOH-positive at *D17S933*, *D17S579*, *D17S588*, and *D17S1290*, respectively. Notably, two patients showed a unique LOH at the *D17S250* locus (Fig. 1). Only one patient may have lost the entire chromosome 17, and two tumors were found to be

TABLE 4. Frequency of LOH in Papillary Renal Cell Carcinoma

Locus	Localization	LOH/Informative	% LOH
<i>D17S969</i>	17p12	3/14	21.4%
<i>D17S1532</i>	17p11.2	0/7	0%
<i>D17S933</i>	17q11.2	1/12	8.3%
<i>D17S250</i>	17q12	7/14	50%
<i>D17S800</i>	17q21.2	4/9	44%
<i>D17S579</i>	17q21.31	5/15	33.3%
<i>D17S588</i>	17q21.33	3/8	37.5%
<i>D17S1290</i>	17q23.2	3/10	30%

positive for microsatellite instability (data not shown). The results for all patients were reproducible. These results enabled us to define a region of minimal deletion around *D17S250* (17q12).

Identification and Characterization of the Human *FBXO47* Gene

By searching for genes within 1 MB of *D17S250*, using several gene prediction programs, we identified a putative candidate with nine exons spanning 30 kb of the AC006441 sequence. We next aligned this genomic region with sequences from the mouse genome database (BLAST). This enabled us to identify a homologous murine sequence on chromosome 11 (GenBank accession number

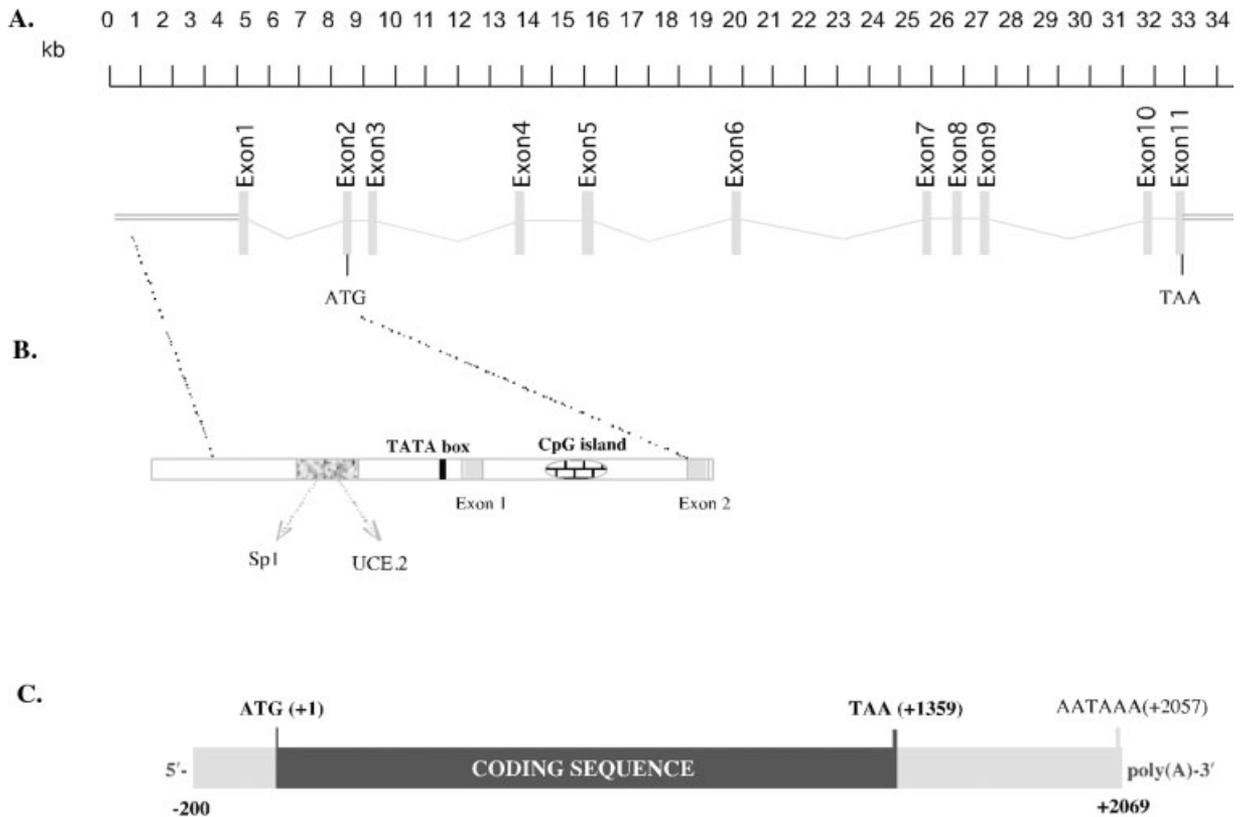


Figure 2. Representation of the *FBXO47* genomic structure. (A) The *FBXO47* gene spans 31 kb and comprises 11 exons, with the first exon noncoding (exons, gray boxes; introns, lines). Translation starts 26 bp from the beginning of exon 2; the stop codon of translation is at the end of exon 11. (B) *FBXO47* comprises a putative promoter region with a TATA box and putative binding sites for Sp1 and UCE.2. A notable

CpG island in the first intron was identified by the Grail program. (C) cDNA of *FBXO47*. The positions of first nucleotides of the start (ATG) and stop (TAA) codons as well as of the polyadenylation signal are given in parentheses. The 5'- and 3' untranslated regions (designated by gray boxes) were obtained by RACE-PCR.

AL569446.11). Strong homology (more than 80%) was observed for the nine putative exons, as well as for two additional exons missed by gene prediction programs. The conservation between species reinforced the relevance of the candidates as exons (Fig. 2A). Connection of the exons revealed an open reading frame (ORF) of 1,359 bp (Fig. 2C).

Using this putative cDNA sequence, we designed primers in order to confirm the genomic structure. We subsequently isolated a transcript of human testis RNA by RT-PCR, using specific primers that were localized 5' and 3' of the ORF. The biocomputerized structure of the gene was confirmed by sequencing of the resulting cDNA. As indicated in Table 5, all splice sites are in accordance with the GT/AG consensus sequence. The codon for the translation initiator methionine is in exon 2 and has a Kozak (1991) consensus sequence (A/GXXATGG). The termination codon, TAA, is in exon 11. On the basis of the SMART RACE

PCR results (data not shown), we concluded that the mature transcript contained 2,269 bp, with a 1,359-bp coding region (GenBank Accession No. AY700775; Fig. 2C), 200 bp of 5'-untranslated region, and 710 bp of 3'-untranslated region. In the search for the presence of promoter elements and transcription factor sites, we restricted the analysis to 3,000 bp upstream of the initiation codon in exon 2. After an NNPromoter search, we detected a putative promoter zone containing a TATA box, localized 200 bp upstream of exon 1 (Fig. 2B). By ProScan, we detected putative binding sites for the Sp1 and UCE.2 transcription regulators.

Moreover, using the Grail program, we identified a CpG island between exons 1 and 2 containing a start codon (Fig. 2B). This GC-rich zone meets the conditions set by Gardiner-Garden and Frommer (1987; Takai and Jones, 2002), with a length of more than 200 nt, a G+C level above 50% (here 54.88), and a CpGo/e ratio higher than 0.6 (0.75 in

TABLE 5. Exon/Intron Junctions of Human *FBXO47* Gene

Exon	Intron-exon	Exon size (bp)	Exon-intron	Intron size (kb)
1	gacgcctggt GGGTAGGGCT	174	AAAATCTGTG gtaaaagagg	4,178
2	tctgaccaag GATAAATTTG	206	TATTTGTCAG gtgaggtttg	0,797
3	cttttggttag TGAAGGATAT	170	AGATCTCTAG gtaacttatg	4,622
4	actttcttag GTCTACTGTT	77	ACTCACAGAA gtaagatcat	2,242
5	ttctttacag GTTTCCTGCT	78	GTTTTTACAG gtacattttt	3,165
6	ttttcctcag ACCTAACAG	108	AGCAAACCAG gtaaaaaac	6,444
7	gtcttcatag GAATTGCCCA	176	CCTCAAGATG gtgagcacc	1,222
8	ttcaattcag GACAGGTGGT	147	GAATATCAG gtaaaaaat	0,668
9	ctaataaaaag TGGTTCCCCG	143	TTTGGCTTTG gtaagtaaaa	4,044
10	tctaaatcag GTGTGTGAGA	160	ATTATGTCTG gtgagtagga	1,282
11	tggcatatag GAGACCGTGA	116		

this case). The 3' untranslated region contained a putative polyadenylation signal (AATAAA) at position +2257 (Fig. 2C). In eukaryotic mRNA, AATAAA is frequently used as the polyadenylation signal.

FBXO47 Protein Analysis

The predicted protein was 453 amino acids long, with a calculated molecular weight of 52 kDa. A systematic search for functional domains in the deduced protein sequence using the PFSCAN algorithm indicated the presence of two conserved motifs (Fig. 3A). The most N-terminal domain showed strong homology with an F-Box motif at amino acids 47–89. This highly degenerate motif is present in F-box protein family members, which belong to the ubiquitin protein ligase complex E3, also known as SCF (Skp1, Cdc53, F-box). SCF is involved in the proteasome-dependent degradation of many proteins in yeast, animals, and plants (Krek, 1998; Yamao, 1999; DeSalle and Pagano, 2001; Kamura et al., 2002). F-box proteins work as adaptors between proteins to be ubiquitinated and the ubiquitination machinery. On the one hand, the link between the F-box protein and the target protein is a specific-target motif. On the other hand, the F-box motif allows a link with the ubiquitination machinery through the help of the SKP1 protein (or its homologue). An alignment of our F-box protein with several other F-Box proteins is shown in Figure 3B. The F-box proteins identified in humans can be divided into three classes: FBXWs, which contain a WD-40 domain; FBXLs, which contain a leucine-rich repeat; and FBXOs, which contain only an F-box motif or another domain (Cenciarelli et al., 1999). Further analysis of the amino acid sequence using ScanProsite revealed the presence of a leucine zipper adjacent

to the F-box motif (Fig. 3A). Leucine zipper domains facilitate the homo- or heterodimerization of polypeptides (Busch and Sassone-Corsi, 1990). The PSORTII program for the prediction of protein localization sites indicated a possible mitochondrial localization signal at residues 1–22, as shown in Figure 3A. This was confirmed by the MitoProtII program.

The C-terminal part of FBXO47 had no significant homology with known proteins. We therefore named this the F-box only protein 47 (*FBXO47*) gene in accordance with the HUGO Nomenclature Committee.

cDNA Expression Cancer Analysis

Some F-box proteins are overexpressed (SKP2) or underexpressed (β TRCP) in several types of tumor tissue (Cenciarelli et al., 1999; Koike et al., 2000; Gstaiger et al., 2001; Latres et al., 2001; Maruyama et al., 2001). We wanted to examine the differential expression of the *FBXO47* gene between tumor and corresponding normal tissues. The results are shown in Figure 4A. The *FBXO47* gene was expressed in all normal tissues tested, with higher expression in kidney, liver, and pancreas. Interestingly, underexpression was observed in the corresponding tumor for these tissues. Focusing on the kidney, phosphor-image quantification revealed significant underexpression (<50% of level intensity) in four renal cell carcinomas and one transitional cell carcinoma compared to the corresponding normal tissue (Fig. 4B). This result was in accordance with *FBXO47* possibly acting as a candidate tumor-suppressor gene in renal carcinoma. Malignancies affecting the pancreas and thyroid gland also revealed underexpression of *FBXO47*, suggesting that this gene may play the same role in other cancers.

A.

```

gggagggtgcagctccggagcctacgggagtgtagcctctcgccctctgagcctcctggcgagaagctcttgaggaga 71
tattttcttttctaactaagaatttgaggatggtggggcaaacgactaactccactttctgtgac 143
ctgtagccggaaaaactgtcaaaatctgtggataaatttgtgagacaagaaga ATG GCA TCC 209
                                     M A S
                                     3
AGA ATA AAT ACA AAT TTC ACT TTG ATT CCC AAC CAG AAA CTT AGA CGT AGT AAT 263
R I N T N F T L I P N Q K L R R S N 21
CGT CAA ACC AGC TGT TAT TCC AAG ACC CTT GGC TCA GGC TTT CAA CCC ATA TCA 317
R Q T S C Y S K T L G S G F Q P I S 39
ACA TTT GGA AAT TTT AAA GCC TTA CCA TTG GAA ATA TTC CAG ATA ATT TTA AAA 371
T F G N F K A L P L E I F Q I I L K 57
TAT TTG TCA GTG AAG GAT ATC AGC ATG CTA AGC ATG GTG TCC AAA ACA GTC AGC 425
Y L S V K D I S M L S M V S K T V S 75
CAA CAC ATT ATT AAT TAT ATC TCA ACC TCA TCA GGA AGC AAA AGA CTT TTA CTA 479
Q H I I N Y I S T S S G S K R L L L 93
CAG GAC TTT CAT AAC CTT GAG CTG CCT GAC AGG AGA CAA GAC TCT GCT ATA CTG 533
Q D F H N L E L P D R R Q D S A I I 111
GAG CAC TAC AGA TCT CTA GGT CTA CTG TTT AAA AGA TGC ACA TTG CTG CTA CCC 587
E H Y R S L G L L F K R C T L I L P 129
ACC AAG GAA AGG CTA AAA TAC ATT CAC AAG ATA CTC ACA GAA GTT TCC TGC TTT 641
T K E R I K Y I H K I I T E V S C F F 147
AAA TTC AAT GGC TGT GCA GCT CCT ATG CAG TGT TTA GGA TTA ACA TGT TAT GGC 695
K F N G C A A P M Q C L G L T C Y G 165
ATG TTT TTA CAG ACC TTA ACA GCA GGT TGG GAT GAA CTT GAG TGC CAT CGC GTT 749
M F L Q T L T A G W D E L E C H R V 183
TAT AAT TTC TTA TGC GAA CTG ACT AAT CTC TGC CGC AAG ATA CAA ATG GCT GTC 803
Y N F L C E L T N L C R K I Q M A V 201
TGC AGC AAA CCA GGA AGT GCC CGA AAA CTG GAG TTA AGA ATC AGA CTC TTC TGT 857
C S K P C G S A R K L E L R I R L F C 219
AGG AAT GTC CTC GAT CAT TGG ACA CAT CGA AGT GAT TCT GCT TTT TGG TTG 911
R N V L L D H W T H R S D S A F W L 237
ACG CGA ATA TTA AAA CCA TGG CCA ATG GTG AAT CAG GCA AGA TTA CTG TAT ATC 965
T R I L K P W P M V N Q A R L L Y I 255
ATC TTT GGG CCA ATA TCT CCT CAA GAT GGA CAG GTG GTT TGG CAG GAA ATG ATA 1019
I F G P I S P Q D G Q V V W Q E M I 273
GAA GAA CCT ACA GAT GAA TTC AGT CTG AAA GGT TTG GCT GAT GCC ATT AAG TTA 1073
E E P T D E F S L K G L A D A I K L 291
CTA TAT GAC GCT AGC ACT AAA GAG TGG ACA GCA GAT GAT GTT ATC AGT CTT GTA 1127
L Y D A S T K E W T A D D V I S L V 309
GAT GAA CTA TCA GTG GTT CCC CGT GAG TGG CTT CTA GAG AAT AAT GCA CGT CTC 1181
D E L S V V P R E W L L E N N A R L 327
CTA ATG CTA AGT GGA AAC AAC ATC TGT TTC AGT TTC ATG GCT AGT AAA GCT GTG 1235
L M L S G N N I C F S F M A S K A V 345
AAT GGA CGC ACC ATT GAA CTG GCA AGG CTC GTA GTC TTT TTG GCT TTG GTG TGT 1289
N G R T I E L A R L V V F L A L V C 363
GAG AAA GAA CTG TAC TGC ATG GAT TGG ACA GTT AAA ATG ATG CAA AAA GTC TGC 1343
E K E L Y C M D W T V K M M Q K V C 381
AAA GTC TTT AGC ACT CCA GTG GAA AGA AAG AAC TTC CTG CAG AAT GTG GCA AAT 1397
K V F S T P V E R K N F L Q N V A N 399
GCA TTT GCA TGT GTT ATA ATG GAA ATG CTG CAA TCA ATT ATG TCT GGA GAC CGT 1451
A F A C V I M E M L Q S I M S G D R 417
GAT GAA GAT GAC AGA AGC TTT TTG AAT TTG TTC CAT CTT GTA CAT GCT CAG GCT 1505
D E D D R S F L N L F H L V H A Q A 435
AAC TTC CAT AAG GAG GTC CTG TAT TTG ACC ATG AAT ACT CCT CTG TCT ACC TAA 1559
N F H K E V L Y L T M N T P L S T * 453
atc ctc aga aag cct tgc ctt tag aga acg tca ctg aac gaa tca tta gag gag 1613
                                     471
tgc 1616
                                     472

```

B.

Consensus	1	LP	x	E	I	L	x	K	I	L	s	y	L	D	x	x	D	L	L	x	L	R	K	V	S	K	R	W	R	x	L	V	D	d	d	x	L	W	K	x	L	41	
FBXO47	1	L	P	L	E	I	F	Q	I	I	L	K	Y	L	S	V	K	D	I	S	M	L	S	M	V	S	K	T	V	S	Q	H	I	I	N	Y	I	S	T	S	S	G	41
SKP2 Hs	1	L	P	D	E	L	L	L	G	I	F	S	C	L	C	L	P	E	L	L	K	V	S	G	V	C	K	R	W	Y	R	L	A	S	D	E	S	L	W	Q	T	L	41
FBG3 Hs	1	L	P	E	N	I	L	L	E	L	F	T	H	V	P	A	R	O	L	L	N	C	R	L	V	C	S	L	W	R	D	L	I	D	L	V	T	L	W	K	R	41	
CyclinF Hs	1	L	P	E	D	V	L	F	H	I	L	K	W	L	S	V	E	D	I	L	A	V	R	A	V	H	S	Q	L	K	D	L	V	D	N	H	A	S	V	W	A	C	41
FBXO10 Mm	1	L	P	L	E	L	W	R	M	I	L	A	Y	L	H	L	P	D	L	G	R	C	S	L	V	C	R	A	W	Y	E	L	I	L	S	L	D	S	T	R	W	R	41
Cdc4p Sc	1	L	P	F	E	I	S	L	K	I	F	N	Y	L	Q	F	E	D	I	I	N	S	L	G	V	S	Q	N	W	N	K	I	I	R	K	S	T	S	L	W	K	K	41

Figure 3.

DISCUSSION

Inactivation of tumor-suppressor genes through LOH and mutation is the hallmark of the development of a variety of human cancers (Marshall, 1991; Weinberg, 1991). LOH on chromosome arm 17q has been observed in a variety of cancers, including melanoma, and breast, ovarian, prostate, renal, thyroid, and pancreas carcinomas (Presti et al., 1993; Gao et al., 1995; Gomez et al., 1996; Crawley and Furge, 2002; Iacobuzio-Donahue et al., 2004). The data emerging from the recent literature concerning the molecular abnormalities in papillary renal cell carcinoma seem unclear and contradictory. pRCC has been characterized by trisomies or tetrasomies in relation to chromosomes 7 and 17 and loss of the Y chromosome, as well as additional gains on chromosome arm 3q and chromosomes 12, 16, and 20 (Kovacs, 1989, 1996; Renshaw et al., 1997). LOH at several loci in pRCC also has been described (Presti et al., 1993; Thrash-Bingham et al., 1995; Bugert and Kovacs, 1996; Salama et al., 2003). To better define the chromosomal alterations in 17q, we analyzed LOH with 7 microsatellites at 17q11–q23 in 15 pRCCs. We defined a region of minimal deletion including *D17S250* (17q12) displaying 50% LOH. This result suggests that a tumor-suppressor gene of pRCC is harbored in this region.

We have identified a new gene, named *FBXO47*, 30 kb upstream of *D17S250*. According to a homology protein database study, this gene belongs to a large family of genes displaying an F-box motif. Although the function of several members of this gene family is still unknown, some [such as *SKP2*, *BTRCP*, and *FBW7* (or *CDC4*)] play a role in carcinogenesis. *SKP2*, for example, has been implicated in the ubiquitin-dependent degradation of p27 (Carrano et al., 1999; Tsvetkov et al., 1999), and acts as an oncogene in many human cancers,

including breast, prostate, and ovarian cancers (Gstaiger et al., 2001; Lu et al., 2002; Signoretti et al., 2002; Shigemasa et al., 2003). *BTRCP*, on the other hand, is responsible for the stability of β -*CATENIN*, a transcription factor for MYC and cyclin D1 (Hart et al., 1999; Polakis, 1999, 2000). *BTRCP* also controls the NF- κ B-dependent apoptotic pathway via ubiquitin-dependent degradation of the I κ B protein (Yaron et al., 1998; Spencer et al., 1999). *FBW7*, which is at 4q32, has been implicated in the degradation of cyclin E (Koepp et al., 2001; Moberg et al., 2001; Strohmaier et al., 2001) and is mutated and/or deleted in a number of primary cancers, such as endometrial, ovarian, and pancreatic cancers (Moberg et al., 2001; Spruck et al., 2002; Calhoun et al., 2003), as well as in a breast cancer cell line (Strohmaier et al., 2001). Interestingly, *FBW7* has been shown to mediate the degradation of MYC, thus explaining its tumor-suppressor function (Yada et al., 2004). Five other genes encoding F-box proteins also have been identified and located in chromosome loci frequently deleted in tumors (Chiari et al., 2000). However, the results of the present study did not show gross alterations in genomic content for 4 of the 5 F-box proteins tested.

According to Knudson's two-hit model, tumor-suppressor genes must be inactivated on both alleles in tumor cells (Knudson, 1971). Genomic loss is always the most common "hit" that inactivates a tumor-suppressor gene. In familial cancer, somatic loss is accompanied by germ-line mutation of the remaining wild-type allele. This model has held true for most tumor-suppressor genes, explaining cancer predisposition in family cancer syndromes. However, in sporadic cancer, with the exception of certain genes such as the retinoblastoma (*RB*) and colon adenomatous polyposis (*APC*) genes, the probability of finding somatic mutations in certain tumor-suppressor genes with a high frequency of LOH has been shown to be extremely small (Shen et al., 2000). This is particularly true for "caretaker" tumor-suppressor genes such as the *BRCA1* gene, which does not show point mutations in sporadic breast cancer (Esteller et al., 2000). The inconsistency between the classical two-hit model and observations made in sporadic cancer has two possible explanations: haploinsufficiency phenomena or epigenetic alterations, resulting in promoter hypermethylation and mutation in the promoter region. With regard to the first explanation, the mechanism of haploinsufficiency shows that inactivating mutations and complete loss of function are not necessary for defective tumor-sup-

Figure 3. (A) cDNA and deduced amino acid sequences of human *FBXO47* gene. The cDNA nucleotide sequences are capitalized. The amino acid identity and position are above the codons. The residues forming the F-box (shown with a gray background) and the amino acids (in bold) were conserved in at least 40% of the F-box protein family (leucine zipper, underlined; the conserved leucine compared to consensus, in clear gray; the predicted cleavage site of exportation to mitochondria, italicized). (B) Multiple alignment of the amino acid sequence of the F-box motif for representative members of the F-box superfamily. The consensus was derived from the alignment of 234 sequences used to create the Pfam F-box profile [<http://pfam.wustl.edu/cgi-bin/getdesc?name=F-box> (Kipreos and Pagano, 2000)]. Alignment was performed with the SeqVu program, followed by manual readjustment (homologous residues, in gray; identical residues, boxed; organisms: Hs, *Homo sapiens*; Mm, *Mus musculus*; Sc, *Saccharomyces cerevisiae*). The GenBank accession number for *SKP2* is AAH07441; for *FBG3*, AAK77940; for *CyclinF*, AAH12349; for *FBXO10*, AAH54731; and for *Cdc4p*, NP-116585.

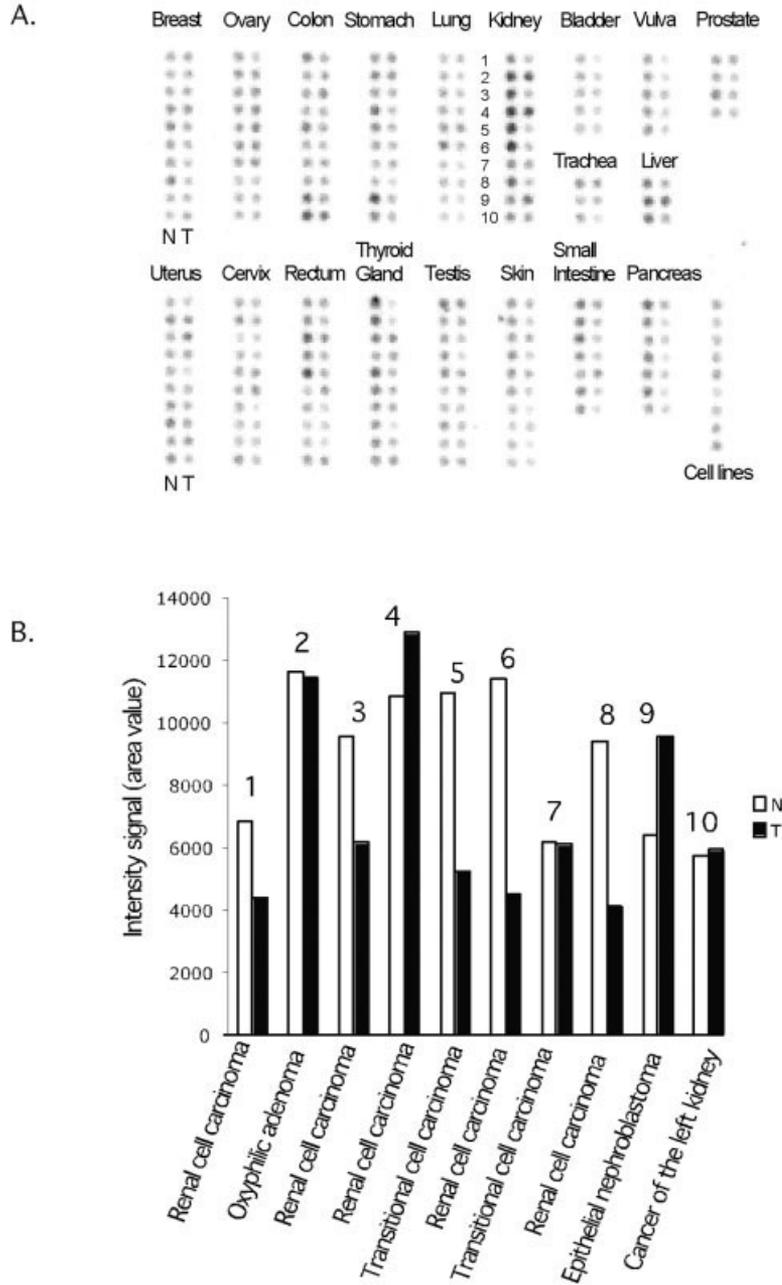


Figure 4. Expression pattern of *FBXO47* gene in healthy and tumoral human tissue. (A) This membrane was hybridized with a ^{32}P -labeled *FBXO47*-specific probe. As a control, it was hybridized with a ^{32}P -labeled ubiquitin probe (data not shown). A diagram of the locations of cDNAs and controls spotted on the Cancer Profiling Array II (Clontech, BDBiosciences, Erembodegem, Belgium) showing the types and positions of cDNAs and controls spotted on the nylon membrane can be found at www.bdbiosciences.com (reference #631777). (B) Quantitative measurements using phosphoimage analysis for normal and tumor tissue in kidney are presented as signal intensity (area value). Numbers indicate the results obtained for each patient and are listed in panel order. The legend on the abscissa indicates the subtype of renal tumor analyzed in this array.

pressor function to occur (Macleod, 2000). Thus, a typical example of haploinsufficiency was observed with p27, a putative human tumor-suppressor protein that inhibits cyclin-dependent kinases and blocks cell proliferation (Fero et al., 1998). With regard to the second explanation, epigenetic alterations such as promoter hypermethylation or mutation in the regulatory region that operate at the transcriptional or translational level can switch genes, thus affecting expression of the normal tumor-suppressor gene rather than totally abolishing expression. Recent studies analyzing methyla-

tion in conjunction with LOH or mutational analysis have shown that the frequency with which promoter methylation contributes to gene inactivation ranges from about 33% for the *VHL* gene in von Hippel–Lindau disease to 84% for *MLH1* in microsatellite-unstable colorectal tumors (Prowse et al., 1997; Herman et al., 1998). Investigations of CpG island methylation would be of interest for *FBXO47*.

In conclusion, we have identified a novel gene encoding an F-box protein (*FBXO47*) that is on the 17q12 band, which often is deleted in pRCC.

These results, together with the underexpression of *FBXO47* in various tumor cells relative to normal tissue, strongly suggest that this gene could play a tumor-suppressor role in renal cell carcinoma and pancreas cancer. Absence of the wild-type allele mutation in tumor DNA could be explained by other mechanisms of inactivation. In addition, the localization of *FBXO47* makes it a good candidate in various sporadic cancers with deletions at 17q12 and in familial Wilms tumor displaying linkage to this region. Further investigations are now needed to clarify the function of *FBXO47* at both the molecular and cellular levels.

ACKNOWLEDGMENTS

The authors thank Fabrice Airaud and Ingrid Maury for technical assistance.

REFERENCES

- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25:3389–3402.
- Bertucci F, Borie N, Ginestier C, Groulet A, Charafe-Jauffret E, Adelaide J, Geneix J, Bachelart L, Finetti P, Koki A, Hermitte F, Hassoun J, Debono S, Viens P, Fert V, Jacquemier J, Birnbaum D. 2004. Identification and validation of an ERBB2 gene expression signature in breast cancers. *Oncogene* 23:2564–2575.
- Bugert P, Kovacs G. 1996. Molecular differential diagnosis of renal cell carcinomas by microsatellite analysis. *Am J Pathol* 149:2081–2088.
- Busch SJ, Sassone-Corsi P. 1990. Dimers, leucine zippers and DNA-binding domains. *Trends Genet* 6:36–40.
- Calhoun ES, Jones JB, Ashfaq R, Adsay V, Baker SJ, Valentine V, Hempen PM, Hilgers W, Yeo CJ, Hruban RH, Kern SE. 2003. BRAF and FBXW7 (CDC4, FBW7, AGO, SEL10) mutations in distinct subsets of pancreatic cancer: potential therapeutic targets. *Am J Pathol* 163:1255–1260.
- Carrano AC, Eytan E, Hershko A, Pagano M. 1999. SKP2 is required for ubiquitin-mediated degradation of the CDK inhibitor p27. *Nat Cell Biol* 1:193–199.
- Cenciarelli C, Chiaur DS, Guardavaccaro D, Parks W, Vidal M, Pagano M. 1999. Identification of a family of human F-box proteins. *Curr Biol* 9:1177–1179.
- Chiaur DS, Murthy S, Cenciarelli C, Parks W, Loda M, Inghirami G, Demetrick D, Pagano M. 2000. Five human genes encoding F-box proteins: chromosome mapping and analysis in human tumors. *Cytogenet Cell Genet* 88:255–258.
- Crawley JJ, Furge KA. 2002. Identification of frequent cytogenetic aberrations in hepatocellular carcinoma using gene-expression microarray data. *Genome Biol* 3:RESEARCH0075.
- DeSalle LM, Pagano M. 2001. Regulation of the G1 to S transition by the ubiquitin pathway. *FEBS Lett* 490:179–189.
- Esteller M, Silva JM, Dominguez G, Bonilla F, Matias-Guiu X, Lerma E, Bussaglia E, Prat J, Harkes IC, Repasky EA, Gabrielson E, Schutte M, Baylin SB, Herman JG. 2000. Promoter hypermethylation and BRCA1 inactivation in sporadic breast and ovarian tumors. *J Natl Cancer Inst* 92:564–569.
- Fero ML, Randel E, Gurley KE, Roberts JM, Kemp CJ. 1998. The murine gene p27Kip1 is haplo-insufficient for tumour suppression. *Nature* 396:177–180.
- Gao X, Zacharek A, Salkowski A, Grignon DJ, Sakr W, Porter AT, Honn KV. 1995. Loss of heterozygosity of the BRCA1 and other loci on chromosome 17q in human prostate cancer. *Cancer Res* 55:1002–1005.
- Garcia A, Bussaglia E, Machin P, Matias-Guiu X, Prat J. 2000. Loss of heterozygosity on chromosome 17q in epithelial ovarian tumors: association with carcinomas with serous differentiation. *Int J Gynecol Pathol* 19:152–157.
- Godwin AK, Vanderveer L, Schultz DC, Lynch HT, Altomare DA, Buetow KH, Daly M, Getts LA, Masny A, Rosenblum N, et al. 1994. A common region of deletion on chromosome 17q in both sporadic and familial epithelial ovarian tumors distal to BRCA1. *Am J Hum Genet* 55:666–677.
- Gomez L, Rubio MP, Martin MT, Vazquez JJ, Idoate M, Pastorfide G, Pestana A, Seizinger BR, Barnhill RL, Castresana JS. 1996. Chromosome 17 allelic loss and NF1-GRD mutations do not play a significant role as molecular mechanisms leading to melanoma tumorigenesis. *J Invest Dermatol* 106:432–436.
- Gstaiger M, Jordan R, Lim M, Catzavelos C, Mestan J, Slingerland J, Krek W. 2001. Skp2 is oncogenic and overexpressed in human cancers. *Proc Natl Acad Sci USA* 98:5043–5048.
- Guardavaccaro D, Pagano M. 2004. Oncogenic aberrations of cullin-dependent ubiquitin ligases. *Oncogene* 23:2037–2049.
- Hart M, Concorde JP, Lassot I, Albert I, del los Santos R, Durand H, Perret C, Rubinfeld B, Margottin F, Benarous R, Polakis P. 1999. The F-box protein beta-TrCP associates with phosphorylated beta-catenin and regulates its activity in the cell. *Curr Biol* 9:207–210.
- Herman JG, Umar A, Polyak K, Graff JR, Ahuja N, Issa JP, Markowitz S, Willson JK, Hamilton SR, Kinzler KW, Kane MF, Kolodner RD, Vogelstein B, Kunkel TA, Baylin SB. 1998. Incidence and functional consequences of hMLH1 promoter hypermethylation in colorectal carcinoma. *Proc Natl Acad Sci USA* 95:6870–6875.
- Iacobuzio-Donahue CA, van der Heijden MS, Baumgartner MR, Troup WJ, Romm JM, Doheny K, Pugh E, Yeo CJ, Goggins MG, Hruban RH, Kern RH. 2004. Large-scale allelotyping of pancreatic carcinoma provides quantitative estimates of genome-wide allelic loss. *Cancer Res* 64:871–875.
- Jarvinen TA, Liu ET. 2003. Topoisomerase IIalpha gene (TOP2A) amplification and deletion in cancer—more common than anticipated. *Cytopathology* 14:309–313.
- Kamura T, Conaway JW, Conaway RC. 2002. Roles of SCF and VHL ubiquitin ligases in regulation of cell growth. *Prog Mol Subcell Biol* 29:1–15.
- Katoh M. 2003. MGC9753 gene, located within PPP1R1B-STAR3-ERBB2-GRB7 amplicon on human chromosome 17q12, encodes the seven-transmembrane receptor with extracellular six-cysteine domain. *Int J Oncol* 22:1369–1374.
- Kiaris H, Spanakis N, Ergazaki M, Sourvinos G, Spandidos DA. 1995. Loss of heterozygosity at 9p and 17q in human laryngeal tumors. *Cancer Lett* 97:129–134.
- Klapper LN, Kirschbaum MH, Sela M, Yarden Y. 2000. Biochemical and clinical implications of the ErbB/HER signaling network of growth factor receptors. *Adv Cancer Res* 77:25–79.
- Knudson AG Jr. 1971. Mutation and cancer: statistical study of retinoblastoma. *Proc Natl Acad Sci USA* 68:820–823.
- Koepp DM, Schaefer LK, Ye X, Keyomarsi K, Chu C, Harper JW, Elledge SJ. 2001. Phosphorylation-dependent ubiquitination of cyclin E by the SCF^{Fbw7} ubiquitin ligase. *Science* 294:173–177.
- Koike J, Sagara N, Kirikoshi H, Takagi A, Miwa T, Hirai M, Katoh M. 2000. Molecular cloning and genomic structure of the betaTRCP2 gene on chromosome 5q35.1. *Biochem Biophys Res Commun* 269:103–109.
- Kovacs G. 1989. Papillary renal cell carcinoma. A morphologic and cytogenetic study of 11 cases. *Am J Pathol* 134(1):27–34.
- Kovacs G. 1996. Molecular genetics of human renal cell tumours. *Nephrol Dial Transplant* 11:62–65.
- Kozak M. 1991. Structural features in eukaryotic mRNAs that modulate the initiation of translation. *J Biol Chem* 266:19867–19870.
- Krek W. 1998. Proteolysis and the G1–S transition: the SCF connection. *Curr Opin Genet Dev* 8:36–42.
- Latres E, Chiarle R, Schulman BA, Pavletich NP, Pellicer A, Inghirami G, Pagano M. 2001. Role of the F-box protein Skp2 in lymphomagenesis. *Proc Natl Acad Sci USA* 98:2515–2520.
- Li Y, Chen Z. 2004. Molecular cloning and characterization of LCRG1 a novel gene localized to the tumor suppressor locus D17S800–D17S930. *Cancer Lett* 209:75–85.
- Lu L, Schulz H, Wolf DA. 2002. The F-box protein SKP2 mediates androgen control of p27 stability in LNCaP human prostate cancer cells. *BMC Cell Biol* 3:22.
- Macleod K. 2000. Tumor suppressor genes. *Curr Opin Genet Dev* 10:81–93.
- Marshall CJ. 1991. Tumor suppressor genes. *Cell* 64:313–326.
- Maruyama S, Hatakeyama S, Nakayama K, Ishida N, Kawakami K. 2001. Characterization of a mouse gene (Fbxw6) that encodes a homologue of *Caenorhabditis elegans* SEL-10. *Genomics* 78:214–222.
- Moberg KH, Bell DW, Wahrer DC, Haber DA, Hariharan IK. 2001. Archipelago regulates cyclin E levels in *Drosophila* and is mutated in human cancer cell lines. *Nature* 413:311–316.

- Polakis P. 1999. The oncogenic activation of beta-catenin. *Curr Opin Genet Dev* 9:15–21.
- Polakis P. 2000. Wnt signaling and cancer. *Genes Dev* 14:1837–1851.
- Presti Jr JC, Reuter VE, Cordon-Cardo C, Mazumdar M, Fair WR, Jhanwar SC. 1993. Allelic deletions in renal tumors: histopathological correlations. *Cancer Res* 53:5780–5783.
- Prowse AH, Webster AR, Richards FM, Richard S, Olschwang S, Resche F, Affara NA, Maher ER. 1997. Somatic inactivation of the VHL gene in Von Hippel–Lindau disease tumors. *Am J Hum Genet* 60:765–771.
- Rahman N, Arbour L, Houlston R, Bonaiti-Pellie C, Abidi F, Tranchemontagne J, Ford D, Narod S, Pritchard-Jones K, Foulkes WD, Schwartz C, Stratton MR. 2000. Penetrance of mutations in the familial Wilms tumor gene FWT1. *J Natl Cancer Inst* 92:650–652.
- Renshaw AA, Zhang H, Corless CL, Fletcher JA, Pins MR. 1997. Solid variants of papillary (chromophil) renal cell carcinoma: clinicopathologic and genetic features. *Am J Surg Pathol* 21:1203–1209.
- Rizos E, Sourvinos G, Spandidos DA. 1998. Loss of heterozygosity at 8p, 9p and 17q in laryngeal cytological specimens. *Oral Oncol* 34:519–523.
- Salama ME, Worsham MJ, DePeralta-Venturina M. 2003. Malignant papillary renal tumors with extensive clear cell change: a molecular analysis by microsatellite analysis and fluorescence in situ hybridization. *Arch Pathol Lab Med* 127:1176–1181.
- Sanders ME, Mick R, Tomaszewski JE, Barr FG. 2002. Unique patterns of allelic imbalance distinguish type 1 from type 2 sporadic papillary renal cell carcinoma. *Am J Pathol* 161:997–1005.
- Sasiadek M, Stembalska-Kozłowska A, Smigiel R, Krecicki T, Blin N, Mirghomizadeh F. 2001. Microsatellite and chromosome instability in squamous cell laryngeal carcinoma. *Int J Oncol* 19:401–405.
- Shen CY, Yu JC, Lo YL, Kuo CH, Yue CT, Jou YS, Huang CS, Lung JC, Wu CW. 2000. Genome-wide search for loss of heterozygosity using laser capture microdissected tissue of breast carcinoma: an implication for mutator phenotype and breast cancer pathogenesis. *Cancer Res* 60:3884–3892.
- Shigemasa K, Gu L, O'Brien TJ, Ohama K. 2003. Skp2 overexpression is a prognostic factor in patients with ovarian adenocarcinoma. *Clin Cancer Res* 9:1756–1763.
- Signoretto S, Di Marcotullio L, Richardson A, Ramaswamy S, Isaac B, Rue M, Monti F, Loda M, Pagano M. 2002. Oncogenic role of the ubiquitin ligase subunit Skp2 in human breast cancer. *J Clin Invest* 110:633–641.
- Slamon DJ, Clark GM, Wong SG, Levin WJ, Ullrich A, McGuire WL. 1987. Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. *Science* 235:177–182.
- Spencer E, Jiang J, Chen ZJ. 1999. Signal-induced ubiquitination of IκappaBα by the F-box protein Slimb/β-TCP. *Genes Dev* 13:284–294.
- Spruck CH, Strohmaier H, Sangfelt O, Muller HM, Hubalek M, Muller-Holzner E, Marth C, Widschwendter M, Reed SI. 2002. hCdc4 gene mutations in endometrial cancer. *Cancer Res* 62:4535–4539.
- Strohmaier H, Spruck CH, Kaiser P, Won KA, Sangfelt O, Reed SI. 2001. Human F-box protein hCdc4 targets cyclin E for proteolysis and is mutated in a breast cancer cell line. *Nature* 413:316–322.
- Takai D, Jones PA. 2002. Comprehensive analysis of CpG islands in human chromosomes 21 and 22. *Proc Natl Acad Sci USA* 99:3740–3745.
- Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22:4673–4680.
- Thrash-Bingham CA, Salazar H, Freed JJ, Greenberg RE, Tartof KD. 1995. Genomic alterations and instabilities in renal cell carcinomas and their relationship to tumor pathology. *Cancer Res* 55:6189–6195.
- Toro JR, Nickerson ML, Wei MH, Warren MB, Glenn GM, Turner ML, Stewart L, Duray P, Tourre O, Sharma N, Choyke P, Stratton P, Merino M, Walther MM, Linehan WM, Schmidt LS, Zbar B. 2003. Mutations in the fumarate hydratase gene cause hereditary leiomyomatosis and renal cell cancer in families in North America. *Am J Hum Genet* 73:95–106.
- Tsvetkov LM, Yeh KH, Lee SJ, Sun H, Zhang H. 1999. p27(Kip1) ubiquitination and degradation is regulated by the SCF(Skp2) complex through phosphorylated Thr187 in p27. *Curr Biol* 9:661–664.
- Varis A, Zaika A, Puolakkainen P, Nagy B, Madrigal I, Kokkola A, Vayrynen A, Karkkainen P, Moskaluk C, El-Rifai W, Knuutila S. 2004. Coamplified and overexpressed genes at ERBB2 locus in gastric cancer. *Int J Cancer* 109:548–553.
- Vernimmen D, Gueders M, Pisvin S, Delvenne P, Winkler R. 2003. Different mechanisms are implicated in ERBB2 gene overexpression in breast and in other cancers. *Br J Cancer* 89:899–906.
- Weinberg RA. 1991. Tumor suppressor genes. *Science* 254:1138–1146.
- Yada M, Hatakeyama S, Kamura T, Nishiyama M, Tsunematsu R, Imaki H, Ishida N, Okumura F, Nakayama K, Nakayama KI. 2004. Phosphorylation-dependent degradation of c-Myc is mediated by the F-box protein Fbw7. *Embo J* 23:2116–2125.
- Yamao F. 1999. Ubiquitin system: selectivity and timing of protein destruction. *J Biochem (Tokyo)* 125:223–229.
- Yaron A, Hatzubai A, Davis M, Lavon I, Amit S, Manning AM, Andersen JS, Mann M, Mercurio F, Ben-Neriah Y. 1998. Identification of the receptor component of the IκappaBα-ubiquitin ligase. *Nature* 396:590–594.
- Zbar B, Glenn G, Lubensky I, Choyke P, Walther MM, Magnusson G, Bergerheim US, Pettersson S, Amin M, Hurley K. 1995. Hereditary papillary renal cell carcinoma: clinical studies in 10 families. *J Urol* 153:907–912.
- Zhao J, Moch H, Scheidweiler AF, Baer A, Schaffer AA, Speel EJ, Roth J, Heitz PU, Komminoth P. 2001. Genomic imbalances in the progression of endocrine pancreatic tumors. *Genes Chromosomes Cancer* 32:364–372.

Identification d'un gène humain localisé en 17q12, codant pour un membre de la famille F-box (Fbxo47).

Étude des caractéristiques de la protéine.

Les altérations chromosomiques dans la région 17q surviennent dans de nombreux cancers, incluant les tumeurs papillaires rénales (pRCC), suggérant la présence d'un gène impliqué dans l'oncogenèse. Une analyse de perte d'hétérozygotie sur 15 cas de pRCC nous a permis de définir une zone minimale de délétion autour du marqueur D17S250 (17q12), dans laquelle nous avons isolé un nouveau gène, FBXO47. Ce transcrit apparaît fortement exprimé dans le testicule. La recherche de domaines protéiques caractéristiques a permis de montrer la présence : d'un domaine F-box, dont nous avons démontré la fonctionnalité; d'un motif Leucine-zipper; d'une séquence d'adressage mitochondrial. Le motif F-box définit une famille de protéines dont la majorité des membres appartient à la voie de dégradation ubiquitine-protéasome dépendante. Afin de déterminer le rôle de Fbxo47 au sein de la cellule, nous avons cherché à identifier ses protéines partenaires. La technique de double-hybride nous a permis d'identifier Gfm1, facteur d'élongation principal de la traduction mitochondriale. Ce travail de thèse représente donc la première description de la présence du système F-box dans la matrice mitochondriale chez l'homme.

Mots Clés : Fbxo47 - 17q12 – F-box – mitochondrie - testicule – Gfm1.

Identification of human gene located in 17q12 band, coding for a novel F-box protein family member (Fbxo47).

Study of protein characteristics.

Genetic alterations of chromosome arm 17q occur in numerous tumor types, including papillary renal cell carcinoma (pRCC). It suggests the presence of a tumor suppressor gene on the long arm of chromosome 17, which is critical for carcinogenesis. In this study, we analyzed 15 cases of pRCC for LOH. We identified a minimal deleted region in which the *D17S250* marker (17q12). We isolated the cDNA of a novel gene named *FBXO47*. *FBXO47* cDNA is preferentially expressed in normal tissue, particularly in the testis. The search of characteristic protein domains showed the presence of : a F-box domain, that we showed the functionality *in vitro* and *in vivo*; a Leucine-zipper; and a sequence of addressing mitochondrial. Most proteins of the family defined by the F-box motif belong to the ubiquitin-proteasome pathway. In order to determine the cellular function of Fbxo47, we sought to identify its protein partners. A screening in two-hybrid system enabled us to identify the protein Gfm1, the principal elongation factor of the mitochondrial translation. This thesis work represents indeed the first description of the presence of F-box system in mitochondrial matrix in human.

Key words: Fbxo47 - 17q12 – F-box – mitochondria - testis – Gfm1.