UNIVERSITE DE NANTES
FACULTE DE MEDECINE

# PORTRAITS MOLECULAIRES DES PATHOLOGIES CARDIAQUES

## THESE DE DOCTORAT

Ecole Doctorale Chimie-Biologie de l'Université de Nantes
Discipline : Sciences de la Vie et de la Santé
Spécialité : Aspects Moléculaires et Cellulaires de la Biologie

*Présentée et soutenue publiquement par*

Guillaume LAMIRAULT

le 15 Octobre 2007, devant le jury ci-dessous

Président du jury :

Rapporteurs : Dr PINET Florence, Directeur de Recherche, Université de Lille II
Pr MERCADIER Jean-Jacques, Professeur, Université de Paris VII

Examinateurs : Dr STEENMAN Marja, Chargé de Recherche, Université de Nantes
Pr BERGER François, Professeur, Université de Grenoble I
Pr LE MAREC Hervé, Professeur, Université de Nantes
Pr MOSSER Jean, Professeur, Université de Rennes I

Directeur de thèse : Dr HOULGATTE Rémi, Directeur de Recherche, Université de Nantes
Codirecteur de thèse : Dr STEENMAN Marja, Chargé de Recherche, Université de Nantes

# TABLE DES MATIERES

# TABLE DES FIGURES ET TABLEAUX

# INTRODUCTION

## I. MALADIES CARDIOVASCULAIRES

### A. Epidémiologie

Les maladies cardiovasculaires représentent la première cause de mortalité en France. Elles sont la cause de 160 000 décès par an (30,2 % des décès en France) dont 25% affectent des personnes de moins de 75 ans. Les cardiopathies ischémiques représentent 27 % des décès, les accidents vasculaires cérébraux 25 % et les insuffisances cardiaques 23 %, de telle sorte que ces trois groupes de pathologies représentent à elles seules 75 % de la mortalité cardio-vasculaire.(DREES, 2005) La prévalence de ces maladies est forte dans la population française. L'insuffisance cardiaque, la fibrillation auriculaire, la cardiopathie ischémique et l'artériopathie oblitérante des membres inférieurs ont chacune une prévalence supérieure à 1%.
Ces pathologies touchent en particulier les personnes âgées. Chez les plus de 65 ans, les maladies cardiovasculaires représentent 63% des motifs de consultation et sont la cause d'une hospitalisation sur 6. Le vieillissement actuel de la population devrait encore renforcer le poids des maladies cardiovasculaires dans les causes de mortalité et morbidité de la population française.

### B. Notion de facteurs de risque

Les affections cardiovasculaires sont d'origines multifactorielles. Leur survenue est dépendante de facteurs de risques associant des caractéristiques ou comportements individuels et des conditions environnementales potentialisant des susceptibilités génétiques familiales (tableau I). L'importance des facteurs de risque dans la genèse des pathologies cardio-vasculaires est fondamentale. L'étude INTERHEART réalisée dans 52 pays a par exemple montré que 9 facteurs de risque

pouvaient rendre compte de 90% du risque d'infarctus du myocarde quelque soit l'âge ou le sexe.(Yusuf et al, 2004) Un des résultat les plus marquant concerne le tabagisme (figure 1). La part attribuable du tabagisme au risque d'infarctus dans la population est d'environ 35%. Le risque d'infarctus est proportionnel à la consommation de tabac. Il est multiplié par 4 chez les fumeurs consommant 20 cigarettes par jour.

Les données épidémiologiques françaises, bien que très segmentées et d'origines très diverses, permettent de mesurer l'importance des populations concernées par les maladies cardiovasculaires : la prévalence de l'hypertension artérielle dans la population de plus de 20 ans est de 16,5%, celle du diabète est de 3,1%, de l'hypercholestérolémie de 30%, du tabagisme de 34,4% et de l'obésité de 10%. Ces quantifications successives ne peuvent s'additionner car certains patients cumulent plusieurs de ces facteurs. Toutefois, au regard de ces données chiffrées, on peut considérer que 20 millions de personnes sont concernées par les risques cardiovasculaires en France.(Ministère de l'emploi et de la solidarité, 2002)

Parmi l'ensemble des pathologies cardiovasculaires, les prévalences de la fibrillation auriculaire et l'insuffisance cardiaque croissent rapidement. Ces 2 pathologies ont ainsi été qualifiées de « nouvelles épidémies cardio-vasculaires ».(Braunwald, 1997) Les travaux présentés dans ce document s'intéresseront particulièrement à ces 2 pathologies.

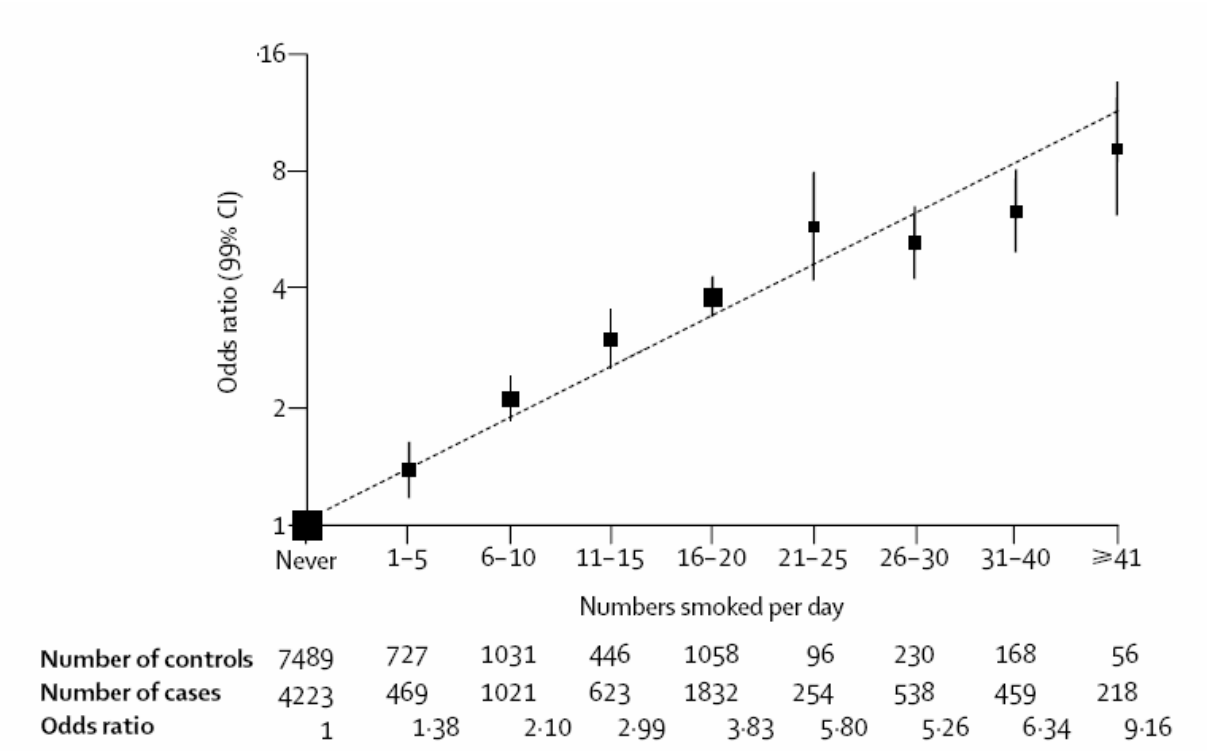



**Figure 1 :** Risque de développer un infarctus du myocarde (ordonnée) en fonction du nombre de cigarettes consommées par jour (abscisse).(Yusuf et al, 2004)

**Tableau I** : Facteurs de risque cardio-vasculaire :(Lusis et al, 2004)

| Genetic and environmental risk factors for CV diseases |
|---|
| Risk factors with a significant genetic component (heritability) |
| Myocardial infarction (25% to 60%) |
| Stroke |
| Total cholesterol (40% to 60%) |
| HDL-cholesterol (45% to 75%) |
| Total triglycerides (40% to 80%) |
| Body mass index (25% to 60%) |
| Systolic blood pressure (50% to 70%) |
| Diastolic blood pressure (50% to 65%) |
| Lp(a) levels (90%) |
| Homocysteine levels (45%) |
| Type 2 diabetes (40% to 80%) |
| Fibrinogen (20% to 50%) |
| C-reactive protein |
| Gender |
| Age |
| Environmental risk factors |
| Smoking |
| Diet |
| Exercise |
| Infection |
| Fetal environment |
| Air pollution (particulates) |

# II. INSUFFISANCE CARDIAQUE

## A. Définition et place de la maladie dans la population

L'insuffisance cardiaque est un état physiopathologique dans lequel le cœur est incapable d'assurer un débit sanguin adapté à la demande métabolique de l'organisme.

Ce syndrome représente le phénotype évolutif de nombreuses pathologies cardio-vasculaires. Dans les pays développés, les étiologies sont aujourd'hui dominées par la maladie coronaire, les cardiomyopathies dilatées, l'hypertension artérielle et les atteintes des valves du cœur gauche. L'insuffisance cardiaque peut également être la conséquence d'une maladie du péricarde, d'une malformation cardiaque congénitale, d'un trouble du rythme ou de la conduction cardiaque. Plus rarement le dysfonctionnement est lié à une anomalie touchant primitivement le muscle cardiaque comme une mutation génétique codant une protéine essentielle au fonctionnement du cardiomyocyte.

Cet élément déclencheur extrinsèque ou intrinsèque est le point de départ d'une évolution le plus souvent lente et progressive de la maladie. Des anomalies structurales et fonctionnelles cardiaques apparaissent, puis secondairement les premiers symptômes sont perçus par le patient (figure 2).(Hunt et al, 2005) Si l'insuffisance cardiaque au sens strict du terme est liée à la présence de symptômes, la compréhension actuelle de la maladie incite à englober les phases précédentes (pathologie causale, apparition d'anomalies structurales du myocarde) comme les éléments premiers du processus évolutif de la maladie. Cette vision de la maladie souligne que le diagnostic est souvent réalisé tardivement après des mois et souvent des années d'évolution silencieuse. Le développement de stratégies de traitement précoce visant à stopper cette progression avant l'apparition des symptômes est aujourd'hui un objectif essentiel.

L'insuffisance cardiaque est associée à une lourde morbidité et mortalité. Les dépenses liées à l'insuffisance cardiaque représentent actuellement plus de 1% des dépenses médicales totales.(Delahaye et al, 2001) En France, 500 000 à 600 000 personnes sont atteintes d'insuffisance cardiaque. Plus de 30 000 décès par an sont liés à la maladie. La mortalité dans l'insuffisance cardiaque est supérieure à la majorité des pathologies cancéreuses.(Stewart et al, 2001) C'est une pathologie du sujet âgé. En effet, 6 à 10% des plus de 65 ans sont touchés par cette maladie et le risque de développer une insuffisance cardiaque est environ de 20% chez une personne de 40 ans.(Lloyd-Jones et al, 2002; Mosterd et al, 1999) 120 000 nouveaux cas sont déclarés chaque année. L'incidence de la maladie devrait augmenter dans les prochaines années du fait du vieillissement de la population et de la meilleure prise en charge des infarctus du myocarde à leur phase initiale.

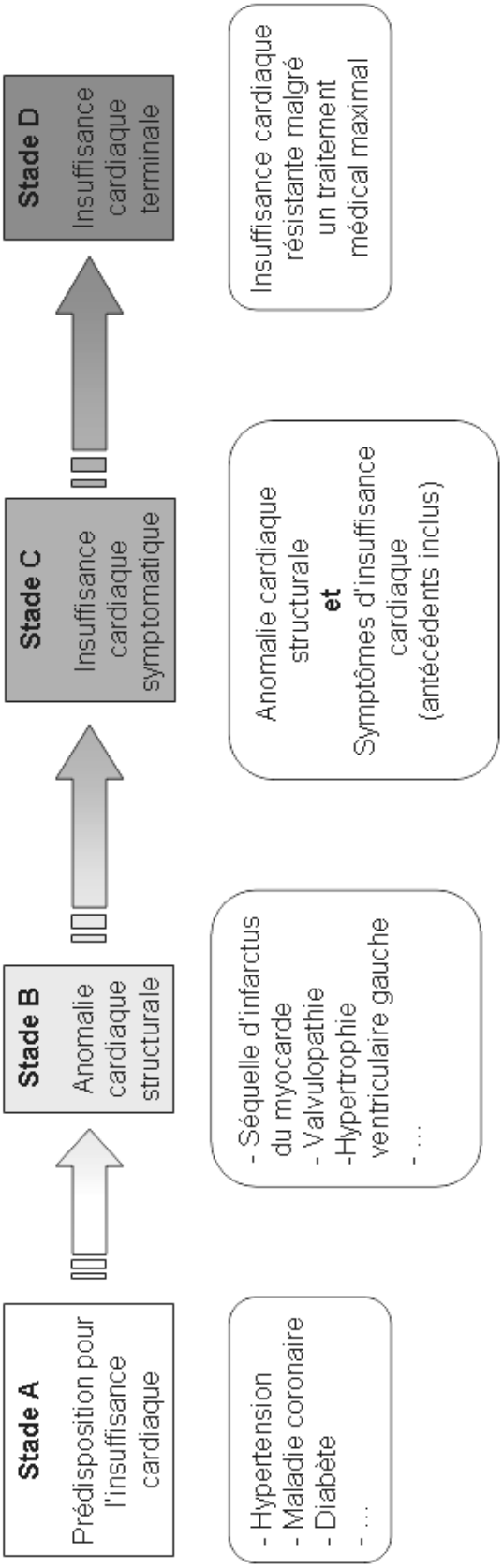



**Figure 2** : Différents stades de l'insuffisance cardiaque.(Hunt et al, 2001)

### B. Dysfonctions systoliques et diastoliques

On distingue classiquement les atteintes de la fonction cardiaque en atteinte systolique et atteinte diastolique suivant que les anomalies observées, essentiellement au niveau du ventricule gauche, prédominent pendant la phase de systole ou de diastole du cycle cardiaque. Cette séparation est en partie théorique car les 2 types de dysfonction coexistent fréquemment. L'insuffisance cardiaque systolique est liée à une perte de la force de contraction (c'est à dire de la capacité d'éjection du sang du cœur vers les vaisseaux périphériques) de tout ou une partie du muscle cardiaque. L'insuffisance cardiaque diastolique est caractérisée par une perte de la capacité du cœur à se remplir normalement pendant la phase de diastole, alors que la fonction systolique est conservée.

La dysfonction diastolique isolée est encore très mal connue. Les données épidémiologiques récentes ayant comparé les 2 types de dysfonction montrent pourtant une gravité proche en terme de mortalité.(Bhatia et al, 2006; Owan et al, 2006) De nombreuses zones d'ombres persistent sur ce sous-groupe qui représenterait près de la moitié des insuffisants cardiaques. Il n'existe pas de vrai critère utilisable en routine pour le diagnostic positif de l'insuffisance cardiaque diastolique. Le diagnostic est porté après élimination d'une dysfonction systolique et on parle plus volontiers d'insuffisance cardiaque à fonction systolique préservée que d'insuffisance cardiaque diastolique. La grande majorité des progrès réalisés dans la compréhension des mécanismes de la maladie et les nouvelles approches thérapeutiques qui en ont découlé sont issus de travaux sur l'insuffisance cardiaque systolique. Il n'est donc pas certain que ces hypothèses physiopathologiques sont également valables pour l'insuffisance cardiaque diastolique. En dehors de rares cas ces nouveaux traitements n'ont pas encore été évalués dans l'insuffisance cardiaque diastolique.(Yusuf et al, 2003) Les travaux de génomique décrits dans la littérature et ceux qui seront présentés dans ce travail sont également globalement centrés sur l'insuffisance cardiaque systolique.

## C. Physiopathologie

Les connaissances sur la physiopathologie de l'insuffisance cardiaque ont considérablement évolué ces dernières années. On estime aujourd'hui que l'insuffisance cardiaque est l'aboutissement d'un long processus au cours duquel le cœur subit un remodelage, essentiellement au niveau du ventricule gauche, qui *in fine* entraîne la défaillance cardiaque.

Ce processus de remodelage peut-être est initié par un grand nombre de facteurs déclencheurs parfois évidents (infarctus du myocarde) ou parfois indéterminés (cardiopathies primitives avec bilan étiologique négatif). Par ailleurs, l'intervention de facteurs d'environnement ou génétiques pour moduler la vitesse de progression du remodelage est très probable.(Feldman et al, 1999) La résultante commune à ces facteurs déclencheurs est l'application sur le muscle cardiaque d'un stress biomécanique dont les déterminants principaux sont une surcharge mécanique systolique et/ou diastolique et l'activation de certains systèmes neuro-hormonaux comme le système sympathique ou le système rénine-angiotensine-aldostérone. Ce stress biomécanique est à l'origine du processus de remodelage. Il est également possible que la persistance de ce stress au cours du temps soit impliquée dans la vitesse d'évolution du remodelage cardiaque vers l'apparition des anomalies structurales puis vers l'insuffisance cardiaque symptomatique.

Le remodelage cardiaque est un processus complexe qui comporte dès le début l'intrication de mécanismes bénéfiques et délétères. Les mécanismes bénéfiques vont tendre à restaurer les capacités globales de travail du myocarde pour maintenir une adaptation du débit cardiaque au besoin de l'organisme. Les mécanismes délétères vont au contraire précipiter le cœur vers la défaillance cardiaque. Schématiquement, tant que cet équilibre entre mécanismes bénéfiques et délétères est préservé, la maladie évolue sur un mode compensé. La rupture de cet équilibre est associée à l'apparition d'une insuffisance cardiaque symptomatique. Historiquement, le processus de remodelage a été décrit au niveau du ventricule gauche dans les suites d'un infarctus du myocarde.(Pfeffer et al, 1990) Le terme est maintenant utilisé quelle que soit l'étiologie de l'insuffisance cardiaque.

Les premières études ont réalisé une description anatomique puis histologique du remodelage. Les principales modifications anatomiques dans le cas d'un trouble de fonction systolique consistent en un amincissement des parois et une dilatation de la cavité du ventricule gauche, parfois associé à un asynchronisme de contraction des différentes parois ventriculaires. Un épaississement des parois du ventricule gauche sans modification systématique du volume de la cavité ventriculaire est observé dans l'insuffisance cardiaque diastolique. Dans tous les cas on note une augmentation de la masse du cœur d'où le nom de remodelage hypertrophique. Au niveau tissulaire, on observe des modifications touchant toutes les cellules du tissu cardiaque, la matrice extracellulaire ainsi que la vascularisation myocardique. Les principales caractéristiques de ce remodelage sont une diminution du nombre de myocytes, une fibrose tissulaire et une hypertrophie du cardiomyocyte dont les propriétés mécaniques sont altérées.

L'avènement des techniques de biologie moléculaire a finalement montré que ce processus de remodelage existait également au niveau moléculaire en association à des modifications de l'expression de nombreux gènes.(Swynghedauw, 1999) Cette modification profonde du programme d'expression génique est quantitative conduisant à une augmentation globale de la synthèse protéique responsable de l'hypertrophie de cellules. Des modifications d'expression géniques plus spécifiques sont également associées comme la réexpression de gènes de la vie fœtale. Des altérations des protéines contractiles du sarcomère et du cytosquelette, des anomalies du cycle du calcium, une modification de la composition de la matrice extracellulaire, la synthèse de facteurs de croissance et de cytokines et enfin une diminution du métabolisme cellulaire sont les principales composantes de ce remodelage moléculaire spécifique.(Swynghedauw, 1999) Parmi toutes ces modifications certaines vont favoriser une adaptation de la cellule à son nouvel environnement comme les modifications des iso-formes de la myosine.(Palmer, 2005) D'autres sont délétères pour le fonctionnement cardiaque comme les perturbations de l'homéostasie calcique de la cellule qui favorisent les troubles du rythme.(Tomaselli et al, 2004) Ce remodelage de l'expression des gènes dans le tissu cardiaque est au centre du processus évolutif de remodelage hypertrophique. Il est donc essentiel de parfaitement identifier au sein de ces altérations celles qui sont délétères pour le fonctionnement cardiaque de manière à

développer des stratégies de traitement ciblées sur ces mécanismes. Il apparaît également fondamental de mieux étudier les signaux cellulaires qui, activés par le stress biomécanique, sont à l'origine des phénomènes délétères et conduisent à l'insuffisance cardiaque.(Hunter et al, 1999)

La biologie moléculaire et en particulier l'utilisation des modèles murins d'invalidations des gènes (knock-out) a commencé d'analyser les voies de signalisation liant stress biomécanique et remodelage cardiaque. Plusieurs modèles transgéniques de souris, invalidées pour une voie de signalisation spécifique et présentant une dysfonction cardiaque atténuée en réponse à un stress biomécanique ont pu ainsi être développés. Ces données renforcent l'idée que l'insuffisance cardiaque est la conséquence de l'activation de certaines voies de signalisation cellulaires délétères activées en réponse à un stress biomécanique.(Esposito et al, 2002) Les différentes voies de signalisation analysées ont pu ainsi être classifiées comme bénéfiques (comme la voie de la protéine kinase C ε)(Dorn et al, 1999; Mochly-Rosen et al, 2000) ou délétères (comme la voie de la protéine kinase A)(Antos et al, 2001) dans l'insuffisance cardiaque ce qui devrait servir de base à de nouveaux développements thérapeutiques.

## D. Diagnostic et traitement actuel de l'insuffisance cardiaque

Le diagnostic d'insuffisance cardiaque est avant tout clinique. Les signes associent principalement une dyspnée, présente initialement à l'effort puis éventuellement au repos, une asthénie et des signes de rétention hydrosodée. L'échographie cardiaque trans-thoracique est un examen central pour poser le diagnostic, identifier une étiologie, préciser le mécanisme de la dysfonction cardiaque (systolique et/ou diastolique). Elle permet aussi d'évaluer le pronostic et peur être réalisée de manière itérative pour évaluer l'effet des traitements instaurés. Des marqueurs biologiques, essentiellement bio-chimiques, de la dysfonction cardiaque sont également de plus en plus utilisés pour le diagnostic et le suivi des patients. Les peptides natriurétiques et en particulier le BNP (B-type Natriuretic Peptide) ont par exemple aujourd'hui une place importante dans la prise en charge des malades.

La prise en charge des malades a fortement changé depuis la mise en évidence du rôle majeur de l'activation neuro-hormonale dans la physiopathologie de l'insuffisance cardiaque. L'utilisation de médicaments comme les inhibiteurs de l'enzyme de conversion ou les bêtabloqueurs utilisés pour bloquer les stimuli neuro-hormonaux délétères ont montré un effet bénéfique sur la mortalité et la morbidité des patients en insuffisance cardiaque.(Bouzamondo et al, 2003; Garg et al, 1995; Jong et al, 2002) En parallèle des approches non pharmacologiques de l'insuffisance cardiaque comme la re-synchronisation cardiaque,(Cleland et al, 2005) l'utilisation des défibrillateurs automatiques implantables pour prévenir la mort subite cardiaque(Moss et al, 2002) et l'utilisation (temporaire(Birks et al, 2006) ou définitive(Rose et al, 2001)) des systèmes d'assistance ventriculaire mécanique font aujourd'hui partie de l'arsenal thérapeutique en association avec le traitement pharmacologique. La transplantation cardiaque reste une alternative efficace en cas d'insuffisance cardiaque réfractaire à tous ces traitements. Malgré tous ces progrès la morbidité et la mortalité de la maladie restent élevées et la fréquence de l'insuffisance cardiaque augmente dans la population justifiant le développement de nouvelles approches.

### E. Vers une prise en charge plus individualisée de l'insuffisance cardiaque

L'identification précoce des patients qui peuvent rapidement progresser vers une insuffisance cardiaque sévère ou qui sont à haut risque de mort subite cardiaque est un objectif central de la prise en charge des malades. Cette stratégie est fondamentale pour adapter la prise en charge thérapeutique avec le meilleur ratio bénéfice/risque pour chaque patient et obtenir une meilleure balance coût/efficacité à l'échelle de la population. Par exemple, à la suite de l'étude MADIT II,(Moss et al, 2002) il est actuellement recommandé l'implantation préventive d'un défibrillateur automatique implantable (DAI) pour les patients porteurs d'une insuffisance cardiaque d'origine ischémique avec une fraction d'éjection inférieure à 30-35%.(Swedberg et al, 2005) En pratique, on sait que sur une année seulement 5% de ces patients vont réellement présenter des troubles du rythme ventriculaire

conduisant au déclenchement du DAI. Par ailleurs, le coût d'un DAI étant de plusieurs milliers d'euros, la taille de population cible est trop importante pour envisager l'implantation systématique de tous les patients. La définition de marqueurs prédictifs fiables du risque de troubles du rythme chez ces patients pourrait donc permettre de mieux cibler les indications de ce traitement. Si l'étude du remodelage et de ces mécanismes est fondamentale pour susciter le développement de nouveaux traitements de l'insuffisance cardiaque, elle est également une piste intéressante pour l'identification de marqueurs biologiques comme facteurs pronostiques de la maladie.

La sévérité de l'insuffisance cardiaque est fortement liée à l'intensité du processus de remodelage cardiaque qui survient dans l'évolution de la maladie.(Francis, 2001) De fait, les marqueurs cliniques utilisés aujourd'hui sont des marqueurs indirects de la progression du remodelage cardiaque. La fraction d'éjection du ventricule gauche est par exemple un marqueur de diminution de la fonction contractile et la natrémie reflète l'activation du système rénine-angiotensine-aldostérone. Ces marqueurs ne sont donc individuellement que le reflet d'un des aspects de la complexité du remodelage cardiaque. Des modèles combinant plusieurs de ces marqueurs ont donc été développés pour intégrer différents composants du remodelage.(Aaronson et al, 1997; Levy et al, 2006; Smits et al, 2003) La prédiction du risque d'évolution péjorative à l'échelle de l'individu reste cependant encore médiocre. Il existe donc un besoin de développer de nouveaux indicateurs pronostiques permettant de mieux prendre en charge les malades et en particulier de détecter plus précocement les formes les plus sévères de la maladie.
De même la possibilité de détecter précocement au niveau moléculaire l'apparition des éléments délétères du remodelage cardiaque avant l'apparition des altérations structurales cardiaques permettrait la mise en place d'une stratégie de dépistage des patients à haut risque d'insuffisance cardiaque et le développement de stratégies thérapeutiques préventives.

# III. FIBRILLATION AURICULAIRE

## A. Définition et place de la maladie dans la population

La fibrillation auriculaire est un trouble du rythme cardiaque caractérisé par une activation rapide et irrégulière des oreillettes, induisant une activité ventriculaire rapide et irrégulière. Il s'agit du trouble du rythme cardiaque le plus fréquent dans la population générale avec une prévalence d'environ 1%. Il touche essentiellement les sujets âgés. La prévalence est de 2,3% chez les plus de 40 ans et de 5,9% chez les plus de 65 ans. Environ 70% des patients en fibrillation auriculaire sont dans la tranche d'âge 65-85 ans.(Feinberg et al, 1995) L'incidence de la fibrillation auriculaire croit également avec l'âge de 0,2 à 0,3% de 55 à 64 ans à 3,5% après 85 ans.(Falk, 2001)

La fibrillation auriculaire constitue un facteur de risque indépendant de mortalité. Le risque relatif de mortalité lié à la fibrillation auriculaire est de 1,5 chez l'homme et 1,9 chez la femme.(Benjamin et al, 1998) La fibrillation auriculaire est fortement associée au risque d'accident vasculaire cérébral par mécanisme thromboembolique à point de départ cardiaque. Elle est retrouvée chez 20% des patients ayant présenté un accident vasculaire cérébral (AVC) et le risque d'AVC est multiplié par 5 chez les patients en fibrillation auriculaire.(Wolf et al, 1991) La fibrillation auriculaire est également un facteur aggravant de l'insuffisance cardiaque dont elle accroît la mortalité.(Middlekauff et al, 1991)

## B. Présentations cliniques de la fibrillation auriculaire

Dans la majorité des cas, la fibrillation auriculaire est associée à une pathologie cardiaque préexistante, telle qu'une pathologie valvulaire, une hypertension artérielle ou une insuffisance cardiaque. Dans 15% des cas cependant, aucune pathologique associée n'est diagnostiquée. On parle alors de fibrillation auriculaire isolée. Des formes génétiques de fibrillation auriculaire dont la prévalence semble faible ont été également décrites.

Selon sa présentation clinique la fibrillation auriculaire est qualifiée de :(Fuster et al, 2006)

1) fibrillation auriculaire paroxystique : les épisodes d'arythmie s'arrêtent spontanément en quelques minutes à quelques jours.

2) fibrillation auriculaire persistante : les épisodes durent plus d'une semaine, ne cèdent pas spontanément mais peuvent être traités par cardioversion.

3) fibrillation auriculaire permanente ou fibrillation auriculaire chronique : l'arythmie se maintient sur le long terme (> 1 an) et/ou la cardioversion est inefficace.

Il existe en fait un réel continuum entre ces 3 formes de la maladie un patient en fibrillation auriculaire paroxystique progressant souvent vers une fibrillation auriculaire chronique.

La tachycardie atriale entraîne une perte de la systole mécanique des oreillettes qui permet de vidanger de manière active le sang des oreillettes vers les ventricules en fin de diastole ventriculaire. La fibrillation auriculaire favorise donc une stase du sang dans les cavités atriales et la formation pathologique d'un caillot sanguin dans l'oreillette gauche. Ceci explique en partie le risque d'AVC par mécanisme thromboembolique associé à la fibrillation auriculaire et la nécessité d'un traitement anticoagulant permanent pour la grande majorité des patients. La perte de la systole auriculaire modifie également le fonctionnement hémodynamique du cœur. La systole auriculaire représente environ 15% du remplissage ventriculaire d'un cœur sain et jusqu'à 40% d'un cœur présentant une dysfonction diastolique. L'apparition d'une fibrillation auriculaire dans l'insuffisance cardiaque entraînera donc une déstabilisation immédiate du fonctionnement d'un cœur défaillant.

### C. Physiopathologie

La fibrillation auriculaire est liée le plus souvent à la présence de circuits de réentrées électriques uniques ou multiples dans le tissu atrial ; la présence de foyers ectopiques auriculaires est également décrite (figure 3).(Nattel, 2002) Il semble que tous ces mécanismes puissent coexister et interagir dans la genèse et le maintient de l'arythmie.

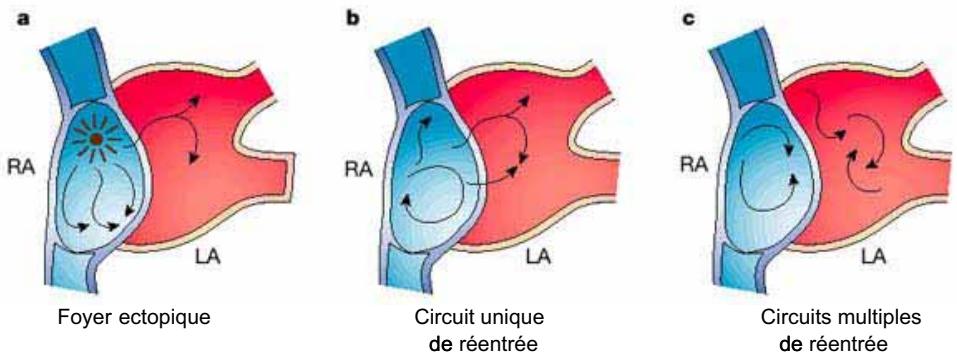
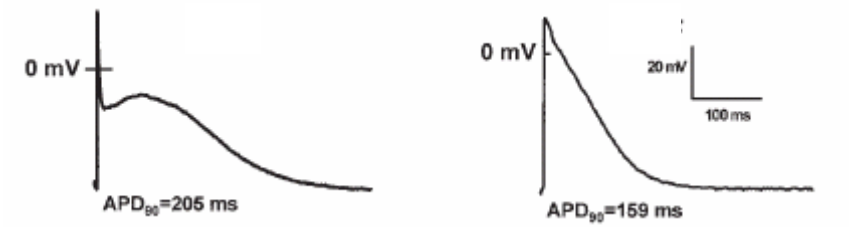



**Figure 3** : Différents Mécanismes de propagation de l'influx électrique dans la fibrillation auriculaire.(Nattel, 2002)
RA, oreillette droite ; LA, oreillette gauche.

Les études électrophysiologiques sur cellules atriales ont montré que la fibrillation auriculaire est associée à une diminution de la durée du potentiel d'action atrial et de sa période réfractaire, ainsi qu'à la perte de sa phase de plateau (Figure 4). La vitesse de conduction est également altérée dans le tissu atrial des patients en fibrillation auriculaire ce qui peut également favoriser les réentrées.




Rythme sinusal Fibrillation auriculaire

**Figure 4** : Potentiels d'action enregistrés à partir de cellules atriales droites de patients en rythme sinusal (gauche) et de patients en fibrillation auriculaire (droite).(Dobrev et al, 2003) On observe clairement le raccourcissement de la durée du potentiel d'action et la perte de la phase de plateau.

La découverte de ces modifications ont conduit à parler de remodelage électrique atrial dans la fibrillation auriculaire. Il s'associe à un remodelage de l'activité des canaux ionique dans la cellule atriale en partie expliqué par des modifications d'expression des gènes codant ces canaux. Le remodelage est non seulement électrique mais touche également les fonctions contractiles et structurales

des oreillettes.(Allessie et al, 2002) Il est à la base du mécanisme d'auto-entretient du trouble du rythme. Ce mécanisme a été mis en évidence grâce à un modèle de fibrillation auriculaire induit par stimulation atriale rapide chez la chèvre. Après stimulation, le temps passé en fibrillation auriculaire était proportionnel au temps de fibrillation imposé par la stimulation, l'augmentation et la répétition des stimulations conduisant à la persistance du trouble du rythme sur des durées de plus de 24h. La fibrillation auriculaire conduit donc elle-même à sa persistance et sa permanence ("Atrial fibrillation begets atrial fibrillation").(Wijffels et al, 1995)

Le déclenchement du trouble du rythme est maintenant vu comme l'association d'éléments déclencheurs (« trigger ») et d'un substrat favorable au maintient de l'arythmie. Le rôle des veines pulmonaires dans l'initiation de la fibrillation auriculaire est maintenant bien identifié.(Haissaguerre et al, 1998) Elles peuvent être le site d'initiation ("gâchette") d'une tachycardie atriale rapide favorisant la transition vers des circuits de réentrée. Le tissu cardiaque présente, au niveau des veines pulmonaires, des propriétés électriques arythmogènes, et peut favoriser des réentrées atriales. Le système nerveux autonome est également un déclencheur bien identifié de fibrillation auriculaire. La présence de fibrose dans le tissu atrial est considérée comme un substrat de la fibrillation auriculaire qui pourrait favoriser les phénomènes de réentrées. Des zones de fibrose sont couramment retrouvées chez les patients en fibrillation auriculaire associée à une pathologie valvulaire mitrale, à une insuffisance cardiaque ainsi que chez les patients en fibrillation auriculaire idiopathique. Chez les patients en fibrillation auriculaire, l'activation de la voie pro-fibrotique des Mitogen Activated Protein-kinases (MAP-kinases) par l'angiotensine II et l'existence de phénomènes apoptotiques dans le tissu atrial favorisent la progression des lésions de fibrose.(Aime-Sempe et al, 1999; Goette et al, 2000)

L'importance des phénomènes thromboemboliques dans la fibrillation auriculaire a entrainé la réalisation d'études visant à identifier des facteurs favorisant la formation de thrombus dans les cavités atriales en association avec la stase sanguine. Ces travaux ont conduit à identifier des modification de l'activation plaquettaire, de la coagulation ou de l'endothélium vasculaire qui favorisent l'existence dans la fibrillation auriculaire d'un état pro-thrombotique systémique.(Kumagai et al, 1990; Li-Saw-Hee et al, 2000) Une meilleure

compréhension de cet état pro-thrombotique et l'identification d'autres mécanismes impliqués dans le risque thromboembolique lié à la fibrillation auriculaire est un objectif majeur qui permettrait l'identification de nouvelles cibles thérapeutiques.

### D. Traitement

Les approches pharmacologiques classiques de la fibrillation auriculaire paroxystique ou persistante ont pour objectif le maintient du rythme sinusal par modification des propriétés électriques cardiaques. Cependant, les traitements antiarythmiques disponibles actuellement ne sont pas spécifiques de l'activité électrique atriale et peuvent donc paradoxalement avoir des effets proarythmiques sur l'activité électrique ventriculaire. A l'heure actuelle, l'amiodarone est l'antiarythmique le plus efficace et le plus utilisé dans le traitement de la fibrillation auriculaire. Son efficacité est tout de même relative puisque seul 69% des patients ne présentent pas de nouvel épisode d'arythmie après un an de traitement.(Roy et al, 2000) Parallèlement, les thérapeutiques non-pharmacologiques prennent une part croissante dans le traitement de la fibrillation auriculaire. Les techniques de radio-fréquence permettent d'isoler des foyers ectopiques du reste des oreillettes (déconnexion des foyers veineux pulmonaires), ou de supprimer le substrat arythmogène en divisant fonctionnellement l'oreillette (segmentation atriale).

Les avancées dans la connaissance des substrats de la fibrillation auriculaire commencent à fournir de nouvelles cibles thérapeutiques potentielles et devraient aboutir à l'élaboration d'approches thérapeutiques plus efficaces et plus sûres. Plusieurs essais cliniques ont par exemple suggéré que l'utilisation des inhibiteurs de l'enzyme de conversion de l'angiotensine II permettrait de prévenir le développement de la fibrillation auriculaire.(Levy, 2006) Cet effet bénéfique potentiel des inhibiteurs de l'enzyme de conversion de l'angiotensine II pourrait par exemple s'exercer sur la voie pro-fibrotique des MAP-kinases activée par l'angiotensine II.(Goette et al, 2000)

# IV. BIOMARQUEURS

## A. Définitions :

Un biomarqueur désigne une caractéristique mesurée objectivement et évaluée comme indicateur de processus physiologiques normaux ou pathologiques, ou d'une réponse à un traitement.(Biomarkers Working Group, 2001) Les biomarqueurs représentent un domaine d'application très vaste et sont aujourd'hui omniprésents tant au niveau des soins que de la recherche médicale. Ils sont utilisés comme :

- Outil diagnostic, pour l'identification des patients porteurs d'une maladie ou d'une anomalie
- Outil d'évaluation du stade évolutif ou du niveau de gravité d'une maladie
- Outil d'évaluation du pronostic d'un malade
- Outil de prédiction et évaluation de la réponse à un traitement
- Outil d'évaluation de la toxicité d'un médicament.

L'intérêt des biomarqueurs est fondamental pour permettre la transition de la médecine actuelle globalement appliquée de manière identique à l'ensemble d'une cohorte de patients vers l'émergence d'une pratique médicale mieux adaptée à chaque individu.

## B. Notion de critère de substitution :

Dans les essais cliniques, les biomarqueurs sont de plus en plus utilisés à titre de critère de jugement de substitution (« surrogate end-point ») à la place d'un critère de jugement clinique (« clinical end-point »). L'utilisation de ce critère de substitution est subordonnée à sa capacité à prédire le critère clinique qu'en général il précède dans le temps. L'utilisation de ces critères permet donc un gain de temps (et donc d'argent) dans la réalisation des études d'intervention thérapeutique. Le taux sérique de cholestérol total ou de LDL-cholestérol est par exemple utilisé comme marqueur de substitution aux évènements coronariens. On peut d'ailleurs noter que les biomarqueurs utilisés comme critères de substitution dans les études cliniques sont

souvent utilisés dans la pratique médicale pour mesurer la réponse au traitement à l'échelle individuelle.

Il faut cependant signaler que s'il existe un lien entre le traitement et le critère de jugement de substitution cela ne signifie pas forcement un lien entre le traitement et le critère de jugement clinique. Les résultats de l'étude CAST sont, sur ce point, démonstratifs.(Echt et al, 1991) Les troubles du rythme ventriculaire non soutenus sont un facteur de risque de mort subite dans les suites d'un infarctus du myocarde. Cette étude a évalué l'utilisation de 3 médicaments anti-arythmiques dans les suites d'un infarctus du myocarde chez des patients présentant des extrasystoles ventriculaires. Après vérification de l'effet « suppresseur » des anti-arythmiques sur les extrasystoles, les patients étaient randomisés entre différents groupes de traitement et un groupe placebo. Contre toute attente, l'étude a été arrêtée prématurément pour un excès de mortalité dans les groupes de traitement par rapport au groupe placebo. Les liens entre le critère de jugement de substitution et le critère clinique et entre le traitement et le critère de jugement de substitution avaient portant été préalablement parfaitement établis.

### C. Marqueurs biologiques en cardiologie

En cardiologie, les biomarqueurs utilisés sont issus de techniques aussi diverses que l'électrocardiogramme, l'échographie cardiaque ou l'épreuve d'effort. L'analyse des tissus biologiques, essentiellement le sang périphérique et pour une moindre part le tissu cardiaque, représente une part très importante des biomarqueurs utilisés en routine. Ces marqueurs biologiques sont donc une sous-catégorie de l'ensemble des biomarqueurs. Les termes « biomarqueurs » et « marqueurs biologiques » ne sont donc pas synonymes.

Le développement des techniques de biologie moléculaire a permis le développement de marqueurs biologiques dans de nombreuses pathologies cardiaques et a déjà fait progresser fortement la prise en charge des malades. L'utilisation des marqueurs biologiques de nécrose myocardique dans la prise en charge des syndromes coronariens aigus est certainement l'exemple le plus parlant. L'identification de protéines cardiaques spécifiques comme les troponines permettent

aujourd'hui de disposer de tests sensibles et fiables de nécrose myocardique à partir d'un échantillon sanguin.(Cummins et al, 1987)

### D. Marqueurs biologiques et insuffisance cardiaque

Les marqueurs biologiques décrits dans l'insuffisance cardiaque sont nombreux. Ils sont essentiellement mesurés dans le sang périphérique. Une liste non exhaustive est présentée dans tableau II.

L'exemple majeur de l'apport des marqueurs biologiques dans l'insuffisance cardiaque est certainement apporté par les peptides natriurétiques et en particulier le BNP (B-type natruretic peptide). La synthèse de pro-BNP par les myocytes ventriculaires droits et gauches est stimulée par l'augmentation de la contrainte mécanique appliquée sur la paroi ventriculaire et par différentes stimulation neuro-hormonales comme celles du système Rénine-Angiotensine-Aldostérone ou du système Arginine-Vasopressine. Sous l'action d'endopeptidases, le pro-BNP est clivé en BNP (32 AA) et NT-proBNP (Nterminal-proBNP - 76 AA). Les 2 peptides BNP et NTpro-BNP sont ensuite sécrétés dans la circulation sanguine. Seul le BNP est un peptide physiologiquement actif. Il est vasodilatateur, natriurétique et diurétique. Des formes recombinantes de BNP sont utilisées dans le traitement de l'insuffisance cardiaque décompensée.(Yancy et al, 2004)

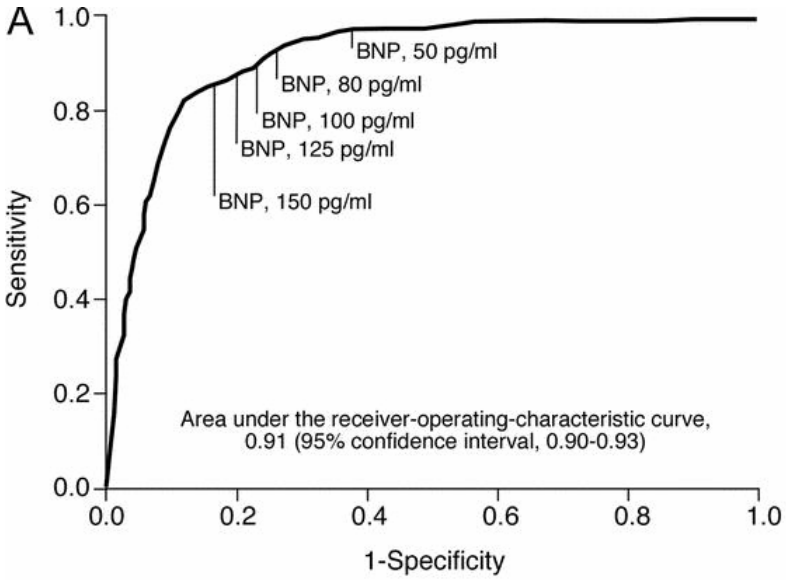
Les dosages sanguins du BNP et du NTpro-BNP ont été évalués comme biomarqueurs dans de nombreuses applications comme le diagnostic de l'insuffisance cardiaque aiguë, la détection de la dysfonction ventriculaire gauche asymptomatique ou le pronostic de l'insuffisance cardiaque symptomatique. Leur utilisation dans les essais cliniques comme critère de jugement de substitution pour évaluer le statut fonctionnel des patients et l'efficacité thérapeutique est également envisagée. Ces marqueurs possèdent une excellente valeur prédictive positive et négative pour le diagnostic d'insuffisance cardiaque dans le contexte d'une dyspnée aiguë et leur usage est maintenant recommandé pour améliorer la fiabilité du diagnostic dans cette situation (figure 5).(Christenson, 2007)

Plusieurs éléments pourraient cependant limiter l'utilisation du BNP chez certains patients. (i) Il existe une variabilité biologique importante du dosage sanguin du BNP et du NTpro-BNP. Les coefficient de variation intra-individuels du BNP sont supérieurs à 40% et sont de 33% pour le NTpro-BNP.(Wu et al, 2003) La conséquence est que dans le cadre de dosages répétés de ces marqueurs chez les patients en insuffisance cardiaque seules des variations de 130% pour le BNP et 90% pour le NTpro-BNP peuvent être considérées comme statistiquement significatives. (ii) Le dosage du BNP varie en fonction de l'indice de masse corporelle. Chez les insuffisants cardiaques obèses les niveaux de BNP sont plus bas et peuvent être à l'origine de faux négatifs dans le diagnostic de décompensation cardiaque.(Mehra et al, 2004) (iii) Les peptides natriurétiques sont des marqueurs reflétant le niveau des pressions de remplissage ventriculaire. Leur spécificité diagnostique dans l'insuffisance cardiaque est donc partiellement limitée car d'autres pathologies comme l'embolie pulmonaire sont associées à une augmentation des pressions de remplissage ventriculaire et donc une augmentation de taux sanguin de peptides natriurétiques.(Kurose et al, 1997) Leur sensibilité diagnostique peut également être limitée par le fait qu'ils ne sont le reflet que d'un des aspects du remodelage cardiaque. L'utilisation de combinaisons de marqueurs reflétant chacun des aspects différents du remodelage est en cours d'évaluation et semble apporter une amélioration sur ce point.

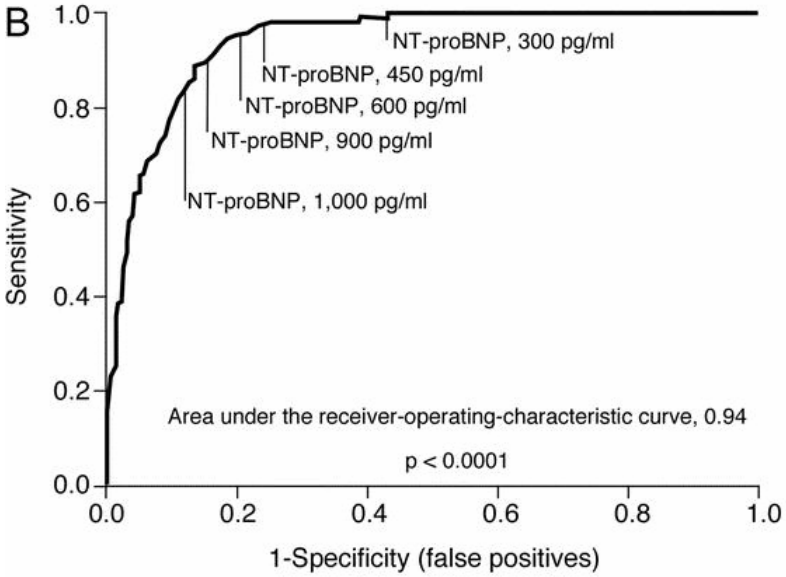
**Tableau II** : Marqueurs biologiques utilisés actuellement dans la prise en charge de l'insuffisance cardiaque humaine.(Christenson, 2007)

| **Selected Biochemical Markers Currently Available or Under Study for Clinical Diagnosis, Management, and Risk Stratification of Heart Failure** |
|---|
| Standard laboratory markers |
| Sodium |
| Blood urea nitrogen |
| Serum creatinine |
| Hemoglobin |
| Leukocyte count |
| Total lymphocyte count |
| Serum albumin |
| Total bilirubin |
| Uric acid |
| Red blood cell distribution width |
| Neurohormones |
| Catecholamines (norepinephrine, epinephrine) |
| Renin, ACE activity, angiotensin II, and aldosterone |
| Natriuretic peptides (ANP, BNP, C-type, N-terminal proANP, N-terminal proBNP, mid-regional pro-ANP) |
| Endothelin-1 |
| Vasopressin/copeptin |
| Cardiotrophin-1 |
| Novel vasodilators (adrenomedullin and mid-regional pro-adrenomedullin, urotensin-II, urocortin) |
| Inflammatory biomarkers |
| High-sensitivity C-reactive protein |
| Myeloperoxidase |
| Galectin-3 |
| Fatty acid binding protein |
| Soluble ST2 receptor |
| Tumor necrosis factor-alpha (TNF-$\alpha$) and receptors |
| Interleukin-6 (IL-6) |
| Growth differentiation factor 15 (GDF-15) |
| Osteopontin |
| Metabolic biomarkers |
| Leptin |
| Adiponectin |
| Ghrelin |
| Apelin |
| Insulin-like growth factor-1 (IGF-1) |
| Other miscellaneous biomarkers |
| G-protein coupled receptor kinase-2 (GRK-2) |
| Cardiac troponin I or troponin T |
| Myotrophin |

**Figure 5** : Courbes ROC (Receiver operator characteristic) du BNP (A) et du NTpro-BNP (B) pour le diagnostic d'insuffisance cardiaque dans un bilan de dyspnée aigue.(Christenson, 2007)




| BNP pg/ml | Sensitivity | Specificity | Positive Predictive Value | Negative Predictive Value | Accuracy |
|---|---|---|---|---|---|
| 50 | 97% | 62% | 71% | 96% | 79% |
| 80 | 93% | 74% | 77% | 92% | 83% |
| 100 | 90% | 76% | 79% | 89% | 83% |
| 125 | 87% | 79% | 80% | 87% | 83% |
| 150 | 85% | 83% | 83% | 85% | 84% |




| Cut Point pg/ml | Sensitivity | Specificity | Positive Predictive Value | Negative Predictive Value | Accuracy |
|---|---|---|---|---|---|
| 300 | 99% | 68% | 62% | 99% | 79% |
| 450 | 98% | 76% | 68% | 99% | 83% |
| 600 | 96% | 81% | 73% | 97% | 86% |
| 900 | 90% | 85% | 86% | 94% | 87% |
| 1000 | 87% | 86% | 78% | 91% | 87% |

## E. Marqueurs biologiques et fibrillation auriculaire :

Le développement de marqueurs biologiques dans la fibrillation auriculaire est aujourd'hui centré sur la prédiction à l'échelle individuelle du risque d'accident vasculaire cérébral par mécanisme thromboembolique, principale cause de morbidité et mortalité liées à cette pathologie. La somme de travaux réalisés est encore modeste et les résultats obtenus sont encore très préliminaires. Ils ne permettent pas actuellement d'envisager une utilisation d'un de ces marqueurs en pratique clinique.

Quelques marqueurs biologiques de l'état pro-thrombotique systémique associé à la FA ont été identifiés dans le sang périphérique.(Becker, 2005) Ils reflètent :

- L'activation de la coagulation
- L'activation plaquettaire
- Des lésions de l'endothélium et une inflammation

Parmi ces altérations, le niveau d'augmentation du facteur von Willebrand ou des D-dimères pourraient à terme s'affirmer comme des marqueurs du risque d'AVC.(Conway et al, 2003; Vene et al, 2003)

# V. PROFILS D'EXPRESSION GENIQUE EN PATHOLOGIE CARDIOVASCULAIRE

## A. Développement de la génomique fonctionnelle

L'essor de procédés permettant le séquençage du génome humain a changé notre approche des gènes et de leur fonctionnement. Les progrès continus des techniques de séquençage, aboutissant à une robotisation quasi complète du procédé, permettent aujourd'hui d'en faire une technique fiable et surtout rapide. Le séquençage complet de virus ou de bactéries a commencé en 1982. Il a constitué une première étape avant le séquençage de génomes plus vastes. La notion de génomique est apparue en 1986.(Hieter et al, 1997) Elle caractérisait initialement une discipline scientifique dont l'objectif était de cartographier, séquencer et analyser les génomes. C'est dans ce contexte que le projet HGP (Human Genome Project) consistant en un séquençage exhaustif du génome humain est né à la fin des années 80. La séquence de plus de 90% du génome humain a été publiée simultanément en 2001 par un groupe privé et un consortium international.(Lander et al, 2001; Venter et al, 2001) Les données ont ensuite été progressivement complétées permettant de disposer d'un inventaire le plus complet possible des gènes humains.(HGP, 2004) Une vue d'ensemble des résultats est présentée dans le tableau III.

**Tableau III** : Vue d'ensemble du génome humain.(HGP, 2004)

| | |
|---|---|
| Taille du génome | 3.08 Gbases |
| Taux d'erreur dans la séquence | 1/100000 bases |
| | |
| Nombre de gènes identifiés | 19438 |
| Nombre de gènes (prédits et identifiés) | 22287 |
| Nombre moyen d'exon par gène | 10,4 |
| | |
| Régions traduites en protéines | 34 Mbases (1,1%) |
| Régions transcrites non traduites | 21 Mbases (0.7%) |

Rapidement, la notion de génomique s'est étendue de la cartographie et séquençage d'un génome à la détermination de la fonction de ses composants. Les termes de génomique structurale et de génomique fonctionnelle ont ainsi été avancés. La génomique structurale représente la phase initiale de l'analyse d'un génome et a un objectif terminal clair : la détermination d'une carte génétique, physique et des transcrits d'un organisme. La séquence linéaire d'Acides DésoxyriboNucléiques (ADN) complète d'une espèce est la base du résultat. La génomique fonctionnelle (ou post-génomique) utilise la formidable ressource de la génomique structurale pour développer une analyse globale du génome permettant d'assigner des fonctions aux gènes. La finalité n'est pas seulement d'attribuer à chaque gène des caractéristiques fonctionnelles, mais de définir l'organisation et le contrôle des processus génétiques qui s'associent pour faire le fonctionnement physiologique ou pathologique d'une cellule, d'un organe ou d'un organisme.(Vukmirovic et al, 2000) La stratégie fondamentale est d'étendre le champ d'investigation biologique de l'étude de quelques gènes ou protéines à l'ensemble des gènes et protéines de manière simultanée et systématique. La génomique fonctionnelle en proposant une vision moins parcellaire des pathologies se différencie d'autres méthodes d'études des relations entre gènes et maladies. Il s'agit d'évaluer le fonctionnement global des gènes dans une affection quelle que soit le ou les évènements déclencheurs.

Avant l'avènement de la génomique, la possibilité d'identifier un ensemble significatif de gènes interagissant dans le développement de pathologies cardiaques complexes et multifactorielles comme l'insuffisance cardiaque restait très limité. L'approche traditionnelle de génétique inverse a permis d'identifier des locus ou gènes impliqués dans les maladies cardiovasculaires à transmission mendéliennes comme les cardiopathies hypertrophiques familiales. Ces maladies pouvant résulter d'une anomalie unique comme une mutation génétique ou une délétion d'une partie du gène. Cependant dans la majorité des cas, Les pathologies cardio-vasculaires résultent de l'accumulation et de la combinaison d'un nombre souvent important de variations génétiques. L'expression de ces phénotypes dépend de plus de l'interaction avec des facteurs d'environnement comme l'alimentation, le style de vie (consommation de tabac, niveau socio-économique,…) et des facteurs de risque cardiovasculaires comme l'hypertension artérielle et le diabète eux-mêmes favorisés

par d'autres variations génétiques. Ces variations génétiques sont souvent des modifications d'une base dans la séquence du gène (comme les « single nucleotide polymorphism » - SNP) qui individuellement sont la cause d'un changement très modeste dans l'abondance et/ou la fonction de la protéine correspondante.(Feuk et al, 2006) Ce n'est que la combinaison des effets de ces variations génétiques en interaction avec l'environnement de l'individu qui va conduire ou non au développement du phénotype malade.(Sing et al, 2003) La faible contribution de chaque variations génétique au phénotype final va rendre difficile la mise évidence gène par gène de la relation gène-phénotype. L'objectif majeur est d'identifier l'ensemble des gènes qui collectivement vont contribuer à la physiopathologie de maladies cardiovasculaires multi-géniques. En parallèle du développement de l'approche de génomique fonctionnelle, le développement de nouvelles technologies comme les puces à ADN permet aujourd'hui d'utiliser les ressources du séquençage du génome humain pour tenter de modéliser la physiologie et la physiopathologie humaine. Parmi ces nouvelles approches les études d'expression génique sont aujourd'hui en fort developpement.

### B. Etudes d'expression génique

Cette approche utilise la séquence primaire de l'ADN pour étudier les conditions de son expression. Pour comprendre la fonction d'un gène, le fait de savoir quand, où et comment il est exprimé est fondamental pour comprendre l'activité et le rôle biologique de sa protéine correspondante. Il existe une connexion étroite entre l'activité biologique du produit d'un gène et son niveau d'expression. Chaque gène est exprimé spécifiquement dans une cellule donnée dans des conditions précises. A un niveau individuel, les données de la biologie moléculaire ont déjà mis en évidence, pour la plupart des gènes, cette relation étroite entre niveau d'expression et fonction.(Brown et al, 1999; Lockhart et al, 2000) L'objectif est donc d'observer les variations d'expression de dizaine de milliers de gènes de manière simultanée dans différentes situations. L'avantage de cette analyse est moins une analyse gène par gène qui peut être réalisée par d'autres approches que d'obtenir une vue d'ensemble de l'expression du génome (profil d'expression génique) et surtout de ces variations dans différentes situations. L'ensemble des

gènes exprimés dans une cellule détermine la composition et la fonction de cette cellule. L'étude des variations d'expression de groupes de gènes définis dans différentes conditions (profil dynamique) va augmenter la connaissance des fonctions de ces gènes sur le plan individuel et aussi celle de leurs interactions. L'observation de gènes dont l'expression varie de manière identique dans différentes situations fait en effet supposer l'existence de mécanismes communs de régulations et/ou de caractéristiques fonctionnelles proches.

Les modifications d'expression de groupes fonctionnels de gènes peuvent ainsi apporter des clés sur les événements moléculaires physiologiques et pathologiques qui surviennent dans une cellule. Dans ce contexte, l'étude des variations d'expression du génome entre des individus sains et malades permettra de décrire l'ensemble des altérations associées au phénotype malade. Elle permettra de mieux préciser les causes et conséquences moléculaires des maladies. De nouveaux produits de gènes pourront être détectés comme des cibles thérapeutiques et le mode d'action des traitements déjà utilisés comme ceux en développement sera également éclairci. L'étude des variations d'expression génique est donc susceptible d'apporter une meilleure prise en charge des maladies cardiovasculaires humaines.

La fonction d'un gène s'exprime souvent par l'intermédiaire d'une protéine. La transcription et la traduction sont les deux étapes majeures qui permettent cette transition du gène vers la protéine (figure 6). Sous l'effet de stimuli appropriés agissant sur les séquences promotrice d'un gène, une première phase de transcription est déclenchée dans le noyau. Des Acides RiboNucléiques messagers (ARNm) sont ainsi produits à partir de la séquence ADN du gène. Plus la stimulation des promoteurs est forte plus le nombre de copies ARNm du même gène sera important. Ces ARNm subissent ensuite un phénomène de maturation. L'épissage a pour but l'élimination des parties introniques du gène et l'assemblage des différents exons pour former l'ARNm définitif. Cet épissage est variable pour un même transcrit initial ; les introns sont toujours éliminés, mais l'assemblage des exons est variable réalisant ainsi, pour un même gène, plusieurs produits différents : il s'agit du phénomène d'épissage alternatif.

L'ensemble des ARNm présents dans une cellule à un moment donné définit le transcriptome de la cellule. A la différence du génome, la « composition » du transcriptome est en constante modification sous l'effet des nouveaux ARNm produits et de la dégradation enzymatique d'autres ARNm. La demi-vie d'un ARNm est brève de quelques minutes à plusieurs jours. Le devenir des ARNm est variable :

- Dégradation précoce par des mécanismes physiologiques cellulaires
- Interaction avec d'autres ARN dans un but de régulation
- Traduction en protéines

La traduction est un mécanisme cytoplasmique par lequel les ARNm sont utilisés pour la production de leur protéine correspondante. Un même ARNm peut être à l'origine de la création de plusieurs protéines identiques. L'ensemble des protéines présentes dans une cellule définit le protéome de cette cellule. Ces protéines subissent souvent des modifications post-traductionelles (phosphorylation par exemple) et acquièrent ainsi leur entière capacité à exercer une fonction précise. L'ensemble de ces fonctions exercées dans une cellule donnée, à un moment précis, définit le physiome de la cellule.

A chaque étape de cette transformation, l'étude du transcriptome, du protéome ou du physiome est le reflet de l'état d'une cellule et de son fonctionnement. Une étude globale et comparative à un de ces trois niveaux, dans le contexte de l'étude d'une pathologie, va apporter des éléments de réponse sur les mécanismes fondamentaux de la maladie et ainsi permettre le développement d'éléments diagnostics, pronostics et thérapeutiques.

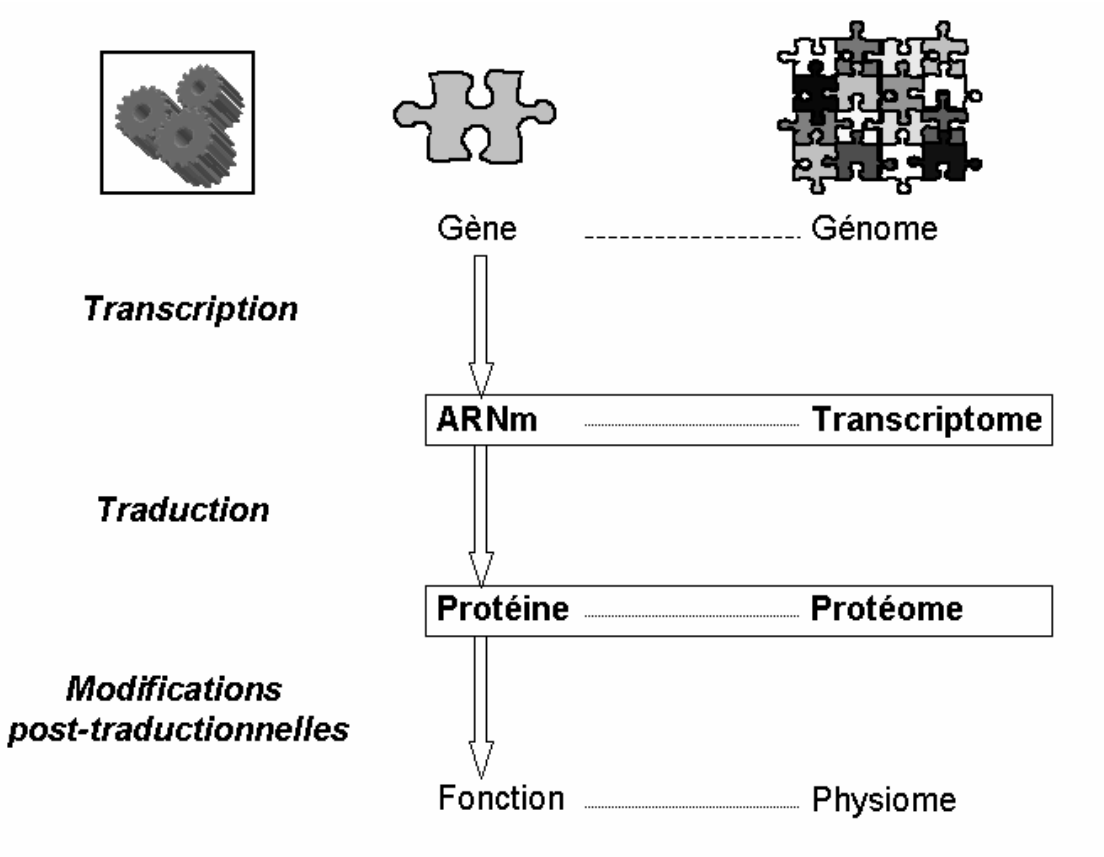



**Figure 6** : Processus intracellulaires impliqués dans l'expression des gènes.

C. Etude du transcriptome

La collection des gènes exprimés ou transcrits de l'ADN génomique, appelée transcriptome est un déterminant majeur du phénotype cellulaire et de la fonction. La transcription est l'étape initiale et indispensable de l'expression génique. Dans un échantillon biologique donné, l'absence d'ARNm pour un gène habituellement traduit en protéine indique que ce gène n'exerce aucune activité dans la cellule ou le tissu étudié. L'analyse du transcriptome est donc d'un intérêt tout particulier dans les études d'expression des gènes.(Brown et al, 1999; Lockhart et al, 2000)

Il est probable qu'il existe pour chaque gène exprimé dans une cellule un niveau d'expression « basal ». Ce niveau d'expression est en équilibre dynamique sous la dépendance des phénomènes de transcription et de dégradation physiologique des ARN. En réponse à des perturbations ou même à des évènements cellulaires physiologiques comme la réplication ou la division cellulaire,

le niveau d'expression va se modifier rapidement et profondément. Le transcriptome est donc une structure hautement dynamique et qui reflète l'adaptation de la cellule à son environnement.

L'expression des gènes en ARNm est régulée par des protéines appelées facteurs de transcription fixées sur une région promotrice qui est une séquence située en général en amont du gène. Ceux-ci agissent comme des transmetteurs d'information en réponse à un stimulus interne ou externe à la cellule. Un facteur de transcription peut en général réguler la transcription de plusieurs gènes. Un gène possède des sites d'interactions pour plusieurs facteurs de transcription. Pour comprendre l'influence des facteurs d'environnements sur une cellule ou un tissu, il est important de définir les relations existantes entre facteurs de transcription et gènes. L'étude des variations transcriptomales consécutives à une perturbation prédéfinie (stimulus externe, surexpression d'un facteur de transcription,…) va apporter des informations sur les gènes impliqués dans la réponse cellulaire. La répétition de ce type d'études dans différentes conditions permettra de mieux comprendre le réseau qui lie stimuli, facteurs de transcription et gènes. Le développement de la technologie du ChIP-on-chip (Chromatin Immuno-Precipitation–on-chip)) a permis en association avec des études d'expression des gènes de disposer d'un outil performant pour l'analyse de l'activation ou la répression de la transcription des gènes par les facteurs de transcription.(Kirmizis et al, 2004) Le ChIP-on-chip permet une identification directe des gènes-cibles d'un facteur de transcription dans une situation biologique donnée

La fonction d'un ARN n'est pas restreinte à la production des ARNm. Parmi les ARN non codant, les micro-ARN sont impliqués dans la régulation de la traduction d'autres ARNm. Seule l'étude du transcriptome peut dans ce cas mettre en évidence le niveau d'expression de ces gènes non traduits en protéines. Quelques études montrent que leur rôle semble important en pathologie cardiovasculaire.(Mann, 2007)

L'analyse du transcriptome, dans le but d'étudier la fonction des gènes, ne mesure pas l'expression du produit final des gènes. Le niveau d'expression de l'ARNm et de la protéine ne sont pas toujours corrélés. De plus, la protéine issue de

la traduction d'un ARNm subit le plus souvent des modifications post-traductionelles comme des phosphorylations ou des coupures enzymatiques qui vont influer sur sa fonction et sa structure. Une variation de niveau d'expression d'un ARNm n'est donc pas forcément en rapport avec une modification de fonction de la protéine correspondante. De même, l'absence de modification de niveau d'expression d'un ARNm n'est pas toujours synonyme de stabilité de fonction ; l'action de la protéine qu'il produit peut être amplifiée par une phosphorylation. De nombreuses méthodes sont développées pour étudier le protéome.(Pandey et al, 2000) Elles mesurent l'expression finale du produit du gène et permettent l'identification de modifications post-traductionnelles. Un autre atout majeur de ces techniques est l'identification de la localisation cellulaire « fonctionnelle » de la protéine. Ces méthodes sont cependant de mise en œuvre moins aisée que l'étude du transcriptome. Bien loin d'être opposées, les études du transcriptome et du protéome sont certainement complémentaires.(Lockhart et al, 2000) Des expériences couplant ces deux méthodes sont déjà réalisées. L'étude du transcriptome permet par exemple de détecter les gènes d'intérêt dans les pathologies étudiées. Une étude protéomique centrée sur ces gènes complète les informations obtenues. Elle détecte les modifications traductionnelles et post traductionnelles de ces gènes associés à la maladie.

## D. Etude du transcriptome par les puces à ADN :

De nombreuses techniques ont déjà été développées pour permettre la mesure de l'expression d'un gène (tableau IV) au niveau de son ARN. Depuis de nombreuses années, le Northern Blot(Alwine et al, 1977) et la RPA (RNase Protection Assay)(Berk et al, 1977) permettent de détecter et quantifier l'expression d'un ARNm spécifique dans un échantillon biologique. Plus récemment des techniques de PCR (Polymerase Chain Reaction) quantitatives ont été adaptées de la PCR classique.(Brand et al, 2002) Ces techniques sont fiables, précises et largement utilisées. L'inconvénient majeur de ces approches est en fait lié au nombre de données d'expression mesurables. On peut estimer qu'au mieux l'expression de quelques dizaines de gènes peut être mesurée de manière simultanée. Ces techniques ne sont donc pas adaptées à l'analyse d'un transcriptome humain.

D’autres approches ont donc été développées afin de caractériser les variations d’expression à l’échelle du transcriptome.(Dempsey et al, 2000) Dans ce groupe de nouvelles techniques, les puces à ADN se détachent nettement.

Les réseaux biologiques d’objets ordonnés (microarrays) appartiennent à un ensemble de nouvelles techniques développées depuis quelques années à l’interface de nombreuses spécialités comme la biologie moléculaire, l’informatique et l’électronique. Dans ce groupe, les puces à ADN (DNA chips) connaissent un essor rapide du fait du grand nombre d’applications proposées comme le séquençage, le génotypage et la mesure de l’expression des gènes.(Hacia, 1999; Lander, 1999) L’adaptation de cette technique à l’analyse du transcriptome permet une évaluation simultanée de l’expression de milliers de gènes dans un échantillon de tissu. L’étude du transcriptome, par ces méthodes, a été réalisée initialement chez la levure(DeRisi et al, 1997; Eisen et al, 1998) puis chez l’homme essentiellement dans les cancers et hémopathies avec des résultats concluants.('t Veer et al, 2002; Alizadeh et al, 2000; Perou et al, 2000; Sorlie et al, 2001) Les applications en pathologies cardiovasculaires apparues plus tardivement sont actuellement en plein développement.

**Tableau IV** : Comparaison de méthodes couramment utilisées pour la détection des ARN (adapté de (Brand et al, 2002)).

| Méthode | Protocole | Avantages de la méthode | Inconvénients de la méthode |
|---|---|---|---|
| **Northern blot** | ● Séparation de l'ARN sur gel<br>● Transfert sur filtre<br>● Hybridation avec la sonde<br>● Détection par autoradiographie | ● Fiable<br>● (Semi) Quantitatif<br>● Réutilisable (limité) | ● Matériel radioactif<br>● Nécessite une grande quantité d'ARN<br>● Sensible à la dégradation |
| **RNase protection assay (RPA)** | ● Hybridation de sondes radio-marquées avec l'ARN<br>● Digestion des sondes non hybridées<br>● Séparation des sondes hybridées sur gel<br>● Détection par autoradiographie | ● Haute sensibilité<br>● (Semi) Quantitatif<br>● Distinction fiable entre séquences proches et entre épissages alternatifs | ● Charge de travail importante<br>● Non réutilisable |
| **PCR (Polymerase Chain Reaction)** | ● Obtention d'ADNc par RT-PCR<br>● Amplification des ADNc par PCR<br>● Visualisation des produits de PCR sur gel | ● Haute sensibilité<br>● Niveau minime d'ARN nécessaire | ● Mauvaise quantification |
| **PCR quantitative (Taqman ou Sybrgreen)** | ● Obtention d'ADNc par RT-PCR<br>● Amplification des ADNc par PCR<br>● Détection des cibles par analyse de la fluorescence interne des sondes | ● Haute sensibilité<br>● Quantitatif<br>● Quantification du produit à chaque cycle de PCR<br>● Quantité d'ARN minime | ● Nécessite un matériel coûteux pour la détection des sondes fluorescentes |
| **microarrays** | ● Dépôt de sondes spécifiques de différents gènes sur un support solide<br>● Hybridation avec des cibles marquées | ● Permet l'analyse simultanée de milliers de gènes<br>● Quantitatif | ● Nécessite du matériel et logiciels coûteux |

Le principe fondateur des puces à ADN est basé sur la technique d'hybridation entre des séquences complémentaires d'acides désoxyribonucléiques, conformément aux observations dues à E. Southern en 1975.(Southern, 1975) La particularité des puces à ADN réside dans la miniaturisation du procédé permettant des milliers d'hybridations simultanées en utilisant une moindre quantité de matériel génétique (figure 7). Une puce à ADN est constituée d'un ensemble de plusieurs centaines ou milliers de sondes nucléotidiques déposées ou synthétisées sur un support solide de quelques centimètres carrés, chaque sonde étant spécifique d'un gène ou fragment de gène déterminé.

Plusieurs procédés techniques d'utilisation sont possibles. Une distinction réside dans le nombre des échantillons analysés sur une puce. Les puces conçues pour l'hybridation d'un seul échantillon sont souvent appelées puces « mono-couleur ». les puces conçues pour la co-hybridation de 2 échantillons sont souvent appelées puces « bi-couleur ».

La description suivante et la figure qui s'y réfère sont donc un exemple de puce bi-couleur, le principe de base restant conservé entre les différentes solutions techniques. Deux cibles complexes d'ARN (une cible test et une cible référence), sont obtenues par méthodes d'extraction à partir d'un tissu défini. Elles sont transformées en ADNc et marquées par des fluorochromes différents. Elles sont ensuite déposées en quantités égales sur la puce permettant une hybridation compétitive et spécifique entre sondes et cibles. Le signal d'hybridation est ensuite recueilli, traité et analysé. Le paramètre interprété est le rapport d'intensité d'hybridation entre les deux cibles (test/référence). L'utilisation des puces à ADN nécessite la définition préalable des sondes utilisées donc des gènes dont on va mesurer l'expression.(Dempsey et al, 2000) L'achèvement du séquençage du génome humain permet de générer des sondes pour l'ensemble du génome humain. Les progrès technologiques dans le domaine des puces permettent aujourd'hui de déposer l'ensemble de ces sondes sur un seul et même support.

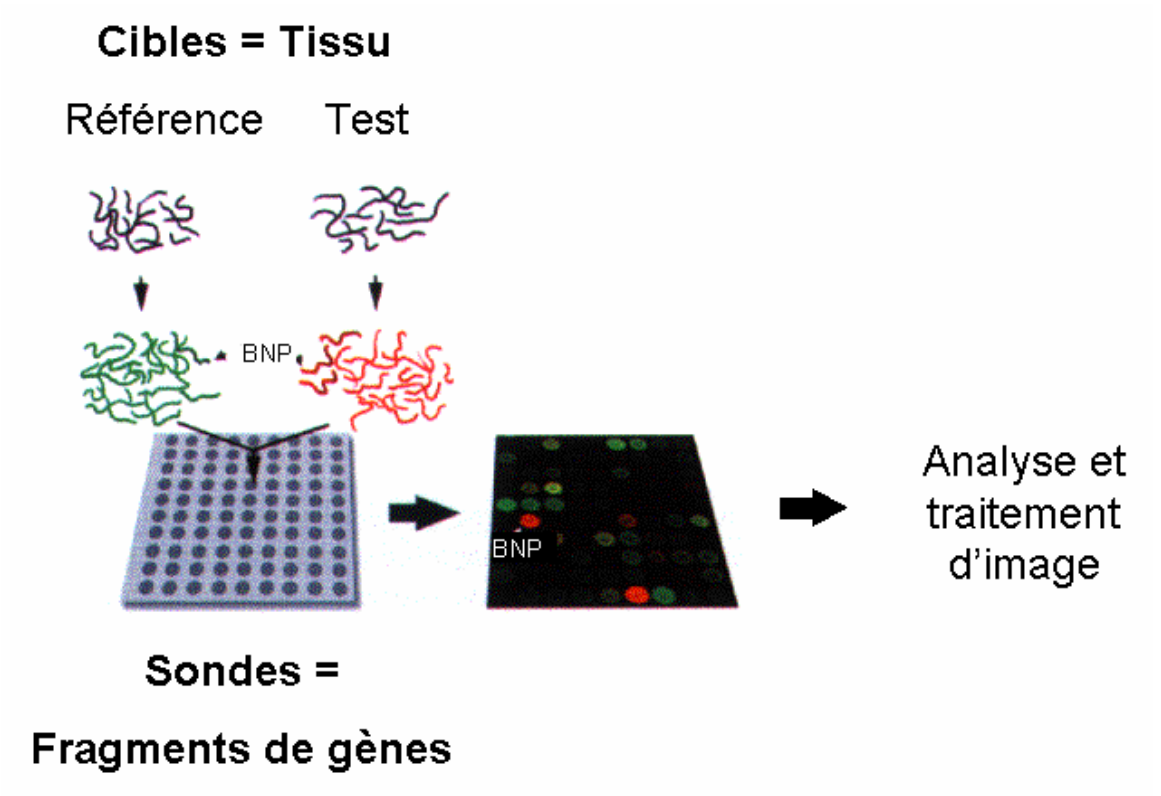



**Figure 7** : Principe général de réalisation des puces à ADN.

E. Autres techniques pour l'étude du transcriptome :

D'autres techniques ont été développées pour mesurer sur un large spectre l'expression des gènes au niveau du transcriptome. Elles sont basées sur les techniques de séquençage.

La stratégie SAGE (Serial Analysis of Gene Expression)(Velculescu et al, 1995) et le séquençage de bases EST (Expressed Sequence Tag)(Adams et al, 1991; Okubo et al, 1992) sont basées sur la génération aléatoire d'un répertoire de fragments d'ADN complémentaire (ADNc) courts voire très courts (SAGE) à partir d'une population d'ARNm (transcriptome). Chaque fragment est ensuite séquencé. Chaque séquence contient suffisamment d'information pour identifier le transcrit et le gène correspondant à ce fragment. Cette identification est réalisée par l'interrogation de bases de données publiques de séquences. Cette approche est devenue moins attractive en pathologie humaine depuis l'achèvement du séquençage du génome humain.

La génération aléatoire de fragments à partir d'une même population d'ARNm va définir un répertoire. Plus ce répertoire sera vaste, plus il sera représentatif du transcriptome étudié sur le plan qualitatif et indirectement sur le plan quantitatif. Une grande majorité des gènes réellement exprimés seront représentés dans le répertoire créé. De plus, leur représentation dans le répertoire sera proportionnelle à leur niveau d'expression. La comparaison des niveaux relatifs de représentation d'un gène entre deux répertoires permettra de déterminer s'il existe une variation du niveau d'expression entre les deux situations étudiées. L'identification des transcrits les plus faiblement exprimés est plus aléatoire avec cette technique. Un gène non représenté dans un répertoire peut en effet tout de même être exprimé.

Un autre moyen de détection de gènes d'expression différentielle est la création de banques différentielles. De nombreux procédés techniques permettent de réaliser une sorte de soustraction entre deux populations d'ARNm. Il s'agit d'identifier les gènes dont les niveaux d'expression sont différents entre ces deux transcriptomes.(Liang et al, 1992) La soustraction aboutit à la création de deux banques de clones ADNc enrichis en gènes différentiellement exprimés entre les deux populations de départ. Elles sont issues des deux soustractions réalisées :

- Population 1 - population 2
- Population 2 - population 1

Cette méthode réalise une présélection des gènes potentiellement intéressant. Elle ne permet pas d'évaluation du niveau d'expression relatif des ARNm dans chaque population.

### F. Etude du transcriptome cardiaque humain utilisant les puces à ADN

L'objectif de ces études d'expression génique à large spectre est d'identifier les modifications de quantité d'ARNm de certains gènes en rapport avec une pathologie donnée, un paramètre de jugement clinique ou la réponse à un traitement. L'idée est que de l'analyse de ces modifications d'expression émergera l'identification des mécanismes impliqués dans la maladie étudiée ou la définition d'une signature

moléculaire spécifique du diagnostic ou du pronostic. Deux stratégies d'utilisation majeures se détachent :

- L'identification de « gènes candidats » :

L'objectif est d'identifier une relation entre un ensemble de gènes et une caractéristique (pathologie, phénotype particulier) sur la base de leurs variations d'expression entre les différentes situations étudiées. Ces gènes sont des candidats valables pour l'identification de variations génétiques associées à un risque morbide accru ou une réponse au traitement anormale. L'identification d'un ensemble de variations génétiques permettant de prédire avec une bonne fiabilité un risque morbide permettrait la mise en place d'un test de dépistage individuel de ces anomalies à l'échelle des populations à partir d'un échantillon sanguin. Ceci aboutira à une meilleure prédiction du risque à l'échelle de l'individu. L'étude de ces gènes et de leur fonction permet également une meilleure compréhension des mécanismes physiopathologiques et l'identification de nouvelles cibles thérapeutiques potentielles.

- L'identification de signatures moléculaires :

L'ensemble des gènes spécifiquement différentiellement exprimés dans une situation clinique ou biologique précise correspondent à une signature moléculaire de cette situation. Cette signature peut être utilisée comme un biomarqueur. Les études d'expression génique peuvent fournir une vision très précise, basée sur des milliers de caractéristiques du phénotype moléculaire d'un individu. L'utilisation d'une signature basée sur plusieurs dizaines ou centaines de caractéristiques moléculaires permettrait une stratification fine du risque morbide ou de la réponse à un traitement à l'échelle de l'individu. Des applications dans le domaine de la cancérologie ont confirmé ce concept.(Alizadeh et al, 2000; Bertucci et al, 2004; Chen et al, 2007)

Depuis la fin des années 90, de nombreuses études se sont intéressées au transcriptome cardiaque humain dans différentes situations physiologiques et pathologiques. Leurs résultats peuvent être retrouvés dans de nombreuses synthèses.(Nanni et al, 2006; Seo et al, 2006; Steenman et al, 2005) Les paragraphes suivants présenteront un état synthétique des résultats obtenus dans l'insuffisance cardiaque et la fibrillation auriculaire.

## 1. Etude du transcriptome humain et insuffisance cardiaque

Plus de 20 travaux sur les variations globales du transcriptome cardiaque humain dans l'insuffisance cardiaque ont été publiées depuis la première étude réalisée en 2000.(Yang et al, 2000) La majorité des échantillons de cœur en insuffisance cardiaque analysés étaient obtenus au moment d'une chirurgie de transplantation cardiaque ou d'implantation d'une assistance mécanique ventriculaire, les tissus contrôles étant en général des cœurs de donneurs prélevés pour une transplantation mais non utilisés pour des causes techniques ou la présence d'anomalies comme des calcifications coronaires. Plusieurs études ont également utilisé des biopsies endomyocardiaques.(Grzeskowiak et al, 2003; Kittleson et al, 2004; Wittchen et al, 2007) Ces travaux ont montré la faisabilité d'études d'expression génique à partir de très petites quantités de matériel biologique, combinant éventuellement des techniques d'amplification des ARN et de puces à ADN. Les champs d'investigations ont essentiellement concerné les variations transcriptomales entre cœurs en insuffisance cardiaque et cœurs contrôles, la recherche de modifications d'expression plus spécifiques d'une étiologie donnée, l'exploration du « remodelage inverse » induit par l'implantation d'un système d'assistance ventriculaire gauche mécanique. En plus de ces approches de type « gène candidat », quelques travaux ont commencé à rechercher des signatures moléculaires utilisables comme biomarqueurs d'une étiologie ou du pronostic.

- Variations transcriptomales entre cœurs en insuffisance cardiaque et cœurs contrôles :

Ces études ont contribué à identifier l'ensemble des gènes impliqués dans le remodelage myocardique associée l'insuffisance cardiaque et permis de disposer d'une vue d'ensemble des différentes fonctions biologiques perturbées dans le tissu cardiaque défaillant. En dépit de l'utilisation de plate-formes technologiques différentes, un ensemble de résultats concordants se dégagent de ces études. La sur-expression de gènes codant des protéines du sarcomère, du cytosquelette ou de la matrice extracellulaire apporte des bases moléculaires à la fibrose, la dysfonction contractile et les altérations de la structure cellulaire observées dans le remodelage myocardique. La sur-expression de nombreux gènes codant des protéines

ribosomales et l'identification de la majorité des études d'un nombre de gènes sur-exprimés plus important que de gènes sous-exprimés témoignent du processus général d'activation de la biosynthèse des protéines associée au développement du phénotype hypertrophique. Parmi les gènes les plus fortement sur-exprimés, on retrouvait les gènes codant les peptides natriurétiques (Atrial Natriuretic Factor, BNP). Un résultat particulièrement intéressant et retrouvé dans différentes études concernait le métabolisme énergétique. La principale source d'énergie du cardiomyocytes est la transformation des acides gras par la β-oxydation. Dans l'insuffisance cardiaque, la production d'énergie par le myocarde est diminuée mais les mécanismes en jeu restent encore discutés.(Huss et al, 2005) Les résultats obtenus avec les puces à ADN montrent pour la majorité des études ayant analysé le transcriptome de patients en défaillance cardiaque avancée, une sur-expression de gènes impliqués dans la phosphorylation oxydative, témoignant d'une altération de la respiration mitochondriale, et une sous-expression de gènes de la glycolyse. Au contraire, dans une étude réalisée sur des sujets à un stade modéré de la maladie, on observait une sur-expression des gènes de la glycolyse.(Grzeskowiak et al, 2003) Ces résultats suggèrent que les altérations du métabolisme énergétique puisse être différent suivant le stade de gravité de la maladie. Des stratégies de traitement ciblées sur le métabolisme cardiaque pour favoriser le métabolisme du glucose sont en cours de développement dans l'insuffisance cardiaque. Cette stratégie pourrait s'avérer particulièrement bénéfique pour les patients les plus sévèrement touchés.

- Variations transcriptomales en rapport avec une étiologie spécifique d'insuffisance cardiaque :

L'insuffisance cardiaque est le stade évolutif d'un grand nombre de pathologies. Plusieurs études ont cherché à identifier dans le remodelage transcriptomal cardiaque un phénotype moléculaire particulièrement associé à une étiologie.(Cunha-Neto et al, 2005; Hwang et al, 2002; Kittleson et al, 2005b; Wittchen et al, 2007) Wittchen et al. ont étudié le remodelage transcriptomal des cardiomyopathies inflammatoires associées aux infections virales.(Wittchen et al, 2007) Cette équipe a montré récemment que ces cardiomyopathies sont favorisées par la persistance du virus dans le myocarde(Kuhl et al, 2005) et que l'utilisation d'un traitement antiviral par interféron-β permet une amélioration de la fonction myocardiaque.(Kuhl et al, 2003) L'objectif de cette étude était d'identifier les réseaux

de gènes impliqués dans le développement de la cardiopathie et d'éventuelles cibles thérapeutiques. L'analyse a été réalisée à partir de biopsies endo-myocardiques prélevés à 8 patients infectés par le parvovirus B19 et présentant une cardiopathie dilatée inflammatoire à fonction systolique modérément altérée. Des biopsies de 4 patients sans cardiopathie connue étaient utilisées comme groupe contrôle. Un profil d'expression génique de 1418 gènes associé à la cardiomyopathie inflammatoire a été identifié. Un premier réseau de gènes était centré sur CYR61 (cysteine-rich angiogenic inducer 61). Un second réseau était centré sur APN (adiponectin). CYR61 était sur-exprimé dans les tissus pathologiques alors que APN était sous-exprimé. Ces modifications d'expression ont pu être reproduites *in vitro* par stimulation de cellules endothéliales avec différentes cytokines (dont l'interféron-β). De manière intéressante des connexions entre ces 2 réseaux ont pu être validées. *In vitro*, la stimulation de APN a inhibé la sur-expresion de CYR61 induite par les cytokines. Ces résultats associés aux rôles déjà connus de APN dans l'ischémie myocardique(Shibata et al, 2005) suggèrent que la balance CYR61-APN est un élément important du phénotype moléculaire des cardiomyopathies inflammatoires et pourrait représenter une cible thérapeutique privilégiée.

- Variations transcriptomales en rapport avec le « remodelage inverse » lié à la décharge mécanique du ventricule gauche de l'insuffisant cardiaque :

Les systèmes d'assistance ventriculaire mécanique permettent la décharge du ventricule gauche et de restaurer un contrôle hémodynamique satisfaisant pour des patients en échec thérapeutique dans l'insuffisance cardiaque chronique ou aigue. L'implantation peut être temporaire (« bridge ») en attendant une transplantation cardiaque ou la récupération d'un fonctionnement cardiaque compatible avec l'ablation du dispositif. Des stratégies d'implantations « à demeure » sont en cours d'évaluation (« destination therapy »). Un phénomène macroscopique et histologique de remodelage « inverse » du myocarde est décrit après implantation du matériel et mise en décharge de la (des) cavité(s) ventriculaire(s). Plusieurs études ont analysé le remodelage du transcriptome ventriculaire gauche survenant chez un même patient pendant cette phase de décharge mécanique avant une transplantation cardiaque.(Blaxall et al, 2003; Chen et al, 2003a; Chen et al, 2003b; Hall et al, 2004; Margulies et al, 2005) Les échantillons analysés étaient issus de l'apex ventriculaire gauche au moment de l'implantation de la machine et au moment de la

transplantation cardiaque. L'étude la plus complète a également étudié des échantillons de ventricule sans dysfonction cardiaque.(Margulies et al, 2005) L'objectif était de rechercher les gènes dont l'expression était altérée par la cardiopathie et normalisée par la décharge mécanique. Sur les 3088 gènes différentiellement exprimés entre la situation insuffisance cardiaque et contrôle, seuls 238 voyaient leur expression modifiée après implantation de l'assistance ventriculaire et pour 75% de ces 238 gènes, les différences d'expression persistaient ou étaient amplifiées. Ces résultats suggèrent que les modifications d'expression génique du ventricule gauche associées à l'implantation d'une assistance ventriculaire ne représentent pas une simple inversion des modifications d'expressions observées durant le processus de remodelage hypertrophique de l'insuffisance cardiaque. Le terme de remodelage « inverse » semble donc impropre. Ces résultats suggèrent par ailleurs que parmi l'ensemble des altérations d'expression génique décrites comme associées à l'insuffisance cardiaque, une proportion importante pourrait représenter des épiphénomènes autour de mécanismes fondamentaux du remodelage, expliquant que leurs variations d'expression ne sont normalisées par la décharge mécanique du ventricule.

- Identifications de signatures moléculaires au sein du transcriptome cardiaque :

Une première étude a mis en évidence la faisabilité de cette stratégie dans l'insuffisance cardiaque. Kittleson *et al.* ont analysé la transcriptome cardiaque de patients insuffisants cardiaques sévères de cause ischémique (n=16) et de cause non ischémique (n=32).(Kittleson et al, 2004) Ils ont pu identifier une signature moléculaire de 90 gènes permettant de séparer les échantillons cardiaques en ischémique et non-ischémique avec une excellente performance quel que soit le niveau de sévérité de l'insuffisance cardiaque ou d'autres paramètres potentiellement confondants comme l'âge ou le sexe. Ces résultats n'ont a priori aucune application clinique mais doivent être vus comme une preuve du concept d'utilisation des signatures moléculaires comme biomarqueur dans l'insuffisance cardiaque humaine.
Une partie de mon travail de thèse a consisté à rechercher si cette même approche pouvait être appliquée au pronostic de l'insuffisance cardiaque. Les résultats sont présentés dans le chapitre suivant.

## 2. Etude du transcriptome humain et fibrillation auriculaire

Plusieurs travaux ont récemment utilisé la technologie des puces à ADN pour comparer les profils d'expressions géniques d'échantillons auriculaires en fibrillation auriculaire et en rythme sinusal. Ces échantillons étaient des auricules droits prélevés lors d'une procédure de chirurgie cardiaque. L'objectif était d'identifier et caractériser les altérations du transcriptome atrial en rapport avec le trouble du rythme.

Utilisant une puce contenant 1152 gènes, Kim *et al.*(Kim et al, 2003) ont comparé les profils d'expression de tissu auriculaire droit de 26 patients en FA opérés pour une intervention de Maze(Cox et al, 1996) et de 26 patients en rythme sinusal prélevés lors de chirurgie de pontages coronaires. Leur analyse était centrée sur l'étude du stress oxydatif. Ils ont pu mettre en évidence une surexpression de gènes liés à la production de radicaux libres même temps qu'une diminution d'expression de gènes impliqués dans une activé anti-oxydante. Ces résultats suggèrent l'existence d'un déséquilibre de la balance oxydant–antioxydant dans le tissu atrial en fibrillation auriculaire, soulignant que le stress oxydant pourrait être un élément significatif dans la fibrillation auriculaire et une cible thérapeutique potentielle.

Dans une deuxième étude, Ohki *et al.*(Ohki et al, 2005) ont comparé le transcriptome atrial de 10 sujets en rythme sinusal et 7 sujets en fibrillation auriculaire. Utilisant une puce couvrant 12000 gènes humains, ils ont pu identifier 33 gènes surexprimés et 63 gènes sous exprimés dans les tissus en fibrillation auriculaire. Leurs résultats suggèrent une surexpression de gènes impliqués dans l'inflammation. Une sous expression de ATP2A2 était également notée.

Dans une troisième étude, Barth *et al.*(Barth et al, 2005) ont analysé le transcriptome atrial de 10 patients en FA permanente et 20 patients en rythme sinusal. 5 échantillons de tissu ventriculaire gauche de cœur de donneur non utilisés étaient également analysés. Les analyses ont été réalisées avec un jeu de puces commerciales pan-génomiques (U133A+B, Affymetrix®) contenant un total de 44928 sondes oligonucléotidiques. Un ensemble de 1434 gènes (452 surexprimés et 982 sous-exprimés) étaient différentiellement exprimés entre les tissus auriculaires des patients en fibrillation auriculaire et en rythme sinusal. L'analyse des fonctions

moléculaires associées à ces gènes a révélé l'importance de phénomènes pro-fibrotiques. Une surexpression de métallo-protéases de la matrice extracellulaire (MMP9 et ADAM19) associée à une sous-expression de leurs inhibiteurs comme TIMP3 était observée. Une activation de voies de signalisation pro-fibrotiques comme celle du TGFβ était également notée. De manière assez étonnante, la signature moléculaire des patients en FA pour l'ensemble de ces 1434 gènes était plus semblable à celle des tissus ventriculaires qu'à celle des tissus atriaux en rythme sinusal, suggérant une perte des caractéristiques transcriptomales spécifiques du tissu atrial lors de la FA. Ces modifications pourraient renforcer le concept d'un processus adaptatif de dédifférenciation du tissu atrial lors de la fibrillation auriculaire.(Thijssen et al, 2001)

Ces études ont pu apporter des informations intéressantes pour mieux comprendre le remodelage atrial lié à la fibrillation auriculaire. Leurs résultats doivent cependant être modérés car un certain nombre de précautions méthodologiques n'ont pas été respectées. Certains patients n'étaient pas en fibrillation auriculaire permanente(Kim et al, 2003) ou le type de FA n'était pas précisé.(Ohki et al, 2005) La pathologie au cours de laquelle survient la fibrillation auriculaire n'a pas été prise en compte dans le plan expérimental.(Barth et al, 2005; Kim et al, 2003; Ohki et al, 2005) En effet, si la survenue d'une fibrillation auriculaire chez un patient porteur d'un valvulopathie provoque un remodelage atrial, la valvulopathie elle-même a certainement profondément modifié le transcriptome atrial. Il en résulte que la comparaison des transcriptomes de sujets en FA et de sujets en rythme sinusal sans substrat atrial connus (per exemple patients opérés pour une maladie coronaire sans dysfonction ventriculaire) identifiera des altérations d'expression génique lié au substrat et/ou la fibrillation auriculaire. Ce point sera clairement mis en évidence dans les résultats présentés dans ce document.

# DE LA CONCEPTION A L'INTERPRETATION DES PUCES A ADN POUR L'ETUDE DU TRANSCRIPTOME CARDIAQUE HUMAIN

L'objectif de ce chapitre est de synthétiser les étapes successives permettant une réalisation optimale d'une étude d'expression génique utilisant les puces à ADN. Les étapes clés de l'expérience seront plus particulièrement détaillées. Il existe un nombre très important de solutions technologiques et toutes ne sont pas détaillées dans ce chapitre. Les méthodes présentées ici sont celles couramment utilisées au sein de la plateforme transcriptome de OUEST-génopole® et qui ont été utilisées pour la réalisation de mes travaux de thèse.

Ce chapitre s'appuie sur la publication suivante (cf annexes) :
Lamirault G, Steenman M, Le Meur N, Demolombe S, Trochu JN, Leger JJ. DNA chip technology in cardiovascular research. Arch Mal Coeur Vaiss. 2004 Dec;97(12):1251-5.

## I. IMPORTANCE DU PLAN EXPERIMENTAL POUR LA QUALITE DES MESURES D'EXPRESSION

Un plan expérimental décrit les expériences et conditions d'expérience à mettre en oeuvre pour répondre à une question. Dans le domaine des puces à ADN, cette étape est cruciale dans l'obtention de mesures de qualité. Ainsi, lorsque deux niveaux d'expression sont comparés, l'enjeu principal est de faire la distinction entre les variabilités techniques et les variations biologiques.

La difficulté de distinguer les variabilités techniques et biologiques d'intérêt réside dans leurs origines multiples. Les variabilités techniques sont des biais expérimentaux, tels que la qualité des lots de lames, des réactifs ou des préparations, qui entraînent des erreurs dans les mesures et nécessitent un traitement (nettoyage) des données. Les variations biologiques sont les variabilités intra et inter-individus.

Les variabilités intra-individus font référence à l'hétérogénéité des mesures obtenues pour différents prélèvements d'un même tissu chez un individu. Les variations inter-individus correspondent aux variations biologiques entre individus et sont le plus souvent l'objectif de la recherche.

L'utilisation de réplicats (résultat d'une mesure qui a été réalisée plusieurs fois sur un même objet) permet d'appréhender et de contrôler ces différentes variabilités. Une des solutions proposées pour contrôler les variabilités techniques et les variations biologiques est de maximiser les réplicats biologiques et optimiser les réplicats techniques.(Lee et al, 2000; Pavlidis, 2003) Réaliser des réplicats biologiques consiste à analyser le plus grand nombre d'échantillons possibles. Si l'objectif de l'analyse est la découverte d'une nouvelle taxonomie au sein d'une maladie, les réplicats biologiques peuvent être intra et inter-individus. Les réplicats intra-individus correspondent à différents prélèvements du même tissu chez un même patient ; tandis que les réplicats inter-individus sont différents prélèvements du même tissu chez différents individus. Cette approche aide à déterminer la consistance des groupes taxonomiques identifiés.(Simon et al, 2003) Les réplicats techniques, quant à eux, peuvent correspondre au dépôt multiple d'une même sonde sur une même lame ou à l'hybridation de plusieurs lames avec les mêmes échantillons. Les réplicats techniques permettent de contrôler la qualité des mesures d'expression obtenues.
Le nombre de réplicats nécessaires et suffisants à l'obtention de données de qualité dépend du nombre de facteurs étudiés, et plus particulièrement du type d'expérience et des modes d'analyses employés. La principale méthode pour estimer ce nombre est l'analyse de puissance.(Pan et al, 2002; Pavlidis, 2003) Une étude de puissance réalisée au cours de mon travail de thèse est présentée dans la partie résultats (Etude 1).

Un autre concept important du plan expérimental est la randomisation (choix au hasard) du plus grand nombre possible de paramètres capables d'influencer les résultats de l'expérience (expérimentateurs, dépôts multiples sur une lame, lot de lames...). Ces paramètres sont souvent qualifiés de facteurs de « nuisance ». Un exemple est l'utilisation de lames provenant de différents lots. En effet, si l'ensemble des échantillons contrôles est hybridé sur les lames d'un lot X et les échantillons traités sur des lames du lot Y, il sera impossible de distinguer les variations

biologiques des variations liées aux lots des lames. Les deux facteurs seront confondus. Un plan d'expérience en « bloc » (block design) permet de gérer l'influence de certains facteurs de « nuisance » sur une expérience.(Draghici S, 2003) Un bloc est un sous-ensemble de conditions expérimentales (par exemple une puce) pour lesquelles il est possible de bloquer l'effet des facteurs de « nuisance » sur la mesure.

Le choix du plan expérimental à mettre en oeuvre dépend essentiellement de la question biologique posée.(Yang et al, 2002b) Les plans expérimentaux se distinguent principalement par le mode de comparaison des échantillons qui peut être direct et/ou indirect. Le plan expérimental employé dans la plupart des travaux de cette thèse est le « Reference design » (figure 8). Ce mode consiste à utiliser pour l'un des échantillons hybridé sur l'ensemble des puces le même ARNm dit de « référence », par exemple un pool des ARN « contrôles ». Généralement, l'échantillon de référence est marqué par le même fluorochrome pour toutes les puces. Les progrès dans les étapes de préparation des cibles permettent aujourd'hui d'éviter la réalisation d'une hybridation répétée des échantillons avec une inversion des fluorochromes ou « *Dye swap* ».

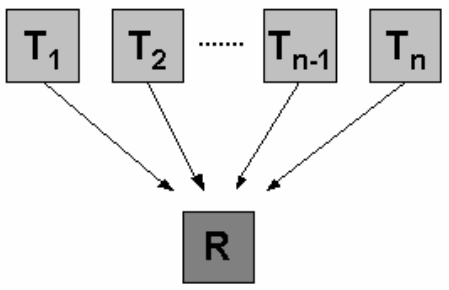



**Figure 8 : Exemple de plan expérimental de type « reference design »**
Tous les échantillons à analyser (T) sont comparés à une même référence (R).

## II. PUCES PANGENOMIQUES ET PUCES DEDIEES

Le choix de la puce utilisée doit être envisagé très tôt dans la conception d'une expérience. Les informations obtenues dépendent de la collection de sondes utilisées. Aujourd'hui la majorité des puces à ADN utilisées pour l'étude du transcriptome humain sont pangénomiques, utilisent des sondes oliogonucléotidiques et sont d'origine commerciale. Les puces pangénomiques dites « généralistes » proposent des jeux de sondes oligonucléotidiques de 50 mers en moyenne représentant l'ensemble des séquences potentiellement transcrites du génome humain. Ainsi, il existe des puces à ADN contenant plus de 40 000 sondes représentant les quelques 25 000 gènes du génome humain. L'avantage des puces pangénomiques réside dans leur exhaustivité. Elles permettent une réelle analyse globale du transcriptome. Sur ce type de puces, chaque sonde n'est en général déposée qu'une seule fois sur le support. Aucune validation statistique des mesures ne peut donc être réalisée à l'intérieur de la puce et toute anomalie physique sur ce dépôt unique rend l'analyse de cette sonde et du gène correspondant impossible De nombreuses études visent à analyser des variations du transcriptome en rapport avec un groupe de gènes d'intérêt. L'incapacité d'obtenir une mesure d'expression d'un ou plus de ces gènes va constituer un obstacle important à l'analyse de toute l'expérience.

Les puces dédiées sont constituées d'une collection de gènes spécifiquement (voire exclusivement) liés à un tissu, une pathologie et/ou une thématique. Elles permettent de mieux cibler les transcrits pertinents pour l'étude d'un type cellulaire donné. En effet, on estime que parmi les 10 000 à 20 000 transcrits potentiellement exprimés dans une cellule spécialisée, seuls 4 000 à 6 000 d'entre eux sont caractéristiques de ce type cellulaire. Les transcrits d'intérêt peuvent être obtenus de différentes manières, souvent complémentaires :

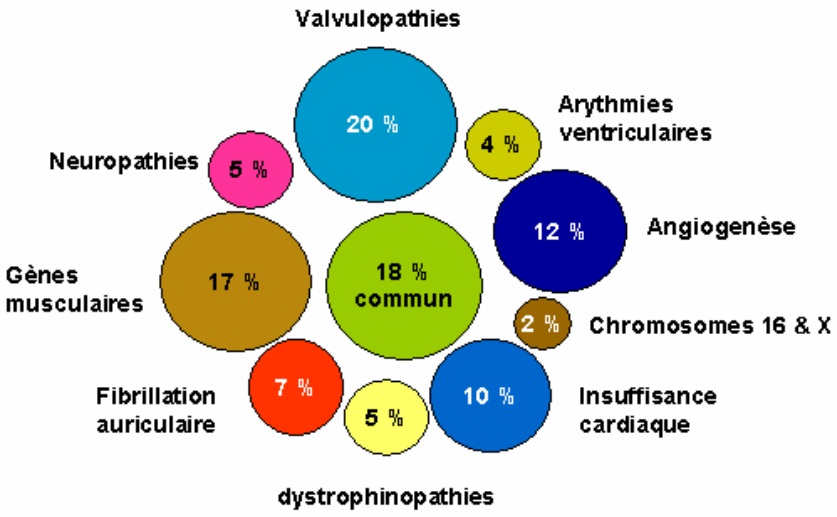
- Sélection expérimentale par criblage (screening) de puces à ADN pangénomiques
- Connaissances biologiques a priori par analyse de la littérature et/ou interrogation des bases de données publiques

- Constitution de banques différentielles par approches soustractives (SSH),(Diatchenko et al, 1996; Tkatchenko et al, 2000) séquençage systématique de banques d'EST et/ou analyse de banques SAGE.(Velculescu et al, 1995)

Par exemple, la comparaison des transcriptomes de patients insuffisants cardiaques ou non, au moyen de puces pangénomiques, nous a permis de montrer que seuls 2031 gènes sur 12626 (soit 16%) s'exprimaient dans au moins un des échantillons.(Steenman et al, 2003)

Les premières générations de puces à ADN dédiées viennent des institutions académiques, les sociétés commerciales n'étant pas convaincues d'une telle approche. Les sondes étaient alors principalement des produits de PCR. Ces puces sont entre autres baptisées CardioChip(Barrans et al, 2001), IonChip(Le Bouter et al, 2003) ou Myochip. La Myochip, développée dans notre laboratoire, est dédiée à l'exploration des pathologies cardiovasculaires et neuromusculaires telles que les valvulopathies, les arythmies, la myogenèse ou les dystrophinopathies (figure 9). La sélection des marqueurs des transcriptomes musculaires et cardiaques a été réalisée par un consortium de laboratoires d'OUEST-Génopole® qui travaillent sur des thématiques voisines. Aujourd'hui, la Myochip se compose de 4217 sondes, représentatives de 4012 gènes.

**Figure 9** : Répartition des gènes entre les différentes thématiques sur la Myochip



Les puces à ADN dédiées offrent la possibilité d'évaluer plusieurs fois la même sonde sur la même puce (réplicats). En effet, les robots de dépôts ont actuellement la capacité de déposer sur une lame environ10.000 sondes par cm² soit près de 40.000 sondes pour une puce. Les puces à ADN dédiées, composées de quelques milliers de sondes, permettent le dépôt d'une même sonde en plusieurs exemplaires. Elles offrent ainsi la possibilité d'une validation statistique de la mesure à l'intérieure de la puce. La prise de conscience par les entreprises commerciales de l'intérêt des puces à ADN dédiées, les a incitées à dessiner leurs propres puces dédiées. Certaines sociétés proposent également de créer des puces à façon et/ou de commercialiser les oligonucléotides à l'unité. Les puces à ADN dédiées (académiques) sont alors devenues des puces à oligonucléotides, telle la Myochip qui est aujourd'hui composée d'oligonucléotides 50 mers.

## III. ECHANTILLONS BIOLOGIQUES

La qualité des échantillons biologiques est essentielle pour l'obtention de données d'expression géniques fiables. Les ARN sont des molécules fragiles très sensibles à la dégradation enzymatique réalisées par les RNAses, enzymes retrouvées de manière ubiquitaire. Ces RNAses rendent l'utilisation de tissus humains obtenus *post-mortem* impossibles. Après prélèvement du tissu, des stratégies de conditionnement et conservation stricte des tissus doivent être employées. La stratégie classique consiste à immerger le prélèvement dans de l'azote liquide après éventuellement un bain dans de l'isopentane refroidi par l'azote liquide. La conservation sera ensuite possible à –80°C pendant plusieurs mois voire années. Une alternative consiste à utiliser un liquide protection des ARN (RNAlater®, Ambion®). Le prélèvement est immédiatement plongé dans cette solution après obtention. Il peut être ainsi conservé à température ambiante plusieurs heures, à 4°C plusieurs jours et à –20°C ou –80°C plusieurs mois ou années.

L'extraction des ARN est réalisée selon la technique de référence de Chomczynski et Sacchi utilisant le Trizol® (GIBCO®).(Chomczynski et al, 1987) Plusieurs kits commerciaux adaptés à la quantité initiale de tissu et à sa résistance mécanique sont commercialisés. Les kits « Rneasy Fibrous tissue® » (Qiagen®) ont été utilisés pour l'étude sur la fibrillation auriculaire présentée dans les résultats.

L'analyse quantitative et qualitative des ARN extraits est indispensable pour le déroulement ultérieur de l'expérience. L'analyse quantitative est réalisée avec un spectrophotomètre UV/visible (220-750 nm) (Nanodrop®, Nyxorbiotech®) qui permet le dosage des ARN/ADN/protéines à partir d'un seul µl de solution d'ARN. 200 ng d'ARN total suffisent pour réaliser une hybridation de puce à ADN via l'utilisation d'un système d'amplification *in vitro* des ARNm. Le Nanodrop® évalue la contamination en protéines et solvants organiques. Le Bioanalyzer® 2100 (Agilent®) permet une analyse de l'intégrité des ARN extraits. Il réalise une électrophorèse capillaire à partir d'1 µl de solution d'ARN. Le rapport des ARN ribosomaux 28S/18S donne une première indication de la qualité des ARN. L'analyse du spectre d'électrophorèse permet une analyse plus complète aboutissant au calcul d'un indice de qualité appelé « RNA Integrity Number – RIN ».(Schroeder et al, 2006) L'obtention d'un RIN idéal de

9 à 10 est exceptionnelle avec des tissus cardiaques humains. Le calcul du RIN permet surtout d'identifier des ARN de qualité moyenne ou limite dont les résultats d'hybridation devront être considérés avec prudence.

Un dernier point mais pas le moins important concernant les échantillons est la détermination du phénotype de l'individu auquel le prélèvement tissulaire correspond. Une rigueur maximale dans le recueil des paramètres cliniques et para-cliniques s'impose avant la réalisation de l'expérience et idéalement avant même le prélèvement. Une sélection préalable des échantillons les mieux adaptés à l'objectif de l'expérience évite l'identification *a posteriori* de biais parfois impossibles à corriger. L'exemple des tissus contrôles pour l'étude du transcriptome dans la fibrillation auriculaire en est un exemple (Etude 5).

## IV. PREPARATION DES CIBLES, HYBRIDATION ET ACQUISITION DES SIGNAUX

Une fois la qualité et quantité des ARN validées, la préparation des cibles et leur hybridation sur la puce est réalisée. Deux étapes successives (synthèse puis marquage des cibles) permettent la préparation des cibles à partir des ARN extraits. Lors de la synthèse de la cible des bases Uraciles couplées à un groupement aminoallyl (aa-dUTP) sont incorporées. La synthèse est réalisée par transcription inverse en présence d'amorces oligo-dT qui vont se fixer sur la queue poly-A des ARNm. Après dégradation de l'ARN, le produit obtenu est de l'ADNc. Si une étape préalable d'amplification des ARN est nécessaire, la transcription inverse est réalisée en présence d'amorces oligo-dT portant une séquence promotrice T7 qui servira à réaliser une transcription in vitro. Dans ce cas, les aa-dUTP sont incorporées lors de la transcription in vitro. Le produit obtenu est de l'aRNA (amplified RNA). Un point essentiel est que l'amplification semble introduire un biais dans la représentativité des ARN. Tous ne sont pas amplifiés avec la même importance. Cependant cette amplification semble reproductible.(Ma et al, 2006) La stratégie actuelle est donc de ne comparer que des cibles issues d'ARN tous amplifiés ou tous non amplifiés et surtout pas issues d'un mélange d'ARN amplifiés et non amplifiés. Ceci doit être bien intégré dans le design expérimental.

Les cibles synthétisées sont marquées par une molécule fluorescente (carbocyanine 3 –Cy3- ou carbocyanine 5 -Cy5-). Les molécules fluorescentes seront greffées sur les groupements aminoallyl incorporés pendant la synthèse. Les produits issus de l'amplification devront être fragmentés pour faciliter leurs hybridation. Des kits commerciaux (Cyscribe post-labelling®, Amersham® - Amino Allyl MessageAmp® II aRNA amplification, Ambion®) sont utilisés pour toutes ces étapes.

Avant hybridation, les cibles sont reprises dans un tampon d'hybridation et les puces subissent éventuellement un traitement de pré-hybridation. Les 2 cibles marquées par des fluorochromes distincts sont mélangées et déposées sur la puce. L'hybridation spécifique entre cibles et sondes se poursuit 12 à 18h. Un lavage stringent des lames est ensuite réalisé permettant d'éliminer les hybridations non spécifiques.

Les lames hybridées sont lues par un scanner à microscopie confocale (ScanArray®, GSI Lumonics) muni de 2 lasers permettant l'excitation de chaque fluorochrome (532 nm pour le Cy3 et 633 nm pour le Cy5). La résolution linéaire est de 10 µm par pixel. Deux images distinctes sont donc obtenues, chacune correspondant à la cartographie de l'intensité de fluorescence recueillie pour un marqueur fluorescent. Des préréglages d'intensité globale pour chaque laser sont réalisés avant l'acquisition des images en modulant la puissance des lasers et d'un photomultiplicateur. L'objectif est d'éviter l'enregistrement des signaux saturés et d'équilibrer l'intensité globale du signal entre les 2 images. Il est cependant fréquent d'obtenir un déséquilibre entre les intensités globales des 2 images. Ce biais technique est par la suite corrigé *in silico*.

Suite à la lecture des puces à ADN par un scanner, les niveaux d'expression sont estimés grâce à des logiciels d'analyse d'images. Ces logiciels superposent les 2 images obtenues (une image par fluorochrome) et extraient des informations qualitatives et semi-quantitatives pour chaque spot (unité d'hybridation) dans chacune des images. Globalement, ces logiciels sont basés sur le même principe et possèdent la même procédure de traitement qui se déroule en trois étapes :

- Localisation des spots sur la puce
- Segmentation de l'image en pixels appartenant aux signaux et aux bruits de fond
- Extraction des données qualitatives et semi-quantitatives

Cette analyse est semi-automatique. L'expérimentateur doit vérifier la segmentation proposée par le logiciel et identifier les spots non analysables du fait de biais physiques majeurs (poussières, spots arrachés,...) que le logiciel n'aurait pas lui-même détecté. Les résultats sont présentés sous forme d'un tableau sauvegardé dans un fichier informatique. Chacune ligne du tableau correspond à une sonde et contient des informations comme les coordonnées physiques de la sonde sur la puce, la taille du spot, le signal d'hybridation (signal dans le spot) et le bruit de fond (signal autour du spot) pour chacun des 2 fluorochromes et de nombreux paramètres permettant une analyse qualitative de l'hybridation. Le logiciel actuellement utilisé au sein de notre équipe est GenepixPro® 5.1 (Axon®).

## V. TRAITEMENT DES DONNEES PRIMAIRES

A chaque étape mise en oeuvre pour obtenir des données d'expression génique des biais expérimentaux sont des sources de variabilité de la mesure finale. Les variations biologiques d'intérêt peuvent donc être masquées par des bruits techniques et biologiques. Aussi, outre la définition d'un plan expérimental adéquat, il est nécessaire d'appliquer une procédure systématique de traitement et de transformation à ces données afin de minimiser (voire corriger) ces variations indésirables.(Quackenbush, 2002; Tseng et al, 2001)

Le traitement des données primaires (preprocessing) vise à minimiser l'effet des biais expérimentaux sur les mesures et écarter les données de mauvaise qualité. Les principales étapes de traitements couramment appliquées sont :

- Filtration, afin d'écarter les spots défectueux (comètes, spots saturés)
- Normalisation
- Standardisation
- Détection des valeurs aberrantes, afin d'éliminer les mesures répétées non reproductibles

La suite logicielle MADSCAN développée par N Le Meur au sein de notre équipe propose l'ensemble de ces traitements en une seule procédure automatisée. Cette procédure peut être appliquée à une puce ou à plusieurs puces répétées (lot) en même temps. (www.madtools.org)

L'étape de filtration des données primaires est complémentaire à la détection des spots non valides par les logiciels d'analyse d'images. La filtration des données primaires se base sur les critères de qualités physiques et géométriques des spots tels que le niveau de saturation du signal ou le rapport signal sur bruit Dans MADSCAN, un arbre de décision oriente la filtration des mesures et leur attribue un score de qualité pour minimiser les biais systématiques

La seconde étape du processus de traitement des données est la normalisation intra-lame. La normalisation des données vise à minimiser (voire

corriger) les biais techniques, systématiques ou liés au hasard. Un déséquilibre des niveaux d'intensités moyens entre les deux fluorochromes est souvent observé. Il témoigne le plus souvent de biais techniques, tels que le biais systématique d'une incorporation hétérogène des fluorochromes ou l'usure inégale (au hasard) des aiguilles du robot servant à déposer les sondes sur la lame ou encore d'un mauvais réglage des paramètres de lecture du scanner. Par ailleurs, les spots de faible intensité présentent une variabilité plus importante que les spots de forte intensité car globalement plus sensibles aux biais expérimentaux. Des approches par régressions non paramétriques ont donc été proposées pour normaliser les données en fonction de leur niveau d'intensité.(Faller et al, 2003; Tseng et al, 2001; Workman et al, 2002) Parmi ces méthodes, l'algorithme de lowess fitness (LOcally WEighted Scatter plot Smoothing) et ses variantes sont actuellement les plus employées. Leur principe est basé sur le découpage des données en fenêtres d'intensités de taille connue (généralement un pourcentage du nombre de données), suivi d'une somme de régressions locales pondérées. L'ensemble des régressions locales est ensuite lissé pour former la courbe d'ajustement des données.

MADSCAN propose une version adaptée de l'algorithme de lowess fitness selon Yang *et al*.(Yang et al, 2001) L'algorithme de lowess fitness peut être sensible aux valeurs extrêmes de ratios et l'utilisation de l'ensemble des sondes de la puce pour normaliser n'est pas toujours efficace. De même, l'utilisation de jeux de sondes sélectionnés a priori c'est avéré inefficace car ces gènes effectivement stables dans certaines conditions peuvent varier dans d'autres. MADSCAN applique l'algorithme de lowess fitness à une sélection de « gènes invariants ».(Tseng et al, 2001) Ces sondes sont sélectionnées a posteriori par une méthode non paramétrique de calcul des rangs des intensités. Les sondes sont dites « invariantes » si et seulement si, d'une part la différence des rangs de leurs niveaux d'intensités (Cy3, Cy5) n'est pas significative (inférieure à un seuil d) et d'autre part que la moyenne de ces rangs soit comprise entre un seuil de signification (l) et le plus haut rang. Enfin, Yang *et al*.(Yang et al, 2002a) ont également mis en évidence une corrélation entre la position des spots sur la lame (liée aux aiguilles du robot de dépôt) et leur niveau d'intensité. Une normalisation spatiale, est alors nécessaire pour corriger cet « effet pointe ». MADSCAN propose différents modes de normalisation spatiale. Le choix de la méthode utilisée est fonction du pourcentage de « gènes invariants » estimé localement.

La normalisation intra-lame permet de centre la distribution des ratios sur une valeur identique pour toutes les puces. Toutefois, pour comparer des puces d'un même expérience, il est parfois nécessaire d'appliquer une normalisation inter-lames (scaling) ou standardisation afin de réduire l'écart de la variance des mesures entre les lames. Dans ce but, la distribution des ratios est standardisée (c'est à dire ramenée aux mêmes paramètres de dispersion). Comme pour la normalisation intra-lame, plusieurs techniques ont été développées.(Quackenbush, 2002; Yang et al, 2002a) MADSCAN propose de réduire la distribution des mesures d'expression de chaque lame répétée à la même déviation absolue de la médiane ou MAD (Median Absolute Deviation), cette mesure étant moins sensible aux valeurs atypiques que l'écart-type.

La dernière correction est la recherche des valeurs aberrantes (outliers) parmi les mesures répétées d'une même sonde. Les valeurs aberrantes ont, par exemple, pour origine la faible spécificité des sondes et/ou un bruit de fond hétérogène à la surface des lames. Aussi, l'application de tests statistiques permet d'évaluer et valider le niveau de reproductibilité des mesures répétées. Toutefois, les expériences de puces à ADN disposent généralement d'un faible nombre de réplicats ce qui limite l'application de la plupart des tests statistiques classiques. MADSCAN propose deux types de tests : un test t modifié par la médiane et le test de Grubb.(Burke, 2001) Ces tests présentent l'avantage de pouvoir être appliqués à un petit nombre de répétitions (n=3). MADSCAN permet également de tester la reproductibilité des mesures à l'intérieure des puces puis entre les puces répétées.

Finalement, une synthèse des niveaux d'expression par sonde à l'intérieur et entre les puces répétées offre une vision globale des résultats et de la qualité de l'expérience. MADSCAN propose l'intégration des données sous la forme d'une matrice de données d'expression « consolidées ». Pour chaque sonde, les moyennes des niveaux d'expression sont calculées par lames et entre les lames, associées à des indices de qualité des mesures, tels que le nombre de valeurs valides et le coefficient de variation.

## VI. TRAITEMENT ET INTERPRETATION DES DONNEES CONSOLIDEES

Une fois les données d'expression extraites, traitées de manière à corriger les fluctuations expérimentales, et validées, l'étape proprement dite d'analyse des données débute. De nombreuses méthodes et outils sont disponibles afin d'extraire de la connaissance à partir de ces données (« knowledge discovery »). C'est dans ce domaine que les efforts et progrès dans l'approche « transcriptome » ont été les plus importants ces dernières années. Différentes méthodes sont utilisées :

- L'identification des gènes différentiels par des méthodes statistiques.
- La mise en évidence de groupes de gènes ou d'échantillons comportant des profils d'expression communs par des méthodes de classifications.
- L'annotation fonctionnelle des gènes et de groupes de gènes sélectionnés (« clusters ») par l'utilisation des ontologies ou par une analyse automatisée des données de la littérature.

La stratégie d'utilisation de ces différentes approches est bien sûr dépendante de la finalité de la recherche, la règle étant souvent de combiner ces différentes méthodes selon un scénario (« script ») adapté à chaque étude. Il n'existe aucun scénario universel qui permettrait d'extraire de manière exhaustive l'ensemble des connaissances d'un jeu de données ou qui pourrait être utilisé quelle que soit l'étude réalisée.

### A. Mise en évidence de gènes différentiellement exprimés :

L'objectif de ces méthodes est l'identification d'un groupe de gènes dont les niveaux d'expression génique sont différents entre 2 (ou plus) groupes d'échantillons biologiques représentatifs de situations biologiques distinctes. L'utilisation de méthodes statistiques est ici essentielle pour vérifier que les différences observées de manière quantitative ne sont pas liées au hasard ou au bruit expérimental.

Toutefois, la structure même des données issues des études « transcriptome » comporte 2 caractéristiques qui imposent une adaptation des stratégies statistiques

classiques.(Draghici S, 2003) Les variations d'expression observées peuvent être autant liées aux fluctuations expérimentales qu'aux variations biologiques d'où l'intérêt d'un design expérimental solide et des étapes de traitement et de validation des données. De plus, fait relativement rare pour les statisticiens, les matrices de données d'expression sont dissymétriques de telle sorte que le nombre de variables (gènes) excède toujours celui des échantillons. Cette seconde constatation a imposé le développement de procédures statistiques spécifiques afin en particulier de mieux traiter le risque d'erreur de type I (présence de faux positifs).

Les méthodes proposées sont trop nombreuses pour être détaillées dans ce chapitre. Il est cependant intéressant d'expliciter leur principe général de construction qui est souvent basé sur 3 éléments :

- Calcul pour chaque gène d'un score permettant de classer les gènes du plus au moins « différentiel ».
- Comparaison du score obtenu pour chaque gène à la distribution des valeurs du score pour l'hypothèse nulle (qui est que le gène n'est pas différentiellement exprimé). Cette distribution est soit fixe (cas d'un test t classique) ou recalculée pour chaque jeu de données par des méthodes de permutation aléatoire (bootstrapping) des échantillons biologiques entre les groupes testés. On en déduit ainsi la probabilité statistique pour le gène d'être différentiellement exprimé.
- Correction de cette probabilité en regard du nombre de tests effectués en simultané. Cette correction est indispensable et est aisément justifiée par l'exemple suivant :10000 tests simultanés sont par exemple réalisés pour une matrice de 10000 gènes (un test par gène). La probabilité qu'au moins un gène soit faussement déclaré différentiel par le test est donc 10000 fois plus forte que pour la réalisation d'un seul test. Ainsi, si on utilise pour chaque test le seuil classique de 5 pour cent on obtient mathématiquement 500 (10000 * 0.05) gènes faussement déclarés comme différentiellement exprimés, ce qui est bien sûr non acceptable.

C'est cette dernière étape qui a essentiellement induit le développement de méthodes statistiques dédiées aux puces à ADN car la méthode de contrôle de l'erreur de type I utilisée jusqu'ici pour les tests multiples (correction de Bonferonni)

minimise fortement le nombre de gènes différentiellement exprimés induisant alors un nombre important de faux négatifs. L'utilisation de méthodes de type FDR (False Discovery Rate) permet un meilleur compromis entre faux positifs et faux negatifs.(Storey et al, 2003) Ces méthodes mesurent la proportion de faux positifs dans les gènes détectés comme différentiellement exprimés au contraire des approches classiques FWER (Family Wise Error Rate) qui mesurent la probabilité qu'il existe au moins un faux positif dans l'ensemble des tests. Il faut bien sûr rappeler que le nombre de faux négatifs est également intiment lié au nombre d'échantillons biologiques analysé par groupe (nombre de réplicats biologiques) soulignant encore une fois l'importance du dessin expérimental.

La méthode statistique SAM (Significance Analysis of Microarrays) est certainement aujourd'hui la plus utilisée.(Tusher et al, 2001) Elle permet, sur la base d'un test t modifié et de l'utilisation de permutations aléatoires des échantillons, de sélectionner un groupe de gènes différentiels en contrôlant le pourcentage estimé de faux positifs par méthode FDR. Des critères supplémentaires comme une amplitude moyenne de variation minimale entre les groupes peuvent être associés lors de la sélection. Le logiciel est accessible gratuitement sous forme d'une macro Excel® et prend en charge la comparaison de 2 ou plus échantillons, la comparaison d'échantillons appariés et les études cinétiques.
Dans le cas d'analyses multifactorielles, des méthodes plus complexes utilisant une approche bayesienne(Smyth, 2004) ou une ANOVA(Kerr et al, 2000) sont souvent nécessaires.

### B. Mise en évidence de groupes de gènes ou d'échantillons comportant des profils d'expression communs

L'intérêt des regroupements des gènes par profil d'expression commun a été démontré dès la fin des années 90. A partir d'une étude de cinétique sur S. Cerevisiae, M. Eisen a montré que les gènes présentant un profil d'expression génique similaire (« cluster » de gènes) sont très souvent impliqués dans une même fonction biologique.(Eisen et al, 1998) De plus, S. Tavazoie a montré la présence, au sein des régions promotrices des gènes d'un cluster, d'une séquence commune

correspondant au site de fixation d'un facteur de transcription, chaque cluster étant ainsi associé à une séquence spécifique.(Tavazoie et al, 1999) Ces données suggèrent que l'expression des gènes d'un cluster soit co-régulée par un même facteur de transcription pour exercer une fonction biologique commune. L'application des techniques de regroupement aux échantillons biologiques selon leur profil d'expression génique a également montré tout son intérêt en permettant par exemple l'identification de nouveaux sous-groupes étiologiques de maladies complexes.(Ballester et al, 2006)

L'objectif général de cette approche est de définir des groupes de gènes ou d'échantillons dont les éléments sont proches les uns des autres et distants des éléments n'appartenant pas au groupe sur la base de leur profil d'expression génique. Par convention, la suite du chapitre décrira l'utilisation de ces méthodes pour le regroupement des gènes. Les mêmes méthodes sont également appliquées aux échantillons biologiques.
La notion de distance entre gènes est au centre de ces méthodes de classification. Il s'agit d'un nombre positif qui reflète la proximité (calcul de similarité) ou l'éloignement (calcul de dissimilarité) entre 2 gènes. Parmi les distances les plus utilisées on peut citer le coefficient de corrélation de Pearson qui est une mesure de similarité (plus la distance est élevée, plus les éléments sont proches), et la distance euclidienne qui est une mesure de dissimilarité (plus la distance est élevée, plus les éléments sont éloignés). Il n'existe pas de distance meilleure qu'une autre, le choix de la distance devant être adapté à l'étude réalisée.

La spécificité de chaque méthode de regroupement est liée d'une part à la méthode de calcul de la distance et d'autre part à la manière dont les groupes sont ensuite créés. Deux grands types de méthodes de regroupement doivent être distingués et opposés. Les approches non-supervisées analysent les données d'expression sans a priori (techniques exploratoires). A l'opposé, les approches supervisées analysent les données d'expression au regard de connaissances préalables.

- Méthodes non supervisées

Au sein des approches non supervisées, les classifications hiérarchiques ascendantes et les méthodes par partitionnement sont les plus utilisées. La classification hiérarchique ascendante popularisée dans le milieu de la transcriptomique par M Eisen (on parle parfois d'« eisenification des données ») est certainement la méthode la plus utilisée actuellement à travers les gratuiciels « Cluster » et « Tree View » (rana.lbl.gov/EisenSoftware.htm). Après le choix d'une distance (« métrique »), qui est souvent le coefficient de corrélation de Pearson normalisé ou la distance euclidienne, les gènes ou groupes de gènes sont agglomérées 2 à 2 en débutant par les éléments les plus proches. Deux gènes agglomérés deviennent alors un groupe indissociable et ainsi de suite. Des méthodes de calcul des distances inter-groupes doivent donc être définies pour pouvoir agglomérer 2 groupes de gènes. Le lien moyen est la méthode la plus utilisée. Il s'agit de la distance moyenne entre toutes les paires d'objets de ces 2 groupes. Le résultat obtenu est présenté sous la forme d'un arbre de classification (dendrogramme) des gènes. La similarité des profils d'expression génique entre 2 gènes ou groupes de gènes est alors représentée par la distance entre le sommet de l'arbre et le premier embranchement commun à tous ces éléments. La matrice des données d'expression à l'origine de la classification est également présentée en face de l'arbre sous forme colorisée (« heat map »), les gènes ayant été ordonnés en fonction de l'organisation du dendrogramme et les valeurs d'expression remplacées par un codage couleur distinguant les valeurs d'expression selon leur intensité. De nombreux exemples de ce mode de classification peuvent être retrouvés dans la partie « résultats ». L'avantage et l'inconvénient de cette technique est l'absence de définition préalable du nombre de groupe de gènes à identifier. Aucune connaissance préalable sur l'organisation du jeu de donnée n'est donc indispensable, en revanche la délimitation de groupes de gènes n'est pas automatique et doit donc être réalisé par l'investigateur incluant donc une part subjective à l'analyse.

Les méthodes de partitionnement au contraire imposent la définition préalable du nombre de groupes de gènes attendus. Les méthodes des k-moyennes (k-means) et SOM (Self-Organizing Maps) sont les plus utilisées. L'objectif global est de minimiser la distance de chaque gène au centre du groupe auquel il appartient. Ces méthodes sont non déterministes c'est-à-dire qu'appliquées plusieurs fois de suite sur le même jeu de données elles ne produiront pas le même résultat. La détermination

du nombre de groupes de gènes est essentielle pour permettre une analyse efficace mais ce nombre est en général inconnu. La stratégie choisie est souvent de démarrer par une analyse hiérarchique pour évaluer le nombre de groupes de gène attendus puis de lancer ensuite des méthodes par partitionnement qui vont permettre une délimitation automatique des clusters.

- Méthodes supervisées

Ces méthodes utilisent des connaissances préalables pour construire des classifications. Elles ont pour l'instant été principalement appliquées aux échantillons biologiques. Leur application à l'analyse d'échantillons biologiques humains dans de nombreuses pathologies a permis l'identification et la validation de nouveaux bio-marqueurs diagnostiques et pronostiques. La classification (class prediction) de nouveaux échantillons biologiques (test set) dans des classes diagnostiques préétablies à partir de leur profil d'expression génique dans un jeu de données d'apprentissage (training set) bien annoté est ainsi réalisée par ces méthodes permettant d'orienter le diagnostic pour ces nouveaux échantillons. Ces méthodes supervisées comprennent donc en une phase initiale d'apprentissage sur des gènes dont la classification est connue, puis une phase de prédiction sur un ensemble d'autres gènes à classer.

De nombreux logiciels académiques (donc gratuits) intègrent un large éventail de méthodes de classification des données et de visualisation des résultats. Ils sont souvent simples d'utilisation et possèdent une interface graphique conviviale. A titre d'exemple on peut citer Genesis.(Sturn et al, 2002)

### C. L'annotation fonctionnelle des gènes et de groupes sélectionnés (« cluster »)

Afin de mieux extraire le sens biologique des expériences de transcriptomique, des méthodes permettent l'intégration aux mesures d'expression des gènes, de données issues d'autres sources d'informations comme les ontologies ou la littérature.

- Utilisation des ontologies

L'aboutissement des projets d'analyse à large spectre du génome humain tant au niveau des gènes, que du transcriptome ou du protéome a produit une énorme masse d'information sur les gènes et leurs produits. L'organisation et la standardisation des ces informations sont ainsi vite devenue une nécessité afin de pouvoir les exploiter pleinement. Le consortium HGNC (HUGO Gene Nomenclature Committee) a défini une nomenclature « officielle » des gènes. En parallèle, Gene Ontology (GO) a été développé pour annoter les gènes par rapport à la fonction de leurs produits.(Ashburner et al, 2000) Ces annotations sont réalisées à partir d'un vocabulaire défini et structuré de « termes GO » décrivant des rôles biologiques, des localisations sub-cellulaires ou des fonctions moléculaires. Pour chaque lien crée entre un « terme GO » et un gène, un niveau de preuve (référence scientifique, prédiction automatique,...) est indiqué. De nombreux outils comme GOMiner,(Zeeberg et al, 2003) Onto-Express(Draghici et al, 2003) ou Fatigo(Al Shahrour et al, 2004) utilisent Gene Ontology pour réaliser une annotation fonctionnelle des groupes de gènes identifés par des approches statistiques ou de clustering. L'objectif est l'identification d'une fonction biologique ou moléculaire sur-représentée dans le cluster analysé versus le reste des gènes analysés sur la puce. Un très grand nombre de clusters d'intérêt peut ainsi être annoté de manière automatique et donc beaucoup plus rapide.

- Utilisation des données de la littérature

D'autres méthodes ont été développées pour utiliser la masse de données que représentent les millions d'articles scientifiques contenus dans les base de données comme Pubmed. Ces méthodes sont basées sur la notion de co-occurrence des gènes dans les publications. Une analyse automatique des abstracts contenus dans Pubmed permet de détecter les gènes co-cités dans un même article. La comptabilisation du nombre de co-citations d'un couple de gènes est utilisée pour déterminer la force du lien unissant ces 2 gènes. On peut ainsi déterminer pour un gène donné ses plus proches « voisins » dans la littérature ou identifier au sein d'un cluster de gènes un réseau de gènes précédemment reliés par les données de la littérature.(Jenssen et al, 2001) Un des logiciels actuellement les plus populaires pour cette application est (Ingenuity Pathways Analysis, Ingenuity®).

# RESULTATS

## I. ETUDE 1

Le Meur N, Lamirault G, Bihouee A, Steenman M, Bedrine-Ferran H, Teusan R, Ramstein G, Leger JJ. A dynamic, web-accessible resource to process raw microarray scan data into consolidated gene expression values: importance of replication. Nucleic Acids Res. 2004 Oct 8;32(18):5349-58.

Un point critique de la technologie des puces à ADN est la suite des étapes mise en oeuvre pour obtenir des données d'expression génique fiables. A chaque étape des biais expérimentaux peuvent représenter sources de variabilité la mesure finale. Les variations biologiques d'intérêt peuvent donc être masquées par des bruits techniques et biologiques.

Le premier objectif de ce travail a été de construire une suite logicielle efficace de traitement de données de puces à ADN permettant d'obtenir à partir d'un ensemble de résultats bruts individuels de puces, une matrice organisée et consolidée de valeurs d'expression génique. Nous avons pu montrer son efficacité dans la correction des biais techniques pour améliorer la qualité des données.

Le second objectif de ce travail a été à partir de cette suite logicielle de réaliser une étude de puissance pour montrer l'importance du dessin expérimental et l'effet de la répétition des mesures sur la fiabilité des résultats en particulier la mise en évidence des gènes différentiellement exprimés entre deux situations biologiques.

**Published online October 8, 2004**

*Nucleic Acids Research, 2004, Vol. 32, No. 18* ***5349–5358***
*doi:10.1093/nar/gkh870*

# A dynamic, web-accessible resource to process raw microarray scan data into consolidated gene expression values: importance of replication

**Nolwenn Le Meur*, Guillaume Lamirault, Audrey Bihouée, Marja Steenman, Hélène Bédrine-Ferran[1], Raluca Teusan, Gérard Ramstein[2] and Jean J. Léger**

Ouest genopole, Institut du Thorax, Institut National de la Santé et de la Recherche Médicale (UMR 533), Faculté de Médecine, 44035 Nantes, France, [1]Centre National à la Recherche Scientifique (UMR 6061), Rennes, France and [2]Laboratoire d'Informatique de Nantes Atlantique, 44322 Nantes, France

Received June 25, 2004; Revised August 17, 2004; Accepted September 17, 2004

## ABSTRACT

**We propose a freely accessible web-based pipeline, which processes raw microarray scan data to obtain experimentally consolidated gene expression values. The tool MADSCAN, which stands for MicroArray Data Suites of Computed ANalysis, makes a practical choice among the numerous methods available for filtering, normalizing and scaling of raw microarray expression data in a dynamic and automatic way. Different statistical methods have been adapted to extract reliable information from replicate gene spots as well as from replicate microarrays for each biological situation under study. A carefully constructed experimental design thus allows to detect outlying expression values and to identify statistically significant expression values, together with a list of quality controls with proposed threshold values. The integrated processing procedure described here, based on multiple measurements per gene, is decisive for reliably monitoring subtle gene expression changes typical for most biological events.**

## INTRODUCTION

During the last decade, cDNA microarray technology has been extensively applied to determine gene expression levels in diverse tissues, animals and diseases, at high throughput levels. As a result of the increasing knowledge of several genomes (especially the human genome), thousands of gene-fragments have been spotted or *in situ* synthesized to globally monitor various gene expression situations. Attention has been paid to mathematical (statistical) methods pertinent for the analysis, organization and handling of the enormous quantities of gene expression data [reviewed in (1)]. When comparing different situations like patients and controls, or when analyzing ontogenic or kinetical events, the challenge is to identify the combinatorial and hierarchical complexity of the gene expression profiles. Parallel to those efforts on interpretation of the data, other studies have aimed to identify the different physical and biological factors which have to be controlled to improve the reliability of massive gene expression studies (2–4). The multiple experimental steps involved in microarray procedures are sources of often badly controlled variation, which is superimposed on the biological variation under study. Experimental variation along the successive steps of preparation, purification and labeling of RNA samples, as well as the hybridization conditions, are inherent to all microarray experiments. Mechanical and optical distortions could locally or globally influence the raw expression values obtained after microarray image scanning. In addition, other factors like intrinsic heterogeneity, conditioning parameters and even erratic contamination of the biological samples may modify the gene expression results. To compensate (and/or better evaluate) the importance of these composite experimental and biological noises in microarray experiments, diverse numerical treatment procedures of the raw microarray scan image values and quality measures have been proposed. These procedures include filtering of bad spots following segmentation methods, normalizing between two channels (or signal scaling within monocolor microarrays), comparative scaling between different chips, and diverse statistical methods for selecting (ranking) differentially expressed genes (1,5). The most reliable way of evaluating the ratios between the different experimental noises and the biological signals is to produce replicate gene measurements within each microarray and to hybridize replicate microarrays with replicate targets obtained from the same biological samples. The metrological importance of such replications in microarray gene expression studies (6–9) casts doubts on the numerous microarray analyses performed with only singular gene spots and/or without sample replicates. The high cost associated with microarrays does not justify the metrological insufficiencies of any experiment. The accessibility to high throughput spotting robots deposing up to 25 000 spots per chip combined with a careful selection of a few thousands of theme-relevant genes now allows the use of such noise-informative chips and the design of corresponding reliable experiments.

*To whom correspondence should be addressed at INSERM U533, Faculté Médecine, BP 53508, 44035 Nantes cedex 1, France. Tel: +33 240412957; Fax: +33 240412950; Email: nolwenn.lemeur@nantes.inserm.fr

Here, we present a freely accessible web-based pipeline which processes raw microarray scan data to obtain experimentally consolidated gene expression values. The proposed module, MADSCAN, MicroArray Data Suites of Computed ANalysis (http://www.madtools.org), makes a practical choice among the numerous methods available for filtering, normalizing and scaling microarray data, in a dynamic and automatic way. Using a careful experimental design with replication information, we present different numerical and statistical methods, detection of outlying expression values and data integration with a list of quality controls, through proposed threshold values. A tutorial for MADSCAN is included on the website.

## MATERIALS AND METHODS

### Biological samples

Cardiac tissue was obtained from the left ventricle of explanted hearts from two male patients who underwent heart transplantation. One patient (experiment 1) was affected by idiopathic dilated cardiomyopathy; the other patient (experiment 2) was affected by valvular heart disease and coronary artery disease. Both samples were compared to a common reference that was obtained from pooled RNAs extracted from the left and right ventricles of explanted hearts from 47 end-stage heart failure patients (G. Lamirault, N. Le Meur, M.F. Le Cunff, C. Chevalier, I. Guisle, A. Bihouée, J.J. Léger and M. Steenman, manuscript in preparation).

### RNA isolation and labeling

Total RNA was isolated using TRIZOL®Reagent (Life Technologies). Two parallel RNA extractions from two different samples (spatially separated) of the same tissue were performed. Poly $A^+$ RNA was isolated using the Oligotex mRNA kit (Qiagen) and quality was assessed using an Agilent 2100 bioanalyzer. Cy3- and Cy5-labeled cDNA was prepared using the CyScribe cDNA Post Labelling Kit (Amersham Pharmacia Biotech). Samples from the two patients under study were each labeled individually with Cy3. The reference pool was labeled with Cy5. No dye-swap was used.

### Microarrays

Microarrays were prepared in-house using 50mer oligonucleotide probes (MWG Biotech). The probes were arrayed onto epoxy-silane-coated glass slides using the Lucidea printer from Amersham. The 4217 genes represented on the microarray were selected for involvement in skeletal muscle and/or cardiovascular normal and pathological functioning (10–13). Gene selection was based on (i) subtractive hybridization experiments, (ii) genome-wide microarray hybridizations, (iii) literature data. Each Cy3-labeled sample was mixed with equal amount of Cy5-labeled sample, pre-incubated with human Cot-I DNA (Gibco-BRL), yeast tRNA and poly $A^+$ RNA, and hybridized to the microarrays. Two independent hybridizations were performed for each RNA sample, leading to four hybridizations per patient. Hybridized arrays were scanned by fluorescence confocal microscopy on a ScanArray 4000XL (GSI Lumonics, Downers Grove, IL) at laser power ranging from 75 to 100% and photo-multiplier tube gain settings ranging from 65 to 100%. Measurements were obtained separately for each fluorochrome at 10 μm/pixel resolution.

### Microarray data acquisition

Signal intensities were extracted with the GenePix Pro 5.0 image analysis software (Axon Instruments, USA). Segmentation of the spots was done using the adaptive approach. Segmentation criteria were optimized visually for each slide. Alternate standard deviation (SD2) was chosen to quantify background SD. This setting uses the median of the background pixels as an estimator of the center of the distribution. This method is less susceptible to skewing by very bright pixels. Our data processing procedure uses background corrected median intensities; the given ratio corresponds to the ratio of background corrected median intensities. For further quality controls (see the preprocessing step in Results and Discussion, and in the tutorial), analytic parameters provided by the GenePix Pro 5.0 image analysis software were used. Other image analysis software like Quantarray (PE. Packard Biochip technologies, USA), Imagene (BioDiscovery, USA) or ScanAlyze (http://rana.lbl.gov/, Stanford University, USA) also deliver the minimal set of parameters required to perform the MADSCAN procedure. For a comparison between different image analysis software, see http://cardioserve.nantes.inserm.fr/ptf-puce/publications.php. Analysis files issued from different types of image software can be reformatted following procedures noted in the MADSCAN tutorial.

### Power study

A power study of a standard *t*-test was performed on the heart expression data set of experiment 1 (four replicate spots for each oligonucleotide and four replicate chips). Only genes with at least two valid *M*-values [$=\log_2$(signal ratio)] for each array were selected for power analysis. We thus selected 3804 genes with reliable *M*-values. The 'power *t*-test', which is implemented in the '*c*test' package of *R* (14) was applied in five replication conditions: 4, 6, 8, 12 or 16 replicate *M*-values. Parameters of the power test were defined as follows:

- For each gene the mean level of differential expression between the two fluorochromes ($\Delta$) was defined as the arithmetic mean of the four arithmetic means of the 4 *M*-values in each of the quadruplicate chips.
- The parameter for data variability (SD) was arbitrarily set as identical for all genes. SD was calculated as the median of the 3804 SDs determined from the replicate *M*-values for each gene.
- Significance level ($\alpha$) was a priori set to 0.05, but a Holm correction (15) was applied to $\alpha$ for each gene in order to account for multiple testing hypothesis. The 3804 genes under analysis are ranked according to their individual *P*-values, by application of a standard one-sample *t*-test. On the basis of the calculated rank of the corresponding gene, the basal $\alpha$-value of 0.05 was then corrected individually.

Using the values fixed for SD and $\alpha$ for each gene, the individual power test was performed on the basis of one sample and a two-sided *t*-test. Power values of $(1 - \beta)$ were deduced for each gene in the five replication conditions analyzed. To evaluate the effect of between-gene differences in sampling

variation on the power test values, the power test was calculated with three different values of SD in two particular replication conditions (4 and 12 replicates). The three different SD values were defined as first quartile, median (as earlier) and third quartile of the 3804 SDs previously calculated. Other parameters of the power test were left unchanged. Power values ($1 - \beta$) were calculated for both replication conditions and the three variability levels.

### Estimation of false positive and false negative rates

The false positive (FP) rate is the proportion of negative cases that were incorrectly classified as positive in the predicted condition compared to the experimental observation. The false negative (FN) rate is the proportion of positive cases that were incorrectly classified as negative in the predicted condition compared to the experimental observation. Genes differentially expressed between the heart expression data sets of experiments 1 and 2 were first identified by a SAM modified two-class *t*-test (16), using 16 (4 within- and 4 between-chip) replicates for each data set. The number of differentially expressed genes was then determined based on six different replication conditions: 4, 8 or 12 replicates with various proportions of within- and between-chip replicates (see Supplementary Material). Six different two-by-two confusion matrices (17) were built to determine the FP and FN rates in the six simulated replication conditions with regard to the experimental situation based on 16 replicates.

### MADSCAN implementation

MADSCAN was written in R (14) and Perl. A user-friendly web-interface was implemented in PhP to allow easy access (http://www.madtools.org) and rapid handling of data on our local server (PowerEdge 4600, Dell, USA). Access is obtained through a password, given to any requester. The raw microarray data are uploaded as compressed tabulated text files. MADSCAN analysis can be done either step by step or from A to Z, i.e. one can either apply one test at a time or all steps in a single and complete procedure. The results can be downloaded from the web-interface, where a window of results displays a summary of the performed procedure. Alternatively, MADSCAN results can be recovered through an e-mail service.

## RESULTS AND DISCUSSION

### Outline of analysis procedure

Our goal was to provide a practical, accessible, integrated suite of different analytic procedures for the handling of raw data issued from two-fluorochrome (color) image scanning of microarray glass slides and to obtain consolidated expression values. According to the MIAME (Minimum Information About a Microarray Experiment) glossary, data processing means 'the set of steps taken to process the data, including the normalization strategy and the algorithm used to allow comparison of all data' (18). Draghiči (5) defined preprocessing as the initial step that extracts and enhances meaningful data characteristics from raw data files from scanned images. Preprocessing prepares the data for the application of successive procedures or analytical methods. Using tabulated text

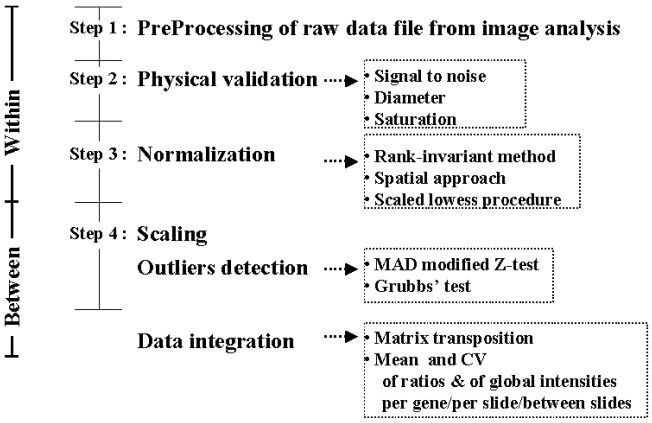



**Figure 1.** MADSCAN procedure steps within each chip and between replicate chips.

files of raw microarray image values issued from widely used scanners and related image analysis software, we have developed a four-step procedure to transform the raw data into consolidated, robust expression values: the first three steps concerned each individual chip, whereas step 4 integrated the expression values issued from replicate chips if available (Figure 1).

The proposed integrated tool has been constructed around a few now well-accepted main analytic steps to numerically handle microarray image values within one chip and between replicate chips. Most of the used algorithms or methods—such as the rank invariant method (19), lowess fitness normalization (20,21) and outlier detection (although never used for microarrays) (22,23)—have been defined and documented by others. The corresponding methods/algorithms have been reformatted in a plug-in architecture system to make the whole numerical procedure reliable and fluent. Algorithmic approaches chosen for each step and modifications or adaptive processes made along the procedure are described in the following subsections. The computational tool, MADSCAN, is freely accessible via a web server (http://www.madtools.org), where detailed information is available in a tutorial.

### The importance of the experimental design

Before describing the different steps of the MADSCAN procedure, we addressed the important issue of the reproducibility of microarray experiments. We proposed a 'reference design' with an experimental procedure based on replicate spots within each microarray, and replicated microarrays for two spatially separated samples from each tissue, compared in a hybridization to a reference sample (Figure 2). The replicate spots are issued from different print-tips and are therefore printed in different array blocks. This procedure allows the evaluation of the importance of the biological noise—due to sample heterogeneity—and numerous experimental noises. The latter could arise from variations in the molecular biology procedures for the extraction and labeling of RNA samples (e.g. dye quality, or possibly dye-swapping), from physical distortions in glass slides and from the scanner (optical irregularities in the laser performances and in the excitation of the fluorochromes). To be able to take into account such

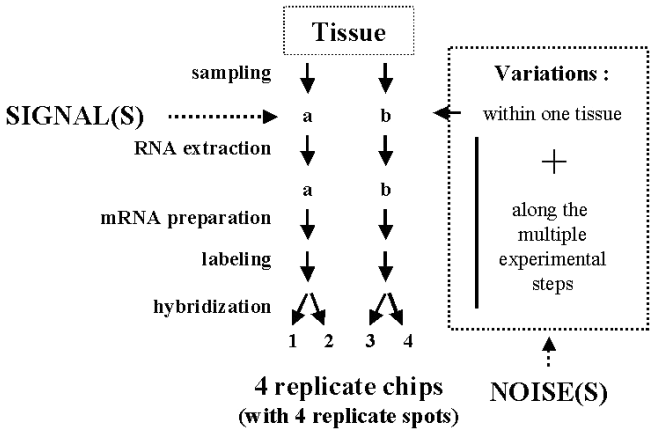



**Figure 2.** Experimental design. Two independent RNA samples (a and b) from the same tissue, replicated spots within one chip and replicated chips for one biological point are necessary to discriminate between the signal under study and those due to the inherent experimental noises.

composite noises from one chip, it is obvious that microarrays need to contain at least triplicate spots. This allows the statistical evaluation of the internal variability of the signals corresponding to each gene (oligonucleotides or PCR products) and the detection of outlying values within one chip. Furthermore, a minimum of four (two pairs) replicate chips is necessary to evaluate the variations between the two independent RNA samples issued from the same tissue (6,8). The within-chip replicates reveal only technical noises, whereas the between-chip replicates give information on both technical and biological noises. Microarray experiments as designed in Figure 2 actually allow simultaneous measurements of the different technical noises, together with that of the biological signal under investigation. In addition, randomized print-tip usage allows a non-uniform distribution of the replicate spots throughout the array. Together with a randomization of the numerous experimental procedures and the use of replicates, this is crucial to obtain statistically significant data.

### Preprocessing of raw data files from image scan analysis

Whatever the type of scanner or related software used, MADSCAN starts with tabulated text files composed of at least eight columns per hybridized microarray. These columns contain information on block designation, gene name, gene ID or annotation, measured intensities in both channels, local (or equivalent) background intensity values for both channels and image analysis software flags for a first determination of spot quality (diameter deviation of the spot, location). Nomenclature and gene annotation have to be carefully formatted. Replicate spots (if present) must be precisely annotated to be identified as such during the data processing procedure.

### Physical validation or quality filtration

The overall quality of the raw image data (before any filtering) is calculated for each print-tip group (block) of spots according to the median values and the variation coefficients of signal and background intensities. Spot diameters and their SD are also determined. In spite of the importance of assessing spot quality, relatively few image analysis software packages provide systematic quality filtration (5). MADSCAN offers physical validation and quality filtration step by step, following a decision tree with a scoring procedure based on successive exclusion thresholds. Each feature is thus tested against a series of quality criteria (image analysis flags, signal-to-noise level, diameter variation and saturation level). Five different arbitrary scores can be attributed according to the spot quality. Score 0 is used for flawed spots whereas most of the good spots have score 2. Scores 3 and 4 are attributed to spots that are partially saturated for one of the channels. For those spots, the expression ratio is calculated from the regression ratios between the intensities of each pixel composing the spot. Score 5 is attributed to features partially saturated in both channels and their expression ratio is calculated as the ratio of their percentages of saturation. Fully saturated spots in both channels are flagged because no reliable information on the pixel values and distribution is available (see tutorial for details). Chips made in-house contain 15 000 to 20 000 spots. Using our conditions for hybridization and image scanning, 5–8% of the spots are flagged (=score 0) whereas 92–95% of the spots pass the quality control criteria. The percentage of partially saturated spots (score 3–5) is generally relatively low (0.05–0.1% of the spots).

### Within-chip normalization step

Normalization issues have been addressed early in the development of microarray data treatment (19–21,24). It is considered an essential step to minimize experimental systematic and random biases, arising from technical variations inherent to the high throughput and complex experiments. The main aspects of any normalization process are whether or not to select a set of reporter (invariant) genes as a reference for the normalization process and whether or not to consider spatial and intensity value-dependent biases. Since most microarrays contain several thousands of spots, and since hybridization values are mostly distributed in an equilibrated (pseudo-gaussian) way in experiments comparing test and control tissues, we chose to adapt the rank invariant method developed by Tseng *et al.* (19) in our procedure. A set of invariant spots or non-differentially expressed genes (if no replicates were spotted) is a posteriori selected from all validated expression ratios for each chip. The rank of Cy3 and Cy5 intensities of each gene on the chip is computed separately. If the ranks of the two intensities for a given gene differ less than a fixed threshold and the rank of their averaged intensity is not among the highest or lowest ranks, this gene is classified as a non-differentially expressed gene. Figure 3A shows an *M–A* plot [$M = \log_2$(signal intensity ratio) and $A = \log_2$(averaged signal intensities) (20)], with a selection of such invariant spots, following the application of the rank invariant method. The invariant spots in Figure 3A are sandwiched between the differentially expressed genes. As has already been described (20), the distribution of expression ratios is intensity dependent and therefore a non-linear normalization method must be used. The *lowess fitness* method, using the set of identified invariant genes, has been incorporated in our MADSCAN procedure. To assess the efficiency (robustness) of coupling

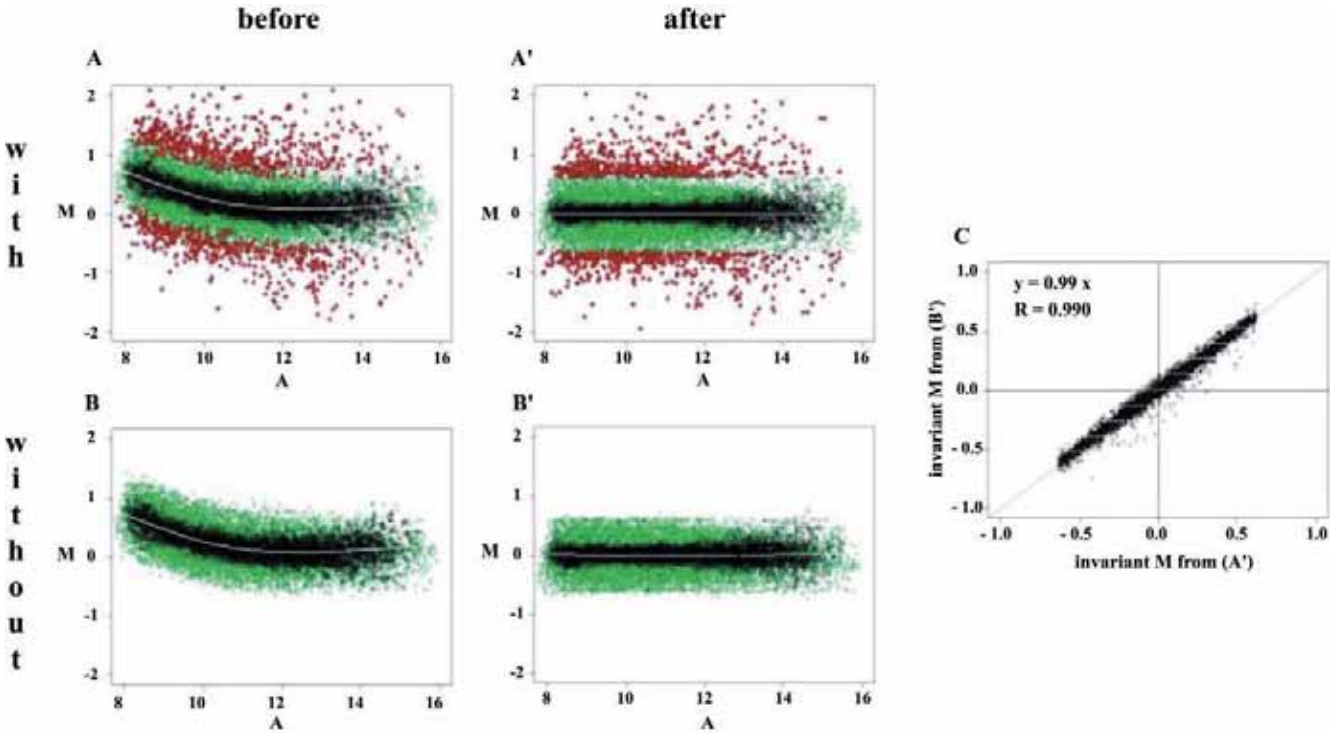



**Figure 3.** *M–A* plots before (**A** and **B**) and after (**A′** and **B′**) global *lowess* normalization, using rank invariant spots. The spots that are potentially differential in graphs (A) and (A′), [$M$ >/< Median($M$) ± 2.5 * MAD($M$)], were eliminated for the determination of invariant spots used for further data normalization in graphs (B) and (B′). 'With' and 'without' refers to the presence or absence, respectively, of potentially differentially expressed genes. The presented expression values were from experiment 1. (**C**) Represents the correlation between the 85% invariant genes, common to the gene populations in graphs A′ and B′.

both rank invariant and *lowess fitness* methods, we removed all identified putative differentially expressed spots (genes) from the original raw expression file [$M$ >/< Median($M$) ± 2.5 * MAD($M$) (median absolute deviation of $M$-values)]. We then applied both complementary methods for normalization on the reduced raw expression files. Figures 3B and 3B′ show the new set of invariant genes and the raw and normalized expression data, respectively. Eighty-five percent of the invariant genes selected before and after the file reduction are identical. A very strong correlation coefficient of 0.99 is observed between both sets of independently normalized expression values (Figure 3C). This is obviously due to the high number of invariant spots (genes) present (~50% of the total amount of spots).

As described by Yang *et al*. (21), the use of a spatial approach is also crucial. The signal as well as the background intensity is often heterogeneous within a slide. This is due to the unavoidable spot dispersion over a relatively large surface, the use of several spotting pins and possible geometrical variations within glass slides. An additional refinement of the normalization procedure thus has to be applied to chips containing more than a few thousand spots. A normalization procedure per zone, usually print-tip group, allows to correct spatially dependent dye biases and probe delivery variations between the different pins as well as other geometrical and optical defects. Practically, in MADSCAN, the normalization is first attempted pin-by-pin (print-tip group), then by proximal or global approach, depending on the number of invariant spots present within individual blocks, contiguous blocks or the whole chip, respectively. We found that at least 50% of invariant genes among all genes under analysis are needed to obtain a robust normalization curve. To illustrate the used procedure and the importance of spatial normalization, *lowess* normalization procedures were applied based on invariant spots selected from either individual blocks (individual print-tip) or proximal blocks or all blocks in a 48-block chip with 420 spots per block. Five individual lowess fitness curves arbitrarily chosen among the 48 different ones obtained in each spatial condition are graphically represented in Figure 4. It is easily seen that the best superposition of the five curves is obtained when the rank invariant method was applied pin-by-pin rather than using proximal blocks or all blocks.

## Scaling and outlier detection

In a metrologically controlled experiment—as described in Figure 2—the presence of replicated features within each slide and of replicated slides for each biological sample allows a statistical validation of the expression results after the three first steps of the procedure (Figure 1). First, scaling procedures have to be applied to bring the variances of filtered and normalized expression values between replicated chips at the same variation level (5,20). Outlying values within the series of expression values obtained for each gene from several spots can then be identified. Because of the low number of replicates in microarray experiments, we propose to apply modified statistical tests. A $z$-score modified by MAD is used to detect outliers within and between slides. In the MADSCAN

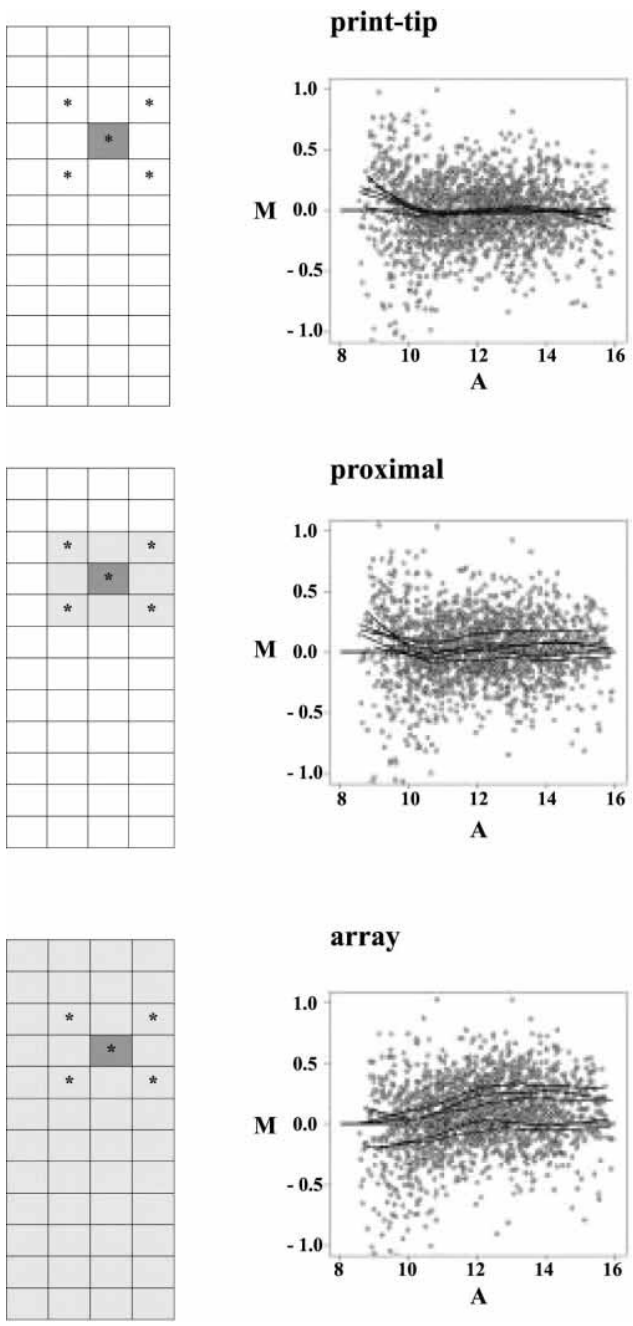



**Figure 4.** Comparisons between global, proximal and local normalization procedures. Five individual lowess fitness curves corresponding to five arbitrarily chosen blocks (asterisks) are represented according to each of the three spatial normalization modes. Light gray blocks represent an example of blocks chosen for selection of invariant genes, to normalize the raw *M*-values in the dark gray block, in each of the three modes. Invariant genes within the dark gray block are part of the invariant population used in each mode. The superposition of the five selected curves shows how uncontrolled local variations may influence the final expression values. The expression values presented here were from experiment 2.

procedure, we have implemented both the MAD and the ESD (Extreme Studentized Deviate or Grubb's test) procedures (22,23). The procedure for detecting outliers requests a minimum of three replicated values. The replicates may be within chips and/or between chips. Outlier detection can be applied iteratively with both tests, until no more outlier is detected.

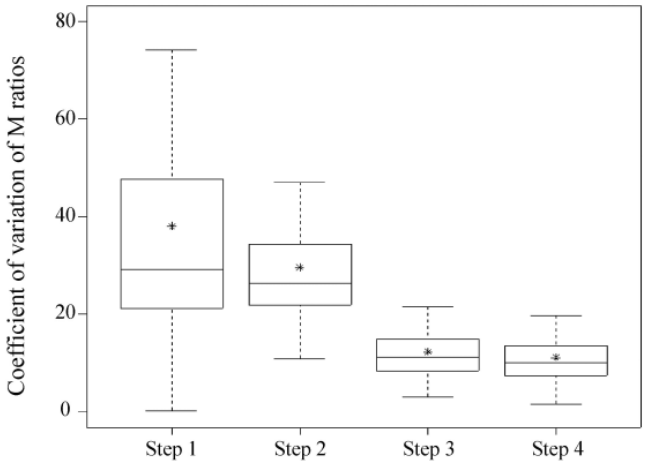



**Figure 5.** Decrease of the coefficients of variation of expression ratios, along the different MADSCAN analysis procedure steps. The expression values were from experiment 2, using 2 × 2 replicate chips with four replicate spots for each gene. The 'box' in a 'box and whisker' plot shows the median of the values as a line, the mean as an asterisk and the first (25th percentile) and third quartile (75th percentile) of the expression values distribution as the lower and upper parts of the box, respectively. The 'whiskers' shown above and below the boxes represent the largest and smallest observed values, respectively, that are less than 1.5 box lengths (interquartile range) from the end of the box. When the box is in the middle of the whiskers, the data are probably more evenly distributed (steps 3 and 4). Steps 1 to 4 are as in Figure 1.

## Crucial steps in microarray data treatment

The presence of replicate spots for each gene on each individual chip and on replicate chips allows the calculation of the within-chip as well as the between-chip coefficients of variation (CV) of the expression ratios ($=2^M$), respectively, at each of the four steps described in Figure 1. Figure 5 shows the variations of the CV calculated from the medians of expression ratios for each gene, in a typical experiment involving 2 × 2 replicate chips with four replicate spots for each gene (Figure 2). Step 3, corresponding to the within-chip normalization procedure, is clearly the most decisive step for reducing the absolute value and the variation of the CV. The CV distributions around their median values are approximatively gaussian, even though they are obviously higher for low intensity values (25). First and third quartile values in each of the four CV distributions are central visual elements for evaluating and controlling the quality of the expression values obtained for each individual chip and for replicate chips in the MADSCAN procedures. In contrast, step 4 does not significantly alter the CV values and their relative variations. This has to be related to the very small number of outlying values usually detected for each gene. However, this does not mean that outlier detection and elimination do not play a role in the CV calculations.

## Spot replicates and the detection of subtle expression changes

The robustness of the proposed 'reference design' with within- and between-chip replicates is illustrated by means of (i) a

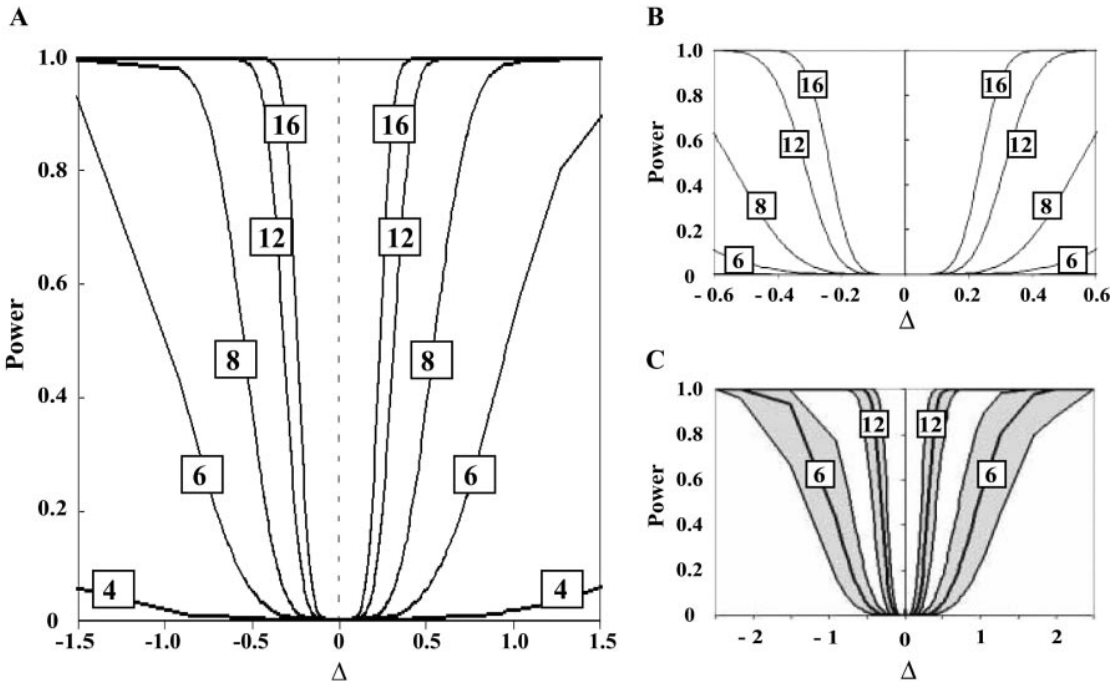



**Figure 6.** Validation experiment and power analysis, using replicate spots and replicate chips. The set of expression values for power calculations were from experiment 2. (**A**) Power values (1 − β) calculated in five replication conditions (4, 6, 8, 12 or 16 replicate *M*-values) were plotted against Δ, the mean level of expression values between the two fluorochromes, which is calculated as the arithmetic mean of the four arithmetic means of the 4 *M*-values in each of the quadruplicate chips, for each gene. (**B**) The same results as in (A) for four replication conditions, but zoomed to a smaller *x*-axis (Δ values ranging from −0.6 to 0.6), to underline the capacity of 12 and 16 replicates to detect small gene expression changes. (**C**) The gray zones around the power values were defined from the power values calculated from the first and third quartiles of all the SD values of the *M*-values in the 6 and 12 replication conditions. The same results as in (A), but zoomed out to a larger *x*-axis (Δ values ranging from −2.5 to 2.5).

power analysis (8) performed using 4, 6, 8, 12 or 16 replicate *M*-values and (ii) an estimation of FP and FN rates at different replication conditions (4, 8 or 12 replicates) as compared to the 16 replications experimentally used.

*Power study*. Power values (1 − β) in each replication condition are plotted against the mean level of differential expression (Δ), which is defined as the arithmetic mean of the four arithmetic means of the 4 *M*-values in each of the quadruplicate chips, for each of the 3804 analyzed genes (Figure 6). Δ represents the most probable (informative) value for the expression ratio for each gene, since it results from the maximum number (in this analysis: 16) of experimental determinations (see Materials and Methods for the definition of the parameters used in the power *t*-test). Figures 6A and 6B show that two-digit replicates (in this analysis: 12 and 16 replicates) allow to detect stable changes in the expression ratios as low as 15% (roughly a variation of 0.20 in *M*) with a probability value lower than 0.05. The methodological sensitivity to detect limited variations in gene expression dramatically decreases when the expression ratios are determined on <6 replicate values. The grayed area between the corresponding power values calculated for the first and the third quartiles (Figure 6C) represent the variations of the SD values of the expression ratios for each gene, deduced from 6 and 12 replicate spots. Only genes with relatively high differential expression levels ($M > \pm 1.5$ at least) show sufficient reproducibility when only six replicates have been used. The present observations on the capacity of replicates to detect limited gene expression changes using DNA chips are in concordance with other studies (6–9). Replicate gene spots, as well as replicate chips, are crucial for reliable monitoring of subtle gene expression changes typical for most biological events. Only large expression changes can be obtained reproducibly from microarray studies performed with chips containing no, or a very limited number of within- and/or between-chip, replicates.

*False positive and false negative rates*. The gain from replications can also be calculated from paired sets of FP and FN rates, determined from differentially expressed gene collections with variable (<16) numbers of mixed within- and between-chip replications. Significantly low FP rates were obtained only with repeated hybridizations (chips) (Figure 7). In parallel, the number of within-chip replicates decreased the FN values. The concomitant use of additional within- and between-chip replicates allowed obtaining balanced values of both FP and FN rates. The simulation of 4 slides with 3 replicates per chip generated a tolerable FP rate of 7% and an FN rate of 14%. This replication pattern, which allowed the evaluation of both technical and biological variations, seems feasible with regards to labor intensity and cost of chips. Both FP and FN rate analysis and power analysis led to coherent conclusions on the importance of replications. This clearly defines the limitations in the use of genome-wide microarrays, which contain many genes but almost no replicates. Any additional experimental variability inherent to other chip

designs—particularly to all 'even designs' including the use of dye-swaps (26)—could be evaluated in the same way.

### Data integration and MADSCAN use

MADSCAN offers multiple processing steps such as filtration, normalization and outlier detection. Raw scan data can be fully analyzed chip by chip or in chip batches. All procedures can be applied independently, i.e. step by step or they can be run in a single and complete procedure, according to the experimental design. Only filtration and normalization procedures are applied to single chips without replicate spots or experiments without replicate chips. All expression data for one experiment are resumed in a consolidated matrix, which allows further comparisons with other data sets in a complete experiment. MADSCAN creates an end file that contains for each gene; its name, the within-slide median of expression ratios in $\log_2(M)$, the within-slide coefficient of variation (CV-M), the between-slide median expression ratio and its CV, and the same data for the geometric mean of the intensities. Intermediate files are accessible for any step. Figure 8A shows the start menu of web-accessible MADSCAN. Figure 8B displays a typical summary of data processing for an example of four replicate chips. A few empirically defined inclusion threshold values for the quality of the chip(s), some statistical parameters on expression ratios and the spatial mode used for normalization are also shown. Detailed information on the quality control parameters is available online (http://www.madtools.org). MADSCAN and its 'A to Z' approach were principally developed to handle replicate chip experiments with a 'reference design'. Other experimental designs, such as 'time series', 'dye-swap' or 'loop'designs, can be analyzed in batches until after the normalization step. The final steps of the current online MADSCAN version, i.e. outlier detection and data consolidation, can easily be performed step by step by reformatting the normalized data files (for details, see tutorial).

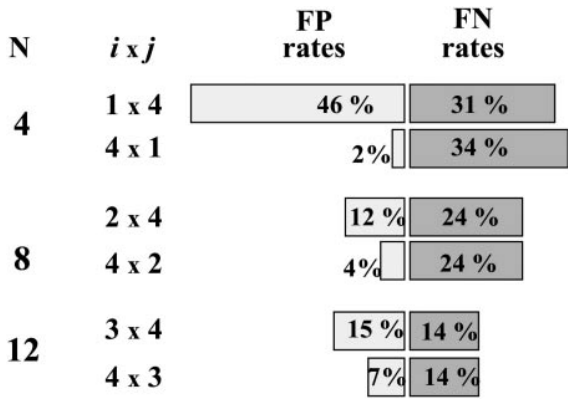
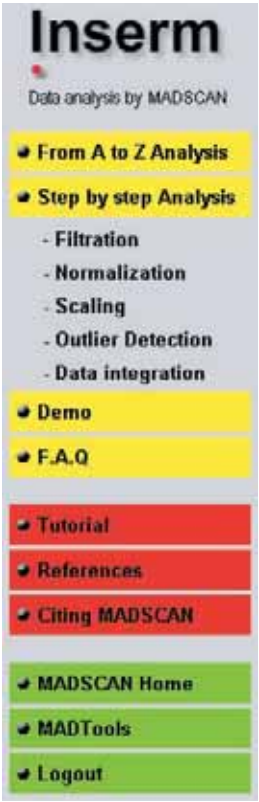



**Figure 7.** Gain from replications. FP and FN rates determined in six simulated replication conditions with regards to the experimental situation based on 16 replicates (*N* replicates obtained by *i* chips with *j* repeated spots).

A




B

| | Slide lmc00035 | Slide lmc00137 | Slide lmc00288 | Slide lmc00378 | Threshold |
|---|---|---|---|---|---|
| **Features:** | | | | | |
| Initial number of features | 16868 | 16868 | 16868 | 16868 | - |
| Validated | 92 % | 92 % | 95 % | 95 % | > 65 |
| Saturated | 0.1 % | 0 % | 0.1 % | 0 % | < 3 |
| Flagged | 8 % | 8 % | 5 % | 5 % | < 35 |
| Statistically validated features | 14357 | 14636 | 15255 | 15228 | - |
| **Ratio:** | | | | | |
| Mean | 0 | 0.01 | 0.02 | 0 | < 0.06 |
| Median | 0 | 0.01 | 0.02 | 0 | < 0.06 |
| Standard Deviation | 0.31 | 0.27 | 0.28 | 0.32 | < 0.50 |
| Median Absolute Deviation | 0.16 | 0.16 | 0.16 | 0.16 | < 0.30 |
| **Normalization mode** | Proximal | Proximal | Proximal | Proximal | Pin |

**Figure 8.** Web-accessible starting menu (**A**) and data summary page (**B**) in the MADSCAN module. For details, see pp. 22 and 27 in the tutorial. Note the definition of threshold values concerning the quality of the chip(s) and related expression measurements.

## CONCLUSION

The challenge of determining thousands of values of gene expression levels in a parallel but unique way using DNA microarrays forces the biologist of today to reliably manage and analyze a deluge of biological data. During the last few years, many alternative algorithms, based on relatively sophisticated and diverse mathematical methods, have been proposed and validated to successfully transform the image scan raw data into consolidated gene expression data. Based on a careful and pragmatic selection among the numerous methods and software available for filtering, normalizing and scaling the raw microarray data, the web-accessible MADSCAN resource presented here offers a dynamic and automatic procedure to obtain a set of reliable gene expression values. The incorporation of methods for within- and between-chip scaling and outlier detection, together with the online access to quality control parameters, complements the proposed plug-in architecture resource in an original way. A careful experimental design—including multiple measurements for each gene under each biological condition—is clearly central to the evaluation of most experimental noises inherently present in high throughput measurements. The significance and quality of any further biological interpretation—gene clustering, coexpression, etc.—are directly dependent on the reliability and significance of the set of consolidated gene expression values derived from image scan values. Obtaining such an initial set of metrologically relevant chip data is the exclusive scope of the MADSCAN procedure.

More or less sophisticated computational tools with various methods for microarray data processing are offered today in many commercially available and/or academic web-accessible software (for a list, see http://ihome.cuhk.edu.hk/~b400559/arraysoft.html or http://genopole.toulouse.inra.fr/bioinfo/microarray/index.php?page=logiciels). Among the available software, the steps corresponding to the initial treatment of raw scan data are either limited to some basic and inadequate transformation algorithms (like a linear normalization based on a few house-keeping genes), or numerous sophisticated, interconnected or independent, algorithmic modules are proposed. In all cases, the biologist has to adjust a series of 'default' parameters, more or less adapted to their own experimental design and the variables measured (27). Some knowledge and even understanding of the details of the algorithms/languages used are necessary to fully appreciate how such changes in the parameters do affect the expression results. To avoid those types of difficulties, we propose MADSCAN. MADSCAN, which has been successfully tested by diverse users on >2000 chips, containing 500 to 24 000 spots, represents an intelligent and powerful tool for the many biologists using DNA chips (12–13,28). The MADSCAN procedure is now plugged into BASE (BioArray Software Environment) (29). Therefore, information on raw image data and their transformation into consolidated expression values is accessible to the entire scientific community, in agreement with the most recent recommendations of the MGED consortium (18).

## SUPPLEMENTARY MATERIAL

Supplementary Material is available at NAR Online.

## ACKNOWLEDGEMENTS

We thank Catherine Chevalier, Isabelle Guisle and Martine Le Cunff for assistance with facilities and experimentation. We are grateful to Geoffroy Golfier, Jean Mosser and Marie-Claude Potier for helpful discussions on methods. We are also grateful to the company Perkin Elmer for their support in the early phase of our study and exchanges concerning the interpretation of microarray data. This work was supported by the 'Institut National de la Santé et de la Recherche Médicale', 'le Centre National à la Recherche Scientifique', 'l'Association Française contre les Myopathies', 'le Conseil Régional des Pays de la Loire' and 'Ouest Génopole'.

## REFERENCES

1. Chipping forecast (2002) *Nature Genet.*, **32** (Suppl.), 461–552.
2. Kerr,M.K., Martin,M. and Churchill,G.A. (2000) Analysis of variance for gene expression microarray data. *J. Comput. Biol.*, **7**, 819–837.
3. Schuchhardt,J., Beule,D., Malik,A., Wolski,E., Eickhoff,H., Lehrach,H. and Herzel,H. (2000) Normalization strategies for cDNA microarrays. *Nucleic Acids Res.*, **28**, E47.
4. Yang,Y.H., Buckley,M.J., Dudoit,S. and Speed,T. (2000) Comparison of methods for image analysis on cDNA microarray data. Technical Report 107, Department of Statistics, University of California, Berkley, CA.
5. Draghiči,S. (2003). *Data Analysis Tools for DNA Microarrays. 1st edn.* Chapman & Hall, Boca Raton, FL.
6. Lee,M.L., Kuo,F.C., Whitmore,G.A. and Sklar,J. (2000) Importance of replication in microarray gene expression studies: statistical methods and evidence from repetitive cDNA hybridizations. *Proc. Natl Acad. Sci. USA*, **97**, 9834–9839.
7. Hwang,D., Schmitt,W.A. and Stephanopoulos,G. (2002) Determination of minimal sample size and discriminatory expression patterns in microarray data. *Bioinformatics*, **18**, 1184–1193.
8. Pan,W., Lin,J. and Le,C.T. (2002) How many replicates of arrays are required to detect gene expression changes in microarray experiments? A mixture model approach. *Genome Biol.*, **3**, research 0022.
9. Pavlidis,P., Li,Q. and NobleW.S. (2003) The effect of replication on gene expression microarray experiments. *Bioinformatics*, **19**, 1620–1627.
10. Tkatchenko,A.V., Le Cam,G., Leger,J.J. and Dechesne,C.A. (2000) Large-scale analysis of differential gene expression in the hindlimb muscles and diaphragm of mdx mouse. *Biochim. Biophys. Acta*, **1500**, 17–30.
11. Cros,N., Tkatchenko,A.V., Pisani,D.F., Leclerc,L., Leger,J.J., Marini,J.F. and Dechesne,C.A. (2001) Analysis of altered gene expression in rat soleus muscle atrophied by disuse. *J. Cell. Biochem.*, **83**, 508–51.
12. Steenman,M., Chen,Y.W., Le Cunff,M., Lamirault,G., Varro,A., Hoffman,E. and Leger,J.J. (2003) Transcriptomal analysis of failing and nonfailing human hearts. *Physiol. Genomics*, **12**, 97–112.
13. Steenman,M., Lamirault,G., Le Meur,N., Le Cunff,M., Escande,D. and Léger,J.J. (2004) Distinct molecular portraits of human failing hearts identified by dedicated cDNA microarrays. *Eur. J. Heart Fail.*, in press, doi:10.1016/j.ejheart.2004.05.008.
14. Ihaka,R. and Gentleman,R. (1996) R: a language for data analysis and graphics. *J. Comput. Graph. Stat.*, **5**, 299–314.
15. Holm,S. (1979) Simple sequentially rejective multiple test procedure. *Scand. J. Stat.*, **6**, 65–70.
16. Tusher,V.G., Tibshirani,R. and Chu,G. (2001) Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl Acad. Sci. USA*, **98**, 5116–5120.
17. Kohavi,R. and Provost,F. (1988) Glossary of terms. *Machine Learning*, **30**, 271–274.
18. Brazma,A., Hingamp,P., Quackenbush,J., Sherlock,G., Spellman,P., Stoeckert,C., Aach,J., Ansorge,W., Ball,C.A., Causton,H.C., Gaasterland,T., Glenisson,P., Holstege,F.C., Kim,I.F., Markowitz,V., Matese,J.C., Parkinson,H., Robinson,A., Sarkans,U., Schulze-Kremer,S., Stewart,J., Taylor,R., Vilo,J. and Vingron,M. (2001) Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nature Genet.*, **29**, 365–371.

19. Tseng,G.C., Oh,M.K., Rohlin,L., Liao,J.C. and Wong,W.H. (2001) Issues in cDNA microarray analysis: quality filtering, channel normalization, models of variations and assessment of gene effects. *Nucleic Acids Res.*, **29**, 2549–2557.
20. Yang,Y.H., Dudoit,S., Luu,P. and Speed,T. (2001) Normalization for cDNA microarray data. *Brief. Bioinformatics*, **2**, 341–349.
21. Yang,Y.H., Dudoit,S., Luu,P., Lin,D.M., Peng,V., Ngai,J. and Speed,T.P. (2002) Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Res.*, **30**, e15.
22. Burke,S. (2001) Missing values, outliers, robust statistics & non-parametric methods. Statistics and data analysis. *Statistics and Data Analysis. LC.GC. Europe Online Supplement*, **59**, 19–24.
23. Müller,J.W. (2000) Possible advantages of a robust evaluation of comparisons. *J. Res. Natl Inst. Stand. Technol.*, **4**, 551–555.
24. Chen,Y., Dougherty,E. and Bittner,M. (1997) Ratio-based decisions and the quantitative analysis of cDNA microarray images. *J. Biomed. Opt.*, **2**, 364–374.
25. Golfier,G., Tran,D.M., Dauphinot,L., Graison,E., Rossier,J. and Potier,M.C. (2004) VARAN: a web server for VARiability ANalysis of DNA microarray experiments. *Bioinformatics*, **20**, 1641–1643.
26. Kerr,M.K. (2003) Experimental design to make the most of microarray studies. *Methods Mol. Biol.*, **224**, 137–147.
27. Bottomley,S. (2004) Bioinformatics: smartest software is still just a tool. *Nature*, **429**, 241.
28. Bédrine-Ferran,H., Le Meur,N., Gicquel,I., Le Cunff,M., Soriano,N., Guisle,I., Mottier,S., Monnier A., Teusan,R., Fergelot,P., Le Gall,J.Y., Leger,J.J. and Mosser,J.. (2004) Transcriptome variations in human Caco-2 cells: a model for enterocyte differentiation and its link to iron absorption. *Genomics*, **83**, 747–950.
29. Saal,L.H., Troein,C., Vallon-Christersson,J., Gruvberger,S., Borg,A. and Peterson,C. (2002) BioArray Software Environment (BASE): a platform for comprehensive management and analysis of microarray data. *Genome Biol.*, **3**, SOFTWARE0003.

## II. ETUDE 2

Steenman M, Chen YW, Le Cunff M, Lamirault G, Varro A, Hoffman E, Leger JJ. Transcriptomal analysis of failing and nonfailing human hearts. Physiol Genomics. 2003 Jan 15;12(2):97-112.

L'objectif de cette étude était de présenter une image globale du transcriptome cardiaque dans des cœur en insuffisance cardiaque et sans dysfonction ventriculaire et de comparer les profils d'expression génique obtenus pour identifier des marqueurs transcriptomaux impliqués dans la fonction cardiaque normale et pathologique.

Les ventricules droits et gauches de 4 patients en insuffisance cardiaque (2 patients atteints de cardiomyopathies dilatées, 2 patients atteints de cardiopathies ischémiques) et d'un patient à fonction cardiaque conservée ont été analysés en utilisant des puces commerciales (Affymetrix®) permettant d'analyser l'expression de 12000 gènes humains.

Les analyses ont identifié 1306 gènes exprimés de manière stable dans toutes les situations étudiées, représentant une partie du profil d'expression génique propre au cœur humain. Parmi les gènes les plus fortement exprimés, des gènes spécifiques de l'activité cardiaque comme ACTC (actine cardiaque), MYL3 (une chaîne légère de la myosine) et PLN (phospholamban) étaient retrouvés. La comparaison de ce profil d'expression cardiaque avec des profils issus d'autres organes pourrait à terme améliorer la compréhension à l'échelle moléculaire du fonctionnement spécifique de chaque tissu.

Notre analyse a identifié 95 gènes différentiellement exprimés entre les cœurs pathologiques et non-pathologiques. Parmi ceux-ci 24 gènes étaient différentiellement exprimés entre les situations insuffisance cardiaque et témoin quelle que soit l'étiologie causale ou la cavité ventriculaire étudiée. Des gènes codant des protéines de signalisation comme les peptides natriurétiques ou impliquées dans la structure et la contraction myocardique comme plusieurs iso-formes de myosine ou l'α-actinine

étaient fortement représentés dans ce sous-groupe. L'analyse a confirmé l'existence d'altérations transcriptomales touchant de nombreux circuits moléculaires et identifié de nouveaux marqueurs de certains de ces circuits.

*Physiol Genomics* 12: 97–112, 2003.
First published November 12, 2002; 10.1152/physiolgenomics.00148.2002.

# Transcriptomal analysis of failing and nonfailing human hearts

M. STEENMAN,[1] Y.-W. CHEN,[2] M. LE CUNFF,[1] G. LAMIRAULT,[1]
A. VARRÓ,[3] E. HOFFMAN,[2] AND J. J. LÉGER[1]
[1]*Institut National de la Santé et de la Recherche Médicale U533, 44035 Nantes, France;* [2]*Children's National Medical Center, Washington, District of Columbia 20010; and* [3]*Department of Pharmacology and Pharmacotherapy, University of Szeged, H-6701 Szeged, Hungary*

Submitted 4 November 2002; accepted in final form 11 November 2002

**Steenman, M., Y.-W. Chen, M. Le Cunff, G. Lamirault, A. Varró, E. Hoffman, and J. J. Léger.** Transcriptomal analysis of failing and nonfailing human hearts. *Physiol Genomics* 12: 97–112, 2003. First published November 12, 2002; 10.1152/physiolgenomics.00148.2002.—Heart failure is a multifactorial disease that may result from different initiating events. To contribute to an improved comprehension of normal cardiac function and the molecular events leading to heart failure, we performed large-scale gene expression analysis of failing and nonfailing human ventricle. Our aim was to define and compare expression profiles of 4 specific pathophysiological cardiac situations: *1*) left ventricle (LV) from nonfailing heart; *2*) LV from failing hearts affected by dilated cardiomyopathy (DCM); *3*) LV from failing hearts affected by ischemic CM (ICM); *4*) right ventricle (RV) from failing hearts affected by DCM or ICM. We used oligonucleotide arrays representing ~12,000 human genes. After stringent numerical analyses using several statistical tests, we identified 1,306 genes with a similar expression profile in all 4 cardiac situations, therefore representative of part of the human cardiac expression profile. A total of 95 genes displayed differential expression between failing and nonfailing heart samples, reflecting a reversal to developmental gene expression, dedifferentiation of failing cardiomyocytes, and involvement of apoptosis. Twenty genes were differentially expressed between failing LV and failing RV, identifying possible candidates for different functioning of both ventricles. Finally, no genes were found to be significantly differentially expressed between failing DCM and failing ICM LV, emphasizing that transcriptomal analysis of explanted hearts results mainly in identification of expression profiles of end-stage heart failure and less in determination of expression profiles of the underlying etiology. Taken together, our data resulted in identification of putative transcriptomal landmarks for normal and disturbed cardiac function.

dilated cardiomyopathy; ischemic cardiomyopathy; differential gene expression; molecular circuits

THE EVOLUTION OF METHODS that enable large-scale expression analyses, like serial analysis of gene expression (SAGE), in silico analysis of expressed sequence tag (EST) databases, cDNA microarrays, and oligonucleotide microarrays, allows researchers to establish organ- or pathology-specific transcriptional profiles. This may lead to the formation of a custom collection of genes, applicable to a specific pathology. These custom collections will allow high-throughput expression analyses of large quantities of samples. Therefore, besides a better understanding of pathological processes, these studies may eventually open the way to improved diagnosis and new treatment strategies for certain human diseases. Most of these methods have been applied to studies of human cancer. Cardiovascular diseases, which also represent a major cause of morbidity and mortality in industrialized countries, have been studied only to a limited extent by these methods. The main reason for this discrepancy is that it is relatively easy to obtain tumor material, whereas human cardiac tissue is, evidently, much harder to come by.

Heart failure is a complex clinical syndrome, which is defined as the development and progression of left ventricle (LV) remodeling (16). It constitutes the endpoint of many different cardiac diseases and is therefore considered to be multifactorial. Two cardiac diseases that may eventually lead to heart failure are dilated (DCM) and ischemic (ICM) cardiomyopathy. Although, or, possibly, resulting from the fact that, clinical treatment for cardiac diseases has improved over the last few decades, heart failure morbidity and mortality have increased (14). To obtain a better understanding of the factors involved in the development of heart failure, we need to improve our knowledge of the disturbed molecular pathways in cardiac pathology. The recently developed techniques, which enable the study of the cardiac transcriptome, may provide us with a tool toward this goal. Expression analyses will allow the identification of a molecular portrait of the human heart, which will indicate the transcriptional requirements for cardiac function. Thus far, studies toward this goal have been based solely on analyses of EST libraries (5, 21). Recently, a report describing the application of SAGE for global expression analysis of mouse heart has been published (2). In addition, large-scale gene expression analyses have been used to detect aberrant expression in cardiac pathologies. The effect of ICM on gene expression levels has been stud-

Article published online before print. See web site for date of publication (http://physiolgenomics.physiology.org).
Address for reprint requests and other correspondence: M. Steenman, INSERM U533, UFR de Médecine, 1 rue Gaston Veil, BP 53508, 44035 Nantes Cedex 1, France (email: marja.steenman@nantes.inserm.fr).

Table 1. *Characteristics of patients analyzed by microarray and real-time PCR analysis*

| Total RNA Pool: | Failing Hearts | | | | | | | | | | Nonfailing Heart |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | DCM LV | | ICM LV | | DCM/ICM RV | | | | | | (NF LV) |
| Patient: | *D1* | *D2* | *I1* | *I2* | *I2* | *I3* | *I4* | *I5* | *D3* | *D4* | NF |
| Age, yr | 45 | 58 | 65 | 66 | 66 | 63 | 56 | 50 | 56 | 35 | 15 |
| Sex | male | male | male | male | male | male | male | male | male | male | male |
| Pathology | DCM | DCM | ICM | ICM | ICM | ICM | ICM | ICM | DCM | DCM | CF |
| Tissue | LV | LV | LV | LV | RV | RV | RV | RV | RV | RV | LV |
| Ejection fraction, % | 23 | 13 | 23 | 21 | 21 | 22 | 24 | 18 | 21 | 20 | NA |
| Treatment | CP, DG, FR | AM, CP, FR | AM, CP, FR | AM, CP, DB, NC, FR | AM, CP, DB, NC, FR | AC, AM | CR, DG, FR, LS, ML | AM, CP, FR, SP | CP, DG, DP, FR | AM, DB, EN, EP, FR, HCT. | none |

DCM, dilated cardiomyopathy; ICM, ischemic cardiomyopathy; LV and RV, left and right ventricle, respectively; NF, nonfailing; NA, not available; AC, acebutolol; AM, amiodarone; CP, captopril; CR, carvedilol; DB, dobutamine; DG, digoxin; DP, dopamine; EN, enoximone; EP, epinephrine; FR, furosemide; HCT, hydrochlorothiazide; LS, losartan; ML, molsidomine; NC, nicorandil; SP, spironolactone; CF, cystic fibrosis.

ied in mouse and rat models using cDNA microarrays containing up to 4,000 genes (29, 38, 40). In humans, differential gene expression associated with heart failure has been studied both by using oligonucleotide microarrays (containing ~7,000 genes) (42, 47), and by using cDNA microarrays (containing up to 10,000 genes) (3, 4, 22). The first oligonucleotide microarray study analyzed tissue from a DCM-affected heart and an ICM-affected heart and identified those genes that showed differential expression in both cardiac pathologies compared with nonfailing hearts. The second oligonucleotide microarray study looked for differences between DCM-affected and nonfailing hearts. The studies using cDNA microarrays analyzed differential gene expression in hypertrophic or dilated hearts vs. nonfailing hearts.

Our goal was to comparatively analyze gene expression in failing and nonfailing human hearts using oligonucleotide microarrays (Affymetrix) containing ~12,000 human genes, representative of a substantial proportion of the human transcriptome. Expression profiles of well-defined cardiac tissues (failing/nonfailing, DCM/ICM, LV/RV) were compared to identify genes expressed at a similar level in all cardiac tissues analyzed and to identify genes differentially expressed between the same tissues. In this way we aimed to identify transcriptomal landmarks for normal and disturbed cardiac function.

## MATERIALS AND METHODS

*Human cardiac tissue.* Left (LV) and right (RV) ventricle tissue was obtained from ICM- and DCM-affected explanted hearts from patients diagnosed with end-stage heart failure, who underwent a heart transplantation. Hearts that were used to obtain RV tissue were selected based on the following criteria: no RV dilation, normal RV ejection fraction, and normal pulmonary blood pressure. Nonfailing LV tissue was obtained from one cystic fibrosis (CF) patient (*patient NF LV*) who underwent a heart-lung transplantation. This patient did not show any sign of heart failure. Chest X-ray, ECG, and echocardiography revealed no cardiac abnormalities. Since both pathological and control hearts were obtained from patients who underwent a heart transplantation, the clinical settings (which may influence gene expression) were comparable for both situations. To determine whether *patient NF LV* represented a valid control for our study, 11 additional NF heart samples were obtained. These samples included LV from two CF patients who underwent a heart-lung transplantation (*patients NF01* and *NF05*), two commercial RNA samples from human adult heart obtained from Stratagene (*NF03* and *NF04*), and seven LV samples from general organ donors (*NF02*, *NF06*, *NF07*, *NF08*, *NF09*, *NF10*, and *NF11*) undergoing pulmonary and aortic valve transplantation surgery. Before explantation of the hearts, the patients did not receive any medication except for dobutamine, furosemide, and plasma expanders. The experimental protocol complied with the Declaration of the World Medical Association proclaimed in Helsinki and was approved by the Ethical Review Board of the Medical Faculty of the University of Szeged (No. 51–57/1997 OEj). See Tables 1 and 2 for patient characteristics.

*Expression profiling.* Total RNA was isolated separately from each cardiac tissue using TRIzol Reagent (Life Technologies). RNA was DNase treated, and quality was assessed by migration on a 1% agarose gel and RT-PCR using primers for β-actin. Absence of DNA contamination was verified by PCR

Table 2. *Characteristics of patients analyzed by real-time PCR only*

| Patient | Age, yr | Sex | Pathology | Cause of Death |
|---|---|---|---|---|
| *NF01* | 20 | male | CF | |
| *NF02* | 53 | female | | CVA |
| *NF03* | 72 | male | | unknown |
| *NF04* | 63 | male | | unknown |
| *NF05* | 32 | male | CF | |
| *NF06* | 57 | male | | CVA |
| *NF07* | 18 | female | | accident |
| *NF08* | 44 | male | | accident |
| *NF09* | 51 | male | | CVA |
| *NF10* | 42 | male | | CVA |
| *NF11* | 43 | male | | CVA |

*NF03* and *NF04* represent commercial human adult heart RNA samples obtained from Stratagene. For these samples no information is available about cause of death or type of tissue used for RNA extraction. All other NF samples were obtained from LV; CF, cystic fibrosis; CVA, cerebrovascular accident.

using the same primers. Three pooled samples were prepared: DCM LV containing equal amounts of total RNA from LV from *patients D1* and *D2*, ICM LV containing equal amounts of total RNA from LV from *patients I1* and *I2*, and DCM/ICM RV containing equal amounts of total RNA from RV from *patients D3*, *D4*, *I2*, *I3*, *I4*, and *I5*. Nonfailing LV was obtained from *patient NF LV* (Table 1). Target preparation, hybridization to the Affymetrix HG-U95A array, and detection was performed as previously described (11). All hybridizations were performed in duplicate on the four samples described above, by dividing the prepared target over two arrays.

*Data analysis: genes similarly expressed in the four cardiac samples.* Analysis of Affymetrix microarrays was performed using the Affymetrix GeneChip software (version 3.3) as described previously (27). Figure 1 displays the level of intra-experimental reproducibility for the four pairs of duplicate experiments. The determination of the transcriptional profile common for the 4 cardiac samples was performed as follows: the average difference (Avg Diff) values on all 8 arrays were normalized by setting the mean Avg Diff of each array at 5,000. All genes (probe sets) that had received an "absent" call for all 8 arrays and all genes with an average Avg Diff value of <500 (for the 8 hybridizations) were not taken into consideration for the transcriptional profile. The 2,031 remaining genes were submitted to 4 statistical tests of increasing stringency using SAS Software (Bonferroni, Newman-Keuls, Ryan-Einot-Gabriel-Welsch, least significant difference, $\alpha = 0.05$) to filter out those genes with a similar expression level in all tissues analyzed, i.e., genes with a similar average Avg Diff value (for the 2 duplicate hybridizations) for the 4 tissue samples. Only those genes that passed at least three of the statistical tests were included in the human cardiac transcriptional profile, since their expression level did not seem to be dependent on the (pathological) origin of the tissue.

*Data analysis: genes differentially expressed between the four cardiac samples.* The identification of genes differentially expressed in the four cardiac samples (DCM LV, ICM LV, DCM/ICM RV, and NF LV) was performed in a fashion similar to that described previously (11). Global scaling was performed to a constant value of 800, as recommended by the manufacturer. The following comparisons of data sets were performed, using the GeneChip software: DCM LV vs. NF LV, ICM LV vs. NF LV, DCM/ICM RV vs. NF LV, DCM LV vs. DCM/ICM RV, ICM LV vs. DCM/ICM RV, and DCM LV vs. ICM LV. Since all hybridizations had been performed in duplicate, we obtained four data sets four each of the six comparisons. Statistically significant differential genes were selected using the following procedure. Genes that had received an "absent" call for all eight chips were eliminated from further analysis. For the remaining genes ($n = 6{,}585$) we transformed the "fold change" values into $\log_2$(ratio) values (4 values for each comparison). On these values we performed significance analysis of microarrays (SAM) (44). Using the "one class" option and setting the false discovery rate at 3%, we tested whether the mean of the four expression ratios within one comparison differed from zero. The genes that where then labeled as "significant" were further analyzed using Affymetrix criteria. We retained those genes that displayed a (marginal) increase or decrease and a fold change >2.0 or less than −2 [or $\log_2$(ratio) >1.0 or less than −1.0] in all four data sets from one comparison. In addition, we retained those genes with a fold change >2.0 or less than −2 when in at least in one of the neighboring comparisons the "difference call" was consistently (marginal) increase or

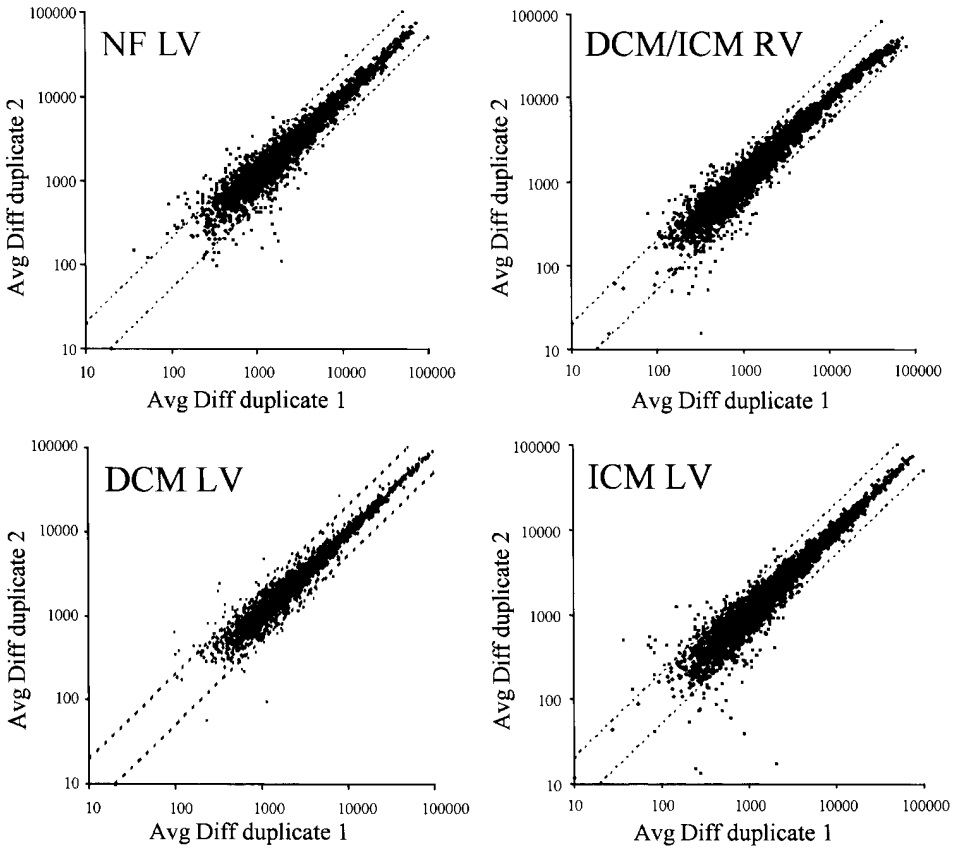



Fig. 1. Intra-experimental reproducibility of normalized signals of hybridization to the 12,626 probe sets on the Affymetrix chips in the duplicate experiments. All points that are located between the dotted lines represent probe sets with an average difference (Avg Diff) ratio between the duplicate experiments of >0.5 and <2.0. DCM, dilated cardiomyopathy; ICM, ischemic cardiomyopathy; LV and RV, left and right ventricle, respectively; NF, nonfailing.

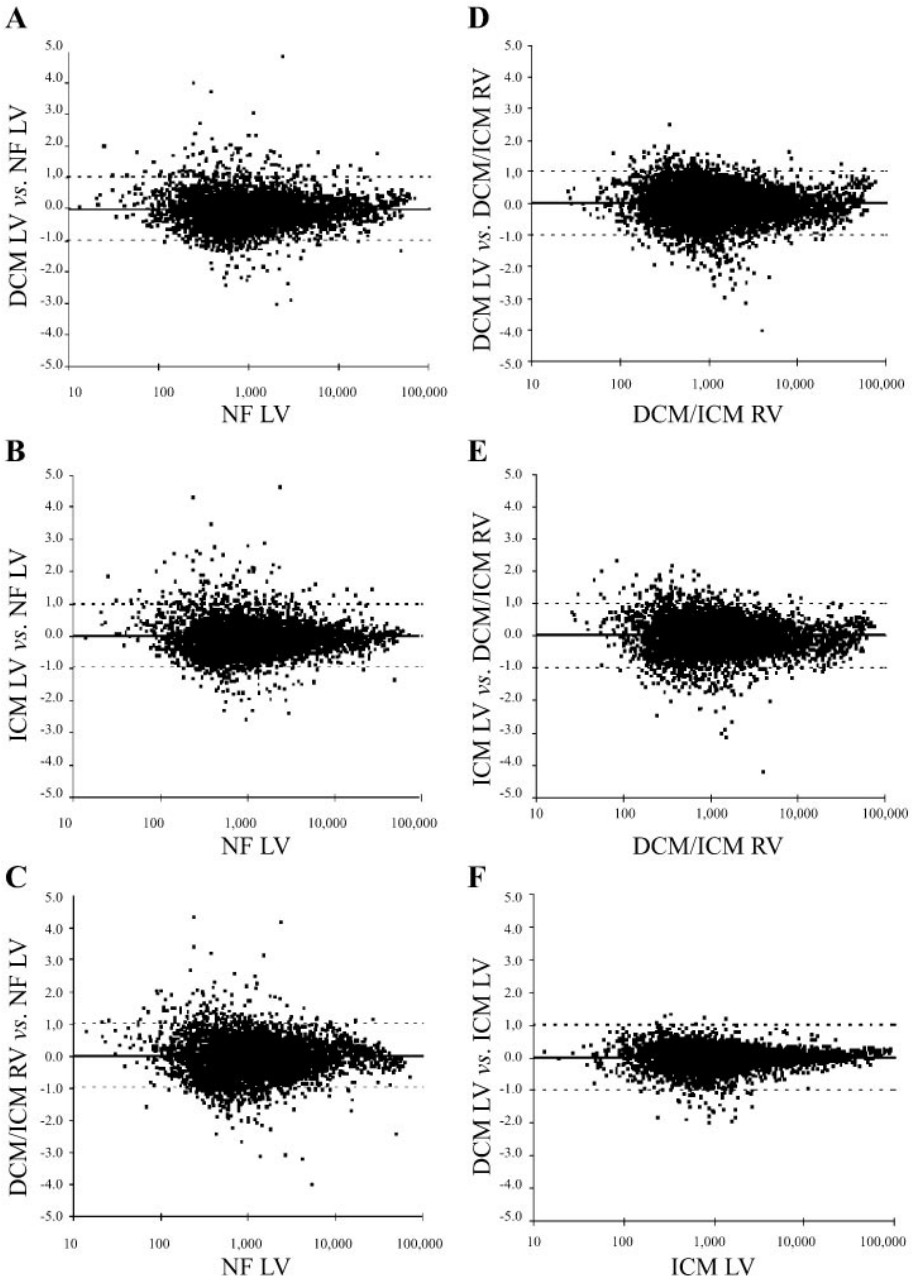



Fig. 2. Representation of expression ratios of the 6,585 genes that were retained for differential analysis based on the fact that they did not receive 8 "absent calls" (see MATERIALS AND METHODS). The expression ratios (in $\log_2$) are plotted against the normalized expression level (mean Avg Diff) of either NF LV (*A–C*), or DCM/ICM RV (*D* and *E*), or ICM LV (*F*).

decrease for the four data sets. Neighboring comparisons are those in which the denominator is identical. Genes with a ratio between the duplicate normalized Avg Diff values $<0.5$ or $>2.0$ were not considered for differential analysis.

*Real-time PCR.* Ten genes (11 transcripts) were selected for quantitative analysis by real-time PCR using SYBR Green I dye (PE Biosystems). Genes and primers were: cytokine inducible nuclear protein (forward CTGTTTATTTATTTACTGCCACGCT, reverse CATTGTTTCCTTTCCTGTCCCT), pyruvate dehydrogenase kinase isoform 4 (forward CTCCAAAACAAACGACAGCAA, reverse ACTCACTCCCTTTCTTATTCTTATCAA), complement protein component C7 (forward ATGCTCGTCTCCAACTCCTGA, reverse TGACACATACTAAAACCCCAAAGAG), δ sleep inducing peptide (forward CCTTGTTCATCCCTCATCCAC, reverse ATGCCAAACCCACCATTCA), ANF (forward CATTTGTGTCATCTTGTTGCCA, reverse GAGGCGAGGAAGTCACCATC), BNP (forward ACCGCAAAATGGTCCTCTACA, reverse TGTGGAATCAGAAGCAGGTGTC), AF1q (forward CCTACTCACTTTACAACTTTGCTCC, reverse TAGCTGAAGAACTGTGCCCACT), SH3BGR (forward GTGAAGACCGTTTATGCATAC, reverse CACAGAGAACTTCCATGCCTTG), FLRG (forward AGACCCAGACTCCAGCCAGAC, reverse TGTGGTTTACACGCAGGCAC), SLIMMER (forward GTCAAGAGTGAGCCACCCAGT, reverse ACAAGTCCTCTCTCCACCCG), SLIM1/SLIMMER (forward GACAATCCTGGCACGACTACTG, reverse GGCACAGTCGGGACAATACAC). hypoxanthine phosphoribosyltransferase (HPRT) was taken as internal standard. Two types of real-time PCR analyses were performed. First, we wanted to confirm the results obtained by microarray analysis. For this we used the same cardiac samples that had been analyzed by the arrays: DCM LV, ICM LV,

DCM/ICM RV, and NF LV. Gene expression ratios were calculated for DCM LV vs. NF LV, ICM LV vs. NF LV, and DCM/ICM RV vs. NF LV. The second type of real-time PCR analysis was performed to evaluate our choice of control tissue in the microarray experiments. We determined expression ratios of the above mentioned genes in 11 additional nonfailing hearts vs. NF LV. All PCR reactions were performed in two experiments, each containing duplicate reactions ($n$ = 4 for each expression value).

## RESULTS

*Human heart sampling and global view.* Our goal was to compare gene expression profiles in human failing and nonfailing LV and RV, using Affymetrix HG-U95A arrays containing 12,626 probe sets. In duplicate experiments we determined expression profiles of LV from one CF patient (*patient NF LV*), of a pooled sample of LV from two DCM-affected failing hearts (DCM LV), of a pooled sample of LV from two ICM-affected failing hearts (ICM LV), and of a pooled sample of RV of two DCM-affected failing hearts and four ICM-affected failing hearts (DCM/ICM RV) (see Table 1 for details). Initial scaling and filtering of the 8 sets of rough expression values corresponding to the 4 specific cardiac physiopathological situations resulted in a list of 6,585 valid expression values. Pairwise comparisons between the different valid expression profiles were performed as follows: *1*) failing heart vs. nonfailing heart, leading to the identification of genes potentially involved in the development of heart failure (Fig. 2, *A–C*); *2*) failing LV vs. failing RV, leading to the identification of genes implicated in the different functional properties of both ventricles or showing a different expression profile in severely (LV) and less severely (RV) affected tissue (Fig. 2, *D* and *E*); *3*) DCM LV vs. ICM LV to look for transcriptomal differences between both pathologies (Fig. 2*F*). It is clear that most differences are found between failing (DCM LV, ICM LV, DCM/ICM RV) and nonfailing (NF LV) hearts (Fig. 2, *A–C*) and least differences are found between DCM LV and ICM LV (Fig. 2*F*). This initial global presentation of gene expression variations between the four human heart samples was followed by a more stringent analysis to identify genes similarly expressed in all samples and genes differentially expressed between the samples. For this purpose we applied robust statistical analyses, Affymetrix criteria and significance analysis methods.

*Genes similarly expressed in the four cardiac samples.* We identified 1,306 probe sets (corresponding to 1,258 unique genes) that displayed similar (i.e., statistically non-different) Avg Diff values in all 4 samples analyzed. The similarity of the Avg Diff values of these probe sets indicated that the expression levels of the corresponding genes were not significantly influenced by pathology, age, or left vs. right ventricular origin. (Complete data are available on **http://www.ifr26.nantes.inserm.fr/composantes/u533/people/team1b/publi/affylogin.htm**) Since the redundant genes were represented on the array by different probe sets, which, in some cases, performed differently, we did not exclude this redundancy (e.g., acid ceramidase, **http://www.ifr26.nantes.inserm.fr/composantes/u533/people/team1b/publi/affylogin.htm**). This explains why some accession numbers appear twice in the profile on **http://www.ifr26.nantes.inserm.fr/composantes/u533/people/team1b/publi/affylogin.htm**. The distribution of the level of expression of these 1,306 genes, which varied within 3 orders of magnitude, is shown in Fig. 3. The vast majority of the genes were expressed at a low level, with only a few genes accounting for the major part of the total amount of RNA in ventricular cardiomyocytes. We divided our population of 1,306 cardiac expressed genes into three groups: *1*) highly expressed genes, whose expression value (Avg Diff) exceeded 25,000. This group contained 40 genes (3.1%) (Table 3); *2*) moderately expressed genes, with an expression value between 5,000 and 25,000, accounting for 213 genes (16.3%); *3*) weakly expressed genes, with an expression value below 5,000, accounting for 1,053 genes (80.6%). All genes belonging to these three groups are listed on **http://www.ifr26.nantes.inserm.fr/composantes/u533/people/team1b/publi/affylogin.htm**. Based on the literature (1, 21) and on gene function descriptions on **http://bioinformatics.weizmann.ac.il/cards/** (GeneCards), we assigned the 1,306 expressed genes to 6 main functional categories; gene/protein expression, metabolism, cell structure/motility, cell signaling/communication, cell/organism defense, and cell division. Of almost half of the genes the function description was either absent or to vague for them to be assigned to one of the functional categories. More than one-third of the classified genes are involved in gene and protein expression. The next largest functional category was that of metabolism, containing one-fifth of the genes, followed by cell signaling and

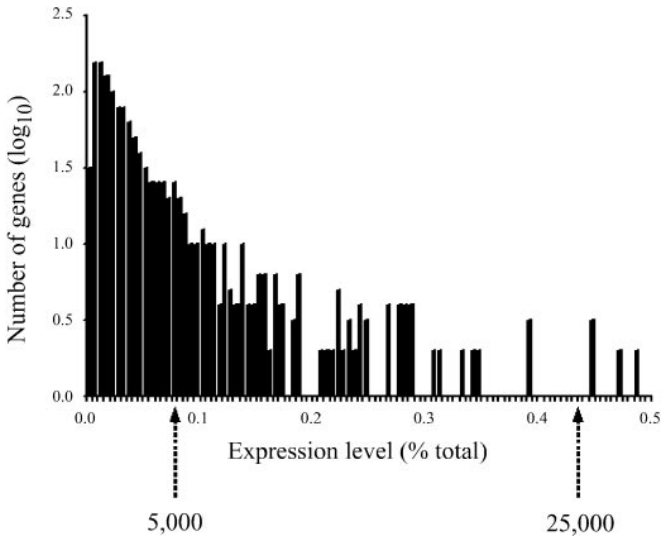



Fig. 3. Distribution of the level of expression of 1,306 genes selected on the basis of similar expression levels in all cardiac samples analyzed. Level of expression is represented as a percentage of the cumulative expression of the 1,306 genes. The Avg Diff values representing the borders between highly, moderately, and weakly expressed genes are indicated by arrows.

Table 3. *Genes highly expressed in human heart*

| Identifier | Description | Expression, % | Functional Category |
|---|---|---|---|
| F27891 | cytochrome *c* oxidase subunit VIa polypeptide 2 | 0.83 | M |
| AA152406 | cytochrome *c* oxidase subunit VIIa polypeptide 1 | 0.81 | M |
| D55654 | cytosolic malate dehydrogenase | 0.72 | M |
| M33197 | glyceraldehyde-3-phosphate dehydrogenase | 0.71 | M |
| AA426364 | ATP synthase, $H^+$ transporting, mitochondrial F0 complex, subunit e | 0.60 | M |
| N47307 | NADH dehydrogenase (ubiquinone) 1α subcomplex, 1 | 0.59 | M |
| U34995 | glyceraldehyde-3-phosphate dehydrogenase | 0.58 | M |
| A1708889 | cytochrome *c* oxidase subunit VIIc | 0.53 | M |
| X05236 | aldolase A | 0.50 | M |
| T79616 | ubiquinol-cytochrome *c* reductase binding protein | 0.49 | M |
| D78361 | omithine decarboxylase antizyme | 0.48 | M |
| T57872 | cytochrome *c* oxidase subunit VIb | 0.47 | M |
| A1207842 | prostaglandin D2 synthase | 0.82 | M-CS/C |
| M98539 | prostaglandin D2 synthase | 0.62 | M-CS/C |
| M26880 | ubiquitin | 0.91 | G/PE |
| L06499 | ribosomal protein L37a | 0.79 | G/PE |
| L06498 | ribosomal protein S20 | 0.71 | G/PE |
| HG1800-HT1823 | ribosomal protein S20 | 0.58 | G/PE |
| M64241 | ribosomal protein L10 | 0.57 | G/PE |
| M17886 | acidic ribosomal phosphoprotein P1 | 0.55 | G/PE |
| A1557852 | ribosomal protein S27 | 0.52 | G/PE |
| Z12962 | homologue to yeast ribosomal protein L41 | 0.51 | G/PE |
| M13932 | ribosomal protein S17 | 0.50 | G/PE |
| X56932 | ribosomal protein L13A | 0.49 | G/PE |
| X70940 | elongation factor 1 α-2 | 0.48 | G/PE |
| U49837 | LIM protein MLP | 0.45 | G/PE |
| J00073 | α-cardiac actin | 0.94 | CS/M |
| M24122 | ventricular myosin alkali light chain | 0.78 | CS/M |
| X69090 | myomesin 1 (skelemin) | 0.73 | CS/M |
| X694940 | titin | 0.66 | CS/M |
| M58018 | β-myosin heavy chain | 0.66 | CS/M |
| AJ010063 | telethonin | 0.57 | CS/M |
| J02854 | myosin light chain 2 | 0.56 | CS/M |
| Z23090 | 28-kDa heat shock protein | 0.68 | C/OD |
| J05401 | sarcomeric mitochondrial creatine kinase | 0.51 | C/OD |
| M63603 | phospholamban | 0.45 | CS/C |
| U14573 | Alu-Sq subfamily consensus sequence | 0.76 | U |
| A1540957 | ubiquinone binding protein | 0.67 | U |
| U70063 | acid ceramidase | 0.49 | U |
| X16064 | translationally controlled tumor protein | 0.46 | U |

Expression is given as a percentage of the cumulative expression of the 1,306 genes present in the transcriptional profile. U, unclassified; CS/M, cell structure and motility; M, metabolism; G/PE, gene and protein expression; C/OD, cell and organism defense; and CS/C, cell signaling and communication.

communication and cell structure and motility. Only a few genes are involved in cell division, and these concern only those weakly or moderately expressed. The highly expressed genes (Table 3) contained several typical cardiac genes, like α-cardiac actin, ventricular myosin alkali light chain, and phospholamban. Most of the ribosomal proteins were moderately expressed, whereas transcription factors were mostly weakly expressed. Figure 4 shows the distribution of the genes over the functional categories in the three groups of genes at different level of expression. This distribution differed between the highly and the weakly expressed genes. Genes involved in metabolism were more strongly represented among the highly expressed genes (Table 3). Furthermore, this group contained relatively more genes involved in cell structure and motility and fewer genes involved in cell signaling and communication.

*Genes differentially expressed between the four cardiac samples.* The sequential application of SAM analysis and Affymetrix criteria to the expression ratios obtained from the comparisons among the four specific cardiac pathophysiological situations led to a stringent, statistically valid selection of subsets of differentially expressed genes. We identified 95 genes differentially expressed in failing vs. nonfailing heart (Tables 4–6), 20 genes differentially expressed between failing LV and failing RV (Table 7), and 0 genes differentially expressed between DCM LV and ICM LV. The subsets of differentially expressed genes were composed as follows.

*Differential genes between failing heart and nonfailing heart.* In Tables 4–6 we grouped together genes showing differential expression in either all three (*group A*, Table 4), or in two of three (*group B*, Table 5), or in only one (*group C*, Table 6) of the failing heart samples compared with NF LV. Some genes that did not show consistent differential expression in all four comparisons according to the Affymetrix criteria, but that were significantly differential according to SAM and had a $\log_2$(ratio) >1.0 or less than −1.0, were

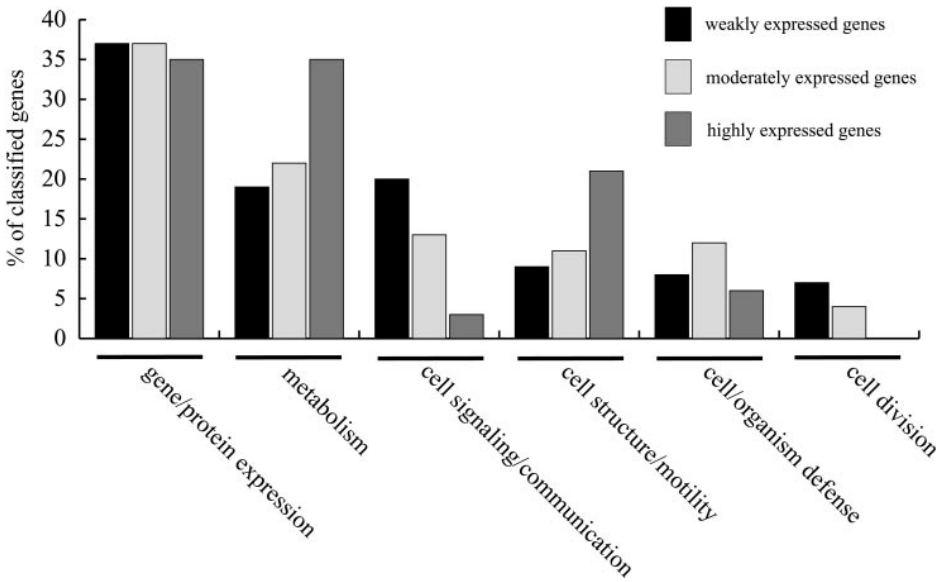



Fig. 4. Distribution of the 6 main functional categories among the weakly, the moderately, and the highly expressed genes.

considered as being differentially expressed. For example, cyclin-dependent kinase inhibitor 1A (CDKN1A), showed an average $\log_2$(ratio) of 4.49, 4.46, and 3.82 in DCM-affected LV, ICM-affected LV, and DCM/ICM-affected RV, respectively. Even though CDKN1A did not show consistent increased expression in DCM-affected LV according to the Affymetrix criteria, we did classify this gene as belonging to *group A*. This approach was only used for those genes that did pass all criteria in at least one of the samples tested. In addition to a classification according to the main functional categories, we attempted to assign the differential genes to the different molecular circuits known to be affected in heart failure. We identified 24 genes (*group A*) showing increased expression in all failing heart samples. In this group we identified eight genes involved in signaling pathways. Among those were the natriuretic factors ANF and BNP, which are involved in vascular homeostasis. Seven genes in this group are involved in biomechanical functions. These genes encode proteins that are components of the sarcomere (e.g., embryonic myosin alkaline light chain), the cytoskeleton (e.g., α-actinin), or the extracellular matrix (e.g., versican). Furthermore, several genes implicated in energy metabolism, oxidative stress, and inflammation showed increased expression in failing LV and RV. In addition, we identified a gene with unknown function in this group (AF1q). *Group A* did not contain any downregulated genes, since they failed to pass the SAM analysis for all three failing heart samples.

The next group of differentially expressed genes, belonging to *group B*, concerns those whose expression changed in the LV of DCM- and ICM-affected hearts but not in the RV. Therefore, the expression of these genes changed only in the part of the heart that is affected the most by the disease. This group, consisting of 19 upregulated genes, also contained a substantial amount of genes (a total of 6) involved in biomechanical functions. Most interesting were those genes whose $\log_2$(ratio) was close to 0 in the failing RV sample and >1.0 in the failing LV samples [e.g., transglutaminase (TGase) and α-tubulin isotype H2-α]. It should be noted that for some of the genes the $\log_2$(ratio) value is not listed because the ratio between the Avg Diff values of the two duplicate hybridizations were outside of the range 0.5–2.0, and results were therefore considered to be unreliable [Tables 5–7; not reliable (NR)]. The classification of those genes in this group should therefore be regarded as ambiguous. Only 5 genes were upregulated in DCM-affected LV and DCM/ICM-affected RV only, whereas 20 genes were differentially expressed in ICM-affected LV and DCM/ICM-affected RV only. A logical explanation for genes assigned to this last group could be that they are differentially expressed in ICM-affected hearts in general, and since the DCM/ICM-affected RV pool consisted mainly of tissue from ICM-affected hearts the same genes would be found to be differentially expressed. However, several genes were upregulated to a higher extent in the RV sample than in the ICM-affected LV sample (e.g., nidogen and TP53BP2) and might therefore represent RV-related events.

We found 27 genes to be differentially expressed in only 1 of the 3 failing heart samples analyzed (*group C*). Two of these were found only in DCM-affected LV (heat shock 70-kDa protein 5 and cytokine inducible nuclear protein). Seventeen genes were deregulated only in ICM LV, with 13 genes up- and 4 genes downregulated. Three of the upregulated genes are involved in signaling pathways: Ras-like protein Tc21, cAMP-responsive element modulator (CREM), and A kinase anchor protein 2 (AKAP2). One of the genes displaying a clear ICM-specific upregulation was pyruvate dehydrogenase kinase isoform 4, a gene involved in glucose metabolism. Of the eight genes differentially expressed specifically in failing RV, two genes were downregulated and six genes were upregulated. The two downregulated genes could be reduced to one nonredundant

Table 4. *Genes differentially expressed in failing hearts vs. nonfailing heart: Group A*

| Identifier | Description | DCM LV vs. NF LV | | | ICM LV vs. NF LV | | | DCM/ICM RV vs. NF LV | | | Functional Category | Molecular Circuit |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Ratio | SAM | Affy | Ratio | SAM | Affy | Ratio | SAM | Affy | | |
| AL021155 | brain natriuretic protein | 7.24 ↑ | + | + | 7.80 ↑ | + | + | 7.41 ↑ | + | + | CS/C | Signaling |
| AL021155 | atrial natriuretic factor | 4.83 ↑ | + | + | 4.59 ↑ | + | + | 4.12 ↑ | + | + | CS/C | Signaling |
| X51345 | JUN-B | 3.40 ↑ | + | + | 3.40 ↑ | + | + | 2.40 ↑ | + | | G/PE | Signaling |
| S37730 | insulin-like growth factor binding protein-2 | 2.87 ↑ | + | | 3.20 ↑ | + | + | 2.27 ↑ | + | | CS/C | Signaling |
| Z24724 | polyA site DNA | 2.06 ↑ | + | + | 2.04 ↑ | + | + | 3.39 ↑ | + | + | G/PE | Signaling |
| AF060568 | promyelocytic leukemia zinc finger protein | 2.33 ↑ | + | + | 2.07 ↑ | + | + | 1.22 ↑ | + | | G/PE | Signaling |
| X78947 | connective tissue growth factor | 1.65 ↑ | + | + | 2.10 ↑ | + | + | 1.75 ↑ | + | + | CS/C | Signaling |
| Z37976 | latent TGF-β binding protein | 1.40 ↑ | + | | 2.53 ↑ | + | + | 1.31 ↑ | + | | CS/C | Signaling |
| X13839 | vascular smooth muscle α-actin | 1.76 ↑ | + | | 1.94 ↑ | + | + | 1.19 ↑ | + | | CS/M | Biomechanics-Cytosk |
| X15804 | α-actinin | 1.53 ↑ | + | | 1.63 ↑ | + | + | 1.63 ↑ | + | + | CS/M | Biomechanics-Cytosk |
| M95178 | non-muscle α-actinin | 1.46 ↑ | + | | 1.89 ↑ | + | + | 1.32 ↑ | + | | CS/M | Biomechanics-Cytosk |
| D17408 | calponin | 2.77 ↑ | + | | 3.95 ↑ | + | + | 3.41 ↑ | + | + | CS/M | Biomechanics-Sarcomere |
| X58851 | embryonic myosin alkaline light chain | 2.79 ↑ | + | + | 3.36 ↑ | + | + | 3.14 ↑ | + | + | CS/M | Biomechanics-Sarcomere |
| AF013570 | smooth muscle myosin heavy chain SM2 | 2.32 ↑ | + | | 2.79 ↑ | + | + | 1.68 ↑ | + | | CS/M | Biomechanics-Sarcomere |
| AF001548 | myosin, heavy polypeptide 11, smooth muscle | 1.71 ↑ | + | | 2.01 ↑ | + | + | 1.18 ↑ | + | | CS/M | Biomechanics-Sarcomere |
| X15998 | chondroitin sulphate proteoglycan versican | 1.69 ↑ | + | | 2.54 ↑ | + | + | 1.44 ↑ | + | | CS/M | Biomechanics-EM |
| Y14737 | immunoglobulin λ heavy chain | 1.68 ↑ | + | | 3.08 ↑ | + | + | 2.87 ↑ | + | + | C/OD | Inflammation |
| M65292 | factor H homolog | 2.02 ↑ | + | + | 2.27 ↑ | + | + | 1.25 ↑ | + | | C/OD | Inflammation |
| U03106 | cyclin-dependent kinase inhibitor 1A | 4.49 ↑ | + | | 4.46 ↑ | + | + | 3.82 ↑ | + | + | CD | Apoptosis |
| AA224832 | metallothionein 1L | 2.59 ↑ | + | + | 2.17 ↑ | + | | 2.93 ↑ | + | + | U | Oxidative stress/Apoptosis |
| X79389 | GSTT1 | 1.84 ↑ | + | | 2.75 ↑ | + | + | 2.28 ↑ | + | + | M | Oxidative stress |
| X16396 | NAD-dependent methylene tetrahydrofolate dehydrogenase cyclohydrolase (EC 1.5.1.15) | 2.06 ↑ | + | + | 1.96 ↑ | + | + | 1.46 ↑ | + | | M | Oxidative/Energy |
| J02611 | apolipoprotein D | 3.02 ↑ | + | + | 2.00 ↑ | + | + | 2.47 ↑ | + | + | M | Energy |
| U16954 | AF1q | 3.98 ↑ | + | + | 4.28 ↑ | + | + | 4.30 ↑ | + | + | U | Unclassified |

*Group A:* genes differential in all 3 failing heart samples. For Tables 4–7, "Ratio" represents the mean value of the 4 $\log_2$(ratio) values available for every comparison between failing and nonfailing tissue (2 duplicates vs. 2 duplicates); up arrows (↑) indicate significantly upregulated genes [note that all genes in *group A* (Table 4) are significantly upregulated], and down arrows (↓) indicate significantly downregulated genes; ratio values with no arrows were not significantly different. Genes that have been assigned a "+" in the "SAM" column, have passed the SAM analysis ('significance analysis of microarrays') for significant differential expression. Genes that have been assigned a "+" in the "Affy" column, have passed the Affymetrix criteria for differential expression, as explained in MATERIALS AND METHODS. For $\log_2$(ratio) values denoted by "NR" (not reliable) no value is given because the ratio between the duplicate average difference (Avg Diff) values was outside of the range 0.5–2.0. See Tables 5 and 6 for *group B* (genes differential in 2 faling heart samples) and *group C* (genes differential in 1 failing heart sample), respectively. Functional categories are as listed in Table 3. Cytosk, cytoskeleton; EM, extracellular matrix.

gene, since both probe sets represented the 18S rRNA gene.

*Differential genes between failing LV and failing RV.* We found relatively fewer genes (20 in total, Table 7) differentially expressed between failing LV and RV than between failing and nonfailing heart. We divided the 20 differential genes into two major groups: *group A*, containing all genes differentially expressed in DCM LV vs. DCM/ICM RV and in ICM LV vs. DCM/ICM RV; and *group B*, containing genes differentially expressed in one of both comparisons. Six genes were upregulated and five genes were downregulated in both DCM LV and ICM LV. These expression changes may reflect a general, nonpathological, difference between LV and RV, but they may also represent a difference in the pathological state of the tissues (in failing hearts, LV function is generally more compromised than RV function). One example is cardiac α-myosin heavy chain (αMHC), a gene that we classified into *group A*. It has already been reported that αMHC expression does not differ between nonfailing LV and RV (28) and that expression decreases in failing hearts (8, 28, 31) (Table 5, this paper). Therefore, a higher expression in DCM and in ICM LV compared with DCM/ICM RV, as we found in our study, implicates that αMHC expression is more strongly affected in failing RV than in failing LV. *Group B* consisted of two genes downregulated in DCM LV and seven genes deregulated in ICM LV. A large difference between DCM and ICM LV was noted for c-fos, which showed a $\log_2$(ratio) of −0.08 for DCM LV vs. DCM/ICM RV and a $\log_2$(ratio) of −1.48 for ICM LV vs. DCM/ICM RV.

Table 5. *Genes differentially expressed in failing hearts vs. nonfailing heart: Group B*

| | Description | DCM LV vs. NF LV | | | ICM LV vs. NF LV | | | DCM/ICM RV vs. NF LV | | | Functional Category | Molecular Circuit |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Ratio | SAM | Affy | Ratio | SAM | Affy | Ratio | SAM | Affy | | |
| U08021 | nicotinamide N-methyltransferase | 2.05 ↑ | + | | 2.07 ↑ | + | + | NR | | | CS/C | Signaling |
| U76702 | follistatin-related protein FLRG | 1.60 ↑ | + | + | 1.68 ↑ | + | + | 0.82 | | | CS/C | Signaling |
| X16302 | insulin-like growth factor binding protein 2 | 1.66 ↑ | + | | 1.24 ↑ | + | + | 1.02 | | | CS/C | Signaling |
| U33632 | two P-domain K+ channel TWIK-1 | 1.62 ↑ | + | + | 1.08 ↑ | + | | 0.90 | | | CS/C | Signaling |
| AJ012737 | filamin, muscle isoform | 1.77 ↑ | + | + | 1.60 ↑ | + | + | 1.11 | | | CS/M | Biomechanics-Cytosk |
| J00068 | adult skeletal muscle α-actin | 1.74 ↑ | + | + | 1.42 ↑ | + | + | 1.12 | | | CS/M | Biomechanics-Cytosk |
| K03460 | α-tubulin isotype H2-α | 1.37 ↑ | + | + | 1.23 ↑ | + | | 0.31 | | | CS/M | Biomechanics-Cytosk |
| U76456 | tissue inhibitor of metalloproteinase 4 | 2.11 ↑ | + | + | 1.83 ↑ | + | + | 0.94 | | | CS/M | Biomechanics-EM |
| M55153 | transglutaminase | 1.51 ↑ | + | | 1.66 ↑ | + | + | 0.07 | | | CS/M | Biomechanics-EM |
| M55153 | transglutaminase | 1.33 ↑ | + | + | 1.23 ↑ | + | + | 0.84 | | | CS/M | Biomechanics-EM |
| D88674 | antizyme inhibitor | 1.69 ↑ | + | | 2.29 ↑ | + | + | 0.93 | | | CS/C | Cell growth |
| X68277 | dual specificity phosphatase 1 | 2.06 ↑ | + | | 2.70 ↑ | + | + | NR | + | | M+CS/C | Oxidative stress |
| L26336 | heat shock protein HSPA2 | 1.34 ↑ | + | | 2.50 ↑ | + | + | NR | + | | C/OD | Stress |
| X07523 | truncated form of complement factor H | 1.84 ↑ | + | | 2.00 ↑ | + | + | NR | | | C/OD | Inflammation |
| Y09836 | 3UTR of unknown protein | 2.37 ↑ | + | | 2.62 ↑ | + | + | NR | + | | U | Unclassified |
| AB020693 | reticulon 4 | 1.90 ↑ | + | + | 2.26 ↑ | + | + | 1.02 | | | U | Unclassified |
| AI651806 | cystein-rich motor neuron 1 | 1.80 ↑ | + | | 2.03 ↑ | + | + | 1.22 | | | U | Unclassified |
| AF025770 | C2H2 zinc finger protein (ZNF 198) | 1.51 ↑ | + | | 1.32 ↑ | + | + | 1.08 | | | G/PE | Unclassified |
| AF063002 | LIM protein SLIMMER | 1.38 ↑ | + | + | 1.42 ↑ | + | + | 0.93 | | | U | Unclassified |
| M18645 | Ig rearranged λ-chain | 1.27 | | | 2.80 ↑ | + | + | 3.08 ↑ | + | + | C/OD | Inflammation |
| M63438 | Ig rearranged γ-chain | NR | + | | 2.76 ↑ | + | + | 2.91 ↑ | + | + | C/OD | Inflammation |
| X57809 | rearranged immunoglobulin λ light chain | NR | | | 2.58 ↑ | + | + | 2.67 ↑ | + | + | C/OD | Inflammation |
| Y09765 | putative GABA receptor ε subunit | 0.88 | | | 1.56 ↑ | + | + | 1.37 ↑ | + | | CS/C | Signaling |
| L26336 | heat shock protein HSPA2 | 0.77 | | | 2.44 ↑ | + | + | 1.84 ↑ | + | | C/OD | Stress |
| U03688 | dioxin-inducible cytochrome P450 | 1.21 | | | 2.42 ↑ | + | + | 1.42 ↑ | + | | M | Energy |
| M30269 | nidogen | 1.10 | | | 1.16 ↑ | + | | 1.62 ↑ | + | + | CS/M | Biomechanics-EM |
| U58334 | tumor protein p53 binding protein, 2 (TP53BP2) | NR | | | 1.19 ↑ | + | | 1.40 ↑ | + | + | CD | Apoptosis |
| Z84718 | DNA sequence from clone 322BI on 22q11–12 | NR | + | | 2.32 ↑ | + | | 3.23 ↑ | + | + | U | Unclassified |
| AB028998 | KIAA1075 | NR | + | | 2.61 ↑ | + | + | 2.45 ↑ | + | | U | Unclassified |
| AF070569 | clone 24659 mRNA sequence | NR | | | 2.26 ↑ | + | + | 1.74 ↑ | + | | U | Unclassified |
| D78014 | dihydropyrimidinase related protein-3 | NR | + | | 2.00 ↑ | + | | 2.00 ↑ | + | + | U | Unclassified |
| AB018283 | KIAA0740 | 1.21 | | | 1.94 ↑ | + | + | 1.55 ↑ | + | | U | Unclassified |
| AJ093511 | cDNA clone IMAGE-1695674 | 0.40 | | | 1.38 ↑ | + | + | 2.07 ↑ | + | + | U | Unclassified |
| D82351 | retropseudogene MSSP-1 | 0.86 | | | 1.76 ↑ | + | + | 1.35 ↑ | + | | U | Unclassified |
| AI337192 | SH3BGR | 1.23 | | | 1.20 ↑ | + | + | 1.53 ↑ | + | + | U | Unclassified |
| AL079314 | cDNA clone EUROIMAGE 469780 | 0.60 | | | 1.05 ↑ | + | | 1.29 ↑ | + | + | U | Unclassified |
| U28686 | RNA binding motif protein 3 | 0.51 | | | 1.05 ↑ | + | | 1.14 ↑ | + | + | U | Unclassified |
| Z20656 | cardiac α-myosin heavy chain | −1.35 | | | −1.36 ↓ | + | + | −2.46 ↓ | + | + | CS/M | Biomechanics-Sarcomere |
| U37408 | phosphoprotein CtBP | −1.42 | | | −1.55 ↓ | + | | −1.45 ↓ | + | + | CD | Signaling |
| M22430 | RASF-A PLA2 | 2.28 ↑ | + | | 1.30 | | | 2.31 ↑ | + | + | M | Energy |
| AB005293 | perilipin | 1.94 ↑ | + | + | NR | | | 1.85 ↑ | + | + | M | Energy |
| AA128249 | fatty acid binding protein 4 | 1.84 ↑ | + | + | 0.85 | | | 1.85 ↑ | + | + | M | Energy |
| V01512 | cellular oncogene c-fos | 1.51 ↑ | + | | 0.72 | | | 1.74 ↑ | + | + | G/PE | Signaling |

*Group B:* genes differential in 2 failing heart samples. For details, see legend to Table 4. Functional categories are as listed in Table 3. Up arrows ( ↑ ) indicate significantly upregulated genes, and down arrows ( ↓ ) indicate significantly downregulated genes; ratio values with no arrows were not significantly different.

*Confirmation of microarray results.* To confirm the validity of our results obtained with the oligonucleotide microarrays, we performed quantitative real-time PCR on 10 genes selected from Tables 4–6: cytokine inducible nuclear protein, pyruvate dehydrogenase kinase isoform 4, complement protein component C7, δ sleep inducing peptide, ANF, BNP, AF1q, FLRG, SH3BGR, and SLIMMER (Fig. 5). For these genes, the expression

Table 6. *Genes differentially expressed in failing hearts vs. nonfailing heart: Group C*

| | Description | DCM LV vs. NF LV | | | ICM LV vs. NF LV | | | DCM/ICM RV vs. NF LV | | | Functional Category | Molecular Circuit |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Ratio | SAM | Affy | Ratio | SAM | Affy | Ratio | SAM | Affy | | |
| X87949 | heat shock protein 5 (HSPA5) | 1.35 ↑ | + | + | 0.54 | | | 0.48 | | | C/OD | Stress |
| X83703 | cytokine inducible nuclear protein | 1.40 ↑ | + | + | 0.65 | | | 0.55 | | | U | Unclassified |
| X15998 | chondroitin sulphate proteoglycan versican | 0.99 | | | 2.02 ↑ | + | + | 1.09 | | | CS/M | Biomechanics-EM |
| M92642 | α-1 type XVI collagen | 0.99 | | | 1.55 ↑ | + | + | 0.83 | | | CS/M | Biomechanics-EM |
| J04599 | bone small proteoglycan I (biglycan) | 0.95 | | | 1.42 ↑ | + | + | 0.92 | | | CS/M | Biomechanics-EM |
| X54304 | myosin regulatory light chain | 1.24 | | | 1.22 ↑ | + | + | 1.09 | | | CS/M | Biomechanics-Sarcomere |
| M95787 | SM22 smooth muscle protein, transgelin | 1.07 | | | 1.27 ↑ | + | + | 0.48 | | | CS/M | Biomechanics-Cytosk |
| HG1111–HT1111 | Ras-like protein Tc21 | 0.86 | | | 1.69 ↑ | + | + | 0.19 | | | CS/C | Signaling |
| S68134 | cAMP-responsive element modulator β isoform | NR | | | 1.52 ↑ | + | + | 0.94 | | | G/PE | Signaling |
| AB023137 | A kinase anchor protein 2 (AKAP2) | 0.73 | | | 1.32 ↑ | + | + | 0.49 | | | CS/C | Signaling |
| U03688 | dioxin-inducible cytochrome P450 | 1.09 | | | 2.26 ↑ | + | + | 1.06 | | | M | Energy |
| U54617 | pyruvate dehydrogenase kinase isoform 4 | 0.13 | | | 1.23 ↑ | + | + | −0.69 | | | M+CS/C | Energy |
| X01060 | transferrin receptor | 1.00 | | | 1.32 ↑ | + | + | 0.98 | | | U | Iron metabolism |
| U51712 | cDNA | 1.13 | | | 1.65 ↑ | + | + | 1.13 | | | U | Unclassified |
| A1635895 | δ sleep inducing peptide, immunoreactor | 1.25 | | | 1.29 ↑ | + | + | 0.26 | | | U | Unclassified |
| X92493 | STM-7 | −0.72 | | | −1.61 ↓ | + | + | −1.08 | | | CS/C | Signaling |
| J03507 | complement protein component C7 | −1.14 | | | −1.46 ↓ | + | + | −0.68 | | | C/OD | Inflammation |
| L25286 | α-1 type XV collagen | −1.56 | | | −1.42 ↓ | + | + | −0.56 | | | CS/M | Biomechanics-EM |
| U87408 | clone IMAGE-74593 | −1.09 | | | −1.76 ↓ | + | + | NR | | | U | Unclassified |
| U96750 | putative tumor suppressor NOEY2 | NR | | | NR | | | 1.86 ↑ | + | + | U | Cell growth |
| X67325 | interferon, α-inducible protein 27 | NR | + | | NR | | | 2.88 ↑ | + | + | U | Unclassified |
| L76259 | 6-pyruvoyltetrahydropterin synthase | 0.29 | | | NR | | | 1.92 ↑ | + | + | U | Unclassified |
| X84195 | acylphosphatase, muscle type isoenzyme | 0.52 | | | 0.81 | + | | 1.86 ↑ | + | + | U | Unclassified |
| Y00317 | liver microsomal UDP-glucuronosyltransferase | NR | | | 0.90 | | | 1.82 ↑ | + | + | U | Unclassified |
| X82103 | β-COP | 0.67 | | | 0.73 | | | 1.30 ↑ | + | + | U | Unclassified |
| U59919 | kinesin-associated protein 3 | 0.76 | | | 0.88 | | | 1.29 ↑ | + | + | U | Unclassified |
| MI0098 | 18S rRNA | 1.00 | | | −0.03 | | | −4.02 ↓ | + | + | G/PE | Translation |
| MI0098 | 18S rRNA | 1.26 | | | 0.11 | | | −3.15 ↓ | + | + | G/PE | Translation |

*Group C:* genes differential in 1 failing heart sample. For details, see legend to Table 4. Functional categories are as listed in Table 3. Up arrows ( ↑ ) indicate significantly upregulated genes, and down arrows ( ↓ ) indicate signifcantly downregulated genes; ratio values with no arrows were not significantly different.

ratios of failing vs. nonfailing heart determined by PCR were compared with the array data in Tables 4–6. Overall, although the expression ratios were not identical, the up- or downregulation of the 10 genes as determined by microarray analysis was comparable to that determined by real-time PCR. For two genes (pyruvate dehydrogenase kinase isoform 4 in DCM LV and δ sleep inducing peptide in DCM/ICM RV) an upregulation was found using microarrays, whereas real-time PCR showed a downregulation. However, in both cases no significant differential expression was detected by our microarray analysis because of failure to pass Affymetrix and SAM criteria. Since SLIMMER is an isoform of the skeletal muscle LIM protein 1 (SLIM1) (9), we designed two pairs of primers; one pair specific for SLIMMER, and one pair from the sequence common to both isoforms. Both pairs identified an upregulation of mRNA in failing ventricle. However, the absolute levels of expression of SLIMMER as determined by microarray (high) and real-time PCR (low, data not shown) made us believe that the SLIMMER probe sets on the array also hybridized with SLIM1 mRNA. Indeed, when we checked the exact sequence of the SLIMMER probe set on **http://www.netaffx.com**, we found that the sequence was in the region common to SLIM1 and SLIMMER.

*Evaluation of the control tissue.* To determine whether NF LV represented a valid control for our microarray study, we measured expression ratios of 10 genes (the same as described above) in 11 additional nonfailing heart samples (*NF01–NF11*, Table 2) vs. NF LV using real-time PCR. All nonfailing heart samples were obtained from individuals without cardiac disease. Two samples were from CF patients, two samples were obtained as commercially available RNA (Stratagene), and seven samples came from individuals who had died from accidents or cerebrovascular accidents. Ages varied from 18 years to 72 yr, and the series

Table 7. *Genes differentially expressed in failing LV vs. failing RV*

| Identifier | Description | DCM LV vs. DCM/ICM RV | | | ICM LV vs. DCM/ICM RV | | | Functional Category | Molecular Circuit |
|---|---|---|---|---|---|---|---|---|---|
| | | Ratio | SAM | Affy | Ratio | SAM | Affy | | |
| | | *Group A* | | | | | | | |
| Z20656 | cardiac α-myosin heavy chain | 1.08 ↑ | + | | 1.17 ↑ | + | + | CS/M | Biomechanics-Sarcomere |
| M83216 | aorta caldesmon | 1.15 ↑ | + | | 1.88 ↑ | + | + | CS/M | Biomechanics-Sarcomere |
| M61906 | phosphoinositide-3-kinase, regulatory subunit, 1 | 1.74 ↑ | + | | 2.00 ↑ | + | + | CS/C | Signaling |
| X75346 | MAP kinase activated protein kinase | 1.39 ↑ | + | | 1.74 ↑ | + | + | CS/C | Signaling |
| M61906 | phosphoinositide-3-kinase, regulatory subunit, 1 | 1.32 ↑ | + | + | 1.35 ↑ | + | + | CS/C | Signaling |
| L78833 | BRCA1, Rho7 and vatl genes, and ipf35 gene | 1.58 ↑ | + | | 1.53 ↑ | + | + | U | Unclassified |
| Y17782 | heat shock protein B3 | −1.26 ↓ | + | + | −1.32 ↓ | + | + | C/OD | Stress |
| U15590 | heat shock protein 3 | −1.14 ↓ | + | + | −1.10 ↓ | + | + | C/OD | Stress |
| I03191 | profilin | −2.35 ↓ | + | + | −2.01 ↓ | + | | CS/M | Biomechanics-Cytosk |
| L36033 | stromal cell-derived factor 1 | −2.10 ↓ | + | | −1.62 ↓ | + | + | C/OD | Inflammation |
| AL096713 | fer-1-like 2, myoferlin | −1.84 ↓ | + | + | −1.30 ↓ | + | | U | Membrane fusion |
| | | *Group B* | | | | | | | |
| Z70276 | fibroblast growth factor 12 | −2.62 ↓ | + | + | NR | + | | CS/C | Signaling |
| S82297 | β2-microglobulin | −1.02 ↓ | + | + | −0.93 | + | + | C/OD | Immune response |
| AE013570 | smooth muscle myosin heavy chain SM2 | 0.72 | + | | 1.31 ↑ | + | + | CS/M | Biomechanics-Sarcomere |
| AF001548 | myosin, heavy polypeptide 11, smooth muscle | 0.82 | + | | 1.25 ↑ | + | + | CS/M | Biomechanics-Sarcomere |
| HG1111-HT1111 | Ras-like protein Tc21 | 0.78 | | | 1.73 ↑ | + | + | CS/C | Signaling |
| U54617 | pyruvate dehydrogenase kinase isoform 4 | 0.85 | + | | 1.98 ↑ | + | + | M-CS/C | Energy |
| AB020693 | reticulon 4 | 0.81 | + | | 1.14 ↑ | + | + | U | Unclassified |
| V01512 | cellular oncogene c-fos | −0.08 | | | −1.48 ↓ | + | + | G/PE | Signaling |
| J02931 | coagulation factor III | −1.00 | + | | −1.84 ↓ | + | + | CS/C | Hemostasis |

For details, see legend to Table 4. Functional categories are as listed in Table 3. Up arrows ( ↑ ) indicate significantly upregulated genes, and arrows ( ↓ ) indicate significantly downregulated genes; ratio values with no arrows were not significantly different.

included both female ($n = 2$) and male ($n = 9$) patients. Figure 6 shows that for 6 of the 10 genes (cytokine inducible nuclear protein, pyruvate dehydrogenase kinase isoform 4, complement protein component C7, SH3BGR, FLRG, and ANF) the expression value of NF LV remained within the variation of the 11 individual expression values of *NF01–NF11* (complete data available on **http://www.ifr26.nantes.inserm.fr/composantes/u533/people/team1b/publi/affylogin.htm**). NF LV clearly showed a lower expression of SLIM1/SLIMMER and of the heart failure marker BNP.

## DISCUSSION

*Tissue sampling.* Our goal was to obtain and compare global images of transcription profiles of four well defined pathophysiological situations in human heart. We believed that this would allow an initial selection of genes of interest and could serve as a basis for follow-up studies of large series of individual patients. For this reason we decided to use pooled samples. The reason for the small sizes of the pools was that we included only those patients with similar clinical characteristics and with a phenotype that could be considered as "representative" for the disease, to exclude any excessive biological heterogeneity. A similar approach was recently used by Hwang et al. (22) in an analysis of gene expression profiles of dilated and hypertrophic failing hearts. For the nonfailing sample we chose LV from an explanted heart of a CF patient, for different reasons. The most evident reason is that this patient had been followed closely by clinicians and did not show any sign of heart failure. Although RV dysfunction has been found in adult patients with severe CF, LV abnormalities have not (or very rarely) been associated with this disease (15). Another advantage of our approach was that both types of tissue, failing and nonfailing heart, were obtained under similar clinical circumstances; both types of hearts came from patients who were undergoing heart transplantation. A disadvantage of our approach is that heart/lung transplantation in CF patients is usually performed at a relatively young age compared with heart transplantation in heart failure patients. Therefore there was a substantial age difference between the failing and nonfailing hearts. We cannot exclude that the CF phenotype was responsible for some of the expression changes we observed. Aging is known to be associated with a number of characteristic morphological and functional changes (reviewed in Ref. 36) that are similar to those seen in failing hearts (fibrosis, loss of myocytes, changes in mitochondrial oxidative metabolism). Therefore, the expression changes we identified in failing heart vs. nonfailing heart comparisons should be regarded as an overestimation for genes involved in those functions. For other functions, however, aging effects have not been proven in human cardiac tissue.

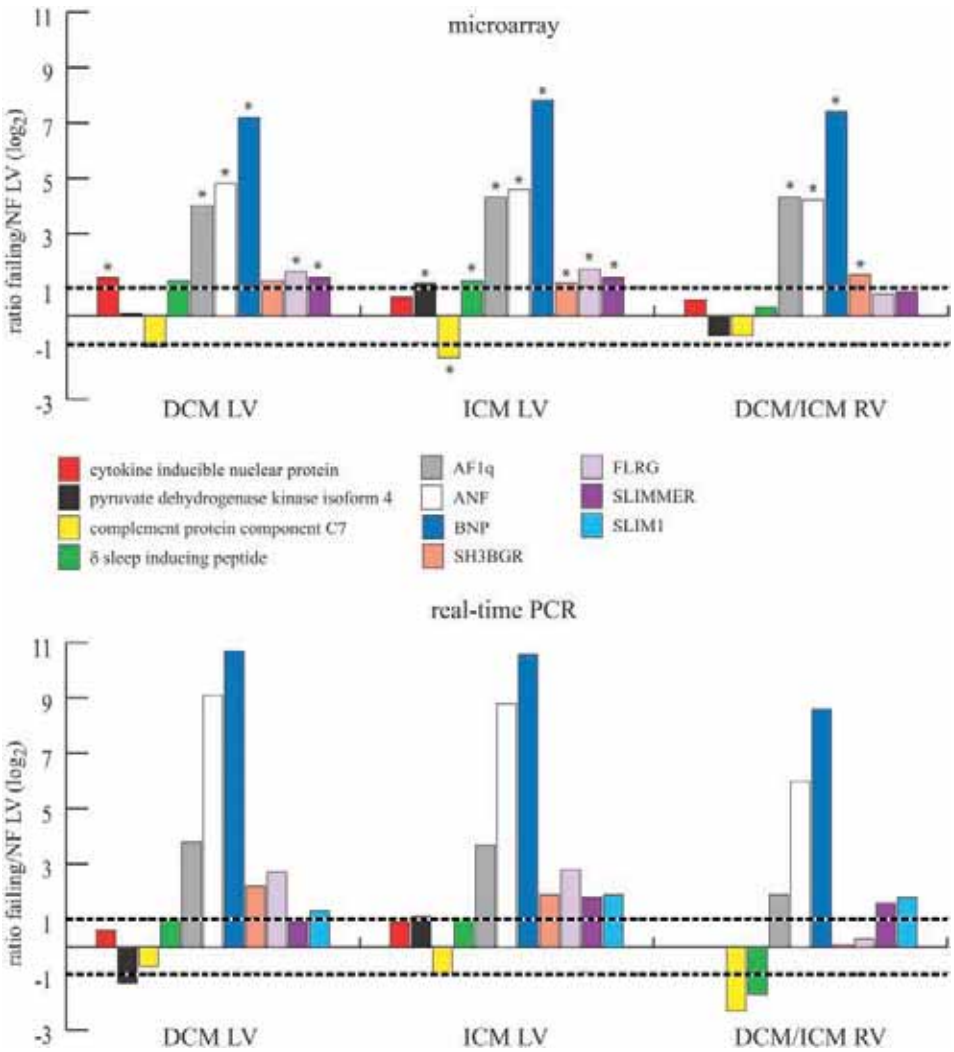



Fig. 5. Expression ratios (displayed in $\log_2$) of 10 genes as determined by microarray analysis and by real-time PCR. All tissues were compared with NF LV. *Genes that were determined to be significantly differential by our microarray analysis.

For example, although it has been suggested, based on animal studies, that apoptosis would be responsible for myocyte loss in aging human hearts, it has now been shown that aging does not influence the percentage of myocyte apoptosis (30) in human cardiac tissue. The same study did show a correlation between gender and apoptosis, but this fact does not affect our study since all samples came from male patients. More generally, we believe that "normal nonfailing" hearts are impossible to obtain. Other researchers have used tissue from donor hearts that could not be transplanted (22, 47). These donor hearts came from individuals that had been declared brain dead. This circumstance in itself or the events leading up to it also might influence gene expression. To address this issue more directly, we compared expression levels of 10 genes in 11 nonfailing hearts to the expression level in our control NF LV. One striking result was that among these nonfailing hearts NF LV displayed the lowest level of BNP expression. BNP is a known marker of heart failure (25), with increasing plasma levels corresponding to the severity of the disease. The lowest expression of BNP in NF LV underlines the "normal aspect" of this heart. Induction of BNP mRNA expression has been shown to occur as rapidly as 4 h after cardiac injury in rats (17). The extremely high expression of BNP in some of the nonfailing heart samples might reflect a deterioration of the hemodynamic status of brain-dead patients occurring shortly before explantation of the heart. The second gene that showed discordance of expression levels in NF LV vs. the additional nonfailing hearts was SLIM1. We found an upregulation of the isoform SLIMMER in failing hearts vs. NF LV using microarrays and an upregulation of both SLIM1 and SLIMMER using real-time PCR. Yang et al. (47) found a downregulation of SLIM1 in failing hearts, whereas Lim et al. (26) found an upregulation of both isoforms in hypertrophied hearts. These different findings may reflect different pathologies, but they may also be the result of the heterogeneity of SLIM1 expression levels in nonfailing hearts. Since for most of the genes tested the expression value of NF LV remained within the gene expression variation of 11 additional nonfailing hearts and since NF LV displayed the lowest level of expression of BNP, we believe that our choice of control tissue is validated. However, we want

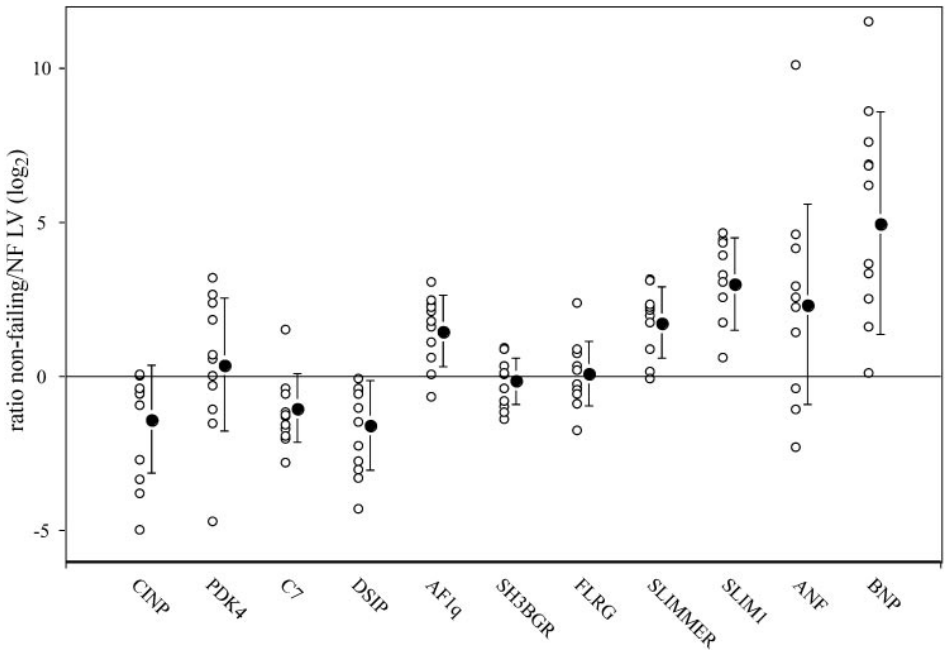



Fig. 6. Expression ratios (displayed in $\log_2$) of 10 genes in 11 nonfailing hearts compared with NF LV; open circles, individual expression ratios of each nonfailing heart; solid circles, mean values. Standard deviations are indicated by vertical bars. CINP, cytokine inducible nuclear protein; PDK4, pyruvate dehydrogenase kinase isoform 4; C7, complement protein component C7; DSIP, δ sleep inducing peptide. The horizontal line represents an expression level identical to that of NF LV [$\log_2$(ratio) = 0].

to stress that as long as different studies are based on different control samples, different results may be obtained.

*Genes similarly expressed in the four cardiac samples.* We have identified 1,306 genes that are significantly representative of a part of the transcriptional profile of the human heart, based on similar expression levels in all 4 pathophysiological situations. This profile therefore does not contain genes whose expression changes in failing hearts. To obtain a more complete profile, "truly normal" cardiac tissue should be analyzed. In the future maybe this could be obtained from cardiac biopsies, although biopsies are generally performed on patients with cardiac problems. Currently the sensitivity of microarrays is insufficient to analyze the minuscule quantities of cardiac biopsies. Another possibility would be to analyze explanted hearts from nonfailing hearts from individuals with different clinical backgrounds and to identify those genes with similar expression levels in all individuals.

It is generally believed that the expression profile characteristic for a certain tissue consists of few genes that are expressed at a high level (Table 3). Several of those genes, identified in this study, are known to be preferentially expressed in heart and/or skeletal muscle and therefore do not represent housekeeping but cardiac-specific genes. The high level of expression of certain genes involved in metabolism and the large representation of metabolic genes in our total cardiac gene population correlated well with the energy demanding function of the heart. The large group of genes involved in gene and protein expression consisted mainly of ribosomal protein encoding genes. This is most likely related to the high demand in metabolic and sarcomeric proteins. Overall, the distribution of our population of cardiac genes over the different functional categories corresponded well with the literature. The strongest represented categories (gene and protein expression, metabolism, and cell signaling and communication) were the same, albeit in a different order, as had been found by Yang et al. (47). This group studied gene expression in failing and nonfailing human LV using oligonucleotide microarrays containing ~7,000 genes, and they listed the expression of 473 genes in nonfailing hearts. The in silico analyses of human cardiac cDNA libraries (20, 21) identified more genes involved in cell structure and motility and less in cell signaling and communication. It is possible that this difference is related to the fact that the cDNA libraries had mainly been prepared from whole heart and not just from ventricle, thereby masking cardiac compartment-specific gene expression. The in silico analyses represented the first studies aimed at the description of a catalog of human cardiac genes (5, 21). We present here the first selection of genes representative of part of the human cardiac transcriptome using oligonucleotide microarrays. Currently, transcriptional profiles have also been assessed for human brain (35) and skeletal muscle (6, 7, 34). In the future, similar data will become available for additional normal human tissues. The comparison of the profiles will serve as a guide in the understanding of the functioning of the human body and in the identification of genes that play key roles in the determination of function.

*Differential gene expression.* In the development of heart failure, several molecular circuits have been shown to be involved (12, 19, 23). Factors implicated in the process of cardiac remodeling include the extracellular matrix, oxidative stress, metabolism, calcium signaling, apoptosis, the cytoskeleton, and the sarcomere. Although our approach was limited in the detection of genes involved in heart failure, we did identify genes associated with each of these factors as being differentially expressed in failing heart vs. nonfailing heart.

The downstream structural phase of the remodeling process was represented by changes in expression of genes encoding cytoskeletal proteins, sarcomeric proteins, and extracellular matrix proteins. These changes are responsible for progressive cytoskeletal stiffness, contractile dysfunction, and fibrosis. Some of these changes had already been described, like the downregulation of αMHC in failing LV and RV (8, 28, 31) and the upregulation of α-tubulin in failing LV (18). We also found additional evidence for a reversal to developmental gene expression, a phenomenon known to occur in cardiac remodeling and heart failure. This evidence consisted of an upregulation of MLC1emb, calponin, and SM22. Some expression changes were indicative of a dedifferentiation process, like the upregulation of other smooth muscle genes (smooth muscle myosin heavy chain and smooth muscle α-actin). Taken together with the data from the literature, our results therefore reinforce the hypothesis that cardiac remodeling involves many structural changes of cardiac tissue.

Structural ventricular remodeling most likely represents the final stage of heart failure. The pathways leading up to this may include disturbed signaling, resulting, for example, in increased apoptosis (33). In our study several apoptosis-related genes were identified as being differentially expressed in failing LV and/or RV, with the one showing the most marked activation being CDKN1A. Apoptosis may be caused by increased oxidative stress (10), a process that marks the transition of hypertrophy to heart failure (39). Stress-inducible metallothionein, which we found to be upregulated in failing LV and RV, is involved in both these processes. This protein functions as an antioxidant, and, in addition, inhibits the production of ANF and its apoptotic effect (24). It has been described that administration of isoproterenol, a β-receptor agonist, induces cardiac metallothionein levels (32). None of our patients received isoproterenol, and we found no association between administration of other β-receptor agonists (dobutamine, epinephrine) and levels of metallothionein RNA upregulation. Therefore, in our patients the increase of metallothionein is not related to medication.

In addition to genes that are assigned to certain pathways, we also detected aberrant expression of genes of which it is not known in which molecular circuit they are involved. Examples are AF1q, a transmembrane protein (43) also overexpressed in muscle tissue affected by α-sarcoglycan deficiency (11), and SH3BGR, which is preferentially expressed in skeletal muscle and heart (37) and maps to the Down syndrome heart critical region (13).

Relatively few genes (20) were differentially expressed between failing LV and failing RV. Two conclusions could be drawn from these results: *1*) LV and RV display similar expression profiles; and *2*) LV and RV are similarly affected by heart failure. Genes belonging to the functional category of cell and organism defense (C/OD) were found only to be lower expressed in failing RV vs. failing LV. Most likely this implies that those genes (heat shock protein B3, stromal cell-derived factor 1, and β2-microglobulin) are more severely upregulated by heart failure in LV than in RV. The gene that showed the most marked expression difference between failing LV and RV was profilin, which displayed a more than fourfold lower expression in failing RV compared with failing LV. Profilin is an actin-binding protein that has been implicated in the control of actin polymerization and cytoskeletal reorganization (41, 45, 46). However, questions remain about the exact in vivo function of the protein. Our findings about differential expression of profilin in failing LV and failing RV may lead to novel hypotheses.

Direct comparison of DCM LV and ICM LV showed that, after application of SAM and Affymetrix criteria, no genes were significantly differentially expressed between dilated and ischemic failing hearts. This result underlined the fact that all analyzed tissue came from end-stage failing hearts. The expression profiles that we determined were probably more representative of end-stage heart failure than of DCM or ICM specifically.

Although the analysis of gene expression using (nucleotide) microarrays is a powerful technique, it does have its limitations. Obviously, not all genes are represented (yet) on the array, and therefore, the knowledge that can be acquired from these experiments remains incomplete. In addition, the outcome of microarray analysis is sensitive to the design of the probes on the array. This point was highlighted in our results by the difficulty of discriminating the alternatively spliced isoforms SLIM and SLIMMER. However, with our study we succeeded to add to the clarification of the cardiac transcriptome in different situations. Part of our data had already been described by others and therefore functioned as a confirmation of those findings and a validation of our method. In addition, we presented novel information resulting in an expanded view of human cardiac gene expression. Taken together, our results provide another step forward on the pathway toward improved understanding and treatment of cardiac pathologies.

We thank Dr. Gilles Lande for advice on the clinical aspects of the study; we acknowledge Drs. Jean-Marie Heslan and Cédric Louvet for assistance with the real-time PCR experiments. We are grateful to the thoracic surgery department of Nantes University Hospital and to Dr. Miklos Opincariu at the Department of Cardiac Surgery in Szeged for participation in providing us with the cardiac tissues.

This work was supported by special grants from Aventis, the Association Française contre les Myopathies, INSERM, CNRS, the Region Pays de Loire, and by Hungarian Ministry of Health Grant ETT 536/200. M. Steenman was recipient of a Poste Vert from INSERM. Y.-W. Chen was supported by a postdoctoral fellowship from Stichting Porticus.

## REFERENCES

1. **Adams MD, Kerlavage AR, Fleischmann RD, Fuldner RA, Bult CJ, Lee NH, Kirkness EF, Weinstock KG, Gocayne JD, White O, and et al.** Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence. *Nature* 377, *Suppl*: 3–174, 1995.
2. **Anisimov SV, Lakatta EG, and Boheler KR.** Discovering altered genomic expression patterns in heart: transcriptome

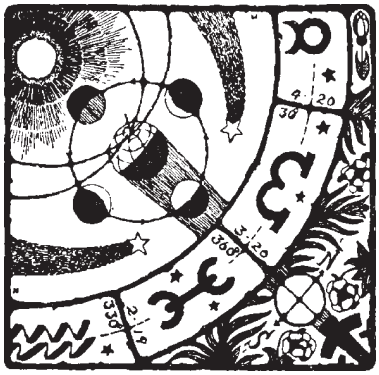
determination by serial analysis of gene expression. *Eur J Heart Fail* 3: 271–281, 2001.
3. **Barrans JD, Allen PD, Stamatiou D, Dzau VJ, and Liew CC.** Global gene expression profiling of end-stage dilated cardiomyopathy using a human cardiovascular-based cDNA microarray. *Am J Pathol* 160: 2035–2043, 2002.
4. **Barrans JD, Stamatiou D, and Liew C.** Construction of a human cardiovascular cDNA microarray: portrait of the failing heart. *Biochem Biophys Res Commun* 280: 964–969, 2001.
5. **Bortoluzzi S, d'Alessi F, and Danieli GA.** A computational reconstruction of the adult human heart transcriptional profile. *J Mol Cell Cardiol* 32: 1931–1938, 2000.
6. **Bortoluzzi S, d'Alessi F, Romualdi C, and Danieli GA.** The human adult skeletal muscle transcriptional profile reconstructed by a novel computational approach. *Genome Res* 10: 344–349, 2000.
7. **Bortoluzzi S, Rampoldi L, Simionati B, Zimbello R, Barbon A, d'Alessi F, Tiso N, Pallavicini A, Toppo S, Cannata N, Valle G, Lanfranchi G, and Danieli GA.** A comprehensive, high-resolution genomic transcript map of human skeletal muscle. *Genome Res* 8: 817–825, 1998.
8. **Bouvagnet P, Mairhofer H, Leger JO, Puech P, and Leger JJ.** Distribution pattern of alpha and beta myosin in normal and diseased human ventricular myocardium. *Basic Res Cardiol* 84: 91–102, 1989.
9. **Brown S, McGrath MJ, Ooms LM, Gurung R, Maimone MM, and Mitchell CA.** Characterization of two isoforms of the skeletal muscle LIM protein 1, SLIM1. Localization of SLIM1 at focal adhesions and the isoform slimmer in the nucleus of myoblasts and cytoplasm of myotubes suggests distinct roles in the cytoskeleton and in nuclear-cytoplasmic communication. *J Biol Chem* 274: 27083–27091, 1999.
10. **Cesselli D, Jakoniuk I, Barlucchi L, Beltrami AP, Hintze TH, Nadal-Ginard B, Kajstura J, Leri A, and Anversa P.** Oxidative stress-mediated cardiac cell death is a major determinant of ventricular dysfunction and failure in dog dilated cardiomyopathy. *Circ Res* 89: 279–286, 2001.
11. **Chen Y, Zhao P, Borup R, and Hoffman EP.** Expression profiling in the muscular dystrophies: identification of novel aspects of molecular pathophysiology. *J Cell Biol* 151: 1321–1336, 2000.
12. **Chien KR.** Genomic circuits and the integrative biology of cardiac diseases. *Nature* 407: 227–232, 2000.
13. **Egeo A, Di L, Sandri C, Mazzocco M, Lapide M, Schiaffino S, and Scartezzini P.** Developmental expression of the SH3BGR gene, mapping to the Down syndrome heart critical region. *Mech Dev* 90: 313–316, 2000.
14. **Eichhorn EJ.** Prognosis determination in heart failure. *Am J Med* 110, *Suppl* 7A: 14S–36S, 2001.
15. **Florea VG, Florea ND, Sharma R, Coats AJ, Gibson DG, Hodson ME, and Henein MY.** Right ventricular dysfunction in adult severe cystic fibrosis. *Chest* 118: 1063–1068, 2000.
16. **Francis GS.** Pathophysiology of chronic heart failure. *Am J Med* 110, *Suppl* 7A: 37S–46S, 2001.
17. **Hama N, Itoh H, Shirakami G, Nakagawa O, Suga S, Ogawa Y, Masuda I, Nakanishi K, Yoshimasa T, Hashimoto Y, Yamaguchi M, Hori R, Yasue H, and Nakao K.** Rapid ventricular induction of brain natriuretic peptide gene expression in experimental acute myocardial infarction. *Circulation* 92: 1558–1564, 1995.
18. **Hein S, Kostin S, Heling A, Maeno Y, and Schaper J.** The role of the cytoskeleton in heart failure. *Cardiovasc Res* 45: 273–278, 2000.
19. **Hunter JJ and Chien KR.** Signaling pathways for cardiac hypertrophy and failure. *N Engl J Med* 341: 1276–1283, 1999.
20. **Hwang DM, Dempsey AA, Lee CY, and Liew CC.** Identification of differentially expressed genes in cardiac hypertrophy by analysis of expressed sequence tags. *Genomics* 66: 1–14, 2000.
21. **Hwang DM, Dempsey AA, Wang R, Rezvani M, Barrans JD, Dai K, Wang H, Ma H, Cukerman E, Liu Y, Gu J, Zhang J, Tsui SKW, Waye MMY, Fung K, Lee C, and Liew C.** A genome-based resource for molecular cardiovascular medicine: toward a compendium of cardiovascular genes. *Circulation* 96: 4146–4203, 1997.
22. **Hwang JJ, Allen PD, Tseng GC, Lam CW, Fananapazir L, Dzau VJ, and Liew CC.** Microarray gene expression profiles in dilated and hypertrophic cardiomyopathic end-stage heart failure. *Physiol Genomics* 10: 31–44, 2002. First published April 30, 2002; 10.1152/physiolgenomics.00122.2001.
23. **Hwang JJ, Dzau VJ, and Liew CC.** Genomics and the pathophysiology of heart failure. *Curr Cardiol Rep* 3: 198–207, 2001.
24. **Kang YJ, Zhou ZX, Wu H, Wang GW, Saari JT, and Klein JB.** Metallothionein inhibits myocardial apoptosis in copper-deficient mice: role of atrial natriuretic peptide. *Lab Invest* 80: 745–757, 2000.
25. **Kjaer A and Hesse B.** Heart failure and neuroendocrine activation: diagnostic, prognostic and therapeutic perspectives. *Clin Physiol* 21: 661–672, 2001.
26. **Lim D, Roberts R, and Marian AJ.** Expression profiling of cardiac genes in human hypertrophic cardiomyopathy: insight into the pathogenesis of phenotypes. *J Am Coll Cardiol* 38: 1175–1180, 2001.
27. **Lockhart DJ, Dong H, Byrne MC, Follettie MT, Gallo MV, Chee MS, Mittmann M, Wang C, Kobayashi M, Horton H, and Brown EL.** Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nat Biotechnol* 14: 1675–1680, 1996.
28. **Lowes BD, Minobe W, Abraham WT, Rizeq MN, Bohlmeyer TJ, Quaife RA, Roden RL, Dutcher DL, Robertson AD, Voelkel NF, Badesch DB, Groves BM, Gilbert EM, and Bristow MR.** Changes in gene expression in the intact human heart. Downregulation of alpha-myosin heavy chain in hypertrophied, failing ventricular myocardium. *J Clin Invest* 100: 2315–2324, 1997.
29. **Lyn D, Liu X, Bennett NA, and Emmett NL.** Gene expression profile in mouse myocardium after ischemia. *Physiol Genomics* 2: 93–100, 2000.
30. **Mallat Z, Fornes P, Costagliola R, Esposito B, Belmin J, Lecomte D, and Tedgui A.** Age and gender effects on cardiomyocyte apoptosis in the normal human heart. *J Gerontol A Biol Sci Med Sci* 56: M719–M723, 2001.
31. **Nakao K, Minobe W, Roden R, Bristow MR, and Leinwand LA.** Myosin heavy chain gene expression in human heart failure. *J Clin Invest* 100: 2362–2370, 1997.
32. **Namikawa K, Okazaki Y, Nishida S, Kimoto S, Akai F, Tomura T, and Hashimoto S.** Changes in myocardial metallothionein on isoproterenol-induced myocardial injury. *Yakugaku Zasshi* 113: 591–595, 1993.
33. **Narula J, Haider N, Virmani R, DiSalvo TG, Kolodgie FD, Hajjar RJ, Schmidt U, Semigran MJ, Dec GW, and Khaw BA.** Apoptosis in myocytes in end-stage heart failure. *N Engl J Med* 335: 1182–1189, 1996.
34. **Pietu G, Eveno E, Soury-Segurens B, Fayein NA, Mariage-Samson R, Matingou C, Leroy E, Dechesne C, Krieger S, Ansorge W, Reguigne-Arnould I, Cox D, Dehejia A, Polymeropoulos MH, Devignes MD, and Auffray C.** The genexpress IMAGE knowledge base of the human muscle transcriptome: a resource of structural, functional, and positional candidate genes for muscle physiology and pathologies. *Genome Res* 9: 1313–1320, 1999.
35. **Pietu G, Mariage-Samson R, Fayein NA, Matingou C, Eveno E, Houlgatte R, Decraene C, Vandenbrouck Y, Tahi F, Devignes MD, Wirkner U, Ansorge W, Cox D, Nagase T, Nomura N, and Auffray C.** The Genexpress IMAGE knowledge base of the human brain transcriptome: a prototype integrated resource for functional and computational genomics. *Genome Res* 9: 195–209, 1999.
36. **Roffe C.** Ageing of the heart. *Br J Biomed Sci* 55: 136–148, 1998.
37. **Scartezzini P, Egeo A, Colella S, Fumagalli P, Arrigo P, Nizetic D, Taramelli R, and Rasore-Quartino A.** Cloning a new human gene from chromosome 21q223 encoding a glutamic acid-rich protein expressed in heart and skeletal muscle. *Hum Genet* 99: 387–392, 1997.
38. **Sehl PD, Tai JT, Hillan KJ, Brown LA, Goddard A, Yang R, Jin H, and Lowe DG.** Application of cDNA microarrays in determining molecular phenotype in cardiac growth, development, and response to injury. *Circulation* 101: 1990–1999, 2000.

39. **Singh N, Dhalla AK, Seneviratne C, and Singal PK.** Oxidative stress and heart failure. *Mol Cell Biochem* 147: 77–81, 1995.
40. **Stanton LW, Garrard LJ, Damm D, Garrick BL, Lam A, Kapoun AM, Zheng Q, Protter AA, Schreiner GF, and White RT.** Altered patterns of gene expression in response to myocardial infarction. *Circ Res* 86: 939–945, 2000.
41. **Suetsugu S, Miki H, and Takenawa T.** The essential role of profilin in the assembly of actin for microspike formation. *EMBO J* 17: 6516–6526, 1998.
42. **Tan FL, Moravec CS, Li J, Apperson-Hansen C, McCarthy PM, Young JB, and Bond M.** The gene expression fingerprint of human heart failure. *Proc Natl Acad Sci USA* 99: 11387–11392, 2002.
43. **Tse W, Zhu W, Chen HS, and Cohen A.** A novel gene, AF1q, fused to MLL in t(1;11) (q21;q23), is specifically expressed in leukemic and immature hematopoietic cells. *Blood* 85: 650–656, 1995.
44. **Tusher VG, Tibshirani R, and Chu G.** Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci USA* 98: 5116–5121, 2001.
45. **Wolven AK, Belmont LD, Mahoney NM, Almo SC, and Drubin DG.** In vivo importance of actin nucleotide exchange catalyzed by profilin. *J Cell Biol* 150: 895–904, 2000.
46. **Yang C, Huang M, DeBiasio J, Pring M, Joyce M, Miki H, Takenawa T, and Zigmond SH.** Profilin enhances Cdc42-induced nucleation of actin polymerization. *J Cell Biol* 150: 1001–1012, 2000.
47. **Yang J, Moravec CS, Sussman MA, DiPaola NR, Fu D, Hawthorn L, Mitchell CA, Young JB, Francis GS, McCarthy PM, and Bond M.** Decreased SLIM1 expression and increased gelsolin expression in failing human hearts measured by high-density oligonucleotide arrays. *Circulation* 102: 3046–3052, 2000.

*Physiol Genomics* 18: 128, 2004;
10.1152/physiolgenomics.00116.2004.

*Online Release 12.2, January 2003*
*Print Volume 12, April 2003*

*Pages 97–112:* Steenman M, Chen YW, Le Cunff M, Lamirault G, Varró A, Hoffman E, and Léger JJ. "Transcriptomal analysis of failing and nonfailing human hearts" (DOI 10.1152/physiolgenomics.00148.2002; http://physiolgenomics.physiology.org/cgi/content/full/12/2/97). The Data Supplement for this article (as mentioned in the second and last paragraphs of the RESULTS; pages 101 and 107) has been moved to a new URL, different from the URL printed in the original article. The Data Supplement may now be accessed through the *Physiological Genomics* web site, via **http://physiolgenomics.physiology.org/cgi/content/full/00148.2002/DC1**. The online version of this article has been corrected to reflect this new URL and thus deviates from the printed journal with regard to this correction.



## III. ETUDE 3

Steenman M, Lamirault G, Le Meur N, Le Cunff M, Escande D, Leger JJ. Distinct molecular portraits of human failing hearts identified by dedicated cDNA microarrays. Eur J Heart Fail. 2005 Mar 2;7(2):157-65.

L'objectif de cette étude était de rechercher si l'utilisation de profils d'expression génique individuels pouvait aboutir à une classification de patients en insuffisance cardiaque.

Des échantillons tissulaires de 15 patients en insuffisance cardiaque avancée et 2 patients sans dysfonction cardiaque ont été analysés avec une puce dédiée construite à partir d'une collection de 1000 séquences nucléotidiques d'intérêt dans l'étude du transcriptome cardiaque. Un ensemble de 159 gènes présentait des variations d'expression discriminantes entre les différents échantillons biologiques. Une classification hiérarchique ascendante a regroupé les 15 patients et les 159 gènes sur la base de leurs profils d'expression.

La classification des patients a permis l'identification de 3 sous-groupes distincts caractérisés chacun par des marqueurs moléculaires spécifiques. La confrontation du phénotype clinique et de la classification moléculaire des patients a révélé une corrélation entre cette classification moléculaire et un indice de gravité clinique de l'insuffisance cardiaque avancée.

Trois clusters de gènes étaient particulièrement impliqués dans la classification des patients. Un cluster contenait essentiellement des sondes représentatives des gènes codant les peptides natriurétiques. Un deuxième cluster contenait principalement des gènes impliqués dans le métabolisme énergétique. Des gènes impliqués dans la structure/contraction cellulaire composaient la majorité du troisième cluster.

Cette étude pilote a démontré qu'une classification de patients en insuffisance cardiaque sur la base de leur profil d'expression génique cardiaque était envisageable. L'existence d'une corrélation entre la classification moléculaire et un paramètre clinique pronostique de l'insuffisance cardiaque avancée a renforcé la stratégie visant à identifier au sein du transcriptome cardiaque de nouveaux marqueurs moléculaires pronostiques de l'insuffisance cardiaque avancée et de stades plus précoces de la maladie.

ELSEVIER

The European Journal of Heart Failure 7 (2005) 157–165

The European Journal of Heart Failure

www.elsevier.com/locate/heafai

# Distinct molecular portraits of human failing hearts identified by dedicated cDNA microarrays

Marja Steenman*, Guillaume Lamirault, Nolwenn Le Meur, Martine Le Cunff, Denis Escande, Jean J. Léger

*Laboratoire de Physiopathologie et de Pharmacologie Cellulaires et Moléculaires, INSERM U533, Faculté de Médecine, 1 Rue Gaston Veil, BP 53508, 44035 Nantes, Cedex 1, France*

Received 31 March 2004; received in revised form 19 April 2004; accepted 17 May 2004
Available online 30 July 2004

**Abstract**

*Aims:* This study aimed to investigate whether a molecular profiling approach should be pursued for the classification of heart failure patients. *Methods and results:* Applying a subtraction strategy we created a cDNA library consisting of cardiac- and heart failure-relevant clones that were used to construct dedicated cDNA microarrays. We measured relative expression levels of the corresponding genes in left ventricle tissue from 17 patients (15 failing hearts and 2 nonfailing hearts). Significance analysis of microarrays was used to select 159 genes that distinguished between all patients. Two-way hierarchical clustering of the 17 patients and the 159 selected genes led to the identification of three major subgroups of patients, each with a specific molecular portrait. The two nonfailing hearts clustered closely together. Interestingly, our classification of patients based on their molecular portraits did not correspond to an identified etiological classification. Remarkably, patients with the highest medical urgency status (United Network for Organ Sharing, Status 1A) clustered together. *Conclusion:* With this pilot feasibility study we demonstrated a novel classification of end-stage heart failure patients, which encourages further development of this approach in prospective studies on heart failure patients at earlier stages of the disease.


*Keywords:* Classification; Idiopathic dilated cardiomyopathy; Coronary artery disease; Gene expression profiling; Cluster analysis

## 1. Introduction

Molecular expression profiling studies conducted in human breast cancer [1], prostate cancer [2], and embryonal tumours of the central nervous system [3] have led to the ultimate utilization of cDNA microarrays to predict clinical outcome based on a tumour's expression profile. Heart failure has recently been compared to cancer [4], in that it involves the same biological principles of cell growth, death, and survival. We thus wondered whether molecular expression profiling could also be used to classify failing hearts, i.e. whether failing hearts show distinct molecular profiles irrespective of their aetiology.

Since human cardiac tissue is evidently less accessible than tumour material, expression profiling has been performed to a lesser extent in human cardiac disease. The first cardiac expression profiles were based on in silico analyses of expressed sequence tags (ESTs) obtained from cardiac cDNA libraries, leading to catalogues of genes expressed in normal or hypertrophied hearts [5–7]. These studies were followed by microarray analyses identifying genes with aberrant expression levels in failing hearts [8–13]. More recently, attempts were conducted to classify small groups of patients with end-stage heart failure based on their expression profile. One study described the classification of seven failing and five nonfailing hearts using the expression ratios of *all* clones on their array [11]. In this study, most of the failing hearts clustered together, which raised a concern as to the usefulness of classification based on molecular portraits. A second study classified eight failing and eight nonfailing hearts, based on the expression ratios of those genes that were differentially expressed between the *group* of failing and the *group* of nonfailing hearts [12]. Their results showed that two patients with an aetiology distinct from the rest (alcoholic and familial cardiomyopathy) clustered away from the other failing hearts.

* Corresponding author. Tel.: +33-240412844; fax: +33-240412950.
*E-mail address:* marja.steenman@nantes.inserm.fr (M. Steenman).

1388-9842/$ - see front matter 
doi:10.1016/j.ejheart.2004.05.008

In the present study, we report expression profiling of failing hearts analyzing the largest group of patients thus far. To achieve our goal, we first constructed a cDNA library consisting only of heart failure and cardiac clones, identified using a subtraction strategy. To obtain molecular portraits, these clones were spotted onto custom microarrays and RNA obtained from 15 failing hearts and 2 nonfailing hearts was analyzed. A subset of genes was selected using a statistical approach that identified genes that showed inter-patient differential expression levels. We anticipated that the genes so selected were relevant for distinguishing between patients. Using this strategy, we identified three subgroups of patients, each with a specific molecular portrait, which could not be unequivocally related to a clinical classification. This pilot study shows that failing hearts can be distinguished according to their molecular profile.

## 2. Methods

### 2.1. Patients

Cardiac tissues were from explanted hearts from male patients who underwent a heart transplantation. The patients were affected by either idiopathic dilated cardiomyopathy (IDCM) or coronary artery disease (CAD). All our CAD patients had a history of myocardial infarction and severe left ventricular systolic dysfunction (mean ejection fraction: $20 \pm 5.4\%$). To establish diagnosis, aetiology, and stage of their disease, all IDCM and CAD patients underwent a complete cardiac evaluation including history, physical exam, coronary angiography, isotopic angiography, cardiac catheterization, and two-dimensional echocardiography (Table 1). None of the IDCM patients had a family history of cardiomyopathy. Since under French law it is not possible to obtain nonfailing hearts from donors for research purposes, nonfailing hearts were obtained from patients affected by cystic fibrosis (CF) who underwent a heart/lung transplantation. These hearts did not show any sign of heart failure. Chest X-ray, ECG, echocardiography, and histology revealed no obvious cardiac abnormalities. For the molecular portrait of individual patients, left ventricle (LV) tissue was used.

To obtain expression ratios, a common reference was prepared. This reference was not used as a control but as a common point of reference that enabled a comparison between the 17 individual patients. Therefore, only two criteria needed to be addressed: (1) we needed a large enough quantity of reference material for 51 hybridizations; (2) the reference needed to hybridize to the cardiac clones on the microarray. A pool of failing and nonfailing right ventricle (RV) met with these criteria. We therefore selected eight patients for the reference pool for microarray analysis (see Table 1) using RV tissue. This reference pool contained cardiac samples from IDCM, CAD, and CF patients. Since the RV reference pool did not consist of normal heart tissue, we obtained relative (not absolute) expression ratios that

Table 1
Patient characteristics

| Patient | Pathology | Age (years) | Sex | Time between onset of heart failure and transplantation (months) | Ejection fraction (%) | MPAP (mm Hg) | Reduced RV function | Complete AV block | VT episodes | Treatment | NYHA | UNOS status |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| D00 | IDCM | 59 | male | 93 | 11 | 39 | yes | no | no | AM, CP, FR | II | 2 |
| D02 | IDCM | 59 | male | 21 | 19 | 29 | yes | no | yes | EL, FR, ST | III | 2 |
| D03[a] | IDCM | 56 | male | 13 | 21 | 38 | no | no | no | CP, DG, DP, FR | IV | 1B |
| D04 | IDCM | 45 | male | 16 | 23 | 35 | yes | yes | no | CP, DG, FR | III | 2 |
| D05 | IDCM | 62 | male | 4 | 15 | 30 | yes | no | no | AM, DB, HCT, MR, SP | IV | 1A |
| D06 | IDCM | 57 | male | 9 | 15 | 30 | yes | yes | yes | AM, CP, FR | IV | 2 |
| D07[a] | IDCM | 34 | male | 10 | 18 | 38 | no | no | no | AM, DB, EN, EP, FR, HCT | IV | 1A |
| D08 | IDCM | 47 | male | 39 | 20 | 44 | no | no | yes | DB, FR, SP | IV | 1A |
| I03 | CAD | 64 | male | 19 | 28 | 34 | yes | no | no | AM, CP, FR | III | 2 |
| I04[a] | CAD | 62 | male | 9 | 22 | 21 | no | no | yes | AC, AM | III | 2 |
| I05 | CAD | 36 | male | 6 | 15 | 46 | yes | no | no | AM, FR, ML, LS | IV | 2 |
| I06[a] | CAD | 65 | male | 18 | 21 | 39 | no | no | no | AM, CP, DB, FR, NC | III | 2 |
| I08[a] | CAD | 56 | male | 44 | 24 | 51 | yes | no | no | CR, DG, FR, ML, LS | III | 2 |
| I09 | CAD | 57 | male | 4 | 12 | 28 | yes | no | no | DB, EN, FR | IV | 1A |
| I10[a] | CAD | 49 | male | 4 | 18 | 33 | no | no | no | AM, CP, FR, SP | IV | 2 |
| NF1[a] | CF | 15 | male | | n.a. | n.a. | no | no | no | none | | |
| NF2[a] | CF | 20 | male | | 59 | n.a. | no | no | no | none | | |

Characteristics of all patients analyzed in the study. MPAP: mean pulmonary arterial pressure; AV: atrioventricular; VT: ventricular tachycardia; n.a.: not available; AC: acebutolol; AM: amiodarone; CP: captopril; CR: carvedilol; DB: dobutamine; DG: digoxin; DL: diltiazem; DP: dopamine; EL: enalapril; EN: enoximone; EP: epinephrine; FR: furosemide; HCT: hydrochlorothiazide; LN: lisinopril; LS: losartan; ML: molsidomine; MR: milrinone; NC: nicorandil; SP: spironolactone; ST: sotalol; NYHA: New York Heart Association functional class; UNOS: United Network for Organ Sharing. The assignment of a UNOS Status takes into account the presence or absence and also the dose of intravenous inotropic treatment at the time of transplantation.

[a] Patients included in the reference pool.

were used to compare expression profiles between patients and to construct molecular portraits.

### 2.2. Construction, hybridization, and validation of cardiac cDNA microarrays

We constructed a cDNA library of cardiac-relevant genes by performing three different suppression subtractive hybridization experiments (SSH) [14]; SSH1: nonfailing atrium (from a CAD patient) vs. nonfailing septum (from patient NF1); SSH2: failing septum (from patient D03) vs. nonfailing septum (from patient NF1); SSH3: failing LV (from patient D07) vs. nonfailing LV (from patient NF2) (Table 1). Details on RNA isolation and labelling, cDNA microarray construction and hybridization, and real-time PCR experiments are available on www.ifr26.nantes.inserm.fr/Francais/Labos/U533/INDEX/index.htm.

### 2.3. Data analysis

For typical reproducibility data using our microarrays, see Ref. [15]. Fluorescence values and ratios were analyzed using GenePix 4000. Low-quality spots were filtered out using a scoring system developed in our laboratory (Le Meur et al., manuscript in preparation), leaving only spots with valid expression values. To normalize Cy3 and Cy5 values, invariant genes were selected using the rank invariant method. A nonlinear regression method (lowess fitness) was applied to the invariant genes to calculate the normalization correction, which was then used to normalize all spots on the array [16]. After normalization, only those clones with valid expression values for at least 2 of the triplicate arrays for all patients were retained for further analysis. Genes with statistically significant differential expression between patients were identified using significance analysis of microarrays (SAM) [17]. SAM was performed using the multiclass option and the delta threshold ($\Delta = 0.00096$) corresponding to the lowest median false discovery rate (FDR = 0.27%). Two-way hierarchical agglomerative clustering was applied to the gene expression matrix consisting of the 17 patients and the statistically significant differentially expressed genes. The input consisted of the average of the triplicate expression values for each clone. We applied average linkage clustering with uncentered correlation using the Cluster program [18]. Clusters were visualized using the Treeview program. Correlations between identified patient subgroups and clinical parameters were analyzed using one-way ANOVA.

## 3. Results

### 3.1. Custom cardiac cDNA microarray

Three separate SSH experiments were performed between failing and nonfailing cardiac tissue and between nonfailing atrium and septum. This strategy allowed us to isolate cardiac cDNA clones that code for genes involved in heart failure and also genes involved in normal ventricular function (as opposed to atrial function). From this collection, 440 genes were selected based on either differential or high-level cardiac expression (Table 2). Many of these genes were previously identified as being differentially expressed between failing and nonfailing human hearts [7,10–12]. We used these 440 genes to construct our cardiac cDNA microarray and spotted all clones minimally as duplicates onto the glass slides. The clones were classified into the six main functional categories: metabolism, gene/protein expression, cell structure/motility, cell signalling/communication, cell/organism defence, and cell division (Table 2). The distribution of the clones into the six functional categories did not differ significantly between the different SSH libraries. Therefore, the representation of the different functional categories in our collection of genes

Table 2
SSH libraries

| Library | Genes preferentially expressed in | | No. of nonredundant genes | Genes in functional categories (%) | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | M | G/PE | CS/M | CS/C | C/OD | CD |
| SSH1 | atrium | | 20 | | | | | | |
| | septum | | 55 | | | | | | |
| | atrium + septum | | 22 | | | | | | |
| | | total: | 97 | 42 | 20 | 17 | 16 | 4 | 0 |
| SSH2 | failing septum | | 43 | | | | | | |
| | nonfailing septum | | 50 | | | | | | |
| | failing + nonfailing septum | | 55 | | | | | | |
| | | total: | 148 | 41 | 21 | 21 | 7 | 10 | 1 |
| SSH3 | failing LV | | 116 | | | | | | |
| | nonfailing LV | | 63 | | | | | | |
| | failing + nonfailing LV | | 16 | | | | | | |
| | | total: | 195 | 47 | 19 | 17 | 10 | 7 | 0 |

Number of genes that were selected from each SSH library to be included on the cDNA microarray. M: metabolism; G/PE: gene/protein expression; CS/M: cell structure/motility; CS/C: cell signalling/communication; C/OD: cell/organism defence; CD: cell division.

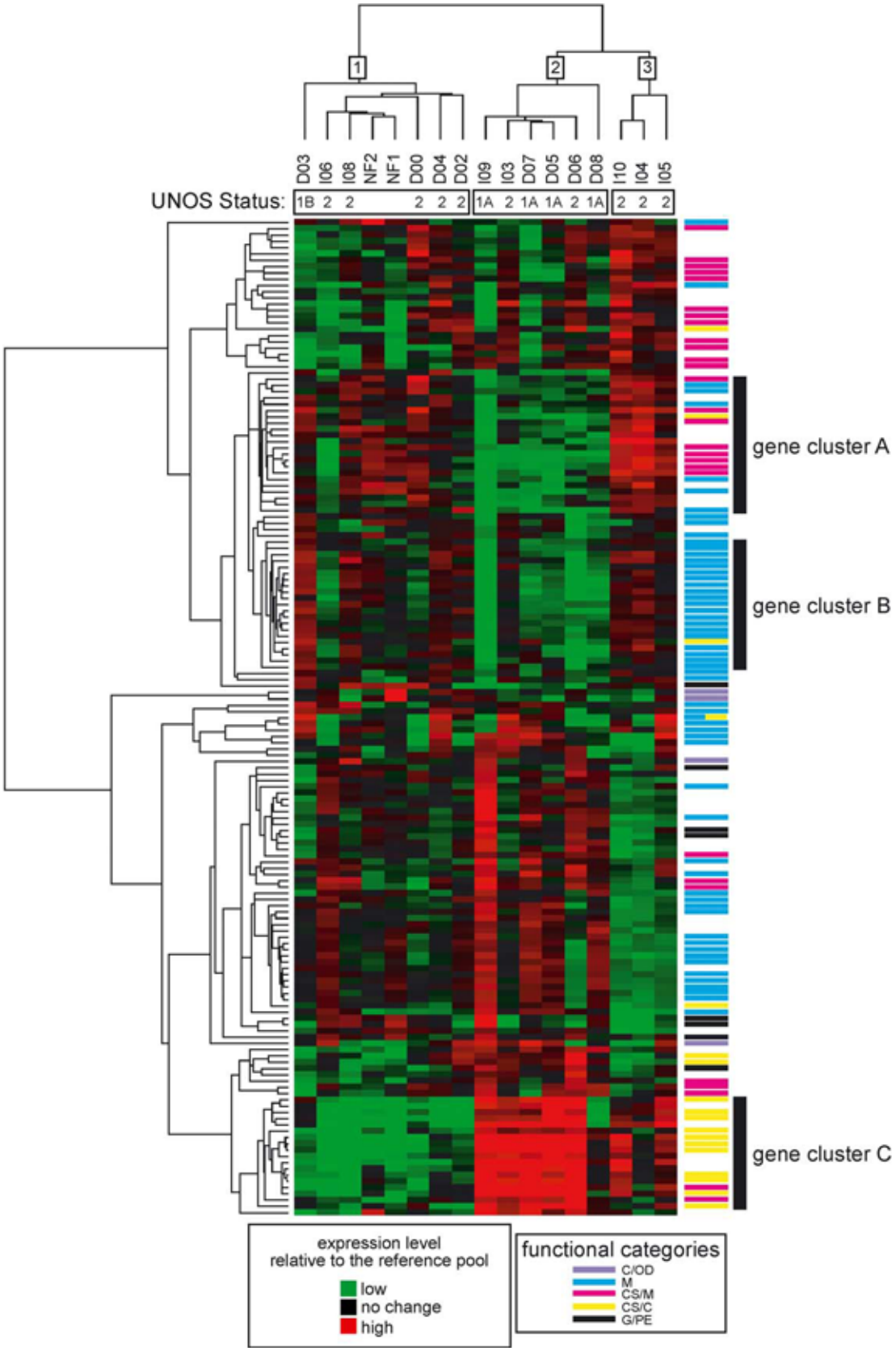



Fig. 1. Two-way hierarchical clustering of 17 patients (horizontally) and 159 statistically significant differential genes (vertically). Patients are classified into three subgroups indicated by "1," "2," and "3." Expression profiles of gene clusters A, B, and C correspond well to the classification of the patients into the three subgroups, and they are highlighted in Fig. 2. "undetermined": sequencing experiments were unsuccessful.

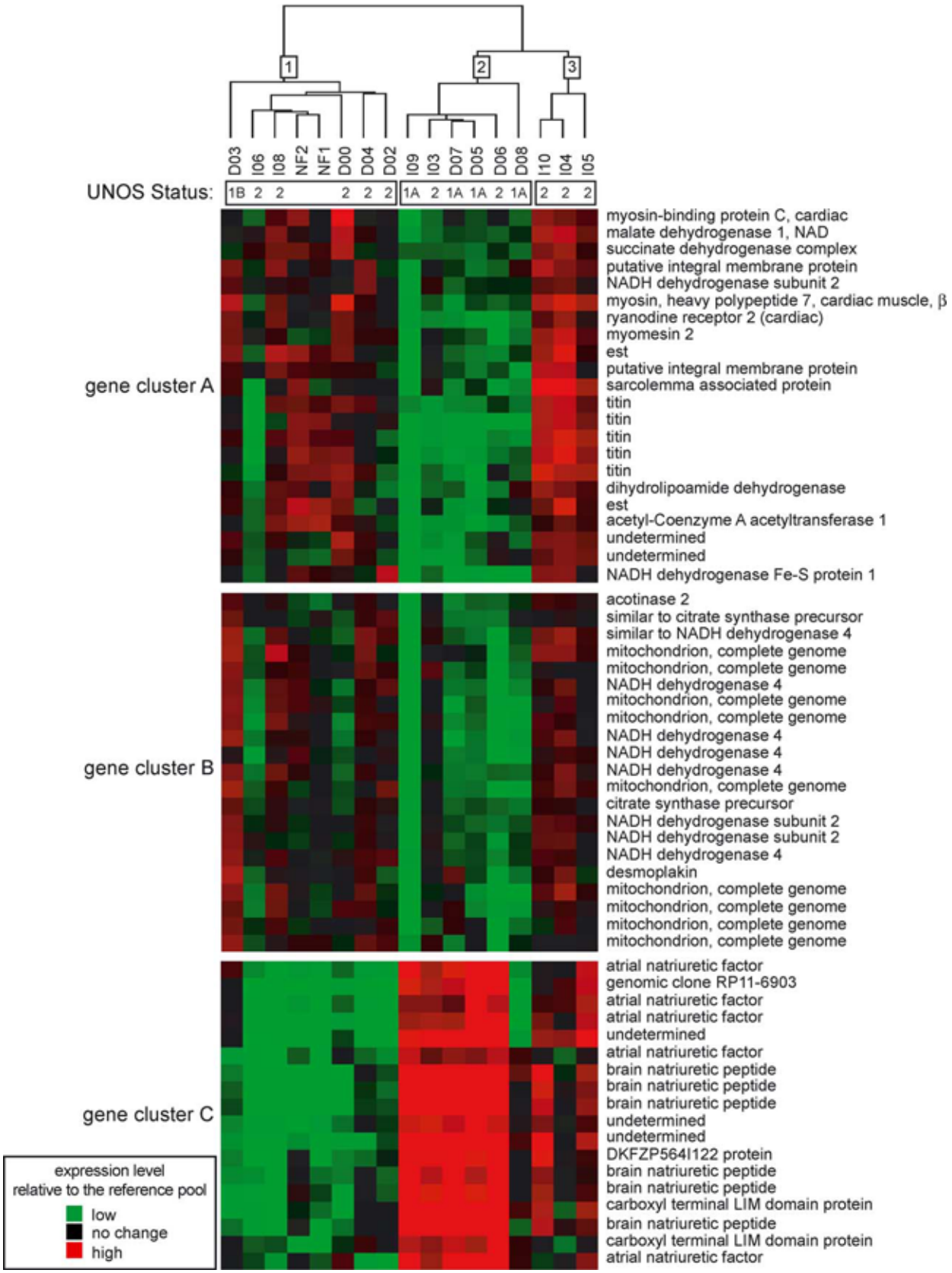



Fig. 2. Enlargement of gene clusters A, B, and C from Fig. 1.

probably reflects the cardiac situation in general. More than 40% of the genes were involved in metabolic pathways. The next largest functional categories were that of gene/protein expression and cell structure/motility, followed by cell signalling/communication and cell/organism defence.

### 3.2. Molecular portrait of failing hearts

Three independent microarray hybridizations were performed for each of the 17 patients. Of the 1034 clones present on the arrays, valid expression values for at least 2 of the triplicate hybridizations were obtained for 648 clones. Nonvalid expression values did not correlate with a specific labelling or hybridization experiment. Multiclass SAM analysis performed on these 648 clones compared their expression levels in each patient to their expression levels in all patients. In this way, we selected 159 clones displaying statistically significant differential expression among the 17 patients. These 159 clones represented at least 81 nonredundant genes.

We characterized expression profiles of 15 failing and 2 nonfailing hearts. To obtain gene expression ratios, all 17 hearts were compared to a common reference that was prepared from a pool of failing and nonfailing RV (Table 1). The use of a common reference sample allowed us to treat the gene expression ratios as measurements of the relative expression level of each gene across all our experimental samples. We compared the expression profiles of the 17 hearts and determined whether all profiles were similar or whether they differed. Fig. 1 shows the results of two-way hierarchical clustering of the 17 patients (including 2 nonfailing hearts) and the 159 cardiac clones that displayed significant differential expression between the patients. A complete figure, containing all gene names, and a clustering reproducibility analysis are available on www.ifr26.nantes.inserm.fr/Francais/Labos/U533/INDEX/index.htm.

Among the 17 patients, we clearly distinguished three subgroups, which are indicated by 1, 2, and 3 in the patient tree in Fig. 1. Within each subgroup, the expression profiles were comparable; therefore, each subgroup was characterized by one molecular portrait. The two nonfailing hearts (NF1 and NF2) clustered closely together within subgroup 1. Interestingly, cardiac expression profiles did not discriminate between IDCM- and CAD-affected patients. Therefore, we did not detect a molecular portrait specific for the IDCM or the CAD phenotype. This finding correlated well with our previous study in which we did not detect genes differentially expressed between failing hearts from IDCM- and CAD-affected patients [13]. However, when we annotated the heart failure patients according to their United Network for Organ Sharing (UNOS) Status [19], all patients with Status 1A clustered in subgroup 2 (Figs. 1 and 2). Fig. 2 highlights three gene clusters that are referred to in Fig. 1 by thick vertical black bars. The bottom cluster (cluster C) differentiates subgroup 1 from subgroups 2 and 3. Cluster C coincides mainly with a relatively low (subgroup 1) or high expression (subgroups 2 and 3) of heart failure markers ANF and BNP. The second branching into the subgroups 2 and 3 reflects differences in relative expression levels of a much larger number of genes. The top gene cluster (cluster A) consists mainly of genes involved in cell structure and motility functions, like β-myosin heavy chain (βMHC) and titin which were expressed at a higher level in subgroup 3 in comparison with subgroup 2. Gene cluster B consists mainly of mitochondrial metabolic genes, which were also expressed at a higher level in subgroup 3 in comparison with subgroup 2. One-way ANOVA analysis did not reveal a significant correlation between the patient subgroups and the following parameters: end-diastolic left ventricular diameter, mean pulmonary arterial pressure, ejection fraction, time between onset of heart failure and transplantation, and NYHA class.

### 3.3. Real-time PCR experiments

Ten genes (ANF, cytochrome *c* oxidase II, FHL2, MLCB, NADH4, titin, PDK4, RPL37A, Riken cDNA 2610318K02, and PEST-containing nuclear protein) were further analyzed by real-time PCR in five heart failure patients (D04, D05, I04, I06, and NF1) and in the reference sample. Except for the duplicate amplification of LV from patient D05 by the RPL37A oligos, none of the duplicate Ct values differed by more than 0.5. We calculated the relative gene expression ratios for each patient vs. the reference sample and compared the values to the microarray-obtained gene expression ratios (Fig. 3). An extensive table containing all ratios is available on www.ifr26.nantes.inserm.fr/Francais/Labos/U533/INDEX/index.htm. We conclude that overall, the data corresponded well. Seventy percent of the real-time PCR-obtained expression ratios differed by less than 0.5 from the microarray-obtained expression ratios.

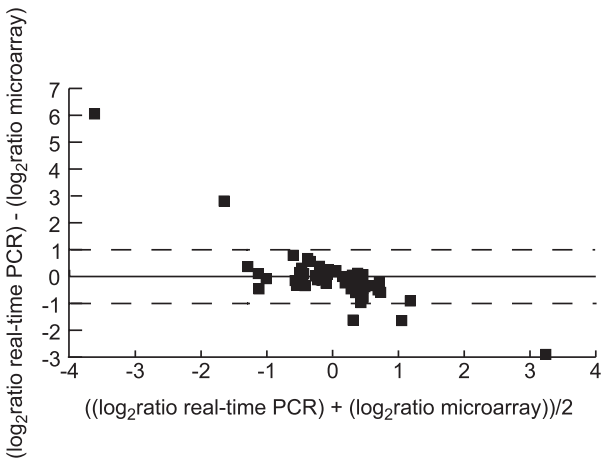



Fig. 3. Bland—Altman plot displaying the correlation between expression ratios obtained with real-time PCR and with microarray analysis. The solid line indicates identical $\log_2$ratios in both techniques. The upper dashed line indicates a twofold higher expression ratio with real-time PCR compared to microarray analysis. The lower dashed line indicates a twofold lower expression ratio with real-time PCR compared to microarray analysis.

The three most outlying values in the Bland–Altman plot in Fig. 3 correspond to the ANF gene. The obtained $\log_2$ratio values for this gene are much more extreme when using real-time PCR than when using microarray analysis. It is a known phenomenon that microarrays are limited in the detection of extreme up- or downregulation of genes.

## 4. Discussion

In this pilot feasibility study, we aimed to establish whether cardiac molecular portraits could be identified that delineate specific subgroups of heart failure patients. Three major subgroups were identified among 17 patients. We also identified groups of genes that clustered closely together and were responsible for the observed classification. We believe that key elements towards these results were the use of dedicated cDNA microarrays and the use of appropriate statistical analysis, together leading to a strict selection of 159 relevant genes that were used for hierarchical clustering. We want to stress that our results should be regarded as a proof of concept, comparable to the first publications of expression profiling in cancer research [20], and that our findings are of preliminary nature.

The first branching of the patient-tree coincided with a relatively low vs. high expression of ANF and BNP (Fig. 2), two genes that are considered as markers for heart failure [21]. BNP in particular has been proposed as an indicator of diagnosis [22] and even prognosis [23] in human heart failure. Interestingly, within our patient subgroup 1—which displayed a relatively low expression of BNP—all heart failure patients had been treated with conversion enzyme inhibitors and/or angiotensin II type 1 receptor antagonists (CEI/$AT_1$-antagonists). This was in contrast to patient subgroup 2, in which only 2/6 patients had received this treatment. Therefore, one might speculate that CEI/$AT_1$-antagonist treatment resulted in an improvement of cardiac function, which is reflected by a lower level of expression of ANF and BNP. Such an effect has already been attributed to this type of treatment [24]. Although ANF and BNP expression levels were responsible for part of the classification, analysis of ANF and BNP alone would not have resulted in the three patient subgroups presented here (data not shown). Genes in gene cluster A displayed a clear difference in relative expression level between patient subgroups 2 and 3. This gene cluster also contained known markers of heart failure, like the structural genes cardiac β-myosin heavy chain (βMHC) and titin. Mutations in βMHC have been associated with (familial) dilated cardiomyopathy [25]. In addition, upregulation of βMHC (coupled to downregulation of αMHC) has been demonstrated in failing human myocardium [26]. Downregulation of titin has been found in human failing heart [27,28], possibly leading to increased ventricular stiffness. We detected a relatively low level of titin in patient subgroup 2 and a relatively high level in patient subgroups 1 and 3. Therefore, a low level of titin corresponded with a high level of ANF and BNP expression. In addition, these unfavourable expression profiles coincided with the highest medical urgency status; all four patients with UNOS Status 1A were clustered in patient subgroup 2. Since all Status 1A patients received high-dose intravenous inotropic treatment, it is also possible that the observed expression profiles reflect an effect of treatment on gene expression. The two nonfailing hearts clustered closely together within patient subgroup 1. This correlates well with the fact that patient subgroup 1 is characterized by a relatively low level of ANF and BNP expression and a relatively high level of titin expression. The fact that the nonfailing hearts did not separate very far from failing hearts might be related to the still limited sample size. Recently, it has been shown that in ischemic human hearts, a titin isoform switch occurs resulting in more compliant myofibrils [29]. The cDNA clones representing titin on our microarrays were not from the region of the gene that is alternatively spliced. Among the 159 genes that were decisive for the molecular classification of the patients, we found many genes involved in energy metabolism (Fig. 2, gene cluster B). This fits well with the general view that mitochondrial dysfunction plays a pivotal role in heart failure. Gene expression changes in NADH dehydrogenase may reflect a certain degree of disturbance of the electron transport chain and therefore of the production of ATP. One might argue that the use of genome-wide microarrays would have revealed more and/or new genes important for the classification of heart failure patients. However, our goal was not to be extensive but to provide a proof of principle. The use of genome-wide microarrays would not have changed our central message which is that failing hearts can be identified by distinct molecular portraits. In addition, recently, a study was published in which the dedicated microarray strategy was directly compared with the genome-wide microarray approach. Based on a minimization of biological noise, the authors concluded in favour of the dedicated microarray strategy [30].

## 5. Study limitations

Some care should be taken while interpreting our data. Firstly, although we described expression profiling of the largest group of heart failure patients thus far, the size ($n = 15$) of our population was still limited. It is important to realize that our group of patients consisted of end-stage heart failure patients. The clinical interest of expression profiling does not lie within the classification of heart failure patients at the end-stage of their disease, but at an earlier stage where treatment strategies need to be improved. This is not within the scope of our study, but represents the direction in which we believe that expression profiling in heart failure should develop. Therefore, our study should be considered as a proof of concept. Even though all heart failure patients were at the end stage of their disease, their molecular portraits differed.

The fact that we identified different molecular portraits in end-stage failing hearts does not automatically imply that similar results will be obtained when analyzing hearts at an earlier stage of the disease. The second issue that should be taken into consideration is the retrospective aspect of our study, which limited us in the analysis of possible associations between clinical parameters and molecular portraits. Thirdly, in our study, all patient RNA samples were compared to a pool of IDCM/CAD/CF patient RNA samples. One could argue that a more appropriate strategy would be to compare patient RNA samples to so-called normal human cardiac tissue. However, we analyzed whether genes were higher, lower, or similarly expressed in *one patient* compared to *other patients*. We did not analyze whether genes were higher, lower, or similarly expressed in *one patient* compared to the *common reference*. Therefore, the use of normal hearts as a reference sample would not have changed our results. A similar approach has successfully been used in expression profiling of human cancer, where pools of cell lines or tumours have been used as a reference.

## 6. Conclusion

Our preliminary findings indicate that heart failure patients can be classified into subgroups based on their cardiac molecular portrait. The present pilot study provides a rationale for further exploration of molecular portraits on endocardial biopsies obtained from heart failure patients at more early stages of their disease. A multicenter investigation should enable the construction of a large-enough collection of biopsies for the prospective determination of prognostic and therapeutic significance of cardiac molecular portraits.

## Acknowledgements

The authors would like to thank the thoracic surgery and cardiology departments of Nantes University hospital for their participation. We thank Isabelle Guisle for spotting of the microarrays and Dr. Jean-Marie Heslan for assistance with the real-time PCR experiments.

This study was supported by special grants from the Aventis, the "Association Française contre les Myopathies," the "Institut National de la Santé et de la Recherche Médicale" (INSERM), the "Centre National de la Recherche Scientifique," and the "Region Pays de Loire." M.S. was recipient of a "Poste Vert" from INSERM.

## References

[1] van't Veer LJ, Dai H, van de Vijver MJ, et al. Gene expression profiling predicts clinical outcome of breast cancer. Nature 2002; 415:530–6.

[2] Dhanasekaran SM, Barrette TR, Ghosh D, et al. Delineation of prognostic biomarkers in prostate cancer. Nature 2001;412:822–6.

[3] Pomeroy SL, Tamayo P, Gaasenbeek M, et al. Prediction of central nervous system embryonal tumour outcome based on gene expression. Nature 2002;415:436–42.

[4] Hoshijima M, Chien KR. Mixed signals in heart failure: cancer rules. J Clin Invest 2002;109:849–55.

[5] Liew CC, Hwang DM, Fung YW, et al. A catalogue of genes in the cardiovascular system as identified by expressed sequence tags. Proc Natl Acad Sci U S A 1994;91:10645–9.

[6] Hwang DM, Dempsey AA, Wang R, et al. A genome-based resource for molecular cardiovascular medicine—toward a compendium of cardiovascular genes. Circulation 1997;96:4146–203.

[7] Hwang DM, Dempsey AA, Lee CY, et al. Identification of differentially expressed genes in cardiac hypertrophy by analysis of expressed sequence tags. Genomics 2000;66:1–14.

[8] Yang J, Moravec CS, Sussman MA, et al. Decreased SLIM1 expression and increased gelsolin expression in failing human hearts measured by high-density oligonucleotide arrays. Circulation 2000; 102:3046–52.

[9] Barrans JD, Stamatiou D, Liew C. Construction of a human cardiovascular cDNA microarray: portrait of the failing heart. Biochem Biophys Res Commun 2001;280:964–9.

[10] Hwang J-J, Allen PD, Tseng GC, et al. Microarray gene expression profiles in dilated and hypertrophic cardiomyopathic end-stage heart failure. Physiol Genomics 2002;10:31–44.

[11] Barrans JD, Allen PD, Stamatiou D, et al. Global gene expression profiling of end-stage dilated cardiomyopathy using a human cardiovascular-based cDNA microarray. Am J Pathol 2002;160:2035–43.

[12] Tan FL, Moravec CS, Li J, et al. The gene expression fingerprint of human heart failure. Proc Natl Acad Sci U S A 2002;99:11387–92.

[13] Steenman M, Chen Y-W, Le Cunff M, et al. Transcriptomal analysis of failing and non-failing human hearts. Physiol Genomics 2003; 12:97–112.

[14] Diatchenko L, Lau YC, Campbell AC, et al. Suppression subtractive hybridization: a method for generating differentially regulated or tissue-specific cDNA probes and libraries. Proc Natl Acad Sci U S A 1996;93:6025–30.

[15] Rouger K, Le Cunff M, Steenman M, et al. Global/temporal gene expression in diaphragm and hindlimb muscles of dystrophin-deficient (mdx) mice. Am J Physiol, Cell Physiol 2002;283:C773–84.

[16] Tseng GC, Oh M-K, Rohlin L, et al. Issues in cDNA microarray analysis: quality filtering, channel normalization, models of variations and assessment of gene effects. Nucleic Acids Res 2001;29:2549–57.

[17] Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. Proc Natl Acad Sci U S A 2001;98:5116–21.

[18] Eisen MB, Spellman PT, Brown PO, et al. Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci U S A 1998;95:14863–8.

[19] Renlund DG, Taylor DO, Kfoury AG, et al. New UNOS rules: historical background and implications for transplantation management. United Network for Organ Sharing. J Heart Lung Transplant 1999; 18:1065–70.

[20] Perou CM, Jeffrey SS, Van de RM, et al. Distinctive gene expression patterns in human mammary epithelial cells and breast cancers. Proc Natl Acad Sci U S A 1999;96:9212–7.

[21] Kjaer A, Hesse B. Heart failure and neuroendocrine activation: diagnostic, prognostic and therapeutic perspectives. Clin Physiol 2001; 21:661–72.

[22] Dao Q, Krishnaswamy P, Kazanegra R, et al. Utility of B-type natriuretic peptide in the diagnosis of congestive heart failure in an urgent-care setting. J Am Coll Cardiol 2001;37:379–85.

[23] Koglin J, Pehlivanli S, Schwaiblmair M, et al. Role of brain natriuretic peptide in risk stratification of patients with congestive heart failure. J Am Coll Cardiol 2001;38:1934–41.

[24] Yoshimura M, Mizuno Y, Nakayama M, et al. B-type natriuretic peptide as a marker of the effects of enalapril in patients with heart failure. Am J Med 2002;112:716–20.

[25] Kamisago M, Sharma SD, DePalma SR, et al. Mutations in sarcomere protein genes as a cause of dilated cardiomyopathy. N Engl J Med 2000;343:1688–96.
[26] Miyata S, Minobe W, Bristow MR, et al. Myosin heavy chain isoform expression in the failing and nonfailing human heart. Circ Res 2000;86:386–90.
[27] Morano I, Hadicke K, Grom S, et al. Titin, myosin light chains and C-protein in the developing and failing human heart. J Mol Cell Cardiol 1994;26:361–8.
[28] Hein S, Scholz D, Fujitani N, et al. Altered expression of titin and contractile proteins in failing human myocardium. J Mol Cell Cardiol 1994;26:1291–306.
[29] Neagoe C, Kulke M, del Monte F, et al. Titin isoform switch in ischemic human heart disease. Circulation 2002;106:1333–41.
[30] Seta KA, Millhorn DE. Functional genomics approach to hypoxia signaling. J Appl Physiol 2004;96:765–73.

## IV. ETUDE 4

Lamirault G, Le Meur N, Chevalier C, Le Cunff MF, Guisle I, Bihouée A, Teusan R, Roussel JC, Trochu JN, Léger JJ, Houlgatte R, Steenman M. Clinical Deterioration of Heart Failure Patients is Associated with a Molecular Deterioration Expression Profile. Article soumis pour publication.

L'objectif de ce travail était de déterminer un profil d'expression génique permettant de discriminer des patients en insuffisance cardiaque avec différents niveaux de sévérité clinique.

L'étude a porté sur une cohorte de 44 patients en attente d'une transplantation cardiaque. Les patients ont pu être classés en 3 groupes reflétant des niveaux de sévérité clinique distincts au moment de la transplantation. Le score utilisé pour classer les patients associait le « United Network for Organ Sharing medical urgency status » et l'occurrence des hospitalisations pour décompensation cardiaque dans les 3 mois précédant la greffe.

Deux échantillons tissulaires de chaque ventricule ont pu être obtenus pour chaque patient, soit un total de 176 échantillons ventriculaires analysés. Une puce (Myochip) contenant un ensemble de 4217 sondes pertinentes pour l'étude du transcriptome cardiaque a été utilisé pour analyser individuellement chaque échantillon.

Des profils d'expression génique associés au niveau de sévérité clinique des malades ont été identifiés au sein des transcriptomes ventriculaires droits (129 gènes) et gauches (170 gènes) des patients. La prédiction de la sévérité clinique des patients à partir de ces profils d'expression génique s'est révélée performante tant pour le ventricule droit que le ventricule gauche. Une bonne corrélation entre les prédictions obtenues pour les ventricules droits et gauches était observée. Les prédictions étaient également reproductibles entre les 2 échantillons tissulaires obtenus du même ventricule pour le même patient.

Ces résultats montrent que l'utilisation des profils d'expression génique pourrait être appliquée à l'évaluation des patients en insuffisance cardiaque avancée et suggèrent un évaluation de cette approche à des stades plus précoce de la maladie.

.

**Clinical Deterioration of Heart Failure Patients is Associated with a Molecular Deterioration Expression Profile**

Guillaume Lamirault, MD[1,2]; Nolwenn Le Meur, PhD[1]; Catherine Chevalier, PhD[3]; Marie-France Le Cunff[1]; Isabelle Guisle[3]; Audrey Bihouée, MSc[3]; Raluca Teusan, MSc[3]; Jean-Christian Roussel, MD[4]; Jean-Noël Trochu, MD, PhD[2]; Jean J. Léger, PhD[1,3]; Rémi Houlgatte, PhD[1,3]; and Marja Steenman, PhD[1]

[1] L'institut du thorax – INSERM U533, Faculté de Médecine, 1, rue Gaston Veil, 44035 Nantes, France. [2] L'institut du thorax - Cardiology Department, Hôpital G&R Laënnec, Nantes, France. [3] West Genopole Transcriptome Facility, Nantes, France. [4] L'institut du thorax - Cardiothoracic and Vascular Surgery Department, Hôpital G&R Laënnec, Nantes, France.

Corresponding author: Dr Guillaume Lamirault

L'Institut du Thorax - INSERM U533, Faculté de Médecine

1, rue Gaston Veil, 44035 Nantes, France

Guillaume.Lamirault@univ-nantes.fr

Tel: +33240412958

Fax: +33240412950

Risk stratification in advanced heart failure (HF) aims at identifying patients who will rapidly progress to refractory myocardial dysfunction or who are at high risk of sudden cardiac death. This stratification is crucial for the individualization of therapeutic strategy, in particular for the listing and prioritization of patients for heart transplantation. Decrease of ventricular function, impairment of functional capacity, and activation of neurohumoral systems are markers of cardiac remodeling progression used in clinical practice to predict outcomes for advanced HF patients. Prediction models combining independent markers have been developed to integrate the distinct components of the remodeling process.[1,2] However, these models have modest predictive capacity, and for the individual patient, outcome prediction still remains to be improved in advanced HF.

It has already been shown that HF severity correlates with the intensity of the cardiac remodeling process occurring during HF progression.[3] This remodeling process is related to transcriptomal alterations affecting numerous molecular pathways and biological functions, modifying tissue and morphological characteristics of the myocardium.[4]One tool that might lead to better outcome prediction is gene expression profiling. This method has successfully predicted patient outcomes in several human diseases (mainly cancer). It may also be useful for the analysis of the cardiac remodeling process and - ultimately - for the development of new predictors in human HF. We and others recently showed that gene expression profiling could distinguish, even in advanced HF, sub-groups of patients with specific cardiac molecular portraits.[5-10]

Here we demonstrate that clinical deterioration of HF patients is associated with a molecular deterioration expression profile.

## METHODS

### Cardiac samples

Cardiac tissue was obtained from explanted hearts from 44 patients with advanced HF who underwent a cardiac transplantation or a total artificial heart placement at the Nantes University Hospital between 1998 and 2002. Pre-transplant evaluation was completed in our institute to validate the diagnosis, etiology and severity of the disease. Coronary artery angiography, cardiac catheterization, and echocardiography were systematically performed. At the time of transplantation, patient evaluation included physical examination, laboratory tests, and echocardiography. Macroscopic and histological examination of explanted hearts confirmed the previously diagnosed etiology for all patients. Extensive individual clinical information was obtained before transcriptome analysis and can be found in the on-line data supplement Table 1.

Patients were classified into 3 severity groups based on their clinical status at the time of transplantation. The clinical status of each patient was defined based on the United Network for Organ Sharing (UNOS) medical urgency status[11] and occurrence of hospitalizations for Acute Decompensated Heart Failure (ADHF) during the 3 months prior to the surgical procedure (defined as recent ADHF). Deteriorating patients were characterized by UNOS 1A status. Stable patients were defined as UNOS 2 patients with no recent ADHF. The remaining patients were classified as Intermediate.

For each of the 44 explanted hearts, two spatially distinct transmural samples were obtained from both left ventricle (LV) and right ventricle (RV) immediately after cardiac explantation, leading to a total of 176 distinct tissue samples. The samples were taken from non-infarcted zones of the ventricular free walls, snap-frozen in

liquid nitrogen, and stored at –80°C. Myocardial tissue obtained from cardiac transplantation procedures is considered discarded tissue. Therefore tissue collection was performed without written informed consent.

**Microarrays**

Microarrays were prepared in-house using human-specific 50-mer oligonucleotide probes (MWG Biotech®). The probes were spotted onto epoxy-silane coated glass slides using the Lucidea Array Spotter (Amersham®). The 4217 human genes that were represented on the microarray had been selected for involvement in cardiovascular and/or skeletal muscle normal and pathological functioning. Selection was based on 1) subtractive hybridization experiments,[12,13] 2) genome-wide microarray hybridizations,[14] 3) literature data. Each probe was spotted in quadruplicate. For more information see:

http://cardioserve.nantes.inserm.fr/ptf-puce/myochips_en.php

**Study design**

Each of the 176 biological samples was compared to a common reference sample consisting of a pool of equal quantities of mRNA from all 176 biological samples. This complex mRNA pool was used as a standard or a common point of measurement that enabled a comparison between the individual mRNA samples.

Control of technical and biological noises inherent to microarray experiments was incorporated in the study design. To account for the technical fluctuation of the expression measurements, eight technical replicate values were obtained for each cardiac sample. To determine the fluctuation due to tissue sampling, two biological replicate samples were obtained from each ventricle. Furthermore, to control for

experimenter bias in sample processing, samples were randomly distributed among the three experimenters in charge of all manipulations, at each step of the experiment.

**RNA isolation, labeling, and hybridization**

Total RNA was isolated using TRIZOL® reagent (Life Technologies). mRNA was isolated using the Oligotex mRNA kit (Qiagen). RNA and mRNA quality was assessed using an Agilent 2100 bioanalyzer. Cy3- and Cy5-labeled cDNA was prepared using the CyScribe cDNA Post Labeling Kit (Amersham Pharmacia Biotech). Each individual mRNA sample was Cy3-labeled and mixed with an equal amount of Cy5-labeled reference sample. The mixture was pre-incubated with human Cot-I DNA (Gibco-BRL), yeast tRNA, and polyA RNA, and hybridized onto duplicate microarrays.

**Raw data extraction and consolidation**

Hybridized arrays were scanned by fluorescence confocal microscopy (Scanarray 4000XL, GSI-Lumonics). Fluorescence signal measurements were obtained separately for each fluorochrome at 10 µm/pixel resolution. Hybridization and background signal intensities, and quality control parameters were measured using GenePix Pro 5.0 (Axon®). For raw expression datasets see: http://cardioserve.nantes.inserm.fr/HF_profiling.

A Lowess normalization procedure was performed to correct for technical biases.[15] The procedure was applied channel-by-channel as described previously.[16] For each microarray, Cy3- and Cy5-signal intensities were separately normalized to a prototype defined as the median profile of all Cy3- or Cy5-signal intensities. Genes of

which all signal intensities were below the background level were filtered out. For each microarray, expression values were calculated as $\log_2$(Cy3/Cy5). For each biological sample, expression values were then consolidated as the median of the eight technical replicate values.

**Data analysis**

Unsupervised hierarchical clustering was applied to the entire data set median-centered on genes, using the Pearson correlation as a similarity metric, and average linkage clustering. Results were displayed using TreeView.[17] Gene clusters were selected using 10 and 0.6 as minimal gene number and minimal correlation respectively. GOMiner was used to identify functional categories that were over- or underrepresented in specific clusters compared to the list of all analyzed genes.[18]

**‘Predictors’ and 'molecular severity score’ values**

LV- and RV-specific data were separated into distinct datasets and analyzed separately using an identical strategy:

The ‘Predictor’ was defined as a list of genes differentially expressed between Stable and Deteriorating patient groups. These genes were identified using ‘Significance Analysis of Microarrays’ (SAM).[19] With this method, each gene is first assigned a score on the basis of a modified t-test and genes with scores greater than a user-defined threshold are selected. Repeated random sample permutations are used to estimate the percentage of genes identified by chance (false discovery rate) among the selected genes. For each analysis we arbitrarily fixed the threshold of statistical significance so that the false-discovery rate < 1% (corresponding to less than one falsely-discovered gene out of 100 selected genes).

LV and RV 'Predictors' and an LV-RV predictor were used to calculate a transcriptome-based 'molecular severity score' (MSS) for each sample. The LV-RV predictor was a combination of the genes of the LV and the RV predictor. First, expression profiles were mean-centered and standard deviation-scaled on genes. The mean profile was calculated for Stable (Ms) and for Deteriorating (Md) samples. The molecular severity score (MSS) of a specific sample was defined as the normalized Euclidean distance (ranging from 0 to 1) between the sample and the stable mean profile and was calculated as described below:

$MSS = \frac{Es}{Es + Ed}$ where $Es = \sum_{i=1}^{n} [X_i - Ms_i]^2$ and $Ed = \sum_{i=1}^{n} [X_i - Md_i]^2$ and X = the expression profile of the specific analyzed sample and i = an index of the n genes included in the 'predictors'.

To define the significance level of the obtained MSS, an unpredictable interval in-between the Stable and Deteriorating profiles was calculated. Using $10^4$ random permutations of the expression profiles, we generated a set of $10^4$ MSS. $2.5^{th}$ and $97.5^{th}$ percentiles of the random-MSS distribution were defined as the cut-off for the unpredictable interval.

Leave-one-out cross validation was performed on Stable and Deteriorating samples. One hundred distinct data sets were produced. Each data set was partitioned into a test set consisting of one sample and a learning set consisting of the 99 others samples. The learning set was used to calculate an LV-RV-MSS using the strategy described. The obtained MSS was employed to predict the MSS value of the test sample. This process was repeated so that the MSS value of each sample was predicted using an MSS estimated from all other samples in the data set.

To test the diagnostic power of our classification, we calculated the number of samples classified in concordance with the clinical classification in 3 distinct sets.

The training set was defined as MSS values obtained for LV samples using the LV predictor and MSS values obtained for RV samples using the RV predictor. The test set consisted of MSS values obtained for RV samples using the LV predictor and MSS values obtained for LV samples using the RV predictor. The cross validation set consisted of MSS values obtained using the leave-one-out method for both LV and RV samples. Sensitivity and specificity of the molecular prediction of Stable status and Deteriorating status were calculated in the test set.

**LV-RV comparison**

To test for between-chamber fluctuations of our molecular prediction, we compared MSS values obtained from the LV and RV samples from the same patient. The correlation coefficient was used as a fluctuation index.

**Reproducibility**

We tested between-sample reproducibility of the MSS values of all biological duplicates. Expression data from biological duplicates were separated to generate 2 comparable data sets. MSS from the duplicate sets were compared using the correlation coefficient. To determine the chamber-specific reproducibility, analyses were performed separately on the LV-specific and RV-specific data sets.

**Drug-related bias**

We tested whether between-group variations in drug treatment could have biased the Predictor discovery. To avoid confounding factors, subgroups of samples from the same chamber and the same severity group were analyzed separately. For each Predictor, mean expression profiles of samples positive and negative for a specific

drug treatment were gene-by-gene compared using a student t-test. Genes with p-value < $10^{-2}$ were considered as significantly influenced by the tested drug treatment.

**Statement of responsibility**

All authors had full access to the data and take responsibility for its integrity. All authors have read and agree to the manuscript as written.

## RESULTS

We profiled cardiac gene expression in a cohort of 44 advanced-HF patients using a 4217-oligonucleotide microarray containing genes selected for their involvement in cardiac (patho)physiology. Based on the analysis of clinical information the 44 patients were classified into three HF-severity groups: Deteriorating (n=12), Intermediate (n=19) and Stable (n=13). After raw data extraction and consolidation, 4035 genes (96% of the total) were validated for further analysis.

### Unsupervised clustering and functional annotation

The 176 cardiac samples and the 4035 selected genes were clustered according to their expression profiles using an unsupervised hierarchical clustering procedure (Figure 1). Samples were grouped in 2 major clusters mainly based on the expression profile of a 387-gene cluster (white bar). This patient molecular clustering was not correlated with the clinical severity classification. However, within each of the 2 major clusters, Stable and Deteriorating samples were preferentially classified into distinct sub-clusters ($p<0.001$ within each major cluster, $\chi^2$ test).

Selected gene clusters were identified by automated analysis of the gene classification. Functional annotation revealed enrichment of genes involved in a specific biological process or tissue-type for most of the clusters. Clusters that were too small to obtain a statistically significant annotation using Gominer software (annotations ‘natriuretic peptides’ and ‘cell metabolism’) were functionally annotated based on literature analysis.

Several of the clusters showed a marked differential expression between Stable and Deteriorating samples for left and/or right ventricle samples. ‘Cell metabolism’ (brown bar), ‘natriuretic peptides’ (red bar), and ‘extracellular matrix’ (grey bar) gene clusters

displayed higher expression for Deteriorating samples than for Stable samples in both left and right ventricles. ‘Cytoskeleton’ (pink bar) and ‘cell death’ (light blue bar) gene clusters displayed higher expression for Stable samples than for Deteriorating samples in both left and right ventricles. Interestingly, the ‘mitochondrion’ (yellow bar) gene cluster displayed higher expression for Stable samples than for Deteriorating samples in right but not left ventricle.

**Prediction of clinical status**

Two-class statistical analysis of gene expression profiles of 12 clinically deteriorating patients and 13 clinically stable patients resulted in the identification of 170 and 129 differentially expressed genes (i.e. predictors) for LV and RV samples respectively. Sixty-six genes were present in both LV and RV predictors. A 233-gene LV-RV predictor was also created by combining the genes of the LV and RV predictors. MSS values of patients were calculated based on their individual expression profile for these 3 predictors, as described in the Methods section.

Figure 2 shows MSS values calculated for the Stable and Deteriorating groups based on the different predictors. In the training set, 95 out of 100 samples were predicted in concordance with the clinical classification, whereas one Stable sample was predicted as Deteriorating and 4 samples were in the unpredictable interval. The samples incorrectly predicted were all LV samples, whereas all RV samples were correctly predicted.

To account for data over-fitting due to classification of samples used to define the predictors, a test set consisting of samples not used in the definition of the predictors was analyzed. In this test set, 89 out of 100 samples were predicted in concordance with the clinical classification, whereas the remaining 11 samples were classified as

unpredictable. No sample was misclassified. For prediction of the Stable status sensitivity was 90% and specificity was 100%, whereas for prediction of the Deteriorating status sensitivity was 87% and specificity was 100%. The proportion of samples predicted in concordance with the clinical classification was not different for LV (44/50) and RV (45/50) samples ($p=1.00$, $\chi^2$ test).

A cross-validation strategy was also employed to test the prediction power of the LV-RV-predictor. Eighty-three samples were predicted in concordance with the clinical classification. One Stable sample was predicted as deteriorating and 16 samples were in the unpredictable interval. The proportion of samples predicted in concordance with the clinical classification was not different for LV (44/50) and RV (40/50) samples ($p=0.41$, $\chi^2$ test).

**Intermediate group analysis**

In agreement with the clinical classification, patients of the Intermediate group - who were not used for the construction of the predictors - were on average classified in-between the two other groups using the LV-RV predictor (Figure 3). Progression of clinical severity for the LV samples was associated with a gradual increase of the MSS mean values from 0.32 for the Stable group to 0.51 and 0.68 for the Intermediate and the Deteriorating group respectively ($p<0.001$ for overall and all pairwise comparisons, one-way analysis of variance on ranks followed by Dunn test). Similar results were obtained for RV samples. However, we observed that 66% of the samples in the Intermediate group exhibited MSS values out of the unpredictable interval and could have been predicted as either Stable or Deteriorating.

## Drug-related effects

Clinical characteristics of the three patient groups are summarized in Table 1. The groups were comparable regarding HF prognostic predictors like sex, age, HF etiology, and LV ejection fraction. As expected, differences in severity levels were associated with significant inter-group variations regarding treatment with adrenergic agonists, phospho-diesterase inhibitors, beta-blockers, and angiotensin converting enzyme inhibitors/angiotensin receptor blockers. Among the 233 selected genes in the LV-RV predictor, significant differences in expression related to medication were found for 3 to 11 genes. Removing these genes from the list of genes included in the predictors did not modify the number of samples predicted in concordance with the clinical classification (data not shown).

## LV-RV comparison and biological reproducibility

Figure 4A shows a comparison of patient MSS values obtained from LV and RV data using the LV-RV predictor. A significant correlation between LV and RV MSS values was observed irrespective of the sample severity group. We also aimed to test whether our classification was reproducible across biological replicates (distinct samples taken from the same chamber in the same patient). A significant correlation between MSS values obtained for the duplicate sets was observed (Figure 4B). When comparing reproducibility among LV and RV samples, a better correlation was observed for RV samples than for LV samples.

## DISCUSSION

We produced and analyzed the largest set to date of transcriptomal profiles of LV and RV samples from a cohort of 44 HF patients. Ventricular samples were analyzed using a dedicated microarray representing genes selected for their contribution to cardiovascular (patho)physiology. Replication at both the biological and the technical level, and control of experimental variations at the different steps of the study allowed detection of even subtle expression changes. We identified a set of genes of which expression changes discriminated between patients with different clinical severity levels and established that clinical deterioration of HF patients was associated with a molecular deterioration expression profile in both LV and RV. Therefore, our study confirms the potential of cardiac gene expression profiling to predict outcomes in human HF.

### Related findings in previous studies

It has previously been shown that gene expression profiling can discriminate between cardiac patients with different clinical characteristics.[20-25] Despite a common final phenotype, etiology-related gene expression profiles have been identified for HF patients affected by Chagas disease, and hypertrophic and dilated cardiomyopathies.[26] A gene expression profile distinguishing between ischemic and non-ischemic cardiomyopathies has also been identified.[27] More recently a gene expression profile associated with inflammatory viral cardiomyopathy has been identified in patients with mild or moderate LV dysfunction.[28] These findings offer valuable information regarding the molecular basis of HF related to distinct etiologies and they could lead to individualized therapeutic strategies in HF.

Other clinical characteristics such as age and sex have also been shown to have an effect on the transcriptomal profile of HF patients.[29] In our study, the molecular severity markers correctly classified HF patients independent of etiology or age (data not shown). Because most of our patients were male, we could not validate our classification in female HF patients.

**Potential clinical significance of findings**

The evaluation of the clinical profile of patients with advanced HF is crucial to detect those who will benefit from cardiac transplantation.[30] Recent developments of alternative therapies in advanced HF[31,32] have improved the non-transplanted patients' life expectancy and quality of life, and may lead to restricted indications for transplantation. Among the patients in our Stable group, some would probably not have been listed for transplantation if evaluated today. Thus, the molecular portrait of these patients in the Stable group may represent a molecular portrait of patients "too well for transplantation". The results of our study suggest that gene expression profiling has the potential to detect HF patients at highest risk of death with high sensitivity and specificity. The use of gene expression profiling could therefore not only improve the selection of candidates for cardiac transplantation but could also improve the prioritization of patients among those listed for transplantation and lead to a better allocation of the inadequate number of available donor hearts.

Transcriptome profiling-based approaches may also have an impact on the clinical management of HF patients at earlier stages of the disease. Usually, HF is diagnosed in patients with no or moderate symptoms. At these stages, numerous therapeutic strategies are available and various disease progressions are observed. Routinely used clinical predictors reflect one but not all aspects of disease

progression in the failing heart.[33] A transcriptome-based approach might better reflect HF deterioration of individual patients, improve prognostic evaluation, and guide therapeutic strategy. Therefore, endomyocardial biopsies (taken preferentially from the RV) may represent interesting samples for global HF severity assessment.

RV transcriptome remodeling in HF has been evaluated to a lesser extend than LV transcriptome. Most of the patients with advanced HF have severe LV dysfunction, whereas RV dysfunction intensity is variable among these patients. It has not been clear whether molecular prediction from RV samples can reflect the clinical status of advanced HF patients. Our results show that most of the molecular processes disturbed in LV are also disturbed in RV. Furthermore, despite the fact that the overlap of genes included in the RV and LV predictors was small, LV predictor could be applied to RV expression profiles to predict clinical status and RV predictor could be applied to LV samples to predict clinical status. These results suggest that molecular prediction using samples taken from RV may have the same prediction power as samples taken from LV. Endomyocardial biopsies obtained from RV are often taken from the intraventricular septum. Predictive power of cardiac samples taken from this zone remains to be evaluated in further studies.

Measurement reproducibility is another crucial point when developing a predictor of patient HF severity. Relatively high variability of widely used biomarkers like B-type Natriuretic peptide (BNP) or Nterminal-proBNP blood levels may be a problem for patient management.[34,35] Our results show that gene repression profiling is reproducible among biological replicates. Reproducibility was higher for RV samples, reinforcing the interest of RV samples utilization to develop a molecular predictor in advanced HF. A hypothesis is that regional tissue heterogeneity may be higher in LV than in RV. One cause may be the presence of infarct scars that preferentially affect

the LV. However, ventricular samples analyzed in this study were obtained after careful dissection of the ventricles excluding infarct scars. We also did not observe a higher variability of MSS values obtained for LV samples in patients affected by coronary artery disease compared to others patients.

**Potential limitations**

Therapeutic interventions, in particular medications, may induce modifications of the cardiac transcriptome.[36] We tested the hypothesis that patient classification may be modified by angiotensin converting enzyme inhibitors/angiotensin receptor blockers, beta-blockers, and inotropic drugs. A very low number of genes included in the distinct predictors were associated with differential expression between patients taking these medications. Removing these genes from the predictors did not modify the patients' classification. Therefore, in advanced HF, medications do not strongly modify the expression level of our predictors.

We compared our molecular predictors to a 2-parameter classification that has not been previously evaluated in advanced HF. Because we used samples taken at the time of cardiac transplantation, it was not possible to compare our predictors to a relevant clinical end-point like mortality or hospitalization for decompensated HF. We decided to combine two established predictors of HF severity to classify patients. The UNOS medical urgency status has been specifically developed for advanced HF patients listed for cardiac transplantation. The UNOS 1A status at the time of listing is associated with a 1 month-mortality > 30% whereas UNOS 2 patients have a 1 month-mortality <10%.[37] The death rate on the UNOS waiting list has been calculated as more than 4 fold higher for UNOS 1A than UNOS 2 patients.[38] To better define our group of Stable patients we combined the UNOS medical urgency

status with the occurrence of ADHF episodes. Frequent rehospitalizations have been recognized as a strong predictor of HF patient mortality.[39] Other HF severity prediction scores have been developed in advanced HF.[37,40] Comparison of one of these HF severity predictors to the UNOS medical urgency status did not reveal a higher predictive power.[37] This predictor included the measurement of a parameter that could not be monitored day-to-day (peak oxygen consumption).[40] Since we obtained the cardiac samples at the time of transplantation, we hypothesized that the use of parameters measured at the time of transplantation would better reflect the clinical phenotype at this time.

We analyzed expression profiles of patients with advanced HF at the time of cardiac transplantation. Further studies are needed to determine whether gene expression profiling of cardiac tissue provides prognostic information for patients at earlier stages of HF. A comparison with clinical end-points like mortality or hospitalization for cardiac failure will be possible in these studies.

**Acknowledgments**

The authors thank the thoracic surgery and cardiology departments of the Nantes University Hospital for their participation. Funding was provided by the "Institut National de la Santé et de la Recherche Médicale" (INSERM), the "Centre National de la Recherche Scientifique" (CNRS), “Ouest Génopole”, the “Association Française contre les Myopathies” (AFM), and the “Region Pays de la Loire”. GL was supported by the “Fondation pour la Recherche Médicale”.

**Conflict of Interest Disclosures**

The authors declare no conflict of interest.

## Reference List


1. Aaronson KD, Schwartz JS, Chen TM et al. Development and prospective validation of a clinical index to predict survival in ambulatory patients referred for cardiac transplant evaluation. *Circulation*. 1997;95:2660-2667.

2. Smits JM, Deng MC, Hummel M et al. A prognostic model for predicting waiting-list mortality for a total national cohort of adult heart-transplant candidates. *Transplantation*. 2003;76:1185-1189.

3. Francis GS. Pathophysiology of chronic heart failure. *Am J Med*. 2001;110 Suppl 7A:37S-46S.

4. Swynghedauw B. Molecular mechanisms of myocardial remodeling. *Physiol Rev*. 1999;79:215-262.

5. Tan FL, Moravec CS, Li J et al. The gene expression fingerprint of human heart failure. *Proc Natl Acad Sci U S A*. 2002;99:11387-11392.

6. Blaxall BC, Tschannen-Moran BM, Milano CA et al. Differential gene expression and genomic patient stratification following left ventricular assist device support. *J Am Coll Cardiol*. 2003;41:1096-1106.

7. Kaynak B, von Heydebreck A, Mebus S et al. Genome-wide array analysis of normal and malformed human hearts. *Circulation*. 2003;107:2467-2474.

8. Liew CC, Dzau VJ. Molecular genetics and genomics of heart failure. *Nat Rev Genet*. 2004;5:811-825.

9. Kittleson MM, Ye SQ, Irizarry RA et al. Identification of a gene expression profile that differentiates between ischemic and nonischemic cardiomyopathy. *Circulation*. 2004;110:3444-3451.

10. Steenman M, Lamirault G, Le Meur N et al. Distinct molecular portraits of human failing hearts identified by dedicated cDNA microarrays. *Eur J Heart Fail*. 2005;7:157-165.

11. Renlund DG, Taylor DO, Kfoury AG et al. New UNOS rules: historical background and implications for transplantation management. United Network for Organ Sharing. *J Heart Lung Transplant*. 1999;18:1065-1070.

12. Rouger K, Le Cunff M, Steenman M et al. Global/temporal gene expression in diaphragm and hindlimb muscles of dystrophin-deficient (mdx) mice. *Am J Physiol Cell Physiol*. 2002;283:C773-C784.

13. Steenman M, Lamirault G, Le Meur N et al. Distinct molecular portraits of human failing hearts identified by dedicated cDNA microarrays. *Eur J Heart Fail*. 2005;7:157-165.

14. Steenman M, Chen YW, Le Cunff M et al. Transcriptomal analysis of failing and nonfailing human hearts. *Physiol Genomics*. 2003;12:97-112.

15. Yang YH, Dudoit S, Luu P et al. Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Res*. 2002;30:e15.

16. Workman C, Jensen LJ, Jarmer H et al. A new non-linear normalization method for reducing variability in DNA microarray experiments. *Genome Biol*. 2002;3:research0048.

17. Eisen MB, Spellman PT, Brown PO et al. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci U S A*. 1998;95:14863-14868.

18. Zeeberg BR, Feng W, Wang G et al. GoMiner: a resource for biological interpretation of genomic and proteomic data. *Genome Biol*. 2003;4:R28.

19. Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci U S A*. 2001;98:5116-5121.

20. Tan FL, Moravec CS, Li J et al. The gene expression fingerprint of human heart failure. *Proc Natl Acad Sci U S A*. 2002;99:11387-11392.

21. Blaxall BC, Tschannen-Moran BM, Milano CA et al. Differential gene expression and genomic patient stratification following left ventricular assist device support. *J Am Coll Cardiol*. 2003;41:1096-1106.

22. Kaynak B, von Heydebreck A, Mebus S et al. Genome-wide array analysis of normal and malformed human hearts. *Circulation*. 2003;107:2467-2474.

23. Liew CC, Dzau VJ. Molecular genetics and genomics of heart failure. *Nat Rev Genet*. 2004;5:811-825.

24. Kittleson MM, Ye SQ, Irizarry RA et al. Identification of a gene expression profile that differentiates between ischemic and nonischemic cardiomyopathy. *Circulation*. 2004;110:3444-3451.

25. Kaab S, Barth AS, Margerie D et al. Global gene expression in human myocardium-oligonucleotide microarray analysis of regional diversity and transcriptional regulation in heart failure. *J Mol Med*. 2004;82:308-316.

26. Liew CC, Dzau VJ. Molecular genetics and genomics of heart failure. *Nat Rev Genet*. 2004;5:811-825.

27. Kittleson MM, Ye SQ, Irizarry RA et al. Identification of a gene expression profile that differentiates between ischemic and nonischemic cardiomyopathy. *Circulation*. 2004;110:3444-3451.

28. Wittchen F, Suckau L, Witt H et al. Genomic expression profiling of human inflammatory cardiomyopathy (DCMi) suggests novel therapeutic targets. *J Mol Med*. 2007;85:253-267.

29. Boheler KR, Volkova M, Morrell C et al. Sex- and age-dependent human transcriptome variability: implications for chronic heart failure. *Proc Natl Acad Sci U S A*. 2003;100:2754-2759.

30. Deng MC, Smits JM, Packer M. Selecting patients for heart transplantation: which patients are too well for transplant? *Curr Opin Cardiol*. 2002;17:137-144.

31. Cleland JG, Daubert JC, Erdmann E et al. The effect of cardiac resynchronization on morbidity and mortality in heart failure. *N Engl J Med*. 2005;352:1539-1549.

32. Bardy GH, Lee KL, Mark DB et al. Amiodarone or an implantable cardioverter-defibrillator for congestive heart failure. *N Engl J Med*. 2005;352:225-237.

33. Mann DL, Bristow MR. Mechanisms and models in heart failure: the biomechanical model and beyond. *Circulation*. 2005;111:2837-2849.

34. Bruins S, Fokkema MR, Romer JW et al. High intraindividual variation of B-type natriuretic peptide (BNP) and amino-terminal proBNP in patients with stable chronic heart failure. *Clin Chem*. 2004;50:2052-2058.

35. Wu AH, Smith A, Wieczorek S et al. Biological variation for N-terminal pro- and B-type natriuretic peptides and implications for therapeutic monitoring of patients with congestive heart failure. *Am J Cardiol*. 2003;92:628-631.

36. Lowes BD, Gilbert EM, Abraham WT et al. Myocardial gene expression in dilated cardiomyopathy treated with beta-blocking agents. *N Engl J Med*. 2002;346:1357-1365.

37. Smits JM, Deng MC, Hummel M et al. A prognostic model for predicting waiting-list mortality for a total national cohort of adult heart-transplant candidates. *Transplantation*. 2003;76:1185-1189.

38. Deng MC, Smits JM, Packer M. Selecting patients for heart transplantation: which patients are too well for transplant? *Curr Opin Cardiol*. 2002;17:137-144.

39. Metra M, Ponikowski P, Dickstein K et al. Advanced chronic heart failure: A position statement from the Study Group on Advanced Heart Failure of the Heart Failure Association of the European Society of Cardiology. *Eur J Heart Fail*. 2007;9:684-694.

40. Aaronson KD, Schwartz JS, Chen TM et al. Development and prospective validation of a clinical index to predict survival in ambulatory patients referred for cardiac transplant evaluation. *Circulation*. 1997;95:2660-2667.

**Table 1: Clinical characteristics of HF severity patient groups**

CAD: Coronary Artery Disease; DCM: Dilated Cardiomyopathy; LVEF: Left Ventricle Ejection Fraction; LVEDD: Left Ventricle End Diastolic Diameter; ACEI: Angiotensin Converting Enzyme Inhibitors; ARB: Angiotensin Receptor Blockers. UNOS: United Network for Organ Sharing; ADHF: Acute Decompensated Heart Failure. Data are presented as 'mean (SD)' when appropriate. P-value indicates the result of a comparison between the three patient groups using Fisher's exact test or Kruskal-Wallis rank sum test. If $p<0.05$, groups were compared two-by-two. *: $p<0.05$ between Deteriorating and Stable; †: $p<0.05$ between Intermediate and Stable; ‡: $p<0.05$ between Deteriorating and Intermediate. An ADHF episode was defined as recent if it occurred during the 3 months before the heart transplantation/total artificial heart placement. Heart failure duration was defined as the delay between onset of heart failure symptoms and heart transplantation/total artificial heart placement. Values for LVEF, LVEDD, blood urea nitrogen, and serum creatinine corresponded to pre-operative measurements. All patients were treated with loop diuretics (furosemide and/or bumetanide). Only medications related to heart failure therapy are presented. The clinical profile was determined based on the patients' medical urgency status in the UNOS classification and the occurrence of recent ADHF episodes.

**Figure -1- 'Clustering': Two-way hierarchical clustering of mRNA expression data for 176 samples and 4035 genes.**

Left: Classification tree of the 176 samples. The dendrogram is based on similarity measurements of 4035 genes among the 176 analyzed samples. Samples were separated into 2 main clusters (A and B) based on the expression profile of a 387-

gene cluster (indicated by a white bar in the middle part of the figure). For clusters A and B, two main sub-clusters (A1, A2 and B1, B2) were identified. Only clusters containing at least 15 samples were considered as significant. Some samples (indicated *) were not included in any cluster. The distribution of clinical statuses associated to the samples was analyzed in each sub-cluster. White, grey, and black boxes on the left side of the dendrogram denote Stable, Intermediate, and Deteriorating clinical status respectively.

Middle: Heat map of expression values for 176 samples and 4035 genes after hierarchical clustering of both genes and samples. Each column represents the 4035-gene expression profile for one sample. Each row represents the 176-sample expression profile for one gene. Two-way hierarchical clustering was applied to the genes and the samples after expression profiles were median-centered on genes. Results are presented using a color code. Green and red represent lower and higher expression levels relative to the median expression level of the gene, respectively.

Right: Selected clusters of correlated genes indicated by colored bars in the middle part of the figure. Intermediate samples were removed and remaining samples were ordered based on their origin (LV: left ventricle, RV: right ventricle) and the clinical status of the patient (S: Stable, D: Deteriorating). On the right side, functional annotation of the clusters obtained using Gominer is shown.

**Figure -2- Prediction of HF severity based on gene expression profiles.**

LV and RV predictors of HF severity were identified after comparison of expression profiles of ‘Deteriorating’ and ‘Stable’ patients in left and right ventricles respectively. The LV-RV predictor was determined as the combination of the LV severity and RV

severity predictors. Three patient classifications were constructed based on LV, RV, and LV-RV severity predictors.

**Top: Gene expression profiles of Stable and Deteriorating samples for the three severity predictors.** Each column represents the gene expression profile for one sample. Each row represents the relative expression level for one gene across the samples. Results are presented using a color code. Green and red represent lower and higher expression relative to the median expression level of the gene, respectively

**Bottom: patient classifications for the three severity predictors**

White and black circles denote Stable and Deteriorating LV samples respectively. White and black triangles denote Stable and Deteriorating RV samples respectively. Dashed lines denote upper and lower limits of the unpredictable interval calculated as $2.5^{th}$ and $97.5^{th}$ percentiles of the random-Molecular Severity Score.

**Figure -3- Prediction of HF severity in all samples**

Individual Molecular Severity Score values obtained for the LV-RV predictor are presented for all 176 analyzed samples. White and black circles denote Stable and Deteriorating LV samples respectively. White and black triangles denote Stable and Deteriorating RV samples respectively. Dashed lines denote upper and lower limits of the unpredictable interval calculated as $2.5^{th}$ and $97.5^{th}$ percentiles of the random-Molecular Severity Score).

**Figure -4- Between-chamber and between-sample reproducibility**

A- Between-chamber reproducibility: Molecular Severity Score values obtained using the LV-RV predictor were used to compare LV and RV samples obtained from the same patient. The between-chamber reproducibility was determined based on the correlation coefficient.

B- Between-sample reproducibility: Molecular Severity Score values obtained using the LV-RV predictor were used to compare samples obtained from the same chamber of a given patient. Subgroup analysis based on the origin of the sample (LV or RV) is shown. Square: LV sample; Triangle: RV sample. Between-sample reproducibility was determined based on the correlation coefficient for overall and subgroup analysis.

**Table 1**

| | **Stable n=13** | **Deteriorating n=12** | **Intermediate n=19** | p-value |
|---|---|---|---|---|
| Male / Female | 12/1 | 10/2 | 16/3 | 0.747 |
| Age, years | 50 (15) | 49 (9) | 48 (12) | 0.559 |
| Initial cardiac disease, CAD / DCM / other | 5 / 6 / 2 | 4 / 7 / 1 | 8 / 7 / 4 | 0.840 |
| HF duration, months | 32 (29) | 24 (33) | 29 (32) | 0.459 |
| Heart rate, $min^{-1}$ | 69 (13) | 100 (16) | 76 (16) | <0.001 * ‡ |
| Systolic arterial pressure, mmHg | 102 (17) | 97 (9) | 103 (12) | 0.509 |
| LVEF, % | 24 (11) | 22 (7) | 24 (7) | 0.829 |
| LVEDD, mm | 73 (13) | 66 (7) | 65 (10) | 0.254 |
| Blood urea nitrogen, mmol/l | 9.1 (5.6) | 9.8 (4.2) | 9.0 (4.1) | 0.884 |
| Serum Creatinine, µmol/l | 107 (26) | 107 (22) | 101 (36) | 0.642 |
| Medications, % of patients | | | | |
| ACEI / ARB | 100 | 58 | 84 | 0.024 * |
| Beta-Blockers | 69 | 0 | 26 | <0.001 * † |
| Adrenergic agonists | 0 | 100 | 42 | <0.001 * † ‡ |
| Phosphodiesterase Inhibitors | 0 | 67 | 0 | <0.001 * ‡ |
| Aldosterone blockers | 77 | 58 | 53 | 0.420 |
| Statin | 46 | 33 | 32 | 0.724 |
| Digoxin / Digitoxin | 46 | 25 | 26 | 0.502 |
| **UNOS medical urgency status** | **2** | **1A** | **1B or 2** | |
| **Number of recent ADHF episodes** | **0** | **1.8 (0.4)** | **1.7 (0.8)** | |

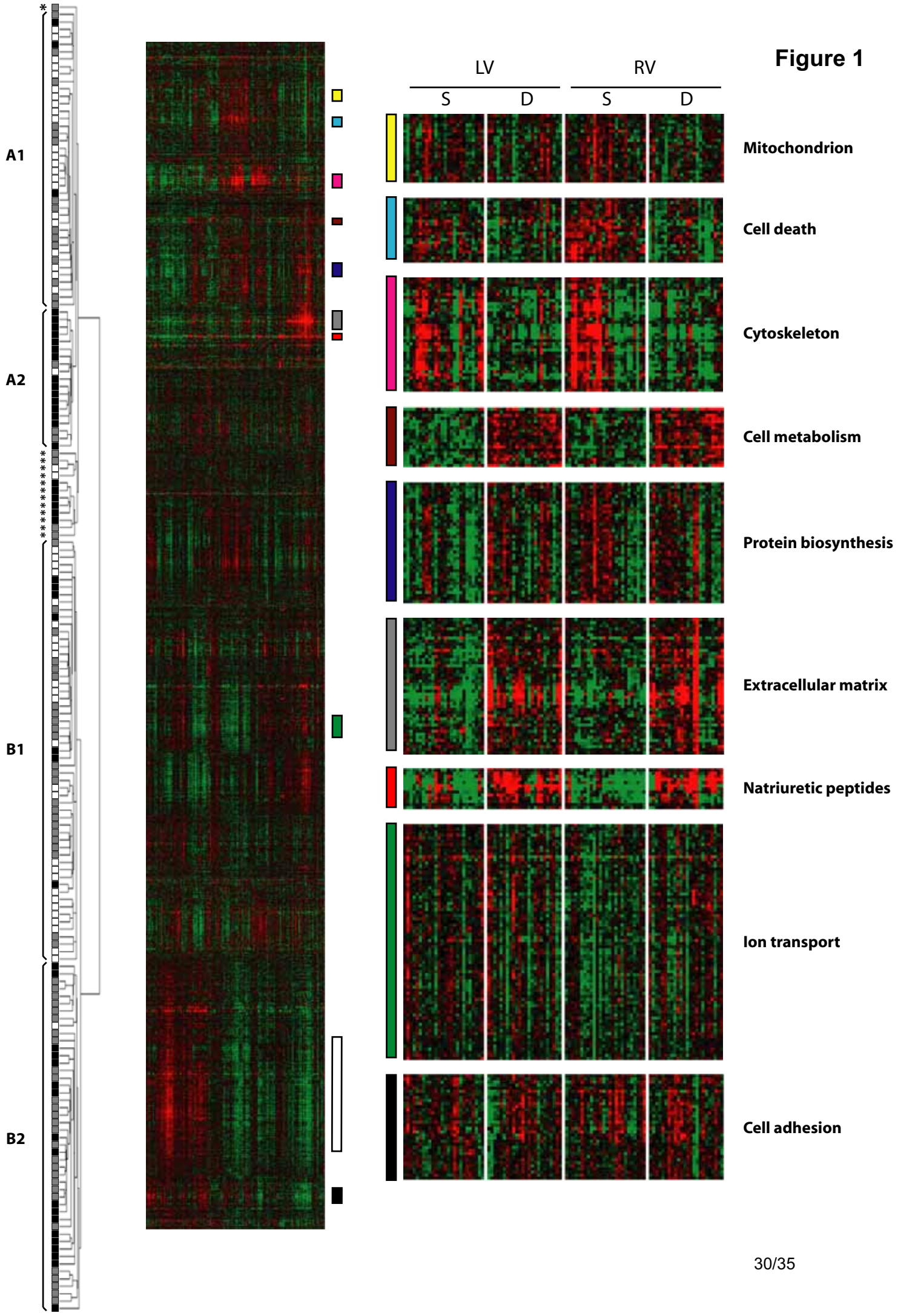
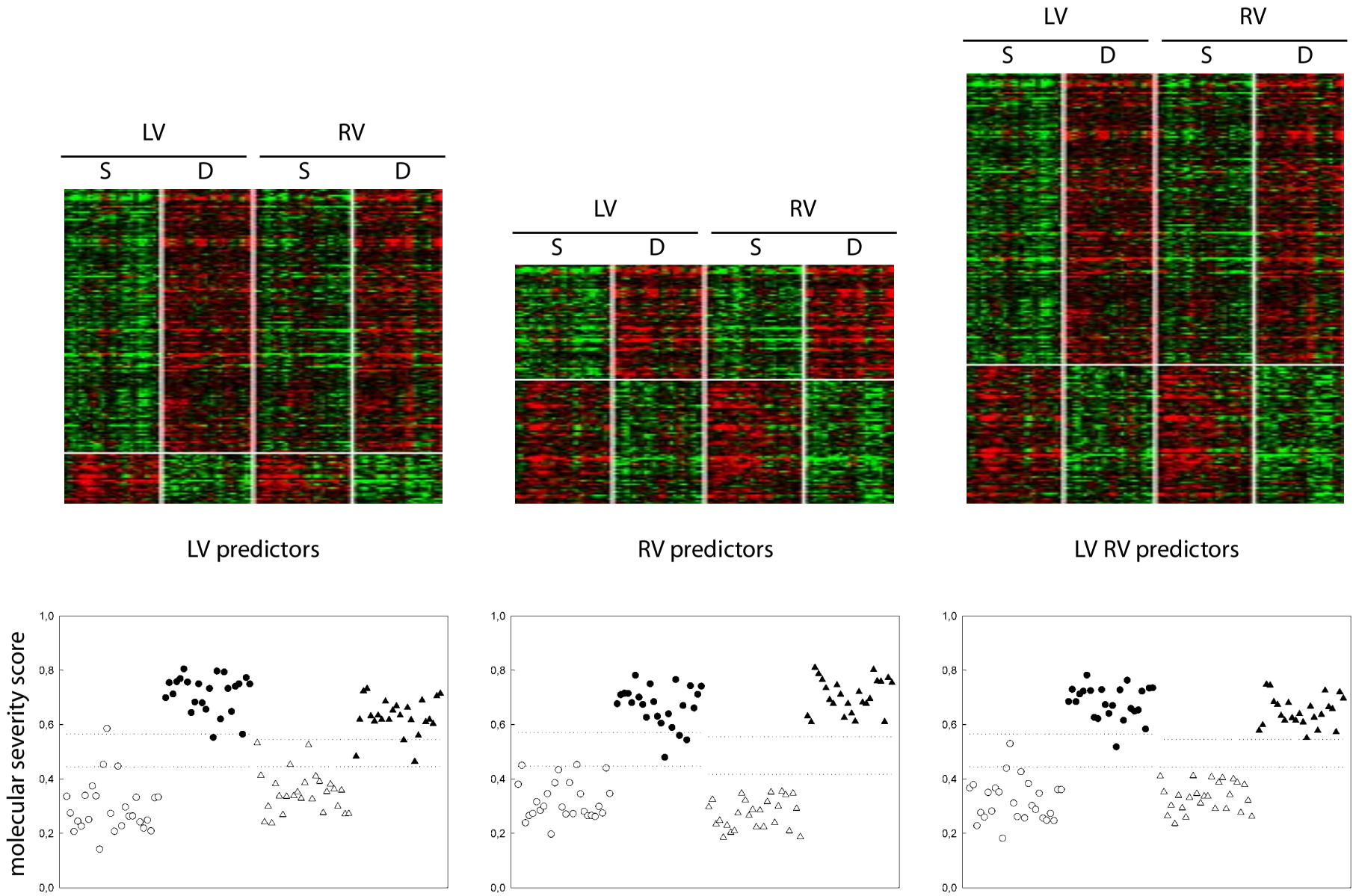
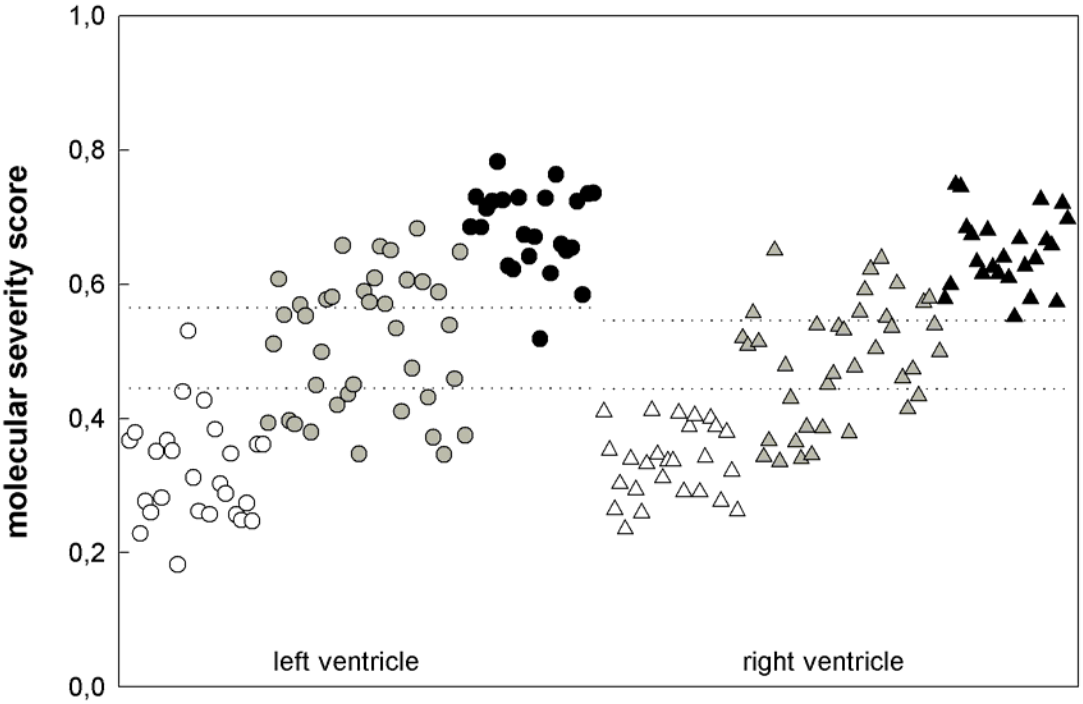
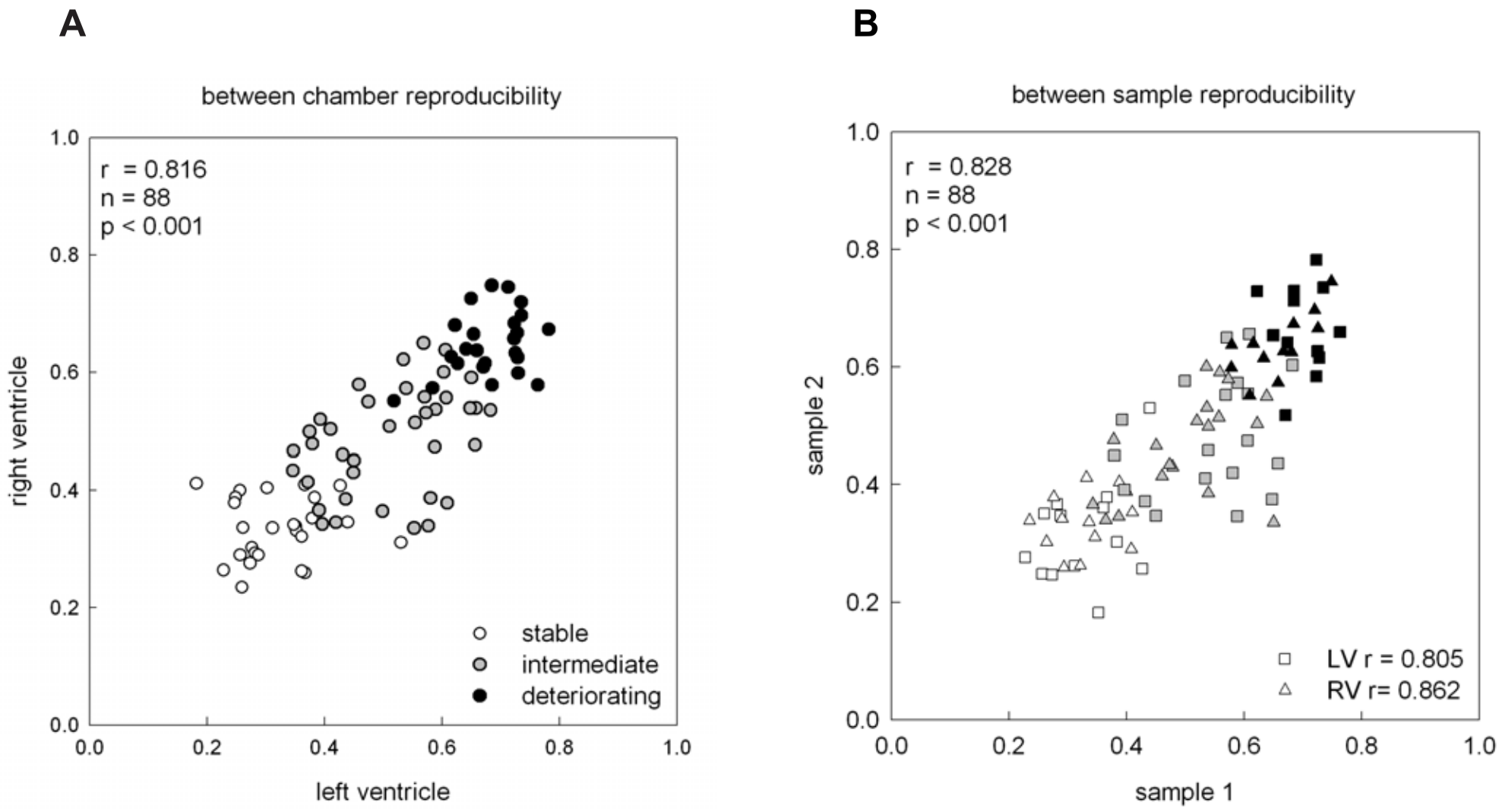

A1
A2
B1
B2
LV
RV
S
D
S
D
Mitochondrion
Cell death
Cytoskeleton
Cell metabolism
Protein biosynthesis
Extracellular matrix
Natriuretic peptides
Ion transport
Cell adhesion

**Figure 1**

## Figure 2


LV
RV
S
D
S
D
LV predictors
RV predictors
LV RV predictors
molecular severity score
1,0
0,8
0,6
0,4
0,2
0,0


1,0
0,8
0,6
0,4
0,2
0,0
molecular severity score
left ventricle
right ventricle


A
between chamber reproducibility
r = 0.816
n = 88
p < 0.001
right ventricle
left ventricle
stable
intermediate
deteriorating
B
between sample reproducibility
r = 0.828
n = 88
p < 0.001
sample 2
sample 1
LV r = 0.805
RV r= 0.862

**Supplemental data – table 1**



| patient | sex | age | severity status | | medical history | | physical examination | | echocardiography | | laboratory tests | | | medications | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | UNOS status | recent ADHF episodes | initial cardiac disease | duration of HF | HR | SAP | LVEF | LVEDD | BNP | BUN | serum creatinin | ACEI ARB | β-blockers | adrenergic agonist | PDE inhibitors | aldosteron blockers | statin | digitoxin digoxin |
| | | years | | number | | months | $min^{-1}$ | mmHg | % | mm | ng/l | mmol/l | µmol/l | | | | | | | |
| **Stable group** | | | | | | | | | | | | | | | | | | | | |
| 01 | M | 46 | 2 | 0 | DCM | 34 | 85 | 120 | 23 | 64 | | 4·1 | 75 | + | - | - | - | - | - | + |
| 02 | M | 60 | 2 | 0 | CAD | 68 | 60 | 92 | 20 | 67 | | 11·5 | 143 | + | + | - | - | + | + | + |
| 03 | M | 63 | 2 | 0 | CAD | 42 | 58 | 99 | 17 | 75 | | 24·3 | 129 | + | + | - | - | + | + | - |
| 04 | M | 50 | 2 | 0 | CAD | 3 | 58 | 100 | 20 | 68 | | 4·5 | 87 | + | + | - | - | - | + | - |
| 05 | M | 50 | 2 | 0 | DCM | 6 | 63 | 100 | 25 | 96 | | 7·1 | 102 | + | + | - | - | + | - | - |
| 06 | M | 64 | 2 | 0 | DCM | 10 | 53 | 113 | 25 | 69 | | 10·1 | 147 | + | + | - | - | + | - | - |
| 07 | F | 24 | 2 | 0 | CHD | 8 | 83 | 91 | 50 | 50 | 666·2 | 5·1 | 72 | + | + | - | - | + | - | - |
| 08 | M | 41 | 2 | 0 | DCM | 48 | 90 | 110 | 7 | 93 | | 7·6 | 89 | + | - | - | - | + | - | + |
| 09 | M | 54 | 2 | 0 | DCM | 59 | 77 | 77 | 30 | 73 | | 5·9 | 125 | + | + | - | - | + | + | + |
| 10 | M | 19 | 2 | 0 | CHD | 6 | 80 | 96 | 40 | na | | 4·6 | 73 | + | - | - | - | + | - | + |
| 11 | M | 64 | 2 | 0 | CAD | 6 | 52 | 80 | 20 | 80 | 58·8 | 12·0 | 115 | + | + | - | - | + | + | - |
| 12 | M | 59 | 2 | 0 | DCM | 91 | 81 | 115 | 13 | 70 | 332·6 | 14·6 | 120 | + | - | - | - | - | - | + |
| 13 | M | 60 | 2 | 0 | CAD | 33 | 61 | 138 | 20 | 66 | | 7·2 | 118 | + | + | - | - | + | + | - |
| **Intermediate group** | | | | | | | | | | | | | | | | | | | | |
| 14 | M | 60 | 2 | 2 | DCM | 26 | 60 | 90 | 19 | 58 | 1016·0 | 11·2 | 134 | + | + | - | - | - | - | - |
| 15 | M | 44 | 1B | 1 | CAD | 3 | 101 | 113 | 30 | 66 | | na | na | | - | + | - | - | - | - |
| 16 | M | 59 | 2 | 2 | DCM | 3 | 78 | 97 | 25 | 69 | 174·9 | 10·3 | 90 | + | + | - | - | + | - | + |
| 17 | F | 38 | 2 | 1 | CHD | 60 | 66 | 101 | 30 | 70 | 794·2 | 8·3 | 98 | + | - | - | - | + | - | - |
| 18 | M | 56 | 1B | 2 | DCM | 18 | 82 | 99 | 14 | 70 | | 7·5 | 147 | + | - | + | - | + | - | - |
| 19 | M | 54 | 2 | 1 | CAD | 31 | 119 | 116 | 20 | 66 | 1261·4 | 16·5 | 157 | + | + | - | - | + | + | - |
| 20 | M | 54 | 2 | 4 | CAD | 63 | 51 | 108 | 35 | 53 | | 21·1 | 184 | + | + | - | - | - | + | - |
| 21 | M | 50 | 2 | 1 | CAD | 3 | 76 | 106 | 16 | 76 | | 7·2 | 92 | + | - | - | - | + | + | - |
| 22 | M | 57 | 1B | 2 | DCM | 4 | 90 | 83 | 21 | 87 | 544·9 | 7·9 | 82 | + | - | + | - | - | - | + |
| 23 | M | 57 | 1B | 2 | DCM | 16 | 82 | 115 | 13 | 69 | | 9·6 | 124 | + | - | + | - | - | - | - |
| 24 | M | 44 | 2 | 1 | RCM | 71 | 60 | 114 | 18 | 49 | | 8·2 | 56 | + | - | - | - | - | - | + |
| 25 | M | 42 | 1B | 1 | CAD | 2 | 75 | 100 | 30 | 60 | 961·3 | 6·3 | 74 | | - | - | - | - | - | - |
| 26 | M | 65 | 1B | 1 | CAD | 16 | 79 | 80 | 20 | 79 | | 4·7 | 94 | + | - | + | - | - | - | - |
| 27 | F | 24 | 1B | 1 | DCM | 1 | 60 | 120 | 25 | 52 | | 3·9 | 58 | | - | - | - | - | - | - |
| 28 | M | 46 | 1B | 1 | CAD | 4 | 89 | 105 | 25 | 63 | 1153·8 | 6·5 | 107 | + | - | + | - | + | + | - |
| 29 | M | 56 | 2 | 2 | CAD | 60 | 61 | 92 | 25 | 72 | | 10·4 | 105 | + | + | - | - | + | + | + |
| 30 | M | 46 | 1B | 3 | DCM | 117 | 70 | 122 | 25 | 69 | 195·2 | 9·6 | 96 | + | - | + | - | + | + | + |
| 31 | F | 16 | 2 | 2 | CHD | 46 | 76 | 98 | 40 | 48 | | 6·2 | 46 | + | - | - | - | + | - | - |
| 32 | M | 48 | 1B | 2 | HCM | 5 | 75 | 93 | 20 | 61 | 599·6 | 6·5 | 79 | + | - | + | - | + | - | - |
| **Deteriorating group** | | | | | | | | | | | | | | | | | | | | |
| 33 | M | 56 | 1A | 1 | CAD | 1 | 97 | 111 | 24 | 54 | | 8·2 | 110 | + | - | + | - | - | + | - |
| 34 | M | 34 | 1A | 2 | DCM | 9 | 122 | 86 | 18 | 64 | | 12·1 | 105 | | - | + | + | - | - | - |
| 35 | M | 53 | 1A | 2 | DCM | 113 | 80 | 95 | 25 | 68 | 517·5 | 10·2 | 133 | + | - | + | + | + | - | - |
| 36 | M | 60 | 1A | 2 | DCM | 20 | 75 | 95 | 20 | 67 | | 5·4 | 84 | + | - | + | + | + | - | + |
| 37 | M | 58 | 1A | 1 | CAD | 4 | 95 | 87 | 12 | 66 | | 18·8 | 110 | | - | + | + | - | + | - |
| 38 | F | 41 | 1A | 2 | VHD | 8 | 102 | 97 | 35 | 59 | | 5·5 | 69 | | - | + | - | + | - | + |
| 39 | M | 63 | 1A | 2 | DCM | 4 | 78 | 111 | 28 | 70 | | 9·4 | 131 | | - | + | + | + | - | - |
| 40 | F | 45 | 1A | 2 | CAD | 2 | 120 | 93 | 22 | 52 | | 5·4 | 83 | | - | + | - | + | + | - |
| 41 | M | 48 | 1A | 2 | DCM | 56 | 110 | 115 | 28 | 71 | | 13·8 | 145 | + | - | + | - | + | - | - |
| 42 | M | 39 | 1A | 2 | DCM | 20 | 100 | 95 | 25 | 70 | | 7·3 | 108 | + | - | + | + | + | - | + |
| 43 | M | 39 | 1A | 2 | DCM | 50 | 120 | 90 | 10 | 71 | | 14·4 | 113 | + | - | + | + | - | - | - |
| 44 | M | 52 | 1A | 2 | CAD | 2 | 98 | 94 | 15 | 75 | | 7·4 | 99 | + | - | + | + | - | + | - |

**Supplemental data – Table 1: Clinical characteristics of advanced HF patients**

UNOS: United Network for Organ Sharing; ADHF episode: Acute Decompensated Heart Failure episode; CHD: Congenital Heart Disease; DCM: Dilated Cardiomyopathy; CAD: Coronary Artery Disease; RCM: Restrictive Cardiomyopathy; VHD: Valvular Heart Disease; ; HCM: Hypertrophic Cardiomyopathy; HR: Heart Rate; SAP: Systolic Arterial Pressure; LVEF: Left Ventricle Ejection Fraction; LVEDD: Left Ventricle End-Diastolic Diameter; BNP: Brain Natriuretic Peptide; BUN: blood urea nitrogen; ACEI: Angiotensin Converting Enzyme Inhibitors; ARB: Angiotensin Receptor Blockers; PDE Inhibitors: Phosphodiesterase Inhibitors. UNOS status corresponds to the medical urgency status as defined.[1] An ADHF episode was defined as recent if it occurred during the 3 months preceding the heart transplantation/total artificial heart placement. Heart failure duration was defined as the delay between onset of heart failure symptoms and heart transplantation/total artificial heart placement. Values for HR, SAP, LVEF, LVEDD, BUN, and serum creatinine corresponded to pre-operative measurements. Values for BNP were obtained within the 2 months before heart transplantation. All patients were treated with loop diuretics (furosemide and/or bumetanide). Only medications related to heart failure therapy are presented.

Reference List

1. Renlund DG, Taylor DO, Kfoury AG, Shaddy RS. New UNOS rules: historical background and implications for transplantation management. United Network for Organ Sharing. *J Heart Lung Transplant*. 1999;18:1065-1070.

# V. ETUDE 5

Lamirault G, Gaborit N, Le Meur N, Chevalier C, Lande G, Demolombe S, Escande D, Nattel S, Leger JJ, Steenman M. Gene expression profile associated with chronic atrial fibrillation and underlying valvular heart disease in man. J Mol Cell Cardiol. 2006 Jan;40(1):173-84.

La fibrillation auriculaire est associée à un remodelage auriculaire au niveau, électrique, contractile, et structural. Les mécanismes moléculaires qui conduisent à ce remodelage ont été étudiés sur la base d'approche individuelle et restent encore mal décrit dans leur globalité, en particulier pour les aspects non-électriques. Plusieurs études ont déjà pu analyser ce remodelage auriculaire sur une approche plus globale et ont apporté des résultats originaux. Cependant ces travaux n'ont pas permis de différencier les aspects du remodelage atrial spécifiquement liés au trouble du rythme de ceux liés à la pathologie sous jacente.

L'objectif de ce travail a été, en utilisant un ensemble de 4000 sondes couvrant largement le transcriptome cardiaque, l'identification des altérations d'expression géniques atriales spécifiques de la fibrillation auriculaire associée à une valvulopathie cardiaque. L'analyse a porté sur des échantillons d'auricules droit de 11 patients en fibrillation auriculaire chronique et opérés pour un remplacement valvulaire mitral ou aortique. Deux groupes contrôles ont été constitués pour permettre de bien distinguer les altérations d'expression génique atriales associées directement à la fibrillation auriculaire de celle associées à son substrat (la valvulopathie). Le premier groupe contrôle était constitué de 7 patients opérés pour un remplacement valvulaire aortique ou mitral et en rythme sinusal. Le second groupe contrôle était constitué de 11 patients en rythme sinusal et sans valvulopathie opérés pour des pontages coronaires. L'absence de facteurs confondants concernant des paramètres majeurs comme l'age ou le sex-ratio ont pu être vérifiés préalablement.

Les résultats ont montre que la présence seule de la valvulopathie cardiaque était associée à un fort remodelage du transcriptome des oreillettes. Le remodelage spécifiquement associé à la fibrillation auriculaire concernait finalement un nombre de

gènes beaucoup plus restreint, mettant l'accent sur l'importance de distinguer sur 2 phénomènes survenant de manière concomitante dans le tissu atrial.

Parmi les altérations du transcriptome associé à la fibrillation auriculaire, des modifications d'expression de gènes impliqués dans la fibrose (par exemple TGFB1 et TGFBI) et la contraction (par exemple PLN et TNNI3) ont été identifiées. Ces modifications représentent les premiers éléments des bases moléculaires du remodelage contractile et structural déjà décrit dans la fibrillation auriculaire. Nous avons également mis en évidence des modifications d'expression de gènes concourant à un phénotype pro-coagulant. Ces modifications sont particulièrement intéressantes car la fibrillation auriculaire est associée à un risque majeur d'accident vasculaire cérébral par mécanisme thrombo-embolique dont le point départ est principalement l'oreillette gauche. La stase sanguine provoquée par la dysfonction mécanique des oreillettes et un état pro-thrombotique systémique favorisent ce risque d'accident vasculaire cérébral. Nos résultats suggèrent qu'un facteur pariétal auriculaire puisse également participer à la formation, de thrombus dans l'oreillette gauche des patients en fibrillation auriculaire.

Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Molecular and Cellular Cardiology 40 (2006) 173–184
Journal of Molecular and Cellular Cardiology
www.elsevier.com/locate/yjmcc

Original article

# Gene expression profile associated with chronic atrial fibrillation and underlying valvular heart disease in man

Guillaume Lamirault [a,b], Nathalie Gaborit [a], Nolwenn Le Meur [a,c], Catherine Chevalier [c], Gilles Lande [a,b], Sophie Demolombe [a], Denis Escande [a], Stanley Nattel [d], Jean J. Léger [a,c], Marja Steenman [a,*]

[a] *L'institut du Thorax, Inserm U533, Faculté de Médecine, 1, rue Gaston Veil, 44035 Nantes, France*
[b] *L'institut du Thorax, Cardiology Department, Hôpital G&R Laennec, Nantes, France*
[c] *West Genopole Transcriptome Facility, Nantes, France*
[d] *Montreal Heart Institute, University of Montreal, Montreal, Canada*

Received 31 August 2005; accepted 8 September 2005

Available online 19 October 2005

**Abstract**

Atrial fibrillation (AF) is the most common sustained cardiac arrhythmia in humans. The pathophysiology of AF involves electrical, structural and contractile remodeling, which is associated with changes in cardiac gene expression. Previous studies of gene-expression changes in clinical AF have mostly been limited to a small number of candidate genes and have not all been well controlled for underlying heart disease. The present study assessed AF-related gene-expression changes in valve-disease patients with microarrays representing the cardiac transcriptome. Right atrial appendages from 11 patients with chronic AF and underlying valvular heart disease (AF-VHD) and seven patients in sinus rhythm with VHD (SR-VHD) were individually compared to an age-matched sinus-rhythm control group (SR-CTRL, 11 patients) using cardiac-specific microarray analysis. One-class statistical analysis was used to identify genes differentially expressed between SR-VHD and SR-CTRL patients. Two-class statistical analysis was used to identify genes differentially expressed between AF-VHD and SR-VHD patients. Out of 3863 analyzed genes, 832 genes were differentially expressed between SR-VHD and SR-CTRL patients, and 169 genes were differentially expressed between AF-VHD and SR-VHD patients. Striking AF-related changes included altered expression of nine genes pointing towards the development of fibrosis (e.g. upregulation of transforming growth factor β1), and changes in eight genes potentially related to an increased risk of thromboembolic events (e.g. upregulation of α2 macroglobulin). Microarray results were confirmed by quantitative PCR. Our results suggest that AF produces a characteristic profile of gene-expression changes that may be related to the pathophysiology of the arrhythmia.


*Keywords:* Atrial fibrillation; Valvular heart disease; Gene expression; Coagulation; Fibrosis

## 1. Introduction

Atrial fibrillation (AF) is the most common sustained cardiac arrhythmia in humans. In most cases, AF is associated with other cardiovascular diseases like heart failure, thyrotoxic heart disease, hypertension, coronary artery disease (CAD) and rheumatic valve disease [1]. The increased risk of stroke and thromboembolism associated with AF has been thought to be mainly due to increased blood stasis in the non-contracting atria [2]. AF causes progressive changes [3] related to electrical, contractile and structural remodeling [4]. Thus far, there has been a major focus on electrical components of the remodeling process, which has been analyzed at the molecular level by candidate gene approaches that have identified expression changes in genes encoding ion channels or calcium-handling proteins [5,6]. Four recent studies characterized the molecular basis of remodeling on a more global scale. Kim et al. [7] analyzed the expression of 1152 genes in 26 AF patients and 26 controls. They focused on genes

* Corresponding author. Tel.: +33 2 40 41 29 59; fax: +33 2 40 41 29 50.
*E-mail address:* marja.steenman@nantes.inserm.fr (M. Steenman).

0022-2828/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.yjmcc.2005.09.004

involved in oxidative stress. Their AF patients did not have permanent AF, and the underlying pathology of the control group was unknown. Lai et al. [8] analyzed the expression of 6035 genes in a porcine model of AF. Ohki-Kaneda et al. [9] analyzed the expression of ~12,000 genes in seven chronic and persistent AF patients and 10 matched controls. They found a very large number of significantly differentially-expressed genes and then focused on genes with a greater than 0.5 U absolute difference in expression. This resulted in the identification of only 11 genes, four of which were of unknown function. Finally, Barth et al. [10] analyzed the expression of ~45,000 genes in 10 permanent AF patients and 20 unmatched controls. They identified 1434 genes deregulated in AF. These studies provided interesting insights but were either directed towards limited components of the cardiac transcriptome or did not contain a disease-matched sinus rhythm group to differentiate changes associated with AF from those related to underlying heart disease.

The present study was designed to pinpoint gene-expression changes in the cardiac transcriptome that are specifically associated with AF in patients with valvular heart disease (VHD), a cardiac disease frequently associated with the arrhythmia. We constructed a long-oligonucleotide microarray representing ~4000 human genes selected on the basis of evidence for involvement in normal or pathological cardiovascular functioning. With these microarrays, we determined the right atrial (RA) expression profile related to VHD and identified genes specifically related to AF. The AF-specific genes were involved in diverse functional categories —particularly in fibrosis and coagulation pathways— revealing potential new components of the pathophysiology of AF.

## 2. Methods

### 2.1. Patients

RA appendage (RAA) samples were obtained from 29 patients undergoing mitral/aortic valve replacement or coronary artery bypass graft surgery. These RAAs were from 11 patients with permanent AF and underlying VHD, constituting the AF-VHD group, seven patients in sinus rhythm with VHD and with no history of AF (SR-VHD) and 11 patients with CAD and no history of AF. The CAD patients taken to constitute a 'control' (SR-CTRL) group for normal RAA gene expression were all undergoing routine aortocoronary bypass surgery. They were selected to have: 1) sinus rhythm and no history of previous cardiac arrhythmias (based on both no clinical history of arrhythmias and the absence of tachyarrhythmias on any ECG recording in the patient's medical record prior to surgery); 2) normal left ventricular function; 3) no dilation of cardiac cavities. The clinical characteristics of individual patients are shown in Table 1. The three groups were comparable in mean age and gender distribution (72 ± 3, 70 ± 9, 69 ± 5 years; 83%, 83%, 82% male for AF-VHD, SR-VHD, SR-CTRL groups, respectively). AF-VHD and SR-VHD had similar left ventricular ejection fractions (59 ± 10%, 63 ± 9%, respectively) and comparable valve-disease distributions (respectively, 45%, 57% aortic valve stenosis; 27%, 43% mitral regurgitation; remainder mixed disease). AF-VHD patients had a greater tendency to have hypertension and atrial dilation, but the differences were not statistically-significant (Fisher exact test). The AF duration was between 1 and 13 years with an average of 7 ± 4 years.

### 2.2. RNA isolation and labeling

Total RNA was isolated using TRIZOL®Reagent (Life Technologies) and treated with DNase using the RNase-Free DNase Set and the RNeasy Mini Kit (Qiagen). RNA quality was assessed using an Agilent 2100 bioanalyzer and RT-PCR with primers for β-actin. Absence of DNA contamination was verified by PCR using the same primers. Cy3- and Cy5-labeled cDNA was prepared using the CyScribe cDNA Post Labeling Kit (Amersham Pharmacia Biotech). RAA samples from AF-VHD patients and from SR-VHD patients were each labeled individually with Cy5. SR-CTRL patients were pooled and then labeled with Cy3.

### 2.3. Microarrays

Microarrays were prepared in-house using 50-mer oligonucleotide probes (MWG Biotech). These oligonucleotides were obtained from the MWG Biotech human genome-wide microarray collection and had all been tested for specificity. The probes were spotted onto epoxy-silane coated glass slides using the Lucidea Array Spotter from Amersham. The 3863 genes that were represented on the microarray had been selected for involvement in skeletal muscle and/or cardiovascular normal and pathological functioning. Selection was based on 1) subtractive hybridization experiments [11]; 2) genome-wide microarray hybridizations [12], including an AF-VHD vs. SR-CTRL hybridization; 3) literature data. Each gene was spotted in quadruplicate on the microarrays. The use of a relevant gene selection—as opposed to the whole genome—allowed us to obtain sufficient replicate expression values per gene to be able to identify statistically significant gene expression changes as small as 15% [13]. For more information see: http://cardioserve.nantes.inserm.fr/ptf-puce/puce_multi_dediee.php?langue=_en&id=0. Each Cy5-labeled sample was mixed with an equal amount of Cy3-labeled SR-CTRL sample, pre-incubated with human Cot-I DNA (Gibco-BRL), yeast tRNA and polyA RNA, and hybridized to the microarrays. Three independent hybridizations were performed for each AF-VHD patient (except patient AF-VHD09, for whom only two hybridizations were possible) and two independent hybridizations were performed for each SR-VHD patient. For each hybridization a different Cy5-labeled sample was used. Hybridized arrays were scanned by fluorescence confocal microscopy (Scanarray 3000, GSI-Lumonics). Measurements were obtained separately for each fluorochrome at 10 μm/pixel resolution.

Table 1
Clinical characteristics of patients

| Patient | Age | Sex | AF duration (years) | Valvular replacement indication | LVEF (%) | Ischemic heart disease | LAD (mm) | HTN | Treatment |
|---|---|---|---|---|---|---|---|---|---|
| *AF-VHD group* | | | | | | | | | |
| AF-VHD02 | 67 | M | 9 | MR | 41 | – | 62 | + | CR, FN, FR, LS, PV |
| AF-VHD03 | 70 | F | 5 | MS + TR | 57 | – | 50 | – | DG, ID, FN |
| AF-VHD08 | 74 | M | 2 | AS | 71 | – | 40 | – | AM, DL, HCT, FN |
| AF-VHD09 | 67 | M | 2 | MR* | 45 | – | 49 | + | FN, RM |
| AF-VHD10 | 76 | M | 12 | AS + MR | 58 | – | 56 | + | FN, FR, PR, PV |
| AF-VHD13 | 74 | F | 1 | AS + MR | 49 | + | 46 | + | DG, DL, EL, FN, FR |
| AF-VHD14 | 72 | F | 10 | MR | 64 | + | 44 | + | CP, DT, FB, FN, FR, LT, NC, PV |
| AF-VHD15 | 73 | M | 6 | AS | 75 | + | 45 | + | DL, EL, FR, PB |
| AF-VHD16 | 73 | M | 13 | AS | 61 | – | 52 | + | DG, EL, FN, FR |
| AF-VHD17 | 69 | M | 7 | AS | 70 | + | 40 | + | AT, FN, SM |
| AF-VHD20 | 75 | M | 10 | AS | 60 | + | 51 | + | DL, EL, FN, FR |
| | 72±3 | 3/8 | 7±4 | | 59±10 | | 49±6 | | |
| *SR-VHD group* | | | | | | | | | |
| SR-VHD31 | 76 | M | | AS | 64 | – | 40 | – | AA, FS |
| SR-VHD32 | 62 | M | | MR | 77 | – | 63 | + | FR, RM |
| SR-VHD34 | 80 | M | | AS | 62 | + | 44 | + | LS, PV |
| SR-VHD37 | 72 | M | | AS | 57 | + | 39 | + | SM, VL |
| SR-VHD39 | 52 | M | | MR | 64 | – | 46 | – | FR, RM |
| SR-VHD46 | 69 | M | | MR | 72 | – | 45 | + | IN, FL, PR, PV, ST |
| SR-VHD47 | 78 | F | | AS | 47 | – | 39 | – | DG, FR, RM, SP |
| | 70±9 | 1/6 | | | 63±9 | | 45±8 | | |
| *SR-CTRL group* | | | | | | | | | |
| CTRL10 | 64 | M | | | 47 | + | | – | AA, AC, MS, SM |
| CTRL13 | 73 | M | | | 50 | + | | – | AT, CD, MS, TT |
| CTRL14 | 68 | M | | | 57 | + | | + | AA, AT, PV |
| CTRL16 | 72 | M | | | 71 | + | | – | AA, MS, MT, TT |
| CTRL17 | 70 | M | | | 52 | + | | – | AO, AT, FL, MS |
| CTRL19 | 68 | M | | | 60 | + | | – | AA, AT, PV, RM |
| CTRL35 | 58 | F | | | 73 | + | | – | AA, AD, AT, MS, TT, PV |
| CTRL36 | 74 | M | | | 73 | + | | + | AA, AC, AD, AZ, PV, SP |
| CTRL48 | 71 | M | | | 72 | + | | – | AA, AD, BT, MS, SM, TT |
| CTRL42 | 73 | F | | | 66 | + | | + | AA, AT, FE, ID, PV, TT |
| CTRL43 | 73 | M | | | 74 | + | | – | AA, DL, TT |
| | 69 ± 5 | 2/9 | | | 63 ± 10 | | | | |

Group data are shown as mean ± S.D. Gender data are summarized as females/males. HTN: arterial hypertension; AS: aortic valve stenosis; MR: mitral regurgitation; *: patient AF-VHD09 has undergone successful aortic valve replacement surgery 12 years ago for treatment of aortic valve stenosis; MS: mitral stenosis; TR: tricuspid regurgitation; LVEF: left ventricular ejection fraction; LAD: left atrium diameter; AA: acetylsalicylic acid; AC: acebutolol; AD: amlodipine; AM: amiloride; AO: atorvastatine; AT: atenolol; AZ: altizide; BT: betaxolol; CD: clopidogrel; CP: captopril; CR: carvedilol; DG: digoxine; DL: diltiazem; DT: digatilin; EL: enalapril; FB: fenobarbital; FE: felodipine; FL: flubriprofen; FN: fluindione; FR: furosemide; FS: fosinopril; HCT: hydrochlorothiazide; ID: isosorbide dinitrate; IN: indapamide; LS: losartan; LT: levothyroxine; MS: molsidomine; MT: metoprolol; NC: nicardipine; PB: pirbuterol; PR: perindopril; PV: pravastatin; RM: ramipril; SM: simvastatin; SP: spironolactone; ST: sotalol; TT: trinitrin; VL: valsartan.

### 2.4. Data analysis and statistics

Fluorescence values and expression ratios were analyzed using GenePix (Axon). Consolidated expression values were obtained using MADSCAN, developed in our laboratory [13,14]. This step includes a physical validation of the spots and a normalization of the Cy3 and Cy5 values, combining the rank invariant and lowest fitness methods with spatial normalization. This strategy excluded the need for dye-swap experiments [15]. All arrays were scaled and outlier values within arrays (based on quadruplicate spot replicates) and between arrays (based on the 'replicate hybridizations' × ''quadruplicate spots') were identified [13] and eliminated from further analysis. For each gene, the median expression ratio within each array was determined. Genes lacking valid expression values for more than one array were eliminated for all patients. One-class Significance Analysis of Microarrays (SAM) [16] and LInear Models for MicroArray data (Limma) [17] were used to identify genes with statistically-significant differential expression between SR-VHD and SR-CTRL patients. Two-class SAM was used to identify genes differentially-expressed between AF-VHD and SR-VHD patients. With this method, each gene is first assigned a score on the basis of a modified *t*-test and genes with scores greater than a user-defined threshold are selected. Repeated random sample permutations are used to estimate the percent-

age of genes identified by chance (false-discovery rate, FDR) among the selected genes. The FDR is similar to a *P* value adjusted for multiple comparisons. For all SAM analyses, the chosen delta threshold corresponded to the lowest median FDR. Two-way hierarchical agglomerative clustering was applied to the gene expression matrix consisting of the 18 patients (AF-VHD and SR-VHD) and the genes selected by the two-class SAM analysis or a selection of these genes. The input consisted of the median of the replicate expression values for each gene. We applied average linkage clustering with uncentered correlation using Cluster [18]. Clusters were visualized using Treeview.

### 2.5. Gene ontology analysis

Genes differentially expressed between SR-VHD and SR-CTRL patients were grouped according to the biological process classification established by the Gene Ontology consortium, using the Biological Process Pie Chart tool of the PANTHER classification system [19]. To obtain a balanced representation of different biological processes between 5% and 25%, the following adjustments were made: 'signal transduction' was divided into 'signal transduction', 'cell communication', 'cell surface receptor mediated signal transduction' and 'intracellular signaling cascade'; 'muscle contraction' was merged with 'cell structure and motility'; 'nitrogen metabolism', 'phospor metabolism' and 'sulfur metabolism' were merged with 'other metabolism'; 'amino acid metabolism' was merged with 'protein metabolism and modification'.

### 2.6. Quantitative PCR

As a control for the results of microarray gene-expression quantification, the relative expression levels of 20 genes were analyzed in all patients using Taqman® Gene Expression Assays (Applied Biosystems) on the ABI PRISM® 7900HT Sequence Detection System, with methods previously described in detail in [20]. Genes were selected from interesting functional categories revealed by the microarray analysis. The identity of all 20 genes is indicated in Table 3 in the on-line data supplement and in Fig. 6. Myo-inositol 1-phosphate synthase A1 (ISYNA1) was taken as an internal standard. This gene was chosen based on ranking according to least expression changes in the 18 microarray experiments (selecting for genes with most constant expression over all subjects and groups) and based on its housekeeping function: ISYNA1 is essential to the organism as the sole supplier of the 6-carbon myo-inositol ring, the precursor of the inositol phospholipids which are important structural components of all membranes [21,22]. Duplicate experiments were performed and each experiment contained duplicate PCR reactions. Mean gene expression ratios were calculated for each AF-VHD and each SR-VHD patient vs. the pooled SR-CTRL patients. The AF-VHD/SR-VHD expression ratio was calculated by dividing the median AF-VHD/SR-CTRL ratio by the median SR-VHD/SR-CTRL ratio.

## 3. Results

The strategy and overall results of our approach are summarized in Fig. 1. RAAs from seven SR-VHD and 11 AF-VHD patients were analyzed by hybridization against pooled control samples on replicate microarrays containing quadruplicate 50mer oligonucleotides corresponding to 3863 non-redundant genes. After analysis of all hybridizations, valid expression values were obtained for 2996 genes in all patients. The advantage of using dedicated as opposed to genome-wide microarrays is reflected by this high percentage (78%) of valid expression values.

### 3.1. Gene-expression changes associated with VHD

One-class SAM analysis of SR-VHD vs. SR-CTRL hybridizations identified 1498 statistically-significant differentially expressed genes (Fig. 1). One-class Limma analysis identified 832 genes differentially expressed in SR-VHD vs. SR-CTRL patients (all 832 genes were present in the SAM gene set). This suggests that VHD requiring valve-replacement surgery produces major changes in cardiac gene expression and that underlying VHD probably could influence the transcriptomal remodeling observed in our AF-VHD patients. Therefore, we set out to further characterize the VHD expression profile. Fig. 2A shows that the distribution of the 832 genes among the Gene Ontology biological processes largely corresponds with the distribution of all analyzed genes. Some functional categories appeared to be somewhat stronger represented among the SR-VHD differential genes. These categories include nucleoside, nucleotide and nucleic acid metabolism (containing many genes involved in RNA transcription), cell cycle and apoptosis. We next classified the 832 genes into two groups; the 'most differentially expressed genes' (ratio > 1.5 or < 0.67, $N = 183$), and the remaining 'least differentially expressed genes' ($N = 649$) and compared their functional-category distribution. Fig. 2B shows that genes involved in cell structure and motility (including

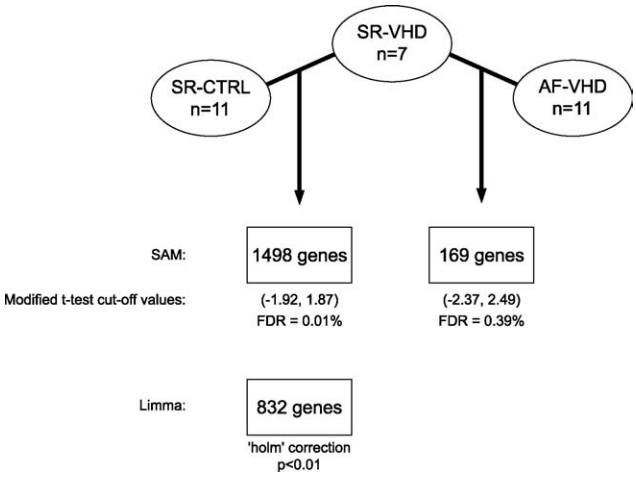



Fig. 1. Experimental set-up and global results. Schematic representation of the experimental set-up and the number of identified differentially expressed genes.

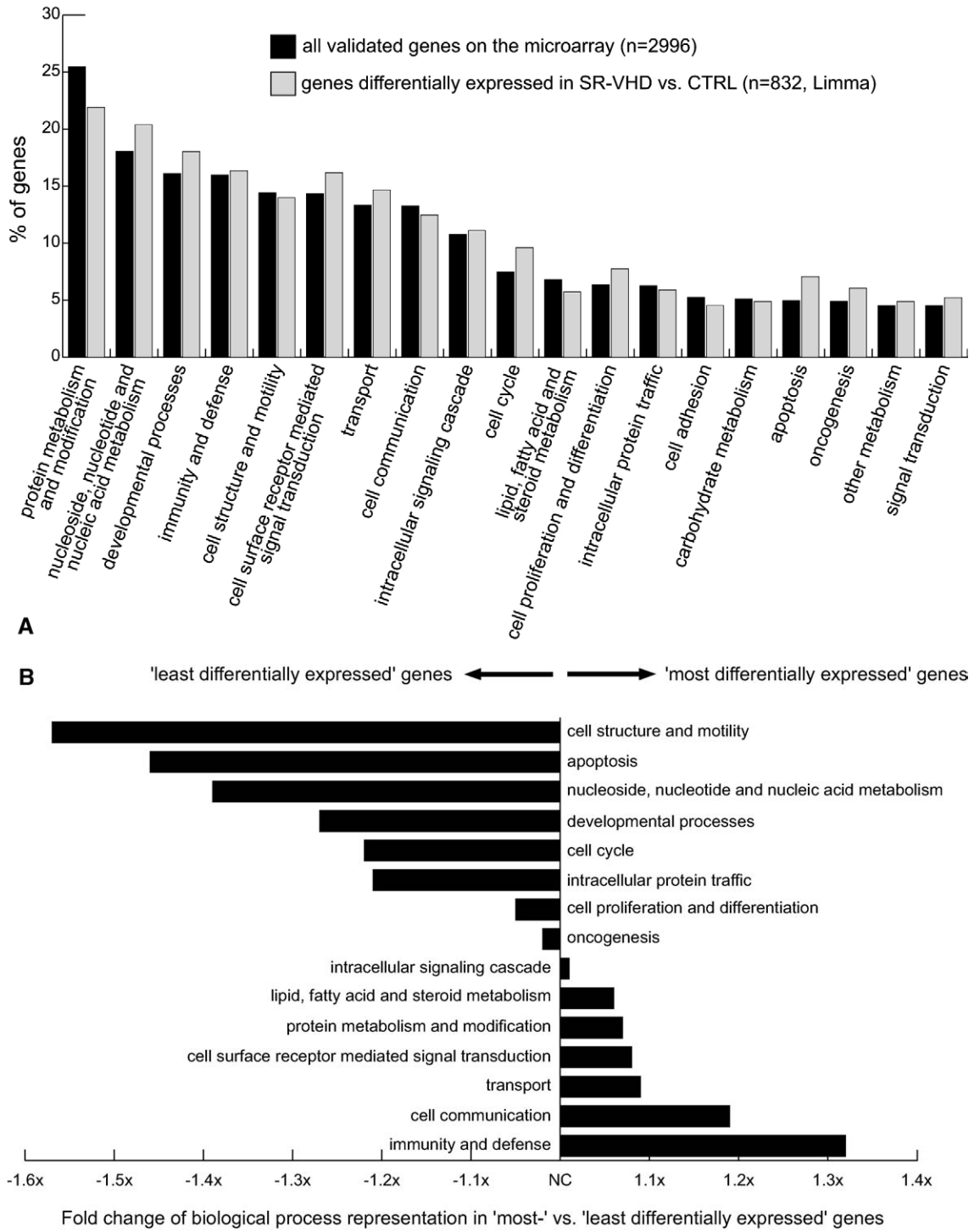



Fig. 2. Gene Ontology classification of genes differentially expressed in SR-VHD vs. SR-CTRL patients.
A: Representation of biological process categories (as defined by the Gene Ontology consortium) among all validated genes on the microarray and among the genes differentially expressed in SR-VHD vs. SR-CTRL patients (according to one-class Limma analysis with 'holm' correction and $P < 0.01$). B: Representation of biological process categories among genes 'most differentially expressed' in SR-VHD vs. SR-CTRL patients (ratio >1.5 or <0.67) as compared to the remaining 'least differentially expressed' genes. The fold change reflects the difference between the number of observed 'most differentially expressed' genes in a biological process and the number of expected 'most differentially expressed genes' based on the representation (percentage) of the biological process among the 'least differentially expressed' genes.

genes involved in muscle contraction) and apoptosis were mostly affected by relatively small expression changes, whereas genes involved in immunity and defense and cell communication were mostly affected by large expression changes. A complete list of the 832 genes is available in the on-line data supplement (Table 2).

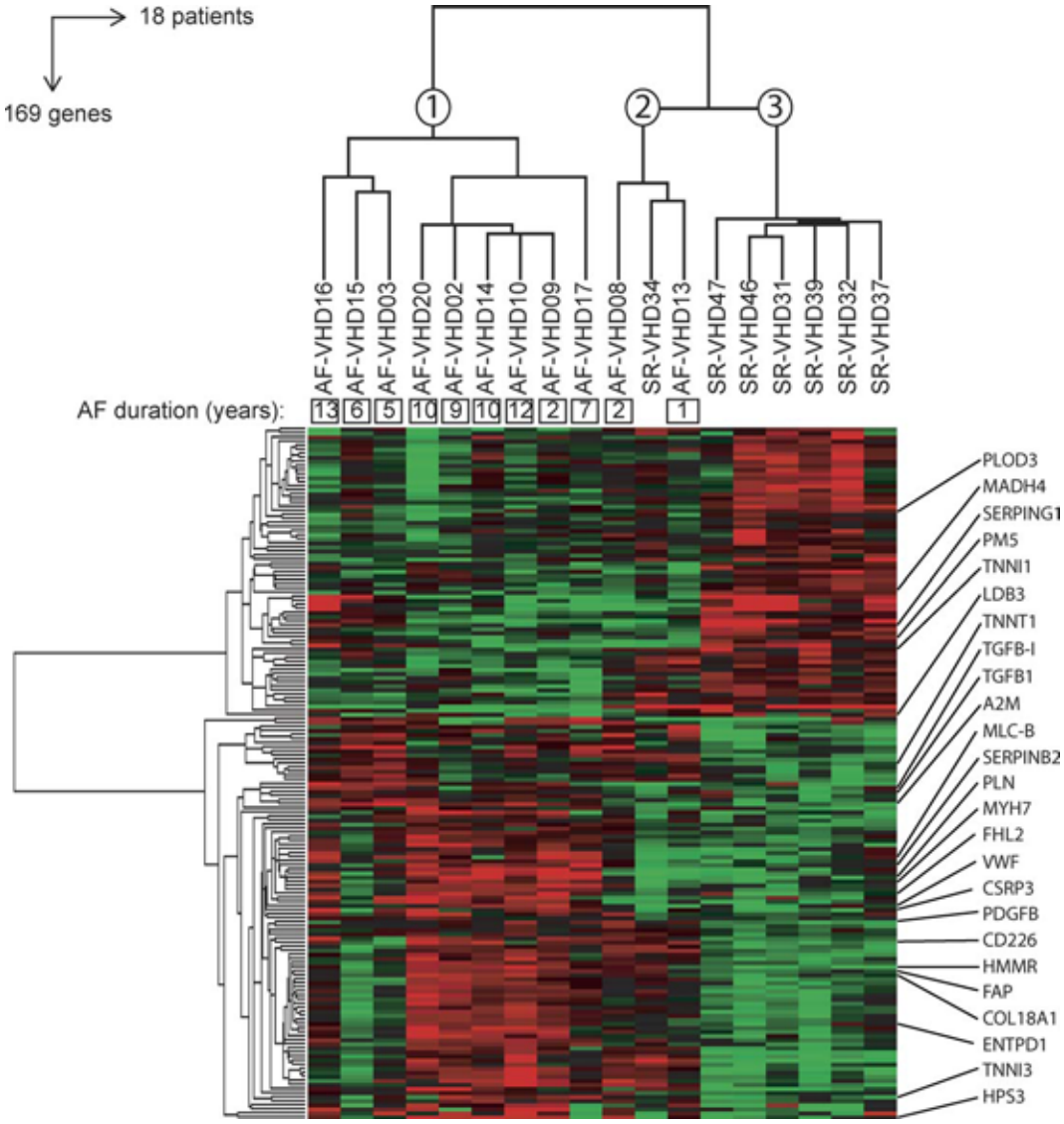
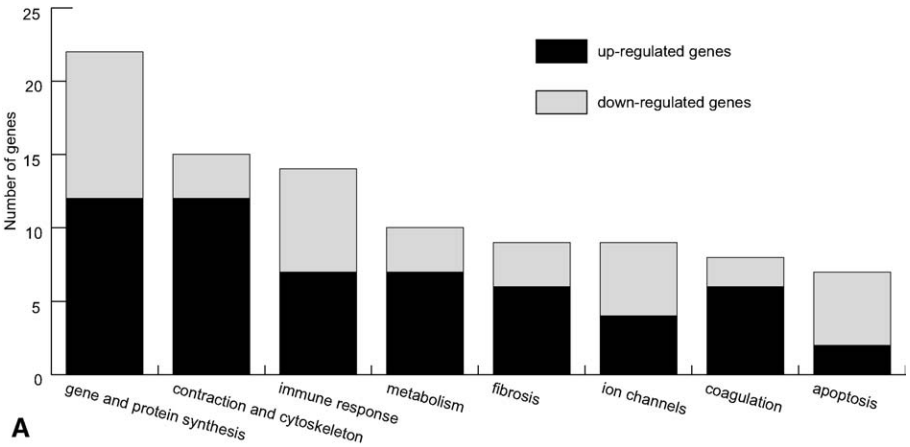



Fig. 3. Two-way hierarchical clustering. Two-way hierarchical clustering of the 18 patients (11 AF-VHD patients and seven SR-VHD patients) and the 169 genes differentially expressed between both patient groups. The duration of AF is indicated. The position of the 25 genes from Fig. 4B is shown. Expression values are indicated by color coding; red > black > green. For each gene, the gene expression values were median centered and therefore represent relative expression ratios.

### *3.2. AF-specific changes in cardiac gene expression*

To identify AF-specific expression changes, we performed two-class SAM analysis between AF-VHD and SR-VHD patients (Fig. 1). This analysis identified 169 genes with statistically-significant differential expression between the two groups. Fig. 3 shows the results of unsupervised two-way hierarchical clustering of the 18 VHD patients (11 AF-VHD patients and seven SR-VHD patients) and the 169 AF-specific genes. Nine of the 11 AF-VHD patients clustered together (in patient cluster 1), as did six of the seven SR-VHD patients (in patient cluster 3). One of the SR-VHD patients (SR-VHD34) had a gene-expression profile closer to two AF-VHD patients (AF-VHD08 and AF-VHD13, patient cluster 2) than to the other SR-VHD patients. These two AF-VHD patients had a relatively short duration of AF (1 and 2 years). After careful analysis of clinical parameters, we discovered that patient SR-VHD34 was the only SR-VHD patient with post-operative AF. Thus, possibly the similar gene-expression profile of SR-VHD34 to the AF group represented a latent AF gene-expression phenotype unmasked by surgery. In any case, the ability of the profile of the 169 genes to divide the patients into relatively distinct AF and SR clusters supports the notion of a selective AF-related gene-expression pattern.

Among the 169 genes displaying AF-specific expression changes, 93 are involved in well-defined functions (Fig. 4A). All genes are displayed in Table 3 (on-line data supplement). Many genes involved in gene and protein expression ($N = 22$) and sarcomeric and cytoskeletal functions ($N = 15$) were differentially expressed. Genes involved in gene and protein expression showed both up- and downregulation, whereas genes involved in sarcomeric and cytoskeletal functions were mostly upregulated (e.g. β myosin heavy chain (MYH7), phospholamban (PLN) and nexilin (NEX)). Fourteen identi-




**B**

| symbol | gene name | ratio AF-VHD/SR-VHD microarray | ratio AF-VHD/SR-VHD Q-PCR | effect on |
|---|---|---|---|---|
| **MYH7*** | myosin, heavy polypeptide 7, cardiac muscle, β | 1.67 | 2.22 † | contraction |
| **PLN*** | phospholamban | 1.55 | | |
| **CSRP3*** | cysteine and glycine-rich protein 3 | 1.53 | 2.11 † | |
| **FHL2** | four and a half lim domains 2 | 1.44 | | |
| MLC-B | myosin regulatory light chain | 1.39 | | |
| **TNNI3** | troponin I, cardiac | 1.23 | 1.64 † | |
| TNNT1 | troponin T1, skeletal, slow | 1.21 | | |
| TNNI1 | troponin I, skeletal, slow | 0.85 | 0.46 † | |
| LDB3* | LIM domain binding 3 | 0.84 | 0.62 † | |
| HMMR | hyaluronan-mediated motility receptor isoform a | 1.46 | | fibrosis |
| FAP | fibroblast activation protein, α subunit | 1.33 | | |
| **TGFBI** | transforming growth factor, β-induced | 1.25 | 1.20 | |
| COL18A1 | collagen, type XVIII, α1 | 1.23 | | |
| **TGFB1** | transforming growth factor, β1 | 1.23 | 1.29 † | |
| PDGFB | platelet-derived growth factor β, isoform 2 | 1.21 | | |
| PM5 | pm5 protein | 0.84 | | |
| PLOD3 | procollagen-lysine, 2-oxoglutarate 5-dioxygenase 3v | 0.82 | | |
| MADH4 | mad, mothers against decapentaplegic homolog 4 | 0.79 | | |
| HPS3 | Hermansky-Pudlak syndrome | 1.76 | | coagulation |
| **A2M** | α2 macroglobulin precursor | 1.49 | 1.45 † | |
| CD226 | adhesion glycoprotein | 1.41 | 1.09 | |
| ENTPD1 | ectonucleoside triphosphate diphosphohydrolase 1 | 1.21 | | |
| **SERPINB2** | serine (or cysteine) proteinase inhibitor, clade b, member 2 | 1.17 | | |
| **VWF** | von willebrand factor | 1.11 | 2.24 † | |
| SERPING1 | complement component 1 inhibitor precursor | 0.85 | 0.71 † | |
| EPCR | endothelial cell protein C receptor | NOC | 0.58 † | |

Fig. 4. Genes differentially expressed between AF-VHD and SR-VHD patients. A: Representation of different functional categories among the 169 AF-specific genes. B: A selection of 26 genes differentially expressed between AF-VHD and SR-VHD patients. Upper: Genes involved in contraction and/or cardiac dilation (*). Middle: Genes involved in the development of fibrosis. Lower: Genes involved in thrombosis. Genes in bold were not differential between SR-VHD and SR-CTRL patients and therefore VHD-independent. Q-PCR: quantitative PCR. NOC: not on chip. †: Genes differentially expressed based on two-class SAM analysis of the Q-PCR results.

fied genes involved in the immune response displayed both up- and downregulation, and the expression changes identified in metabolic genes ($N = 10$) were directed mainly towards increased energy production (e.g. upregulation of cytochrome *c* oxidase assembly protein (COX17) and adenylate kinase 1 (AK1)). Nine genes are potentially involved in the development of fibrosis, and 7 gene expression changes are indicative of a hypercoagulable state (e.g. upregulation of α2 macroglobulin (A2M) and von Willebrand factor (VWF)). Based on this interesting finding related to coagulation we analyzed four additional genes—either not present on our microarray or not differential in our analysis—by quantitative PCR: endothelial protein C receptor (EPCR), thrombomodulin (THBD), P selectin (SELP) and tissue factor (F3). Of these, EPCR (not present on our microarray) was found to be downregulated in AF-VHD patients. Finally, seven genes were found with functions in apoptosis. No specific direction (pro- or anti-apoptotic) was indicated by the gene changes. Expression changes were also identified in a few ion channel encoding genes. Since ion channel genes generally display low levels of expression and are therefore difficult to assess with broad-spectrum gene microarrays like those used in the present study, they were analyzed in a separate investigation with ion channel-selective analysis tools [23]. Changes in con-

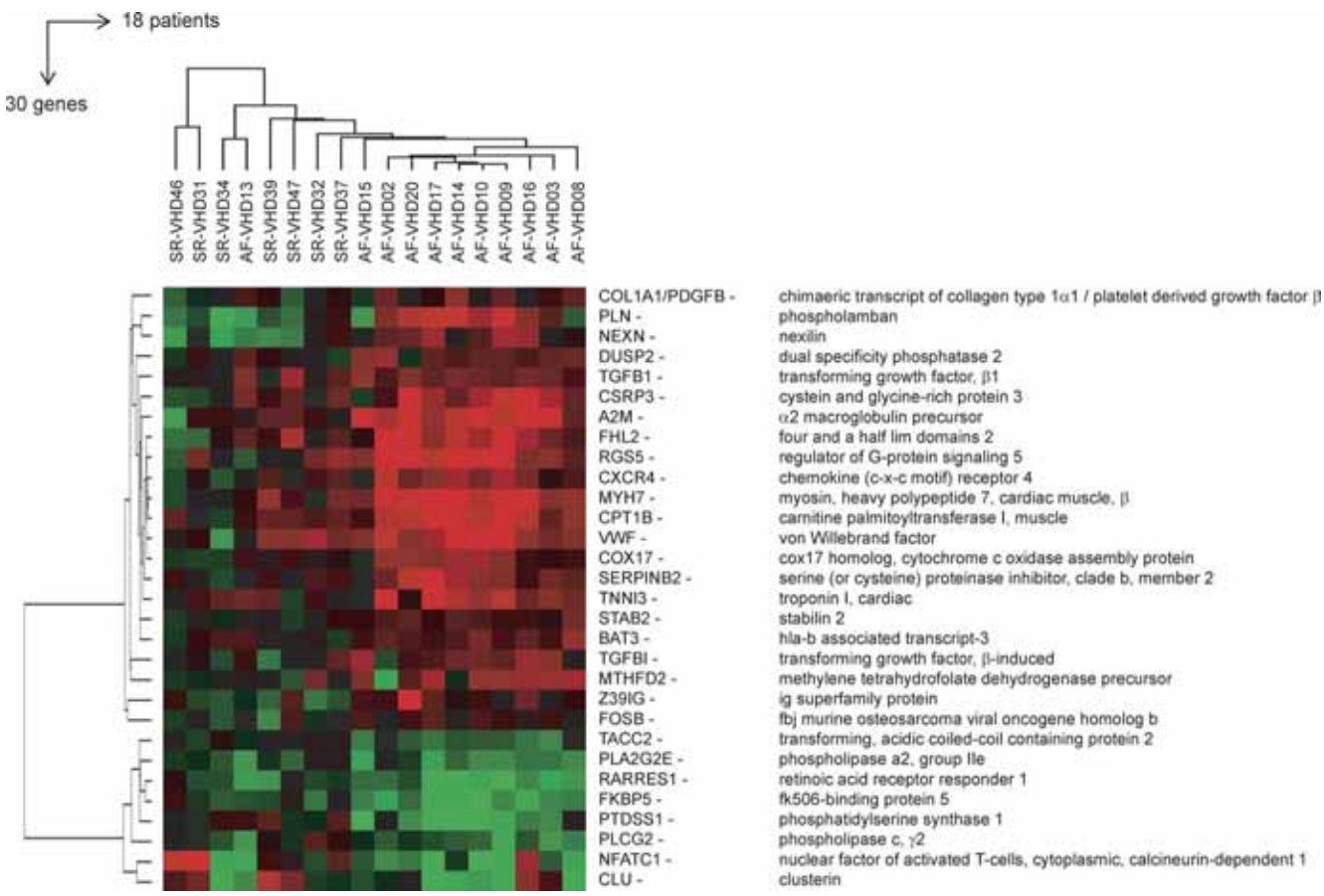



Fig. 5. Two-way hierarchical clustering. Two-way hierarchical clustering of the 18 patients (11 AF-VHD patients and seven SR-VHD patients) and the 30 VHD-independent, AF-specific genes. The gene expression values were not median centered and therefore represent absolute expression ratios.

traction, coagulation and fibrosis-related genes are illustrated in Fig. 4B.

As stated in the previous section, VHD has a substantial effect on the cardiac expression profile. However, among the 169 genes differentially expressed between AF-VHD and SR-VHD patients, 30 genes were not affected by VHD. These genes (indicated in bold in Table 3 in the on-line data supplement and (partly) in Fig. 4B) were identified based on the fact that they were not differentially expressed between SR-VHD and SR-CTRL patients, but were differentially expressed between AF-VHD and SR-VHD patients (based on SAM analysis). Fig. 5 shows the unsupervised hierarchical clustering of these 30 VHD-independent, AF-specific genes and all 18 patients. This clustering provided a comparable separation of the AF-VHD and SR-VHD patients to the clustering based on the 169 genes shown in Fig. 3. Patient AF-VHD13 (with the shortest duration of AF) still clusters together with patient SR-VHD34, whereas patient AF-VHD08 now clusters with the AF-VHD patients. This observation suggests that some gene changes independent of VHD may be particularly characteristic of AF. Some of the prominent differential genes in this analysis include genes involved in inflammation and fibrosis (TGFB1), coagulation (VWF) and contraction (cardiac troponin I, TNNI3).

Although valve-disease contributions in the AF-VHD and SR-VHD patient groups were comparable, small differences did exist. Since mitral valve and aortic valve disease may have different effects on the atria, two-class SAM analysis was performed between patients with pure aortic valve disease ($N = 9$) and patients with pure mitral valve disease ($N = 6$). Of the 79 identified genes, only 4 were also found to be differential between AF-VHD and SR-VHD patients (identified in Table 3, on-line data supplement). Therefore, the small difference in valve disease distribution did not contribute significantly to the AF-specific gene expression profile. Similarly, gender differences did not contribute significantly to the AF-specific gene expression profile (since only two of 62 genes that were differentially expressed between female and male AF-VHD patients were also differentially expressed between AF-VHD and SR-VHD patients).

### 3.3. Quantitative PCR

A comparison of data obtained with microarrays and with quantitative PCR is shown in Fig. 6. Exact values can be found in Table 3 (on-line data supplement). Out of 20 genes analyzed by quantitative PCR, microarray data were obtained for 17. The remaining three genes either failed to provide valid microarray values (atrial natriuretic factor, NPPA) or were not present on the microarray (EPCR and SELP). Quantitative PCR results showed EPCR to be significantly down-regulated. Of the 17 genes analyzed by both methods,

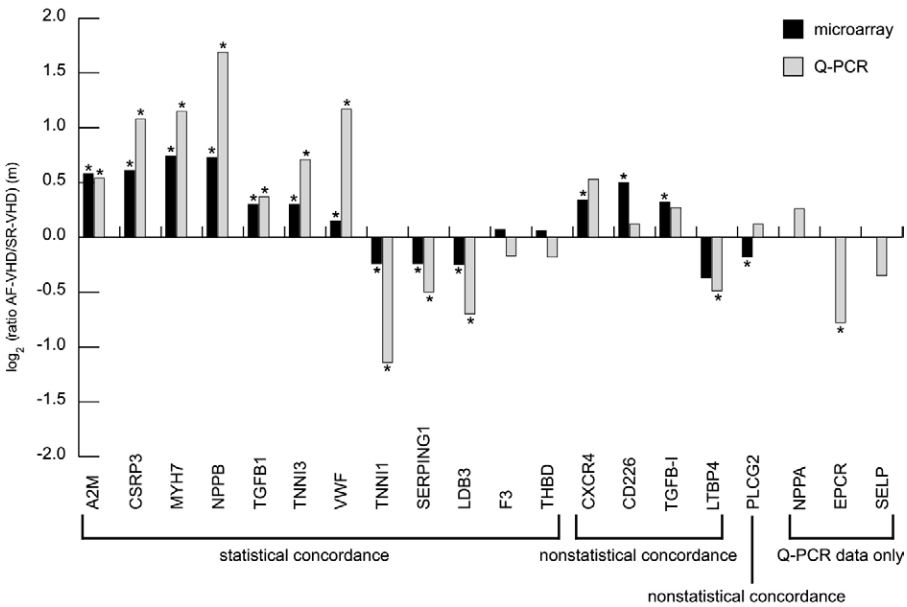



Fig. 6. Quantitative PCR vs. microarray. Comparison of results obtained with microarray and quantitative PCR analysis. Statistical concordance: Genes displaying similar results with both methods and similar results after statistical analysis. Non-statistical concordance: Genes displaying similar results with both methods, but only showing statistical significance with one method. Non-statistical discordance: gene displaying different results with both methods, the result of only one method is validated by statistical analysis. Q-PCR data only: Genes without available microarray data. *: Statistically significant according to two-class SAM analysis.

11 showed concordant significant up- or down-regulation or concordant unchanged expression levels. Five genes displayed significant up- or down-regulation by one method, and a similar (same direction) non-significant expression change by the other. One gene (phospholipase C gamma 2, PLCG2) showed discordant results.

## 4. Discussion

With this study, we aimed to provide a global view of cardiac gene expression changes related to permanent AF. We have succeeded in obtaining an extensive transcriptomal analysis, the largest to date on AF patients. The use of replicate spots on our microarrays and replicate hybridizations allowed us to reliably identify even small expression changes [13]. This is the great advantage of using dedicated microarrays as opposed to genome-wide microarrays, which do not contain replicate spots. Striking points included the extent to which valve disease per se remodeled cardiac gene expression, and the prominent AF-specific changes in genes associated with fibrosis and thrombosis.

### *4.1. Comparison with previous genomic analyses in AF and VHD patients*

Thus far, four other studies have used a genomic approach in the analysis of AF, three of which were on human samples. In contrast to those studies, we analyzed a more defined AF population by assessing only patients with chronic AF and we were careful to include a sinus-rhythm group with similar heart disease to AF patients. We did not identify the same expression changes as Kim et al. [7]. This may be related to the fact that their patients were not in chronic AF or to the fact that they did not match their control group in underlying cardiac disease. The latter is underlined by the fact that we identified differentially-expressed genes from their study as differently expressed between our SR-VHD and SR-CTRL patients but none were differentially expressed between our SR-VHD and AF-VHD patients. The analysis by Ohki-Kaneda et al. [9]—on a mixed population of chronic and persistent AF patients—resulted in the identification of only 11 AF-related genes (four of which were ESTs). These genes were either not present on our microarray or not differential between our SR-VHD and AF-VHD patients. A primary objective of their study was to create a regression relating gene expression to cardiac function. In concordance with the study by Lai et al. [8] we identified up-regulation of the myosin regulatory light chain. The main finding of the study by Barth et al. [10] was that the atrial myocardium of AF patients showed a ventricular-like genomic signature. In concordance with their results we identified an upregulation of transforming growth factor β1 (TGFB1) in human AF patients. Studies on transgenic mice have shown that increased cardiac expression of TGFB1 leads to atrial fibrosis [24] and increased vulnerability to AF [25]. The upregulation of TGFB1 identified in our AF patients may therefore play a significant functional role in promoting atrial fibrosis, a known risk factor for AF [26–28]. In addition, we identified increased expression of two genes regulated by TGFB1: the hyaluronan mediated motility receptor HMMR, whose function is related to the extracellular matrix and whose mRNA stability is induced by TGFB1 [29]; and the TGFB1-induced gene TGFB-I, a component of the extracellular matrix [30]. The upregulation of

platelet-derived growth factor β (PDGFB) that we noted in AF may also lead to the activation of cardiovascular fibroblasts and increased collagen deposition [31]. This gene was also identified by Barth et al. [10].

Our finding that VHD is associated with substantial cardiac transcriptomal remodeling shows the importance of matching AF and control patients in underlying cardiac pathology.

### 4.2. Novel elements and potential significance

Numerous functional categories were affected by AF. Expression levels of genes involved in transcription and genes involved in protein synthesis or modification were found to be both up- and downregulated in AF. Two of these genes are known regulators of cardiac transcription: Gata binding protein 4 (GATA4), which was upregulated, and nuclear factor of activated T cells (NFATC1), which was downregulated. GATA4 plays a crucial role in cardiogenesis [32] and regulates a set of cardiac-specific genes. Among those genes are brain natriuretic peptide (BNP) [33,34] and cardiac troponin I (TNNI3) [35–37], which we found to be upregulated in AF. NFATC1—as part of the NFAT transcription complex—also has been found to be involved in cardiac development [38,39]. The differential expression of many genes involved in gene and protein synthesis is consistent with the extensive changes in electrical, contractile and structural properties associated with AF.

AF has already been associated with a hypercoagulable state by the following observations [40]: First, the loss of atrial systole leads to increased blood stasis. Secondly, atrial endocardial damage has been associated with AF. Thirdly, abnormalities of procoagulant blood constituents have been found in AF patients. In our study, the AF-specific differential expression of several genes was suggestive of a hypercoagulable state, a result of potential importance to understanding the thromboembolic diathesis in AF. Of those genes, only VWF has been previously associated with AF. Several studies identified an increased plasma level [41–43], increased endocardial protein expression [44,45], or increased endocardial RNA expression [45]. VWF is a marker of damaged endothelium and plays a role in platelet aggregation. We also identified the adhesion glycoprotein CD226, a gene involved in platelet activation and aggregation [46], as being upregulated in AF. A third upregulated gene with a potential role in platelet function is Hermansky–Pudlak syndrome 3 (HPS3), a gene with loss-of-function mutation in patients with the HPS [47,48]. One of the consequences of this syndrome is impaired platelet function, related to the absence of platelet dense bodies and decreased aggregation responses. Upregulation of the gene encoding ectonucleoside triphosphate diphosphohydrolase 1 (ENTPD1) might, on the other hand, lead to inhibition of platelet aggregation [49]. The differential expression of two other factors may be linked to increased coagulation directly related to protein C activation. A2M, which was upregulated in our AF patients, inhibits the anticoagulant action of activated protein C [50]. EPCR, which was downregulated in our AF patients, augments protein C activation [51]. Finally, downregulation of the coagulation inhibitory complement component 1 inhibitor precursor (SERPING1) and upregulation of the plasminogen activator inhibitor (SERPINB2) may also be associated with increased coagulation. Thus far, the search for prothrombotic markers has been mainly performed by quantification of candidate proteins in blood samples [52]. Only two studies analyzed potential markers on a local level in atrial endocardium [44,45]. Our study identified novel potential prothrombotic markers in AF patients. These findings may be relevant to both the pathophysiology and prevention of thromboembolic complications in AF. Since the risk of stroke in AF is not homogeneous, better risk stratification for thromboprophylaxis is needed.

Overall, we have provided the largest and most detailed view thus far of gene expression changes related to VHD and AF in humans. The results of our study may contribute to a better understanding of the pathobiology of AF and the potential identification of new therapeutic targets.

### 4.3. Potential limitations

Since it is not ethically possible to obtain atrial samples from patients with normal hearts, our SR-CTRL patients did not have truly normal hearts. This may have affected the SR-VHD vs. SR-CTRL comparison. However, this did not affect the discovery of AF-specific genes, since these were identified by comparing AF-VHD to underlying-disease matched SR-VHD patients. Because of sample size limitations it was not possible to perform multivariate analysis to explore the combined effect of AF and potential other variables like age, mean ventricular heart rate prior to surgery, the timing of atrial biopsy in relationship to bypass, etc. on gene expression variations. We were limited to the use of RAA tissue for our analysis. Our results may therefore not be representative of transcriptomal remodeling processes in the atria as a whole. Had we attempted to obtain samples from other right and left atrial regions, it would have been virtually impossible to obtain a large enough patient population, particularly of patients with sinus rhythm and no valvular disease.

It is now well-recognized that many episodes of AF can be asymptomatic [53]. In fact, in patients with known AF, asymptomatic episodes are more common than symptomatic ones [54]. Therefore, although our SR-CTRL and SR-VHD patients had no history of AF nor any pre-operative ECGs showing AF, we cannot be sure that they had never had episodes of asymptomatic AF prior to surgery. We reviewed all patient files in detail to determine whether 24-hour Holter monitoring recordings had been obtained in the SR groups. Recordings were available for three SR-CTRL patients and one SR-VHD patient and none showed supraventricular tachyarrhythmias. On 6-month follow-up, only one patient had an episode of AF (SR-VHD34, see Section 3), which was

restricted to the immediate post-operative period. The possibility of asymptomatic AF in sinus-rhythm control patients has not been considered in previous studies comparing atrial samples from AF vs. SR patients [7–10,23,28], and is potentially a significant limitation. The effect of AF episodes in SR patients would be to diminish the differences between them and an AF group, and could have reduced our ability to detect AF-related differences.

Most AF-VHD patients (10 of 11) received anticoagulation treatment (fluindone), which might conceivably affect the expression of genes involved in coagulation, although there is no known evidence for such regulation. Some clinical heterogeneity exists within each patient group. If this heterogeneity is reflected by a heterogeneous atrial gene expression profile, the affected genes may not have reached statistical significance in the SAM one- and two-class analyses. Therefore, some genes may have been missed in our study.

Our list of genes involved in AF is probably not exhaustive since we did not use genome-wide microarrays. We preferred statistical robustness over completeness by analyzing a subset of the human genome with replicate experiments. Furthermore, we used a set of genes that we have developed on the basis of known or strongly suspected involvement in cardiac physiology and/or pathophysiology (a fact underlined by the high percentage of valid expression values in our study). The risk of missing potentially significant genes not presently known to be involved in the heart is offset by the large number of false-positive changes that would inevitably be detected by adding the remaining ~20,000 genes in the human transcriptome to the study.

## Acknowledgments

The authors thank the Thoracic Surgery and Cardiology Departments of Nantes University Hospital for their participation, Audrey Bihouee for development of the local microarray database and Raluca Teusan for help with the Gene Ontology analyses. Funding was provided by the "Institut National de la Santé et de la Recherche Médicale" (Inserm), the "Centre National de la Recherche Scientifique", the "Association Française contre les Myopathies", the "Region Pays de Loire" and the Canadian Institutes of Health Research.

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.yjmcc.2005.09.004.

## References

[1] Nattel S. Therapeutic implications of atrial fibrillation mechanisms: can mechanistic insights be used to improve AF management? Cardiovasc Res 2002;54:347–60.

[2] Wolf PA, Dawber TR, Thomas Jr. HE, Kannel WB. Epidemiologic assessment of chronic atrial fibrillation and risk of stroke: the Framingham study. Neurology 1978;28:973–7.

[3] Wijffels MC, Kirchhof CJ, Dorland R, Allessie MA. Atrial fibrillation begets atrial fibrillation. A study in awake chronically instrumented goats. Circulation 1995;92:1954–68.

[4] Allessie M, Ausma J, Schotten U. Electrical, contractile and structural remodeling during atrial fibrillation. Cardiovasc Res 2002;54:230–46.

[5] Brundel BJ, Henning RH, Kampinga HH, Van Gelder IC, Crijns HJ. Molecular mechanisms of remodeling in human atrial fibrillation. Cardiovasc Res 2002;54:315–24.

[6] Yue L, Melnyk P, Gaspo R, Wang Z, Nattel S. Molecular mechanisms underlying ionic remodeling in a dog model of atrial fibrillation. Circ Res 1999;84:776–84.

[7] Kim YH, Lee JH, Lim DS, Shim WJ, Ro YM, Park GH, et al. Gene expression profiling of oxidative stress on atrial fibrillation in humans. Exp Mol Med 2003;35:336–49.

[8] Lai LP, Lin JL, Lin CS, Yeh HM, Tsay YG, Lee CF, et al. Functional genomic study on atrial fibrillation using cDNA microarray and two-dimensional protein electrophoresis techniques and identification of the myosin regulatory light chain isoform reprogramming in atrial fibrillation. J Cardiovasc Electrophysiol 2004;15:214–23.

[9] Ohki-Kaneda R, Ohashi J, Yamamoto K, Ueno S, Ota J, Choi YL, et al. Cardiac function-related gene expression profiles in human atrial myocytes. Biochem Biophys Res Commun 2004;320:1328–36.

[10] Barth AS, Merk S, Arnoldi E, Zwermann L, Kloos P, Gebauer M, et al. Reprogramming of the human atrial transcriptome in permanent atrial fibrillation: expression of a ventricular-like genomic signature. Circ Res 2005;96:1022–9.

[11] Steenman M, Lamirault G, Le Meur N, Le Cunff M, Escande D, Léger JJ. Distinct molecular portraits of human failing hearts identified by dedicated cDNA microarrays. Eur J Heart Fail 2005;7:157–65.

[12] Steenman M, Chen Y-W, Le Cunff M, Lamirault G, Varro A, Hoffman E, et al. Transcriptomal analysis of failing and non-failing human hearts. Physiol Genomics 2003;12:97–112.

[13] Le Meur N, Lamirault G, Bihouée A, Steenman M, Bédrine-Ferran H, Teusan R, et al. A dynamic, web-accessible resource to process raw microarray scan data into consolidated gene expression values: importance of replication. Nucleic Acids Res 2004;32:5349–58.

[14] Tseng GC, Oh M-K, Rohlin L, Liao JC, Wong WH. Issues in cDNA microarray analysis: quality filtering, channel normalization, models of variations and assessment of gene effects. Nucleic Acids Res 2001;29:2549–57.

[15] Workman C, Jensen LJ, Jarmer H, Berka R, Gautier L, Nielser HB, et al. A new non-linear normalization method for reducing variability in DNA microarray experiments. Genome Biol 2002;3 research 0048.1–0048.16.

[16] Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. Proc Natl Acad Sci USA 2001;98:5116–21.

[17] Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat Appl Genet Mol Biol 2004;3:1–26.

[18] Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci USA 1998;95:14863–8.

[19] Thomas PD, Kejariwal A, Campbell MJ, Mi H, Diemer K, Guo N, et al. PANTHER: a browsable database of gene products organized by biological function, using curated protein family and subfamily classification. Nucleic Acids Res 2003;31:334–41.

[20] Winer J, Jung CK, Shackel I, Williams PM. Development and validation of real-time quantitative reverse transcriptase-polymerase chain reaction for monitoring gene expression in cardiac myocytes in vitro. Anal Biochem 1999;270:41–9.

[21] Eisenberg Jr. F. D-Myoinositol 1-phosphate as product of cyclization of glucose 6-phosphate and substrate for a specific phosphatase in rat testis. J Biol Chem 1967;242:1375–82.

[22] Maeda T, Eisenberg Jr. F. Purification, structure, and catalytic properties of L-myo-inositol-1-phosphate synthase from rat testis. J Biol Chem 1980;255:8458–64.

[23] Gaborit N, Steenman M, Lamirault G, Le Meur N, Le Bouter S, Lande G, et al. Human atrial ion channel and transporter subunit gene-expression remodeling associated with valvular heart disease and atrial fibrillation. Circulation 2005;112:471–81.

[24] Nakajima H, Nakajima HO, Salcher O, Dittie AS, Dembowsky K, Jing S, et al. Atrial but not ventricular fibrosis in mice expressing a mutant transforming growth factor-beta(1) transgene in the heart. Circ Res 2000;86:571–9.

[25] Verheule S, Sato T, Everett T, Engle SK, Otten D, Rubart-von der LM, et al. Increased vulnerability to atrial fibrillation in transgenic mice with selective atrial fibrosis caused by overexpression of TGF-beta1. Circ Res 2004;94:1458–65.

[26] Kostin S, Klein G, Szalay Z, Hein S, Bauer EP, Schaper J. Structural correlate of atrial fibrillation in human patients. Cardiovasc Res 2002; 54:361–79.

[27] Cha TJ, Ehrlich JR, Zhang L, Shi YF, Tardif JC, Leung TK, et al. Dissociation between ionic remodeling and ability to sustain atrial fibrillation during recovery from experimental congestive heart failure. Circulation 2004;109:412–8.

[28] Xu J, Cui G, Esmailian F, Plunkett M, Marelli D, Ardehali A, et al. Atrial extracellular matrix remodeling and the maintenance of atrial fibrillation. Circulation 2004;109:363–8.

[29] Amara FM, Entwistle J, Kuschak TI, Turley EA, Wright JA. Transforming growth factor-beta1 stimulates multiple protein interactions at a unique cis-element in the 3′-untranslated region of the hyaluronan receptor RHAMM mRNA. J Biol Chem 1996;271:15279–84.

[30] Billings PC, Whitbeck JC, Adams CS, Abrams WR, Cohen AJ, Engelsberg BN, et al. The transforming growth factor-beta-inducible matrix protein (beta)ig-h3 interacts with fibronectin. J Biol Chem 2002;277:28003–9.

[31] Bishop JE. Regulation of cardiovascular collagen deposition by mechanical forces. Mol Med Today 1998;4:69–75.

[32] Garg V, Kathiriya IS, Barnes R, Schluterman MK, King IN, Butler CA, et al. GATA4 mutations cause human congenital heart defects and reveal an interaction with TBX5. Nature 2003;424:443–7.

[33] Grepin C, Dagnino L, Robitaille L, Haberstroh L, Antakly T, Nemer M. A hormone-encoding gene identifies a pathway for cardiac but not skeletal muscle gene transcription. Mol Cell Biol 1994;14: 3115–29.

[34] Thuerauf DJ, Hanford DS, Glembotski CC. Regulation of rat brain natriuretic peptide transcription. A potential role for GATA-related transcription factors in myocardial cell gene expression. J Biol Chem 1994;269:17772–5.

[35] Murphy AM, Thompson WR, Peng LF, Jones L. Regulation of the rat cardiac troponin I gene by the transcription factor GATA-4. Biochem J 1997;322(Pt. 2):393–401.

[36] Di Lisi R, Millino C, Calabria E, Altruda F, Schiaffino S, Ausoni S. Combinatorial cis-acting elements control tissue-specific activation of the cardiac troponin I gene in vitro and in vivo. J Biol Chem 1998; 273:25371–80.

[37] Bhavsar PK, Dellow KA, Yacoub MH, Brand NJ, Barton PJ. Identification of cis-acting DNA elements required for expression of the human cardiac troponin I gene promoter. J Mol Cell Cardiol 2000;32: 95–108.

[38] De la Pompa JL, Timmerman LA, Takimoto H, Yoshida H, Elia AJ, Samper E, et al. Role of the NF-ATc transcription factor in morphogenesis of cardiac valves and septum. Nature 1998;392:182–6.

[39] Schubert W, Yang XY, Yang TT, Factor SM, Lisanti MP, Molkentin JD, et al. Requirement of transcription factor NFAT in developing atrial myocardium. J Cell Biol 2003;161:861–74.

[40] Choudhury A, Lip GY. Atrial fibrillation and the hypercoagulable state: from basic science to clinical practice. Pathophysiol Haemost Thromb 2003;33:282–9.

[41] Heppell RM, Berkin KE, McLenachan JM, Davies JA. Haemostatic and haemodynamic abnormalities associated with left atrial thrombosis in non-rheumatic atrial fibrillation. Heart 1997;77:407–11.

[42] Li-Saw-Hee FL, Blann AD, Lip GY. A cross-sectional and diurnal study of thrombogenesis among patients with chronic atrial fibrillation. J Am Coll Cardiol 2000;35:1926–31.

[43] Li-Saw-Hee FL, Blann AD, Gurney D, Lip GY. Plasma von Willebrand factor, fibrinogen and soluble P-selectin levels in paroxysmal, persistent and permanent atrial fibrillation. Effects of cardioversion and return of left atrial function. Eur Heart J 2001;22:1741–7.

[44] Nakamura Y, Nakamura K, Fukushima-Kusano K, Ohta K, Matsubara H, Hamuro T, et al. Tissue factor expression in atrial endothelia associated with nonvalvular atrial fibrillation: possible involvement in intracardiac thrombogenesis. Thromb Res 2003;111:137–42.

[45] Kumagai K, Fukuchi M, Ohta J, Baba S, Oda K, Akimoto H, et al. Expression of the von Willebrand factor in atrial endocardium is increased in atrial fibrillation depending on the extent of structural remodeling. Circ J 2004;68:321–7.

[46] Scott JL, Dunn SM, Jin B, Hillam AJ, Walton S, Berndt MC, et al. Characterization of a novel membrane glycoprotein involved in platelet activation. J Biol Chem 1989;264:13475–82.

[47] Anikster Y, Huizing M, White J, Shevchenko YO, Fitzpatrick DL, Touchman JW, et al. Mutation of a new gene causes a unique form of Hermansky–Pudlak syndrome in a genetic isolate of central Puerto Rico. Nat Genet 2001;28:376–80.

[48] Huizing M, Anikster Y, Fitzpatrick DL, Jeong AB, D'Souza M, Rausche M, et al. Hermansky–Pudlak syndrome type 3 in Ashkenazi Jews and other non-Puerto Rican patients with hypopigmentation and platelet storage-pool deficiency. Am J Hum Genet 2001;69:1022–32.

[49] Kawashima Y, Nagasawa T, Ninomiya H. Contribution of ecto-5′-nucleotidase to the inhibition of platelet aggregation by human endothelial cells. Blood 2000;96:2157–62.

[50] Cvirn G, Gallistl S, Muntean W. Alpha-2-macroglobulin inhibits the anticoagulant action of activated protein C in cord and adult plasma. Haemostasis 2001;31:1–11.

[51] Stearns-Kurosawa DJ, Kurosawa S, Mollica JS, Ferrell GL, Esmon CT. The endothelial cell protein C receptor augments protein C activation by the thrombin–thrombomodulin complex. Proc Natl Acad Sci USA 1996;93:10212–6.

[52] Lip GY. Does atrial fibrillation confer a hypercoagulable state? Lancet 1995;346:1313–4.

[53] Flaker GC, Belew K, Beckman K, Vidaillet H, Kron J, Safford R, et al. Asymptomatic atrial fibrillation: demographic features and prognostic information from the Atrial Fibrillation Follow-up Investigation of Rhythm Management (AFFIRM) study. Am Heart J 2005;149:657–63.

[54] Page RL, Wilkinson WE, Clair WK, McCarthy EA, Pritchett EL. Asymptomatic arrhythmias in patients with symptomatic paroxysmal atrial fibrillation and paroxysmal supraventricular tachycardia. Circulation 1994;89:224–7.

# CONCLUSION ET PERSPECTIVES

## I. Synthèse des résultats obtenus

### A. Insuffisance cardiaque

Le premier travail présenté dans l'insuffisance cardiaque a apporté une vision globale du remodelage du transcriptome cardiaque dans l'insuffisance cardiaque avancée. De nouveaux gènes différentiellement exprimés dans l'insuffisance cardiaque ont pu être identifiés. De manière plus large, une identification des circuits/fonctions moléculaires perturbés dans cette pathologie a été proposée. Ce travail a également identifié des profils d'expression génique différents entre les ventricules gauches et droits mais pas entre cardiopathies ischémiques et non-ischémiques. La taille réduite des populations dans notre étude explique probablement que nous n'ayons pas retrouvés de gènes différentiellement exprimés entre les différents étiologies de l'insuffisance cardiaque. Ces altérations transcriptomales liées à l'étiologie ont été depuis largement retrouvées dans différents formes de cardiopathies et à des stades plus précoces de la maladie.(Cunha-Neto et al, 2005; Hwang et al, 2002; Kittleson et al, 2005a; Wittchen et al, 2007)

Le second travail a montré l'existence pour chaque insuffisant cardiaque d'un portrait moléculaire individuel spécifique. Ces portraits ont pu être utilisés pour classifier les patients sur la base de leurs portraits individuels permettant d'individualiser 3 sous-groupes de patients. Les marqueurs moléculaires impliqués dans la classification des patients représentaient des éléments importants de la physiopathologie de la maladie comme les peptides natriurétiques, le métabolisme énergétique ou la structure/contraction cellulaire. De manière très intéressante, la classification des patients était corrélée avec leur statut d'urgence médicale UNOS (United Network for Organ Sharing), un paramètre de sévérité clinique de la maladie.

Ces résultats ont motivé la réalisation du troisième travail dans l'insuffisance cardiaque. Une signature moléculaire associée au niveau de sévérité clinique des

malades a pu être identifiée au sein des transcriptomes ventriculaires droits (129 gènes) et gauches (170 gènes) de 44 patients en insuffisance cardiaque avancée. La prédiction moléculaire du niveau de sévérité clinique du patient s'est révélée performante et reproductible pour différents échantillons de tissu prélevés au même patient. Une bonne corrélation entre les prédictions obtenues pour les ventricules droits et gauches était également notée.

L'ensemble de ces résultats et des résultats obtenus par d'autres équipes au cours de ces dernières années a montré le potentiel des études d'expression des gènes utilisant les puces à ADN pour l'étude de l'insuffisance cardiaque. Au niveau physiopathologique, une meilleure description des mécanismes moléculaires à la base du développement de l'insuffisance cardiaque pourra permettre l'identification de nouveaux mécanismes de la maladie, de mieux intégrer les distinctes composantes du remodelage myocardique et de définir de nouvelles cibles thérapeutiques. Au niveau clinique, ces résultats permettent d'espérer le développement d'outils moléculaires à but diagnostic pour identifier des étiologies spécifiques d'insuffisance cardiaque qui peuvent et pourront justifier d'un traitement spécifique. Enfin, nos résultats concernant le pronostic incitent à la réalisation d'études prospectives pour valider l'utilisation d'une signature moléculaire comme biomarqueur du pronostic dans l'insuffisance cardiaque humaine. Ces études démarrent dans l'insuffisance cardiaque avancée et seront certainement étendues à des stades plus précoces de la maladie.

### B. Fibrillation auriculaire

L'étude réalisée dans la fibrillation auriculaire a permis une description globale des mécanismes moléculaires associés à la fibrillation auriculaire chronique des patients atteints d'une cardiopathie valvulaire. Elle a bien distingué les parts respectives des altérations transcriptomales liées au substrat (valvulopathie) et au trouble du rythme, la majeure partie du remodelage correspondant au substrat.

L'identification des altérations transcriptomales associées à la fibrillation auriculaire permet de proposer des bases moléculaires à certains aspects du

remodelage auriculaire déjà bien identifiés au niveau fonctionnel et cellulaire. Les modifications d'expression de la voie du TGFβ (TGFβ1 et TGFβi) pourraient être impliquées dans le remodelage structural de l'oreillette. En effet, il a déjà été montré dans un modèle murin que la sur-expression de TGFB1 favorise le développement de fibrose dans le tissu auriculaire et l'apparition de la fibrillation auriculaire.(Verheule et al, 2004) Les altérations transcriptomales de PLN (phospholamban) ou TNNI3 (troponine i cardiaque) pourraient être associées au remodelage contractile du tissu auriculaire. Ces bases moléculaires permettront de développer des approches thérapeutiques nouvelles de la fibrillation auriculaire notamment centrées sur le remodelage non-électrique. Les inhibiteurs de l'enzyme de conversion pourraient être le premier exemple de ces nouvelles thérapeutiques.

Un résultat original de cette étude est la mise en évidence de modification d'expressions de gènes impliquées dans la coagulation plasmatique, la fibrinolyse et l'activation endothéliale. L'étude du remodelage transcriptomal global de ces gènes oriente vers l'existence d'un phénotype pro-coagulant dans le tissu atrial. Ce résultat est particulièrement intéressant car la fibrillation auriculaire est associée à un risque majeur d'accident vasculaire cérébral par mécanisme thrombo-embolique dont le point départ est principalement l'oreillette gauche. La stase sanguine provoquée par la dilatation des oreillettes(Cabin et al, 1990) et un état pro-thrombotique systémique sont connus pour favoriser ce risque d'accident vasculaire cérébral.(Becker, 2005) Nos résultats suggèrent qu'un troisième facteur local d'origine pariétal puisse également participer à la formation de thrombus dans l'oreillette gauche des patients en fibrillation auriculaire.

Pour poursuivre dans cette voie, nous avons initié une étude visant à mesurer les taux circulants des différentes protéines pro-thrombotiques produites par ces gènes, dans une population de patients valvulaires en fibrillation auriculaire permanente, une population de patients valvulaires en rythme sinusal et une population de patients sans cardiopathie. Les objectifs de cette étude sont de conforter l'hypothèse d'un rôle du remodelage atrial dans l'établissement de l'état pro-thrombotique associé à la fibrillation auriculaire et de mettre en évidence de nouveaux marqueurs moléculaires de cet état pro-thrombotique. Ces nouveaux marqueurs

pourraient ensuite être évalués comme biomarqueurs pronostiques du risque d'accident vasculaire cérébral chez les patients en arythmie.

## II. Perspectives

### A. Appréhender l'hétérogénéité du tissu cardiaque

L'ARN extrait d'un tissu ou même d'un organe entier reflète l'expression d'une grande variété de types cellulaires. Cette variété inclus parfois des tissus malades et sains. Les conclusions de ce type d'étude sont donc limitées par l'hétérogénéité du matériel de départ. Le signal d'expression altéré d'un groupe cellulaire donné peut ainsi être masqué par l'expression des autres groupes de cellules. Différentes solutions sont envisageables pour remédier à ce problème. L'utilisation de techniques de microdissection par capture laser sous microscope permet d'isoler spécifiquement des cellules sélectionnées.(Bonner et al, 1997) Une analyse transcriptomale spécifique de différents types cellulaires ou de différentes zones d'un tissu peut ainsi être réalisée.(Gabrielson et al, 2001) Il s'agit d'un travail fastidieux car il faut obtenir un grand nombre de cellules et la qualité de l'ARN doit être conservée. Une approche *in silico* est également possible. La microdissection virtuelle utilise les profils d'expression génique des lignées cellulaires correspondant le plus aux types cellulaires composant le tissu à analyser. Un gène dont le niveau d'expression est clairement plus important dans une des lignées par rapport aux autres sera supposé être lié du type cellulaire correspondant à cette lignée. La stratégie et de séparer la matrice d'expression produite sur le tissu entier en grands sous-groupes correspondant aux différents types cellulaires composant le tissu analysé. Le point crucial est de disposer de lignées cellulaires humaines représentatives des types cellulaires étudiés. L'application de cette méthode en cancérologie(Ross et al, 2000) et en immunologie(Shaffer et al, 2001) a apporté des résultats convaincants.

### B. Quels échantillons biologiques pour analyser le transcriptome cardiaque humain ?

La réalisation de grandes collections d'objets biologiques à des fins de recherche est actuellement en cours à l'échelle des nations (www.ukbiobank.ac.uk) et de l'Europe (www.eurobiobank.org/index.htm). La résolution de problèmes d'organisation, de coordination multidisciplinaire et surtout d'annotation des prélèvements est essentielle. Le choix des échantillons à inclure dans ces banques pour la recherche cardiovasculaire est une question qui doit être abordée au plus vite. L'accès à des échantillons biologiques humains à but de recherche soulève des obstacles éthiques et de faisabilité dont l'importance est assez variable suivant le tissu étudié et son mode de prélèvement. Les prélèvements de tissu cardiaque sont certainement parmi ceux qui posent le plus de problèmes. Le fonctionnement particulier de cet organe et sa localisation interne en sont les raisons principales.

L'exérèse tissulaire chirurgicale à titre thérapeutique est un des moyens d'obtenir des échantillons biologiques humains. S'il est assez peu problématique en cancérologie de recueillir des extraits tumoraux, l'obtention d'échantillons cardiaques ou vasculaires peut sembler beaucoup moins facile et moins pertinente. La chirurgie cardiaque et vasculaire réalise des exérèses tissulaires. Les tissus cardiaques ainsi collectés expriment souvent un phénotype final des maladies ne permettant pas toujours la distinction entre causes et conséquences du processus pathologique au sein des altérations d'expression géniques observées. Par ailleurs, des prélèvements à ce stade ultime des maladies présentent peu d'intérêt en pratique clinique. L'étude de tissu à un stade moins évolué des maladies apparaît comme un élément important. Elle permettrait de s'approcher d'une vision des mécanismes « initiateurs » des maladies et d'apporter des éléments de physiopathologie non encore analysés au plan transcriptomal. Cette approche physiopathologique pourrait surtout être associée à un objectif de recherche clinique. Concernant le développement de biomarqueurs d'intérêt thérapeutique, l'étude du profil d'expression génique par un prélèvement tissulaire au moment du diagnostic d'une pathologie pourrait avoir un impact réel sur les choix thérapeutiques. Ceux-ci pourraient ainsi être adaptés au phénotype moléculaire propre à chaque patient.

En cardiologie, l'utilisation de biopsies endomyocardiques est une possibilité. Plusieurs études ont démontré que l'utilisation de biopsies permettait d'obtenir assez d'ARN pour la réalisation d'études d'expression génique.(Grzeskowiak et al, 2003; Kittleson et al, 2004) Cependant, le prélèvement de biopsies endomyocardiques dans l'évaluation des patients en insuffisance cardiaque n'est pas un geste systématique. Sa réalisation par une équipe entraînée présent un risque de complications équivalent aux autres procédures de cathétérisme cardiaque.(Felker et al, 1999) Cette stratégie semble intéressante pour une approche de la physiopathologie des cardiomyopathies. Elle n'est en revanche pas idéale pour le développement de biomarqueurs des maladies cardiaques. Il est en effet difficile d'envisager des prélèvements myocardiques itératifs pour la surveillance des patients.

A contrario, la réalisation de prélèvement de sang périphérique est réalisée en routine dans toute structure hospitalière comme au domicile des patients. L'utilisation de biomarqueurs sanguins parait donc une solution plus applicable que l'utilisation de biopsies endomyocardiques pour le développement d'outils pronostiques de pathologies cardiaques. Cette alternative d'analyse de profil d'expression génique dans le sang représente aujourd'hui une alternative prometteuse pour le développement d'outils cliniques. En 1999, Liew *et al.* ont proposé que le sang circulant puisse être un 'tissu sentinelle' reflétant le fonctionnement normal ou pathologique des différentes structures du corps humain.(Liew et al, 2006) Le sang, à travers la circulation coronaire, est en contact étroit avec la majorité des différents organes dont le cœur. Du fait de ces interactions permanentes, le remodelage myocardique associé à l'insuffisance cardiaque pourrait entraîner dans les cellules sanguines des modifications d'expression génique spécifiques de la maladie.

Sur la base de cette approche sentinelle, des études ont été réalisées sur le niveau d'expression des gènes dans les cellules sanguines concernant différentes pathologies chez l'homme. Ces études ont montré la présence de « profils d'expression génique sanguins » qui caractérisent des patients atteints de maladies aussi différentes que les cancers,(Twine et al, 2003) les maladies mentales(Tsuang et al, 2005) et, concernant les pathologies cardiovasculaires, l'hypertension artérielle pulmonaire(Bull et al, 2004) et la cardiopathie ischémique.(Ma et al, 2003) Aucun résultat n'est encore publié dans l'insuffisance cardiaque ou la fibrillation auriculaire.

Cependant il a déjà été montré dans l'insuffisance cardiaque que la transcription de l'échangeur sodium/calcium augmente dans les globules blancs parallèlement à celle des tissus cardiaques.(Seiler et al, 2004) Cette piste un peu inattendue mais tout à fait sérieuse est en cours d'évaluation. Deux études d'expression génique des globules blancs sont actuellement en cours dans notre laboratoire sur l'insuffisance cardiaque.

### C. Analyse des micro-RNA

Les ARN messager ne sont pas les seuls ARN impliqués dans l'expression des gènes. D'autres ARN qui ne donnent pas lieu à la traduction en protéines peuvent réguler de manière spécifique l'expression des gènes. Parmi ceux-ci, les micro ARN (miRNA) sont des petits ARN qui ne codent pas de protéines mais interagissent spécifiquement avec certains ARNm pour réguler leur traduction en protéines. Leur taille moyenne est de 22 nucléotides. Ile se fixent sur des zones spécifiques de l'ARNm selon le principe d'hybridation spécifique de 2 séquences nucléotidiques complémentaires. Cette liaison est responsable soit de la dégradation de l'ARNm soit de l'inhibition de la traduction en protéines de l'ARNm. La régulation du devenir des ARNm par les miRNA est complexe car plusieurs miRNA peuvent se fixer sur un même ARNm et un miRNA peut se fixer sur plusieurs ARNm. Le rôle des miRNA dans les pathologies cardiovasculaires restent encore largement inexploré. Quelques publications récentes rapportent cependant des données intéressantes dans le remodelage hypertrophique et l'insuffisance cardiaque. Le rôle d'un premier miRNA a été montré dans le contrôle du processus de remodelage hypertrophique.(van Rooij et al, 2007) Une autre publication a étudié le niveau d'expression de profil d'expression génique des miRNA humains associé à l'insuffisance cardiaque.(Thum et al, 2007) Un ensemble de 110 miRNA était différentiellement exprimé dans les tissus cardiaque défaillants. Ce profil d'expression était par ailleurs très proche du profil d'expression des miRNA observé dans des cardiomyocytes fœtaux. Ces résultats suggèrent d'une part un lien entre remodelage transcriptomal des miRNA et réexpression des gènes de la vie fœtale dans l'insuffisance cardiaque humaine ce qui pourrait déboucher sur de nouvelles pistes physiopathologiques et thérapeutiques si ce remodelage au niveau des miRNA apparaissait être un phénomène initiateur

majeur du remodelage ventriculaire. D'autre part, l'intérêt de cette signature moléculaire comme biomarqueur pourrait également être évalué.

### D. Analyse des facteurs de transcription :

Les facteurs de transcription ont été clairement impliqués dans le développement cardiaque ou le remodelage hypertrophique chez l'animal.(Olson, 2004) Cependant le manque d'outils moléculaires a jusqu'ici bridé une étude plus large du rôle des facteurs de transcription dans les maladies cardiovasculaires chez l'homme. Tirant parti du séquençage et de l'annotation du génome humain et de la production de profils d'expression génique dans l'insuffisance cardiaque, une première étude a confirmé *in silico* l'importance de facteurs de transcription des familles NF-AT, MEF2, NKX, GATA ou FOX dans l'insuffisance cardiaque humaine.(Hannenhalli et al, 2006) Le développement de la méthode du ChIP-on-chip a représenté un saut technologique important permettant aujourd'hui une identification directe des gènes-cibles d'un facteur de transcription dans une situation biologique donnée.(Kirmizis et al, 2004)

La première étape de cette technique est une immuno-précipitation de la chromatine (**Ch**romatin **I**mmuno-**P**recipitation) issue du tissu d'intérêt avec un anticorps dirigé contre le facteur de transcription analysé. La seconde étape consiste à considérer les séquences de nucléotides obtenues par la première étape comme une cible qui sera hybridée sur une puce à ADN (**chip**) contenant un ensemble de sondes représentatives des régions promotrices du génome. La réalisation de ces études dans différentes situations pathologiques pourra permettre d'identifier des facteurs de transcription liés aux cibles (gènes) d'intérêt dans la maladie. L'inactivation de ces facteurs de transcription clés devrait se révéler une stratégie thérapeutique intéressante car directement ciblée sur le déclenchement des mécanismes pathologiques.

E. Analyse de la méthylation des histones

La régulation de l'expression des gènes est sous la dépendance de nombreux mécanismes comme l'action des facteurs de transcription ou les modifications des histones. Les histones sont des protéines organisées en complexes octamériques. Elles se lient à l'ADN pour former un nucléosome. Les histones subissent de nombreuses modifications post-traductionnelles qui vont influencer l'expression des gènes. Deux articles publiés récemment ont étudiés à l'échelle pangénomique la relation entre différents types de méthylations des histones et l'organisation du génome ainsi que l'expression des gènes.(Barski et al, 2007; Mikkelsen et al, 2007) Parmi les résultats obtenus, les auteurs ont pu identifier des états de méthylation liés aux régions promotrices des gènes. Ce résultat est particulièrement intéressant pour l'indentification parfois difficile des séquences promotrices de certains gènes ou de promoteurs alternatifs. Ils ont également mis en évidence des états de méthylation correspondant à différents états d'activité des promoteurs (activés, réprimés, ou en attente). Ces résultats vont permettre d'identifier et de comparer dans différents situations physiopathologiques les gènes dont l'expression est activée ou réprimée. Il ne s'agira pas d'observer les différences entre deux états du transcriptome mais la dynamique de changement du transcriptome survenant en temps réel dans une situation biologique donnée. Cette nouvelle approche pourrait aboutir à une meilleure définition des variations transcriptomales pertinentes dans une situation biologique donnée.

L'ensemble des travaux présentés dans cette thèse a confirmé l'intérêt de l'étude des variations transcriptomales dans les maladies cardiaques. La mise en place d'approches méthodologiques rigoureuses et la poursuite du développement des outils bio-informatiques d'analyse des masses de données produites sont des objectifs essentiels pour améliorer l'utilisation des puces à ADN. La poursuite des travaux déjà engagés devrait non seulement permettre une meilleur compréhension des mécanismes moléculaires des pathologies mais également l'identification de biomarqueurs à but diagnostic, pronostic ou d'orientation thérapeutique.

# REFERENCES BIBLIOGRAPHIQUES


't Veer LJ, Dai H, van de Vijver MJ et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature*. 2002;415:530-536.

Aaronson KD, Schwartz JS, Chen TM et al. Development and prospective validation of a clinical index to predict survival in ambulatory patients referred for cardiac transplant evaluation. *Circulation*. 1997;95:2660-2667.

Adams MD, Kelley JM, Gocayne JD et al. Complementary DNA sequencing: expressed sequence tags and human genome project. *Science*. 1991;252:1651-1656.

Aime-Sempe C, Folliguet T, Rucker-Martin C et al. Myocardial cell death in fibrillating and dilated human right atria. *J Am Coll Cardiol*. 1999;34:1577-1586.

Al Shahrour F, Diaz-Uriarte R, Dopazo J. FatiGO: a web tool for finding significant associations of Gene Ontology terms with groups of genes. *Bioinformatics*. 2004;20:578-580.

Alizadeh AA, Eisen MB, Davis RE et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature*. 2000;403:503-511.

Allessie M, Ausma J, Schotten U. Electrical, contractile and structural remodeling during atrial fibrillation. *Cardiovasc Res*. 2002;54:230-246.

Alwine JC, Kemp DJ, Stark GR. Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes. *Proc Natl Acad Sci U S A*. 1977;74:5350-5354.

Antos CL, Frey N, Marx SO et al. Dilated cardiomyopathy and sudden death resulting from constitutive activation of protein kinase a. *Circ Res*. 2001;89:997-1004.

Ashburner M, Ball CA, Blake JA et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet*. 2000;25:25-29.

Ballester B, Ramuz O, Gisselbrecht C et al. Gene expression profiling identifies molecular subgroups among nodal peripheral T-cell lymphomas. *Oncogene*. 2006;25:1560-1570.

Barrans JD, Stamatiou D, Liew C. Construction of a human cardiovascular cDNA microarray: portrait of the failing heart. *Biochem Biophys Res Commun*. 2001;280:964-969.

Barski A, Cuddapah S, Cui K et al. High-resolution profiling of histone methylations in the human genome. *Cell*. 2007;129:823-837.

Barth AS, Merk S, Arnoldi E et al. Reprogramming of the human atrial transcriptome in permanent atrial fibrillation: expression of a ventricular-like genomic signature. *Circ Res*. 2005;96:1022-1029.

Becker RC. Biomarkers in atrial fibrillation: investigating biologic plausibility, cause, and effect. *J Thromb Thrombolysis*. 2005;19:71-75.

Benjamin EJ, Wolf PA, D'Agostino RB et al. Impact of atrial fibrillation on the risk of death: the Framingham Heart Study. *Circulation*. 1998;98:946-952.

Berk AJ, Sharp PA. Sizing and mapping of early adenovirus mRNAs by gel electrophoresis of S1 endonuclease-digested hybrids. *Cell*. 1977;12:721-732.

Bertucci F, Salas S, Eysteries S et al. Gene expression profiling of colon cancer by DNA microarrays and correlation with histoclinical parameters. *Oncogene*. 2004;23:1377-1391.

Bhatia RS, Tu JV, Lee DS et al. Outcome of heart failure with preserved ejection fraction in a population-based study. *N Engl J Med*. 2006;355:260-269.

Biomarkers Working Group. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clin Pharmacol Ther*. 2001;69:89-95.

Birks EJ, Tansley PD, Hardy J et al. Left ventricular assist device and drug therapy for the reversal of heart failure. *N Engl J Med*. 2006;355:1873-1884.

Blaxall BC, Tschannen-Moran BM, Milano CA et al. Differential gene expression and genomic patient stratification following left ventricular assist device support. *J Am Coll Cardiol*. 2003;41:1096-1106.

Bonner RF, Emmert-Buck M, Cole K et al. Laser capture microdissection: molecular analysis of tissue. *Science*. 1997;278:1481,1483.

Bouzamondo A, Hulot JS, Sanchez P et al. Beta-blocker benefit according to severity of heart failure. *Eur J Heart Fail*. 2003;5:281-289.

Brand NJ, Barton PJ. Myocardial molecular biology: an introduction. *Heart*. 2002;87:284-293.

Braunwald E. Shattuck lecture--cardiovascular medicine at the turn of the millennium: triumphs, concerns, and opportunities. *N Engl J Med*. 1997;337:1360-1369.

Brown PO, Botstein D. Exploring the new world of the genome with DNA microarrays. *Nat Genet*. 1999;21:33-37.

Bull TM, Coldren CD, Moore M et al. Gene microarray analysis of peripheral blood cells in pulmonary arterial hypertension. *Am J Respir Crit Care Med*. 2004;170:911-919.

Burke S. Missing Values, Outliers, Robust Statistics & Non-parametric Methods. Statistics and data analysis. 2001.

Cabin HS, Clubb KS, Hall C et al. Risk for systemic embolization of atrial fibrillation without mitral stenosis. *Am J Cardiol*. 1990;65:1112-1116.

Chen HY, Yu SL, Chen CH et al. A five-gene signature and clinical outcome in non-small-cell lung cancer. *N Engl J Med*. 2007;356:11-20.

Chen MM, Ashley EA, Deng DX et al. Novel role for the potent endogenous inotrope apelin in human cardiac dysfunction. *Circulation*. 2003a;108:1432-1439.

Chen Y, Park S, Li Y et al. Alterations of gene expression in failing myocardium following left ventricular assist device support. *Physiol Genomics*. 2003b;14:251-260.

Chomczynski P, Sacchi N. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal Biochem*. 1987;162:156-159.

Christenson RH. National Academy of Clinical Biochemistry Laboratory Medicine Practice Guidelines for Utilization of Biochemical Markers in Heart Failure: Preamble. *Circulation*. 2007;116:e94.

Cleland JG, Daubert JC, Erdmann E et al. The effect of cardiac resynchronization on morbidity and mortality in heart failure. *N Engl J Med*. 2005;352:1539-1549.

Conway DS, Pearce LA, Chin BS et al. Prognostic value of plasma von Willebrand factor and soluble P-selectin as indices of endothelial damage and platelet activation in 994 patients with nonvalvular atrial fibrillation. *Circulation*. 2003;107:3141-3145.

Cox JL, Schuessler RB, Lappas DG et al. An 8 1/2-year clinical experience with surgery for atrial fibrillation. *Ann Surg*. 1996;224:267-273.

Cummins B, Auckland ML, Cummins P. Cardiac-specific troponin-I radioimmunoassay in the diagnosis of acute myocardial infarction. *Am Heart J*. 1987;113:1333-1344.

Cunha-Neto E, Dzau VJ, Allen PD et al. Cardiac gene expression profiling provides evidence for cytokinopathy as a molecular mechanism in Chagas' disease cardiomyopathy. *Am J Pathol*. 2005;167:305-313.

Delahaye F, Roth O, Aupetit JF et al. [Epidemiology and prognosis of cardiac insufficiency]. *Arch Mal Coeur Vaiss*. 2001;94:1393-1403.

Dempsey AA, Ton C, Liew CC. A cardiovascular EST repertoire: progress and promise for understanding cardiovascular disease. *Mol Med Today*. 2000;6:231-237.

DeRisi JL, Iyer VR, Brown PO. Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science*. 1997;278:680-686.

Diatchenko L, Lau YF, Campbell AP et al. Suppression subtractive hybridization: a method for generating differentially regulated or tissue-specific cDNA probes and libraries. *Proc Natl Acad Sci U S A*. 1996;93:6025-6030.

Dobrev D, Ravens U. Remodeling of cardiomyocyte ion channels in human atrial fibrillation. *Basic Res Cardiol*. 2003;98:137-148.

Dorn GW, Souroujon MC, Liron T et al. Sustained in vivo cardiac protection by a rationally designed peptide that causes epsilon protein kinase C translocation. *Proc Natl Acad Sci U S A*. 1999;96:12798-12803.

Draghici S. Data Analysis Tools for DNA Microarrays. First Edition. Chapman & Hall, Boca Raton, Florida, 2003.

Draghici S, Khatri P, Bhavsar P et al. Onto-Tools, the toolkit of the modern biologist: Onto-Express, Onto-Compare, Onto-Design and Onto-Translate. *Nucleic Acids Res*. 2003;31:3775-3781.

DREES. Données sur la situation sanitaire et sociale en France en 2005. 2005.

Echt DS, Liebson PR, Mitchell LB et al. Mortality and morbidity in patients receiving encainide, flecainide, or placebo. The Cardiac Arrhythmia Suppression Trial. *N Engl J Med*. 1991;324:781-788.

Eisen MB, Spellman PT, Brown PO et al. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci U S A*. 1998;95:14863-14868.

Esposito G, Rapacciuolo A, Naga Prasad SV et al. Genetic alterations that inhibit in vivo pressure-overload hypertrophy prevent cardiac dysfunction despite increased wall stress. *Circulation*. 2002;105:85-92.

Falk RH. Atrial fibrillation. *N Engl J Med*. 2001;344:1067-1078.

Faller D, Voss HU, Timmer J et al. Normalization of DNA-microarray data by nonlinear correlation maximization. *J Comput Biol*. 2003;10:751-762.

Feinberg WM, Blackshear JL, Laupacis A et al. Prevalence, age distribution, and gender of patients with atrial fibrillation. Analysis and implications. *Arch Intern Med*. 1995;155:469-473.

Feldman AM, Wagner DR, McNamara DM. AMPD1 gene mutation in congestive heart failure: new insights into the pathobiology of disease progression. *Circulation*. 1999;99:1397-1399.

Felker GM, Hu W, Hare JM et al. The spectrum of dilated cardiomyopathy. The Johns Hopkins experience with 1,278 patients. *Medicine (Baltimore)*. 1999;78:270-283.

Feuk L, Carson AR, Scherer SW. Structural variation in the human genome. *Nat Rev Genet*. 2006;7:85-97.

Francis GS. Pathophysiology of chronic heart failure. *Am J Med*. 2001;110 Suppl 7A:37S-46S.

Fuster V, Ryden LE, Cannom DS et al. ACC/AHA/ESC 2006 guidelines for the management of patients with atrial fibrillation-executive summary: a report of the American College of Cardiology/American Heart Association Task Force on practice guidelines and the European Society of Cardiology Committee for Practice Guidelines (Writing Committee to Revise the 2001 Guidelines for the Management of Patients with Atrial Fibrillation). *Eur Heart J*. 2006;27:1979-2030.

Gabrielson E, Berg K, Anbazhagan R. Functional genomics, gene arrays, and the future of pathology. *Mod Pathol*. 2001;14:1294-1299.

Garg R, Yusuf S. Overview of randomized trials of angiotensin-converting enzyme inhibitors on mortality and morbidity in patients with heart failure. Collaborative Group on ACE Inhibitor Trials. *JAMA*. 1995;273:1450-1456.

Goette A, Staack T, Rocken C et al. Increased expression of extracellular signal-regulated kinase and angiotensin-converting enzyme in human atria during atrial fibrillation. *J Am Coll Cardiol*. 2000;35:1669-1677.

Grzeskowiak R, Witt H, Drungowski M et al. Expression profiling of human idiopathic dilated cardiomyopathy. *Cardiovasc Res*. 2003;59:400-411.

Hacia JG. Resequencing and mutational analysis using oligonucleotide microarrays. *Nat Genet*. 1999;21:42-47.

Haissaguerre M, Jais P, Shah DC et al. Spontaneous initiation of atrial fibrillation by ectopic beats originating in the pulmonary veins. *N Engl J Med*. 1998;339:659-666.

Hall JL, Grindle S, Han X et al. Genomic profiling of the human heart before and after mechanical support with a ventricular assist device reveals alterations in vascular signaling networks. *Physiol Genomics*. 2004;17:283-291.

Hannenhalli S, Putt ME, Gilmore JM et al. Transcriptional genomics associates FOX transcription factors with human heart failure. *Circulation*. 2006;114:1269-1276.

HGP. Finishing the euchromatic sequence of the human genome. *Nature*. 2004;431:931-945.

Hieter P, Boguski M. Functional genomics: it's all how you read it. *Science*. 1997;278:601-602.

Hunt SA, Abraham WT, Chin MH et al. ACC/AHA 2005 Guideline Update for the Diagnosis and Management of Chronic Heart Failure in the Adult: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Writing Committee to Update the 2001 Guidelines for the Evaluation and Management of Heart Failure): developed in collaboration with the American College

of Chest Physicians and the International Society for Heart and Lung Transplantation: endorsed by the Heart Rhythm Society. *Circulation*. 2005;112:e154-e235.

Hunt SA, Baker DW, Chin MH et al. ACC/AHA guidelines for the evaluation and management of chronic heart failure in the adult: executive summary. A report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Committee to revise the 1995 Guidelines for the Evaluation and Management of Heart Failure). *J Am Coll Cardiol*. 2001;38:2101-2113.

Hunter JJ, Chien KR. Signaling pathways for cardiac hypertrophy and failure. *N Engl J Med*. 1999;341:1276-1283.

Huss JM, Kelly DP. Mitochondrial energy metabolism in heart failure: a question of balance. *J Clin Invest*. 2005;115:547-555.

Hwang JJ, Allen PD, Tseng GC et al. Microarray gene expression profiles in dilated and hypertrophic cardiomyopathic end-stage heart failure. *Physiol Genomics*. 2002;10:31-44.

Jenssen TK, Laegreid A, Komorowski J et al. A literature network of human genes for high-throughput analysis of gene expression. *Nat Genet*. 2001;28:21-28.

Jong P, Demers C, McKelvie RS et al. Angiotensin receptor blockers in heart failure: meta-analysis of randomized controlled trials. *J Am Coll Cardiol*. 2002;39:463-470.

Kerr MK, Martin M, Churchill GA. Analysis of variance for gene expression microarray data. *J Comput Biol*. 2000;7:819-837.

Kim YH, Lim DS, Lee JH et al. Gene expression profiling of oxidative stress on atrial fibrillation in humans. *Exp Mol Med*. 2003;35:336-349.

Kirmizis A, Farnham PJ. Genomic approaches that aid in the identification of transcription factor target genes. *Exp Biol Med (Maywood )*. 2004;229:705-721.

Kittleson MM, Minhas KM, Irizarry RA et al. Gene expression analysis of ischemic and nonischemic cardiomyopathy: shared and distinct genes in the development of heart failure. *Physiol Genomics*. 2005a;21:299-307.

Kittleson MM, Minhas KM, Irizarry RA et al. Gene expression in giant cell myocarditis: Altered expression of immune response genes. *Int J Cardiol*. 2005b;102:333-340.

Kittleson MM, Ye SQ, Irizarry RA et al. Identification of a gene expression profile that differentiates between ischemic and nonischemic cardiomyopathy. *Circulation*. 2004;110:3444-3451.

Kuhl U, Pauschinger M, Schwimmbeck PL et al. Interferon-beta treatment eliminates cardiotropic viruses and improves left ventricular function in patients with myocardial persistence of viral genomes and left ventricular dysfunction. *Circulation*. 2003;107:2793-2798.

Kuhl U, Pauschinger M, Seeberg B et al. Viral persistence in the myocardium is associated with progressive cardiac dysfunction. *Circulation*. 2005;112:1965-1970.

Kumagai K, Fukunami M, Ohmori M et al. Increased intracardiovascular clotting in patients with chronic atrial fibrillation. *J Am Coll Cardiol*. 1990;16:377-380.

Kurose M, Yoshimura M, Yasue H. Raised plasma BNP in a patient with acute pulmonary thromboembolism. *Heart*. 1997;78:320-321.

Lander ES. Array of hope. *Nat Genet*. 1999;21:3-4.

Lander ES, Linton LM, Birren B et al. Initial sequencing and analysis of the human genome. *Nature*. 2001;409:860-921.

Le Bouter S, Demolombe S, Chambellan A et al. Microarray analysis reveals complex remodeling of cardiac ion channel expression with altered thyroid status: relation to cellular and integrated electrophysiology. *Circ Res*. 2003;92:234-242.

Lee ML, Kuo FC, Whitmore GA et al. Importance of replication in microarray gene expression studies: statistical methods and evidence from repetitive cDNA hybridizations. *Proc Natl Acad Sci U S A*. 2000;97:9834-9839.

Levy S. Drug Insight: angiotensin-converting-enzyme inhibitors and atrial fibrillation--indications and contraindications. *Nat Clin Pract Cardiovasc Med*. 2006;3:220-225.

Levy WC, Mozaffarian D, Linker DT et al. The Seattle Heart Failure Model: prediction of survival in heart failure. *Circulation*. 2006;113:1424-1433.

Li-Saw-Hee FL, Blann AD, Lip GY. A cross-sectional and diurnal study of thrombogenesis among patients with chronic atrial fibrillation. *J Am Coll Cardiol*. 2000;35:1926-1931.

Liang P, Pardee AB. Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. *Science*. 1992;257:967-971.

Liew CC, Ma J, Tang HC et al. The peripheral blood transcriptome dynamically reflects system wide biology: a potential diagnostic tool. *J Lab Clin Med*. 2006;147:126-132.

Lloyd-Jones DM, Larson MG, Leip EP et al. Lifetime risk for developing congestive heart failure: the Framingham Heart Study. *Circulation*. 2002;106:3068-3072.

Lockhart DJ, Winzeler EA. Genomics, gene expression and DNA arrays. *Nature*. 2000;405:827-836.

Lusis AJ, Mar R, Pajukanta P. Genetics of atherosclerosis. *Annu Rev Genomics Hum Genet*. 2004;5:189-218.

Ma C, Lyons-Weiler M, Liang W et al. In vitro transcription amplification and labeling methods contribute to the variability of gene expression profiling with DNA microarrays. *J Mol Diagn*. 2006;8:183-192.

Ma J, Liew CC. Gene profiling identifies secreted protein transcripts from peripheral blood cells in coronary artery disease. *J Mol Cell Cardiol*. 2003;35:993-998.

Mann DL. MicroRNAs and the failing heart. *N Engl J Med*. 2007;356:2644-2645.

Margulies KB, Matiwala S, Cornejo C et al. Mixed messages: transcription patterns in failing and recovering human myocardium. *Circ Res*. 2005;96:592-599.

Mehra MR, Uber PA, Park MH et al. Obesity and suppressed B-type natriuretic peptide levels in heart failure. *J Am Coll Cardiol*. 2004;43:1590-1595.

Middlekauff HR, Stevenson WG, Stevenson LW. Prognostic significance of atrial fibrillation in advanced heart failure. A study of 390 patients. *Circulation*. 1991;84:40-48.

Mikkelsen TS, Ku M, Jaffe DB et al. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature*. 2007;448:553-560.

Ministère de l'emploi et de la solidarité. Programme national de réduction des risques cardiovasculaires 2002-2005. 2002.

Mochly-Rosen D, Wu G, Hahn H et al. Cardiotrophic effects of protein kinase C epsilon: analysis by in vivo modulation of PKCepsilon translocation. *Circ Res*. 2000;86:1173-1179.

Moss AJ, Zareba W, Hall WJ et al. Prophylactic implantation of a defibrillator in patients with myocardial infarction and reduced ejection fraction. *N Engl J Med*. 2002;346:877-883.

Mosterd A, Hoes AW, de Bruyne MC et al. Prevalence of heart failure and left ventricular dysfunction in the general population; The Rotterdam Study. *Eur Heart J*. 1999;20:447-455.

Nanni L, Romualdi C, Maseri A et al. Differential gene expression profiling in genetic and multifactorial cardiovascular diseases. *J Mol Cell Cardiol*. 2006;41:934-948.

Nattel S. New ideas about atrial fibrillation 50 years on. *Nature*. 2002;415:219-226.

Ohki R, Yamamoto K, Ueno S et al. Gene expression profiling of human atrial myocardium with atrial fibrillation by DNA microarray analysis. *Int J Cardiol*. 2005;102:233-238.

Okubo K, Hori N, Matoba R et al. Large scale cDNA sequencing for analysis of quantitative and qualitative aspects of gene expression. *Nat Genet*. 1992;2:173-179.

Olson EN. A decade of discoveries in cardiac biology. *Nat Med*. 2004;10:467-474.

Owan TE, Hodge DO, Herges RM et al. Trends in prevalence and outcome of heart failure with preserved ejection fraction. *N Engl J Med*. 2006;355:251-259.

Palmer BM. Thick filament proteins and performance in human heart failure. *Heart Fail Rev*. 2005;10:187-197.

Pan W, Lin J, Le CT. How many replicates of arrays are required to detect gene expression changes in microarray experiments? A mixture model approach. *Genome Biol*. 2002;3:research0022.

Pandey A, Mann M. Proteomics to study genes and genomes. *Nature*. 2000;405:837-846.

Pavlidis P. Using ANOVA for gene selection from microarray studies of the nervous system. *Methods*. 2003;31:282-289.

Perou CM, Sorlie T, Eisen MB et al. Molecular portraits of human breast tumours. *Nature*. 2000;406:747-752.

Pfeffer MA, Braunwald E. Ventricular remodeling after myocardial infarction. Experimental observations and clinical implications. *Circulation*. 1990;81:1161-1172.

Quackenbush J. Microarray data normalization and transformation. *Nat Genet*. 2002;32 Suppl:496-501.

Rose EA, Gelijns AC, Moskowitz AJ et al. Long-term mechanical left ventricular assistance for end-stage heart failure. *N Engl J Med*. 2001;345:1435-1443.

Ross DT, Scherf U, Eisen MB et al. Systematic variation in gene expression patterns in human cancer cell lines. *Nat Genet*. 2000;24:227-235.

Roy D, Talajic M, Dorian P et al. Amiodarone to prevent recurrence of atrial fibrillation. Canadian Trial of Atrial Fibrillation Investigators. *N Engl J Med*. 2000;342:913-920.

Schroeder A, Mueller O, Stocker S et al. The RIN: an RNA integrity number for assigning integrity values to RNA measurements. *BMC Mol Biol*. 2006;7:3.

Seiler PU, Stypmann J, Breithardt G et al. Real-time RT-PCR for gene expression profiling in blood of heart failure patients-a pilot study: gene expression in blood of heart failure patients. *Basic Res Cardiol*. 2004;99:230-238.

Seo D, Ginsburg GS, Goldschmidt-Clermont PJ. Gene expression analysis of cardiovascular diseases: novel insights into biology and clinical applications. *J Am Coll Cardiol*. 2006;48:227-235.

Shaffer AL, Rosenwald A, Hurt EM et al. Signatures of the immune response. *Immunity*. 2001;15:375-385.

Shibata R, Sato K, Pimentel DR et al. Adiponectin protects against myocardial ischemia-reperfusion injury through. *Nat Med*. 2005;11:1096-1103.

Simon RM, Dobbin K. Experimental design of DNA microarray experiments. *Biotechniques*. 2003;Suppl:16-21.

Sing CF, Stengard JH, Kardia SL. Genes, environment, and cardiovascular disease. *Arterioscler Thromb Vasc Biol*. 2003;23:1190-1196.

Smits JM, Deng MC, Hummel M et al. A prognostic model for predicting waiting-list mortality for a total national cohort of adult heart-transplant candidates. *Transplantation*. 2003;76:1185-1189.

Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol*. 2004;3:Article3.

Sorlie T, Perou CM, Tibshirani R et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A*. 2001;98:10869-10874.

Southern EM. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J Mol Biol*. 1975;98:503-517.

Steenman M, Chen YW, Le Cunff M et al. Transcriptomal analysis of failing and nonfailing human hearts. *Physiol Genomics*. 2003;12:97-112.

Steenman M, Lamirault G, Le Meur N et al. Gene expression profiling in human cardiovascular disease. *Clin Chem Lab Med*. 2005;43:696-701.

Stewart S, MacIntyre K, Hole DJ et al. More 'malignant' than cancer? Five-year survival following a first admission for heart failure. *Eur J Heart Fail*. 2001;3:315-322.

Storey JD, Tibshirani R. Statistical methods for identifying differentially expressed genes in DNA microarrays. *Methods Mol Biol*. 2003;224:149-157.

Sturn A, Quackenbush J, Trajanoski Z. Genesis: cluster analysis of microarray data. *Bioinformatics*. 2002;18:207-208.

Swedberg K, Cleland J, Dargie H et al. Guidelines for the diagnosis and treatment of chronic heart failure: executive summary (update 2005): The Task Force for the Diagnosis and Treatment of Chronic Heart Failure of the European Society of Cardiology. *Eur Heart J*. 2005;26:1115-1140.

Swynghedauw B. Molecular mechanisms of myocardial remodeling. *Physiol Rev*. 1999;79:215-262.

Tavazoie S, Hughes JD, Campbell MJ et al. Systematic determination of genetic network architecture. *Nat Genet*. 1999;22:281-285.

Thijssen VL, Ausma J, Borgers M. Structural remodelling during chronic atrial fibrillation: act of programmed cell survival. *Cardiovasc Res*. 2001;52:14-24.

Thum T, Galuppo P, Wolf C et al. MicroRNAs in the human heart: a clue to fetal gene reprogramming in heart failure. *Circulation*. 2007;116:258-267.

Tkatchenko AV, Le Cam G, Leger JJ et al. Large-scale analysis of differential gene expression in the hindlimb muscles and diaphragm of mdx mouse. *Biochim Biophys Acta*. 2000;1500:17-30.

Tomaselli GF, Zipes DP. What causes sudden death in heart failure? *Circ Res*. 2004;95:754-763.

Tseng GC, Oh MK, Rohlin L et al. Issues in cDNA microarray analysis: quality filtering, channel normalization, models of variations and assessment of gene effects. *Nucleic Acids Res*. 2001;29:2549-2557.

Tsuang MT, Nossova N, Yager T et al. Assessing the validity of blood-based gene expression profiles for the classification of schizophrenia and bipolar disorder: a preliminary report. *Am J Med Genet B Neuropsychiatr Genet*. 2005;133:1-5.

Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci U S A*. 2001;98:5116-5121.

Twine NC, Stover JA, Marshall B et al. Disease-associated expression profiles in peripheral blood mononuclear cells from patients with advanced renal cell carcinoma. *Cancer Res*. 2003;63:6069-6075.

van Rooij E, Sutherland LB, Qi X et al. Control of stress-dependent cardiac growth and gene expression by a microRNA. *Science*. 2007;316:575-579.

Velculescu VE, Zhang L, Vogelstein B et al. Serial analysis of gene expression. *Science*. 1995;270:484-487.

Vene N, Mavri A, Kosmelj K et al. High D-dimer levels predict cardiovascular events in patients with chronic atrial fibrillation during oral anticoagulant therapy. *Thromb Haemost*. 2003;90:1163-1172.

Venter JC, Adams MD, Myers EW et al. The sequence of the human genome. *Science*. 2001;291:1304-1351.

Verheule S, Sato T, Everett T et al. Increased vulnerability to atrial fibrillation in transgenic mice with selective atrial fibrosis caused by overexpression of TGF-beta1. *Circ Res*. 2004;94:1458-1465.

Vukmirovic OG, Tilghman SM. Exploring genome space. *Nature*. 2000;405:820-822.

Wijffels MC, Kirchhof CJ, Dorland R et al. Atrial fibrillation begets atrial fibrillation. A study in awake chronically instrumented goats. *Circulation*. 1995;92:1954-1968.

Wittchen F, Suckau L, Witt H et al. Genomic expression profiling of human inflammatory cardiomyopathy (DCMi) suggests novel therapeutic targets. *J Mol Med*. 2007;85:253-267.

Wolf PA, Abbott RD, Kannel WB. Atrial fibrillation as an independent risk factor for stroke: the Framingham Study. *Stroke*. 1991;22:983-988.

Workman C, Jensen LJ, Jarmer H et al. A new non-linear normalization method for reducing variability in DNA microarray experiments. *Genome Biol*. 2002;3:research0048.

Wu AH, Smith A, Wieczorek S et al. Biological variation for N-terminal pro- and B-type natriuretic peptides and implications for therapeutic monitoring of patients with congestive heart failure. *Am J Cardiol*. 2003;92:628-631.

Yancy CW, Saltzberg MT, Berkowitz RL et al. Safety and feasibility of using serial infusions of nesiritide for heart failure in an outpatient setting (from the FUSION I trial). *Am J Cardiol*. 2004;94:595-601.

Yang J, Moravec CS, Sussman MA et al. Decreased SLIM1 expression and increased gelsolin expression in failing human hearts measured by high-density oligonucleotide arrays. *Circulation*. 2000;102:3046-3052.

Yang YH, Dudoit S, Luu P et al. Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Res*. 2002a;30:e15.

Yang YH, Dudoit S, Luu P et al. Normalization for cDNA Microarray data. Technical report #589. 2001.

Yang YH, Speed T. Design issues for cDNA microarray experiments. *Nat Rev Genet*. 2002b;3:579-588.

Yusuf S, Hawken S, Ounpuu S et al. Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the INTERHEART study): case-control study. *Lancet*. 2004;364:937-952.

Yusuf S, Pfeffer MA, Swedberg K et al. Effects of candesartan in patients with chronic heart failure and preserved left-ventricular ejection fraction: the CHARM-Preserved Trial. *Lancet*. 2003;362:777-781.

Zeeberg BR, Feng W, Wang G et al. GoMiner: a resource for biological interpretation of genomic and proteomic data. *Genome Biol*. 2003;4:R28.

# ANNEXES

Article cité dans le chapitre « conception et interprétation des puces à ADN »

Lamirault G, Steenman M, Le Meur N, Demolombe S, Trochu JN, Leger JJ. DNA chip technology in cardiovascular research. Arch Mal Coeur Vaiss. 2004 Dec;97(12):1251-5.

Article associé à l'étude 5

Gaborit N, Steenman M, Lamirault G, Le Meur N, Le Bouter S, Lande G, Leger J, Charpentier F, Christ T, Dobrev D, Escande D, Nattel S, Demolombe S. Human atrial ion channel and transporter subunit gene-expression remodeling associated with valvular heart disease and atrial fibrillation. Circulation. 2005 Jul 26;112(4):471-81.

# DNA chip technology in cardiovascular research

G. Lamirault* and **,
M. Steenman* and ***,
N. Le Meur***,
S. Demolombe*, JN. Trochu**
and J.J. Léger* and ***

**Summary**

Global and/or dynamic analysis of the cardiac transcriptome may improve our understanding of the adaptation of cardiac tissue or cells to different physiological or pathological conditions.

The achievement of sequencing projects on mammalian genomes and the development of DNA chip technology have dramatically extended the scale of gene expression studies from a candidate gene approach to a system approach. In current DNA chip experiments, expression levels of thousands of genes can be determined simultaneously. Obviously, the huge quantities of objects and information generated by these experiments require a computational management of the expression data with adequate mathematical (mostly statistical) algorithms.

Here, we will discuss the principle and experimental key points of DNA chips. Four examples will be cited to illustrate applications in the cardiovascular system. Arch Mal Cœur 2004 ; 97 : 1251-5.

**Résumé**

**Une analyse globale du transcriptome cardiaque et de ses modifications peut permettre de mieux déterminer les phénomènes d'adaptation du tissu ou des cellules cardiaques dans différentes situations physiologiques et pathologiques.**

**L'aboutissement des projets de séquençage de nombreux génomes, dont le génome humain, et le développement de la technologie des puces à ADN ont fait basculer les études d'expression des gènes d'une approche gène-candidat vers une approche globale des systèmes biologiques. L'utilisation des puces à ADN permet aujourd'hui de déterminer de manière simultanée les niveaux d'expression de milliers de gènes dans un échantillon biologique. La masse importante de données et d'informations générées par ces travaux doit bien évidemment être organisée et analysée par un ensemble d'outils informatiques spécifiques utilisant des algorithmes mathématiques (le plus souvent statistiques).**

**Cette revue décrira le principe des puces à ADN et les points expérimentaux capitaux dans les études utilisant cette technologie. Quatre exemples illustreront les applications de ces travaux en recherche cardiovasculaire. Arch Mal Cœur 2004 ; 97 : 1251-5.**

The adaptation of cardiac tissue or cells to different physiological or pathological conditions is to a large extent mediated by gene expression modifications. Dynamic analysis of the cardiac transcriptome (all RNA transcripts present in a cardiac tissue) may improve our understanding of these multiple and complex adaptive processes within the cardiac cell. The achievement of sequencing projects on diverse mammalian genomes, together with the introduction of robotics in biology laboratories, has dramatically changed the scale of gene expression studies [1]. Today, exploitation of DNA chips composed of short or long coding sequences representative of thousands of genes permits a global monitoring of the transcriptional activity of any specialized cell, like cardiomyocytes. This extends previously developed methods such as Northern analysis, or quantitative RT-PCR where only a limited number of candidate genes can be analyzed. In current DNA chip experiments, expression levels of either selections of genes of interest or of all genes putatively present in a genome can be determined simultaneously (a complete overview of microarray principles and applications is provided in [2]). Obviously, the huge quantities of objects and information generated by the DNA chip experiments require a computational management of the data with adequate

(*) Institut du thorax-INSERM U533, faculté de médecine, Nantes.

(**) Institut du thorax, clinique cardiologique et des maladies vasculaires, CHU de Nantes.

(***) Plate-forme transcriptome Ouest génopole, faculté de médecine, 1, rue G. Veil, 44035 Nantes Cedex 1.

E-mail : Guillaume.Lamirault@nantes.inserm.fr

(*Tirés à part* : Dr G. Lamirault).

mathematical (often statistical) algorithms. The aim is not only storage of the objects and validation of the data but also knowledge discovery to extract meaningful information from the large amount of expression data. During the last few years, gene expression profiling experiments in cardiovascular research have been providing ever more promising results.

Here, we will first summarize the principles of a DNA chip experiment and key points towards successful outcomes. Next, four applications in the cardiovascular system will be detailed.

## PRINCIPLES OF A MICROARRAY EXPERIMENT

A DNA chip (DNA microarray) is a solid support on which cDNA fragments of different genes (called probes or reporters) are robotically deposited (or synthesized *in situ*) in a grid-like manner. The principle of DNA chip experiments is based on hybridization between these spotted cDNA fragments and a complex mixture of labeled complementary cDNA sequences (called targets) representative of the RNA transcripts isolated from a tissue sample [3]. The hybridization signal intensity measured at each reporter spot is related to the corresponding RNA transcript abundance in the tissue sample. A standard DNA chip experiment aims to detect whether genes are differentially or equally expressed between test and control RNA samples. Depending on the technology used, the two targets are either hybridized separately on two DNA chips to provide two sets of absolute expression values or competitively cohybridized on the same DNA chip to provide a set of relative expression ratios (fig. 1).

## KEY POINTS TOWARDS SUCCESSFUL EXPERIMENTS

### Selection of cardiac-pertinent reporters/probes

Only a part of the ~35 000 sequences of the human genome assumed to encode genes are preferentially, if not exclusively, expressed in cardiovascular tissues. Cardiac-pertinent reporter sequences spotted on microarrays are cDNA sequences representative of the 4 000 to 6 000 genes that control the normal and pathological phenotype of the human cardiovascular system. Using DNA chips with 12 626 human genes partially covering the whole genome, we analyzed gene expression profiles of failing and non-failing human hearts [4]; 2 031 genes (16%) were found to be expressed in the different cardiac samples and are therefore representative of part of the human cardiac transcriptome. As an alternative to such a global approach, we [5, 6] and others [7] have also performed successive subtractive hybridizations between normal and pathological hearts and between different heart chambers. A few thousand cardiac-relevant cDNA sequences were thus identified. Other targeted methods such as EST (Expressed Sequence Tag) analysis [8], and SAGE (Serial Analysis of Gene Expression) [9] have also been used to identify cardiac-relevant genes.

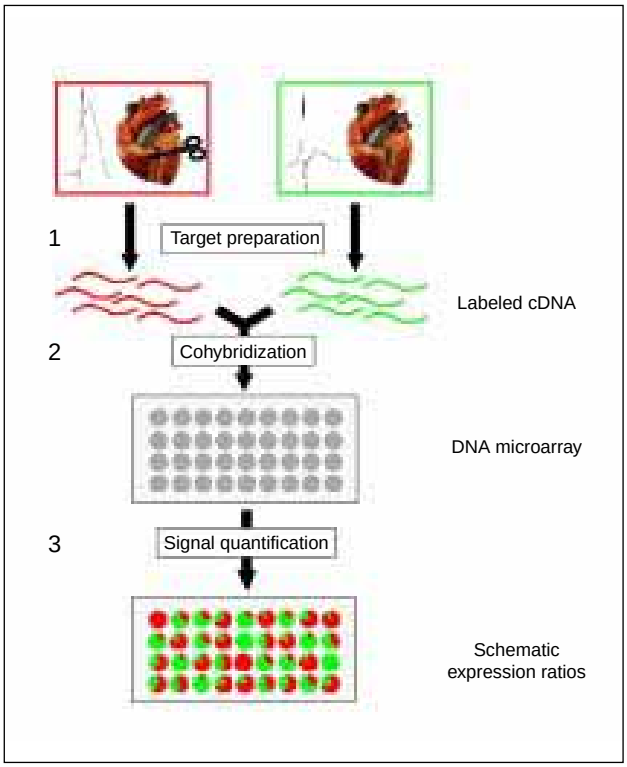



**FIG. 1 – Étude comparative (tissu ischémique vs tissu contrôle) à partir de puces à ADN. 1 : à partir de tissu ischémique (rouge) et contrôle (vert), deux pools de cibles complexes (cDNA) sont générés et marqués avec deux fluorochromes distincts ; 2 : des quantités identiques des cibles complexes marquées sont mélangées et cohybridées sur la puce à ADN ; 3 : pour chaque gène, identifié par sa position sur la puce, le ratio d'expression entre les tissus ischémiques et contrôles est défini comme le ratio d'intensité de fluorescence entre les deux fluorochromes.**

FIG. 1 – Comparative DNA chip experiment (ischemic vs. control). 1: two distinct pools of cDNA targets are generated from ischemic (red) and control (green) RNA samples and labeled with two different fluorescent dyes; 2: equal quantities of the two labeled cDNA targets are mixed and cohybridized onto the DNA microarray; 3: for each reporter sequence, identified by its position on the array, the expression ratio between ischemic and control samples is calculated as the ratio of fluorescent signal intensity (measured by a chip scanner) between the two dyes.

### Complexity of functional annotation of cardiac gene collections

Functional annotation aims to collect all essential information on the gene collections spotted on the DNA chips. Numerous complementary information sources on the genes and their related products (RNA, proteins) are freely available *via* the Internet. Various web-accessible bioinformatic tools have been developed to automatically extract gene-specific data from the literature and ontology classification. However, these data - chromosomal localization, molecular function(s), subcellular localization, interactions with other genes, the presence of known mutations or SNPs (single nucleotide polymorphisms), etc - are stored in different databases presented in different formats. Therefore, genes are often documented

under several aliases. Use of the official gene nomenclature (HUGO: human genome organization http://www.gene.ucl.ac.uk/nomenclature/), together with a computational reformatting of all web-available information, permits to instantly provide an identity card of each gene under study. Such an information processing system using a plug-in architecture model has been developed in our laboratory. This freely accessible system helps microarray users to obtain knowledge on their genes of interest (see http://cardioserve.nantes.inserm.fr/mad/madsense/).

### Assessing quantity and quality of cardiac samples

Since DNA chip experiments measure subtle quantitative expression differences between complex gene sets of RNA samples, both RNA quality and quantity assessments are crucial to ensure production of reliable data. Powerful nanoscale methods currently evaluate RNA sample integrity and determine its quantity at the different steps of the experiment [10]. The ~20 µg of total RNA per sample required to perform DNA chip gene expression experiments are not readily available from samples such as human cardiac biopsies or cells from laser microdissections. RNA amplification methods have to be used in the target preparation to overcome this difficulty. It has been shown that the sequential use of RNA amplification and microarray techniques provides reproducible and reliable results [11].

### Metrological considerations and knowledge discovery

Various sources of experimental noise are present at the different steps of DNA chip experiments and can lead to false knowledge discovery. Technical noises arise from uncontrolled variations in the RNA quality and labeling as well as from mechanical or optical distortions of the support and the chip scanner. Biological noises arise from possible contamination, storage conditions, and heterogeneity of the tissue samples. Spotting of replicates for each gene within one chip and hybridization of replicate chip experiments with the same RNA extract and with different extracts of the same biological sample have to be carried out. The hereby obtained replicate values allow statistical evaluation of the 'true' biological signal amidst numerous experimental noises. They also allow the identification of outlier values and the calculation of the variance of the expression value for each gene. Power analyses show that at least eight true replicate values are required to significantly detect expression variations of 20% between two distinct biological samples [12].

Using such consolidated expression values, different numerical procedures or graphical representations are available to detect and visualize genes of interest in the different pathophysiological situations under investigation. Variance-based statistical analyses are simultaneously applied to thousands of hypotheses (genes) in order to detect genes differentially expressed between two or more biological samples. The crucial point is to determine the threshold of significance of the statistical test to efficiently control the balance between false positive and false negative rates [13]. In addition, diverse clustering techniques aim to classify genes with similar expression patterns under multiple experimental conditions and to identify genes that encode proteins with common molecular functions [14]. Combination of expression values, functional annotation, and clinical information provides a global view of complex disease tissues [15].

## EXAMPLES OF GENE EXPRESSION STUDIES IN THE CARDIOVASCULAR SYSTEM

The hereafter described four examples illustrate the present contribution of cardiovascular DNA microarray studies to physiology, pharmacology, pathophysiology, and clinical research respectively.

### A network of developmental transcription factors maintains the phenotype of cardiac chambers

The identification of chamber–specific gene expression patterns could significantly improve our understanding of the adaptation of the cardiac tissue in different physiological and pathological conditions. Different groups performed regional gene expression profiling studies on human and murine hearts [16-18]. Expression profiles of transcription factors in atria and ventricles were thus determined and compared in adult mice [17]. Based on previously known signaling pathways and on their gene expression profiling results, the authors proposed a transcriptional regulatory network that may play an important role in cardiac chamber specialization. Interestingly, their results suggest that key transcription factors like GATA4, NKX2-5, or TBX5 that control cardiac morphogenesis, and in which mutations cause congenital heart disease, are also involved in the maintenance of the ventricular or atrial phenotype of the cardiac chambers in adult hearts.

### Amiodarone-induced electrophysiological remodeling of cardiac tissue is mediated by targeted gene expression modifications

Transcriptional effects of pharmacological interventions can be analyzed on a global scale with microarrays. Amiodarone is an important drug for the treatment of cardiac arrhythmias. To better clarify the mechanism of its efficacy, pharmacogenomic effects of chronic amiodarone treatment were recently evaluated in the mouse heart using channel-dedicated chips [19]. The obtained expression patterns indicate that the remodeling of the transcriptional profile of cardiac ion channels contributes to the chronic pharmacological effect of amiodarone. The dose-dependent amiodarone-induced prolonged repolarization correlated with progressive downregulation of some $K^+$ channels (KCND2, KCNB1, and KCNA5), whereas conduction delays increased in parallel to downregulation of $Na^+$ channels (SCN4A, SCN5A, and SCN1B) and connexin 43 (GJA1). An improved understanding of the mechanisms of action of cardiovascular drugs, based on pharmacogenomic studies, will help to define the conditions of their utilization.

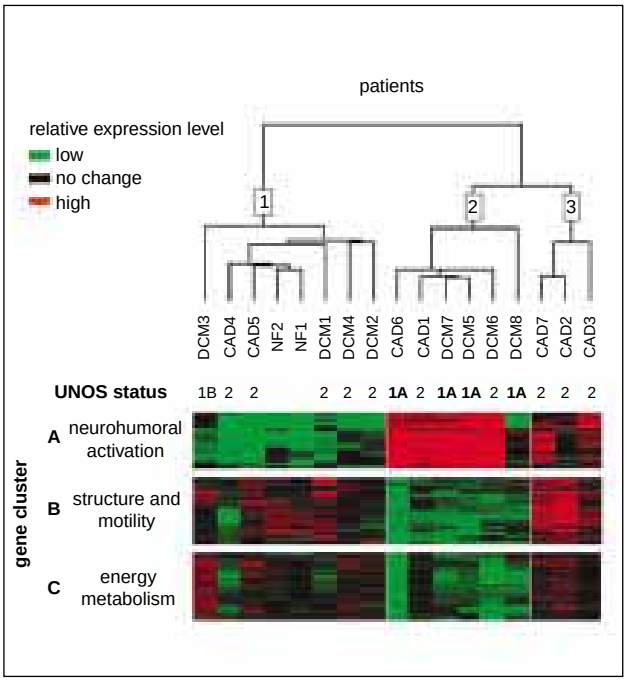



**FIG. 2 – Classification des patients insuffisants cardiaques à partir des profils d'expression génique. Pour chaque gène (ligne) et chaque patient (colonne), la valeur d'expression est ajustée de telle sorte que l'expression médiane de chaque gène pour les 17 patients corresponde à « no change » (noir). Les malades sont classés en 3 groupes repérés par « 1 », « 2 » et « 3 ». Les profils d'expression des groupes de gènes A, B et C sont fortement corrélés avec la classification des malades en 3 groupes (pour une classification complète voir réf. 6). Pour chacun des groupes, une fonction moléculaire prépondérante est mise en évidence. Le groupe A est associé aux fonctions d'activation neurohormonale, le groupe B aux fonctions de structure et de contraction et le groupe C aux fonctions du métabolisme énergétique. DCM : cardiomyopathie dilatée ; CAD : cardiopathie ischémique ; NF : absence d'insuffisance cardiaque.**

FIG. 2 – Gene-expression based classification of heart failure patients. For each gene (row) and patient (column), the expression value was adjusted so that the median expression value in the 17 patients corresponded to "no change" (black). Patients are classified into three subgroups indicated by "1", "2", and "3". Expression profiles of gene clusters A, B, and C correspond well to the classification of the patients into the three subgroups (for complete cluster see [6]). A specific prevalent molecular function was noted in each of the three gene clusters. Cluster A was associated with neurohumoral activation, cluster B with structure and motility functions, and cluster C with energy metabolism. DCM: dilated cardiomyopathy; CAD: coronary artery disease; NF: non-failing heart.

### Analysis of genes up-regulated during the development of cardiac hypertrophy reveals reactivation of the cell proliferation program

A common set of key genes probably controls the transcriptional modifications related to the development of cardiac hypertrophy. Identification of these key genes and their functions will allow development of new therapies to prevent transition from cardiac hypertrophy to heart failure. Recently, gene expression profiles were compared between two models of cardiac hypertrophy initiated by either pressure overload (transverse aortic constriction) or ventricular dysfunction (coronary artery ligation) [20]. Using genome-wide DNA chips, 92 genes were detected as concomitantly regulated in both models. A few of these hypertrophy-dependent genes are involved in cell proliferation. These cell cycle genes (CCNB1, CCND3, CDK4, GAS1, GAS2) were all upregulated, revealing a significant reactivation of the cardiomyocyte proliferation program during the development of cardiac hypertrophy. Recently, cardiac-targeted reactivation of expression of another cell cycle gene (CCNA2) in an *in vivo* murine model induced proliferation of differentiated cardiomyocytes [21]. Such proliferation-inducing genes represent potential therapeutic targets for the restoration of cardiac function in damaged myocardium.

### Gene expression profiles are potential disease markers in human heart failure

Analysis of individual gene expression profiles can help investigators to identify a new gene expression-based patient classification, which could have multiple implications at diagnostic, prognostic, and therapeutic levels. To investigate whether molecular profiling could be used to discriminate between heart failure patients, we determined individual expression profiles of failing and non-failing left ventricles from human explanted hearts using lab-made chips (see http://cardioserve.nantes.inserm.fr/ptf-puce/puce_multi_dediee.php) representing a few hundred cardiac-relevant genes [6]; 159 genes were significantly discriminating between the 17 patients in the study. A hierarchical clustering algorithm classified the individual patient expression profiles into three distinct subgroups, each with a specific gene expression profile or molecular portrait (fig. 2). Interestingly, this gene expression-based patient classification coincided with the UNOS medical urgency status (United Network for Organ Sharing classification [22]) of the same patients. Patients with the highest medical urgency status (UNOS 1A) clustered together in group "2". Gene expression patterns of this group associated high relative expression levels of known neurohumoral activation markers (ANF, BNP) and low relative expression level of sarcomeric genes (β-MHC, titin) and of genes involved in energy metabolism (NADH dehydrogenase). The observed correlation between the UNOS prognostic parameter and the individual gene expression profiles underlines the potential value of an approach based on multiple molecular disease markers in heart failure. It also provides a rationale for future prospective prognostic molecular profiling studies at earlier stages of the disease.

## CONCLUSION

Microarray gene expression studies are used more and more in cardiovascular research. Involvement of nanotechnologies and novel optical methods will most likely enhance the sensitivity and reliability of the detection of gene expression levels, even from minute samples. Progressive sophistication of biological tools

and experimental design, together with an incessant development of knowledge discovery methods, will increase our capacity to generate more and more expression data and to obtain reliable information on transcriptome activity from any cell or tissue. Such improvements urge to allow tangible applications of DNA chips in clinical practice [23]. In the cardiovascular field, like in any other medical domain, pertinent determination of gene expression patterns will certainly lead to future advances in diagnostic, prognostic, and therapeutic management.

*Acknowledgements*
*Our laboratory is supported by the* Institut national de la santé et de la recherche médicale, Centre national de la recherche scientifique, Association française contre les myopathies, Conseil régional des Pays de la Loire *and* Ouest génopole.

KEY WORDS : DNA microarray, gene expression profiling, computational biology, cardiovascular physiology, cardiovascular diseases.

## References

1. **Swynghedauw B.** Physiologie cardiovasculaire : les nouvelles données. Arch Mal Cœur 2002 ; 95 : 299-304.
2. **The Chipping Forecast II.** Nat Genet 2002 ; 32 : 465-552.
3. **Southern EM.** Detection of specific sequences among DNA fragments separated by gel electrophoresis. J Mol Biol 1975 ; 98 : 503-17.
4. **Steenman M, Chen YW, Le Cunff M et al.** Transcriptomal analysis of failing and nonfailing human hearts. Physiol Genomics 2003 ; 12 : 97-112.
5. **Rouger K, Le Cunff M, Steenman M et al.** Global/temporal gene expression in diaphragm and hindlimb muscles of dystrophin-deficient (mdx) mice. Am J Physiol Cell Physiol 2002 ; 283 : C773-84.
6. **Steenman M, Lamirault G, Le Meur N, Le Cunff M, Escande D, Leger JJ.** Distinct molecular portraits of human failing hearts identified by dedicated cDNA microarrays. Eur J Heart Fail 2004 (in press).
7. **Philip-Couderc P, Smih F, Pelat M et al.** Modifications précoces du transcriptome auriculaire dans l'hypertension artérielle associée à l'obésité chez le chien. Arch Mal Cœur 2002 ; 95 : 695-9.
8. **Hwang DM, Dempsey AA, Wang RX et al.** A genome-based resource for molecular cardiovascular medicine: toward a compendium of cardiovascular genes. Circulation 1997 ; 96 : 4146-203.
9. **Anisimov SV, Tarasov KV, Stern MD, Lakatta EG, Boheler KR.** A quantitative and validated SAGE transcriptome reference for adult mouse heart. Genomics 2002 ; 80 : 213-22.
10. **Auer H, Lyianarachchi S, Newsom D et al.** Chipping away at the chip bias: RNA degradation in microarray analysis. Nat Genet 2003 ; 35 : 292-3.
11. **Wang J, Hu L, Hamilton SR, Coombes KR, Zhang W.** RNA amplification strategies for cDNA microarray experiments. Biotechniques 2003 ; 34 : 394-400.
12. **Le Meur N, Lamirault G, Bihouée A et al.** A dynamic, web-accessible resource to process raw microarray scan data into consolidated gene expression values. Importance of replication. Nucleic Acids Res 2004 (in press).
13. **Storey JD, Tibshirani R.** Statistical significance for genomewide studies. Proc Natl Acad Sci USA 2003 ; 100 : 9440-5.
14. **Quackenbush J.** Computational analysis of microarray data. Nat Rev Genet 2001 ; 2 : 418-27.
15. **Chaussabel D, Semnani RT, McDowell MA, Sacks D, Sher A, Nutman TB.** Unique gene expression profiles of human macrophages and dendritic cells to phylogenetically distinct parasites. Blood 2003 ; 102 : 672-81.
16. **Kaynak B, von Heydebreck A, Mebus S et al.** Genome-wide array analysis of normal and malformed human hearts. Circulation 2003 ; 107 : 2467-74.
17. **Tabibiazar R, Wagner RA, Liao A, Quertermous T.** Transcriptional profiling of the heart reveals chamber-specific gene expression patterns. Circ Res 2003 ; 93 : 1193-201.
18. **Kaab S, Barth AS, Margerie D et al.** Global gene expression in human myocardium-oligonucleotide microarray analysis of regional diversity and transcriptional regulation in heart failure. J Mol Med 2004 ; 82 : 308-16.
19. **Le Bouter S, El Archi A, Marionneau C et al.** Chronic amiodarone remodels expression of ion channel transcripts in the mouse heart. Circulation 2004 (in press).
20. **Mirotsou M, Watanabe CM, Schultz PG, Pratt RE, Dzau VJ.** Elucidating the molecular mechanism of cardiac remodeling using a comparative genomic approach. Physiol Genomics 2003 ; 15 : 115-26
21. **Chaudhry HW, Dashoush NH, Tang H et al.** Cyclin A2 mediates cardiomyocyte mitosis in the postmitotic myocardium. J Biol Chem 2004 ; 279 : 35858-66.
22. **Renlund DG, Taylor DO, Kfoury AG, Shaddy RS.** New UNOS rules: historical background and implications for transplantation management. United Network for Organ Sharing. J Heart Lung Transplant 1999 ; 18 : 1065-70.
23. **Petricoin EF, Hackett JL, Lesko LJ et al.** Medical applications of microarray technologies: a regulatory science perspective. Nat Genet 2002 ; 32 (Suppl. 2) : 474-9.

# Human Atrial Ion Channel and Transporter Subunit Gene-Expression Remodeling Associated With Valvular Heart Disease and Atrial Fibrillation

Nathalie Gaborit, MSc*; Marja Steenman, PhD*; Guillaume Lamirault, MD; Nolwenn Le Meur, MSc; Sabrina Le Bouter, PhD; Gilles Lande, MD; Jean Léger, PhD; Flavien Charpentier, PhD; Torsten Christ, MD; Dobromir Dobrev, MD; Denis Escande, MD, PhD; Stanley Nattel, MD; Sophie Demolombe, PhD

***Background***—Valvular heart disease (VHD), which often leads to atrial fibrillation (AF), and AF both cause ion-channel remodeling. We evaluated the ion-channel gene expression profile of VHD patients, in permanent AF (AF-VHD) or in sinus rhythm (SR-VHD), in comparison with patients without AF or VHD, respectively.

***Methods and Results***—We used microarrays containing probes for human ion-channel and $Ca^{2+}$-regulator genes to quantify mRNA expression in atrial tissues from 7 SR-VHD patients and 11 AF-VHD patients relative to 11 control patients in SR without structural heart disease (SR-CAD). From the data set, we selected for detailed analysis 59 transcripts expressed in the human heart. SR-VHD patients differentially expressed 24/59 ion-channel and $Ca^{2+}$-regulator transcripts. There was significant overlap between VHD groups, with 66% of genes altered in SR-VHD patients being similarly modified in AF-VHD. Statistical differences between the AF- and SR-VHD groups identified the specific molecular portrait of AF, which involved 12 genes that were further confirmed by real-time reverse transcription–polymerase chain reaction. For example, phospholamban, the $\beta$-subunit MinK (*KCNE1*) and MIRP2 (*KCNE3*), and the 2-pore potassium channel *TWIK-1* were upregulated in AF-VHD compared with SR-VHD, whereas the T-type calcium-channel Cav3.1 and the transient-outward potassium channel Kv4.3 were downregulated. Two-way hierarchical clustering separated SR-VHD from AF-VHD patients. AF-related changes in L-type $Ca^{2+}$-current and inward-rectifier current were confirmed at protein and functional levels. Finally, for 13 selected genes, SR restoration reversed ion-channel remodeling.

***Conclusions***—VHD extensively remodels cardiac ion-channel and transporter expression, and AF alters ion-channel expression in VHD patients. (***Circulation*. 2005;112:471-481.**)

**Key Words:** remodeling ■ ion channels ■ fibrillation ■ valves ■ atrium

Atrial fibrillation (AF), the most common cardiac arrhythmia, contributes substantially to cardiac morbidity and mortality.[1] The cellular mechanisms underlying AF have been the subject of extensive studies both in human and in animal models (reviewed in reference 2). Ion-channel remodeling clearly plays an important role in the pathophysiology of AF, contributing to its initiation and perpetuation.[2] At the functional level, decreased L-type $Ca^{2+}$-current ($I_{Ca,L}$) density is central to the shortening of the effective refractory period characteristic of AF.[3] Downregulation of the transient outward current, $I_{to}$, has also been reported,[3–6] whereas inward-rectifier currents can be increased.[3,7] Functional alterations in ionic currents have been associated with corresponding alterations in mRNA and ion-channel protein expression (see Brundel et al,[8] Van Wagoner and Nerbonne,[9] and Nattel and Li[10] for reviews).

Valvular heart disease (VHD) is a frequent clinical cause of AF.[1,11] There is limited information available about ion-channel remodeling due to VHD, although VHD-related atrial dilation is associated with reduced $I_{Ca,L}$ and $I_{to}$.[12]

Previous studies of ion-transporter subunit remodeling associated with AF have been limited to the evaluation of a small number of candidate subunits believed to be important, with none examining >6 subunits. An additional limitation has been the fact that the independent contributions of AF and underlying heart disease have not been adequately addressed, because in studies of cardiac disease–related remodeling, many patients also have AF[12]

Received September 14, 2004; revision received April 7, 2005; accepted April 14, 2005.

From l'Institut du Thorax, INSERM U533, Faculté de Médecine, Nantes, France (N.G., M.S., G. Lamirault, N.L.M., S.L.B., G. Lande, J.L., F.C., D.E., S.D.); the Montreal Heart Institute and University of Montreal, Montreal, Canada (S.N.); and Institut für Pharmakologie und Toxikologie, Dresden, Germany (T.C., D.D.).

*The first 2 authors contributed equally to this work.

**Raw data for DNA chips can be found online with this article at http://www.circulationaha.org.**

Correspondence to Sophie Demolombe, PhD, INSERM U533, Faculté de Médecine, 1 rue G. Veil, 44035 Nantes cedex, France. E-mail sophie.demolombe@nantes.inserm.fr



*Circulation* **is available at http://www.circulationaha.org** **DOI: 10.1161/CIRCULATIONAHA.104.506857**

**Clinical Characteristics of Patients in Normal SR (VHD and SR-CAD Groups) and in AF at the Time of Cardiac Surgery**

| | | | Underlying Cardiac Disease | | | | | | PMH | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Patients | Age, y | Sex | Ischemic Heart Disease | Valvular Replacement Indication | Duration of AF, y | LV Ejection Fraction, % | LAD, mm | P Wave, ms | $\beta$-Blockers | Digitalis | Statins | Diuretics | ACE Inhibitors+ $AT_1$ Blockers | $Ca^{2+}$ Antagonists | Nitrates | Other Medication |
| SR-VHD group | | | | | | | | | | | | | | | | |
| SR-VHD1 | 76 | M | − | AS | | 64 | 40 | 118 | | | | | FS | | | AA |
| SR-VHD2 | 62 | M | − | MR | | 77 | 63 | 116 | | | | FR | RM | | | |
| SR-VHD3 | 80 | M | + | AS | | 62 | 44 | 100 | | | PV | | LS | | | |
| SR-VHD4 | 72 | M | + | AS | | 57 | 39 | 105 | | | SM | | VL | | | |
| SR-VHD5 | 52 | M | − | MR | | 64 | 46 | 118 | | | | FR | RM | | | |
| SR-VHD6 | 69 | M | − | AR/MR | | 71.5 | 45 | 111 | ST | | PV | IN | PR | | | FL |
| SR-VHD7 | 78 | F | − | AS | | 47 | 39 | 87 | | DG | | SP+FR | RM | | | |
| | **70±9** | | | | | **63±10** | **45±8** | **108±11** | | | | | | | | |
| AF-VHD group | | | | | | | | | | | | | | | | |
| AF-VHD1 | 67 | M | − | MR | 9 | 41 | 62 | | CR | | PV | FR | LS | | | FN |
| AF-VHD2 | 70 | F | − | MS | 5 | 57 | 50 | | | DG | | | | | ID | FN |
| AF-VHD3 | 74 | M | − | AS | 1 | 71 | 40 | | | | | AM+HCT | | DL | | FN |
| AF-VHD4 | 67 | M | − | AS | 2 | 45 | 49 | | | | | | RM | | | FN |
| AF-VHD5 | 76 | M | − | AS/MR | 12 | 58 | 56 | | | | | FR | PR | | | FN |
| AF-VHD6 | 72 | F | + | MR | 10 | 64 | 44 | | | DT | PV | FR | CP | NC | | FN+LT |
| AF-VHD7 | 73 | M | + | AS | 6 | 75 | 45 | | | | | FR | EL | DL | | |
| AF-VHD8 | 73 | M | − | AS | 13 | 61 | 52 | | | DG | | FR | EL | | | FN |
| AF-VHD9 | 69 | M | + | AS | 7 | 70 | 40 | | AT | | SM | | | | | FN |
| AF-VHD10 | 77 | M | + | AS | 2 | 70 | 40 | | AT | | | | | NC | TT+MS | AA+FN |
| AF-VHD11 | 75 | M | + | AS | 10 | 60 | 51 | | | | | FR | EL | DL | | FN |
| | **72±3** | | | | **7±4** | **61±10** | **48±7** | | | | | | | | | |
| SR-CAD group | | | | | | | | | | | | | | | | |
| SR-CAD1 | 64 | M | + | no | | 47 | | 80 | AC | | SM | | | | MS | AA |
| SR-CAD2 | 73 | M | + | no | | 50 | | 105 | AT | | | | | | TT+MS | CD |
| SR-CAD3 | 68 | M | + | no | | 57 | | 73 | AT | | PV | | | | | AA |
| SR-CAD4 | 72 | M | + | no | | 71 | | 110 | MT | | | | | | TT+MS | AA |
| SR-CAD5 | 70 | M | + | no | | 52 | | 100 | AT | | AO | | | | MS | FL |
| SR-CAD6 | 68 | M | + | no | | 60 | | 98 | AT | | PV | | RM | | | AA |
| SR-CAD7 | 58 | F | + | no | | 73 | | 97 | AT | | PV | | | AD | TT+MS | AA |
| SR-CAD8 | 74 | M | + | no | | 73 | | 112 | AC | | PV | AZ+SP | | AD | | AA |
| SR-CAD9 | 71 | M | + | no | | 72 | | 92 | BT | | SM | | | AD | TT+MS | AA |
| SR-CAD10 | 73 | F | + | no | | 66 | | 89 | AT | | PV | | | FE | TT+ID | AA |
| SR-CAD11 | 73 | M | + | no | | 74 | | 107 | | | | | | DL | TT | AA |
| | **69±5** | | | | | **63±10** | | **97±12** | | | | | | | | |

LAD indicates left atrium diameter; PMH, previous medical history; ACE, angiotensin-converting enzyme; $AT_1$, angiotensin II type 1; AS indicates aortic stenosis; MR, mitral regurgitation; AR, aortic regurgitation; MS, mitral stenosis; ST, sotalol; CR, carvedilol; AT, atenolol; AC, acebutolol; MT, metoprolol; BT, betaxolol; DG, digoxin; DT, digitalis; PV, pravastatin; SM, simvastatin; AO, atorvastatin; FR, furosemide; IN, indapamide; SP, spironolactone; AZ, altizide; AM, amiloride; HCT, hydrochlorothiazide; ID, isosorbide dinitrate; TT, trinitroglycerin; MS, molsidomine; AA, acetylsalicylic acid; AD, amlodipine; CD, clopidogrel; CP, captopril; DL, diltiazem; EL, enalapril; FE, felodipine; FL, flurbiprofen; FN, fluindione; FS, fosinopril; LS, losartan; LT, levothyroxine; NC, nicardipine; PR, perindopril; RM, ramipril; VL, valsartan; +, present; and −, absent. Data in bold are presented as mean±SD.

and in studies of AF, many patients also have significant cardiac disease, often VHD.[6,13] The present study was designed to use a genomic approach to address the following hypotheses: (1) Ion-channel subunit expression in patients with VHD differs from that in patients without VHD, reflecting ionic remodeling and (2) AF induces ion-channel remodeling beyond that induced by VHD alone, as reflected by different ion-channel subunit expression patterns in patients with AF and VHD compared with VHD patients in sinus rhythm (SR).

# Methods

## Human Samples

For molecular biology studies, right atrial appendages (RAAs) were obtained during open-heart surgery from 7 VHD patients in SR (SR-VHD), 11 patients with VHD and permanent (>1 year) AF (AF-VHD), and 11 patients undergoing revascularization surgery for coronary artery disease (CAD), constituting the control group (SR-CAD). SR-VHD and AF-VHD groups were similar in terms of age, sex distribution, types of valve disease, left ventricular (LV) function, and atrial dimensions (the Table). Differences in drug therapy followed underlying heart disease: VHD patients received more

angiotensin-system suppressants and diuretics, which were less common in SR-CAD patients. SR-CAD patients commonly took $\beta$-blockers, statins, and $Ca^{2+}$ antagonists. $Ca^{2+}$ antagonists were prescribed for about half of the AF-VHD and none of the SR-VHD patients. To assess potential reversibility, we obtained data from an additional 6 patients with VHD and persistent AF, who were cardioverted and kept in SR for at least 1 month preoperatively (sinus-rhythm reversion [SRR-VHD] group). These patients were similar to other VHD groups in age (mean±SD, 74±1 years), types of valve disease (4 aortic, 1 mitral, 1 combined), LV ejection fraction (60±2%), and left atrial dimension (50±1 mm). Their drug therapy was also similar, except that they all received oral amiodarone. Finally, RAA samples were obtained from 7 SR-CAD, 6 SR-VHD, and 6 AF-VHD patients for functional studies. These additional patients showed comparable clinical characteristics to other groups. After excision, RAAs were immediately snap-frozen in $LN_2$ and stored at −80°C.

## RNA Isolation and Labeling

Total RNA was isolated with Trizol reagent (Life Technologies). RNA was DNase-treated (RNeasy fibrous tissue mini kit, Qiagen). The quality of isolated RNA was assessed by microelectrophoresis on polyacrylamide gels (Agilent 2100 Bioanalyzer). The absence of DNA contamination was verified by reverse transcription (RT)–negative polymerase chain reaction (PCR). A control pool was prepared, containing equal amounts of total RNA from each SR-CAD sample. RT was performed on this pool to obtain Cy3-labeled cDNA with the CyScribe cDNA postlabeling kit (Amersham Biosciences). For every SR-VHD and AF-VHD patient, 2 or 3 distinct RT reactions (depending on tissue availability) were performed to obtain Cy5-labeled cDNA. Each RT was processed independently.

## Oligonucleotides and Microarrays

Fifty-mer, 5′ amino-modified oligonucleotide probes were synthesized at MWG Biotech AG. Lyophilized oligonucleotides were dissolved at 25 $\mu$mol/L in 1× spotting buffer A (MWG). The microarray stand was epoxysilane-coated slides. Oligonucleotide spotting was performed with the Lucidea array spotter (Amersham).

Each microarray contained 4116 oligonucleotides (spotted in quadruplicate on each slide), including 315 targets for genes encoding human ion-channel subunits and proteins involved in calcium homeostasis.

## Hybridization

Slides were treated with 50 mmol/L ethanolamine for oligonucleotide fixation. Each Cy5-labeled cDNA was mixed with an equal amount of Cy3-labeled control-pool cDNA, preincubated with yeast tRNA and polyA RNA (Gibco-BRL), and hybridized onto microarrays. After overnight hybridization, slides were washed with successively more stringent standard saline citrate solutions.

Forty-six independent hybridizations were performed in total: 2 for each SR-VHD patient and 2 or 3 for each AF-VHD patient. Hybridized arrays were scanned by fluorescence confocal microscopy (ScanArray 3000, GSI-Lumonics). Measurements were obtained separately for each fluorochrome at 10 $\mu$m/pixel resolution.

## Microarray-Data Analysis

Fluorescence values were analyzed with GenePixPro 5.0. Raw data were processed with software (http://cardioserve.nantes.inserm.fr/mad/madscan/login.php) developed internally.[14] Low-quality spots were filtered, leaving only spots with valid expression values. Invariant genes were selected with the rank-invariant method.[15] To normalize Cy3 and Cy5 values, a nonlinear regression method (Lowess fitness) was applied to invariant genes to calculate the correction for every spot of the microarray.[16] Outliers were excluded by the median absolute deviation–modified $Z$ test.[17] To compensate for experimental variability and compare SR-VHD with AF-VHD data sets, expression-values of the 46 slides were scaled to the same median absolute deviation.[16]

Two statistical analyses were performed on the 4116-oligonucleotide data-points: (1) comparison of SR-VHD and AF-VHD groups based on data normalized to the SR-CAD control patient pool and (2) direct comparison of SR-VHD versus AF-VHD patients. Genes with statistically significant differential expression were identified with 1- and 2-class significance analysis of microarrays (SAMs)[18] and linear models for microarray data (LIMMA).[19] For SAM, genes with a $q$ value <0.24% were selected. For LIMMA, genes with a Holm-corrected probability value <0.01 were selected. Genes were considered differentially expressed when they met both SAM and LIMMA criteria. For each gene, the median of replicate expression values was calculated. Oligonucleotides representing 59 genes encoding cardiac ion channels and calcium homeostasis proteins were retained for further study.

Two-way hierarchical agglomerative clustering was applied to the gene expression matrix consisting of 18 VHD patients and 59 selected genes. The input consisted of median replicate expression values (displayed as log-normal) for each gene and patient. Clustering with uncentered correlation with the CLUSTER program[20] was visualized with Treeview software.

## TaqMan Real-Time RT-PCR

First-strand cDNA was synthesized from 2 $\mu$g total RNA with Super Script III first-strand for RT-PCR (Invitrogen). Real-time PCR was performed with predesigned 6-carboxy-fluorescein (FAM)-labeled fluorogenic TaqMan probes and primers and 1× TaqMan universal master mix (Applied Biosystems). After 2 minutes at 50°C and 10 minutes at 95°C, 40 cycles of amplification were performed with the ABI PRISM 7900HT sequence-detection system (Applied Biosystems). Data were collected with instrument spectral compensation by Applied Biosystems SDS 2.1 software. Fluorescence signals were normalized to the housekeeping gene myoinositol 1-phosphate synthetase A1 (*ISYNA1*). The comparative threshold-cycle (Ct) relative-quantification method was used.[21] For each patient, each gene was quantified in duplicate in 3 separate experiments. The values were averaged and then used for the $2^{-\Delta CT}\times 10$ calculation, where $2^{-\Delta CT}$ corresponds to expression relative to *ISYNA1*. Statistical analysis was performed with ANOVA followed by Tukey's test, with $P<0.05$ considered significant.

## Western Blotting

Membrane proteins were extracted (n=6 in each group) and processed as previously described.[22] Antibodies were obtained from Alomone Laboratories (Kir2.1 and Cav1.2) or Chemicon (Cx40).

## $I_{K1}$ and $I_{Ca,L}$ Measurements

Human atrial myocytes were isolated[23] and suspended in storage solution (mmol/L): KCl 20, $KH_2PO_4$ 10, glucose 10, potassium glutamate 70, $\beta$-hydroxybutyrate 10, taurine 10, EGTA 10, and albumin 1, pH 7.4. Whole-cell voltage-clamp $I_{K1}$ and $I_{Ca,L}$ recordings were performed and analyzed as previously described.[7,24] Cell capacitances averaged 89.4±8.8 (n=17) for SR-CAD, 94.3±7.6 (n=18) for SR-VHD, and 112.1±9.4 (n=22) pF for AF-VHD myocytes ($P$=NS).

# Results

## Gene Expression Changes of SR-VHD and AF-VHD Compared With SR-CAD Group

Figures 1 and 2 show ion-transport gene expression values in each VHD group. Results for calcium channels and regulators, $Na^+$,$K^+$-ATPase, and sodium-channel data are shown in Figure 1 and chloride, pacemaker, and potassium-channel genes and connexins in Figure 2. Results are given as percentage change from the control pool (SR-CAD group), with open circles representing individual patient values and horizontal lines representing medians. Red indicates statistically significant changes; black indicates results not signifi-

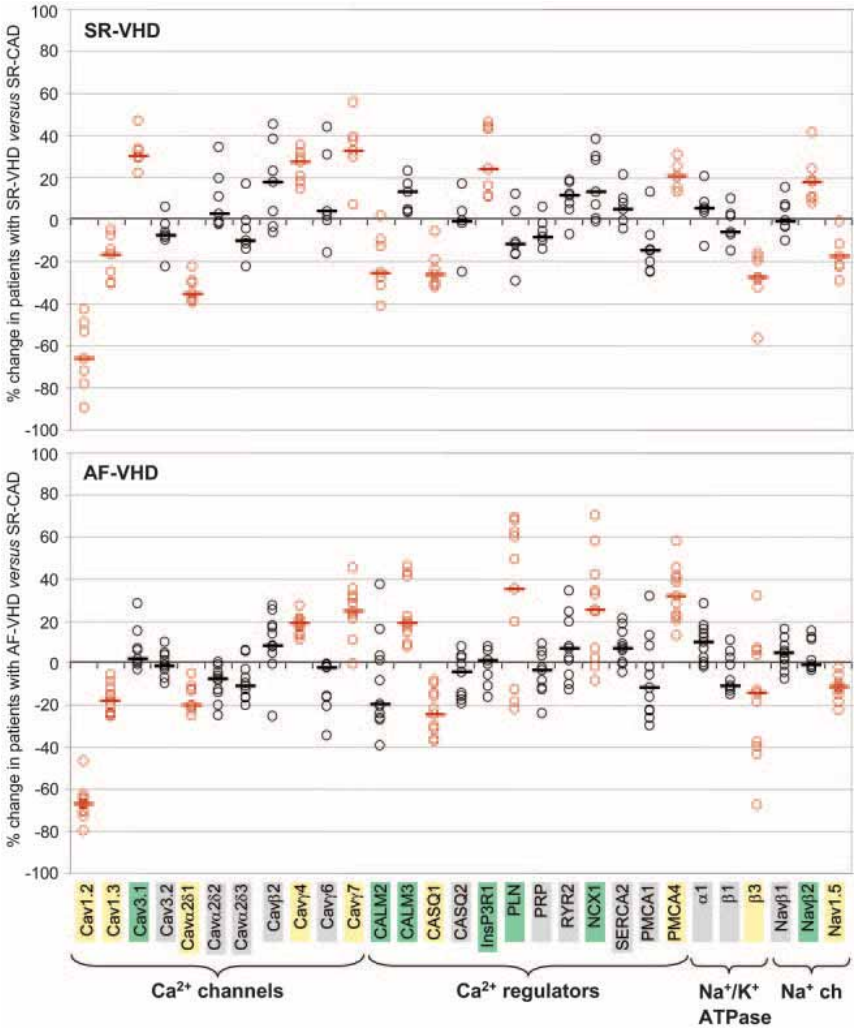



**Figure 1.** Cardiac ion-channel remodeling associated with VHD and AF. Graphs represent percentage change in expression vs control-pool (*y* axis) evaluated by microarrays for genes selected for cardiac expression (*x* axis). Data points indicate medians of repeated measures for each patient. Bars represent median of per-patient values for each gene. Red circles and bars represent genes with statistically significant differential expression in SR-VHD or/and AF-VHD patients vs SR-CAD by 1-class SAM and LIMMA analyses. Genes in gray boxes are unchanged in either SR-VHD or AF-VHD; genes in yellow boxes are statistically significantly changed in same direction; genes in green boxes are changed in only 1 of the VHD groups.

cantly different from the control pool. Of the 59 genes, 28 were unchanged in either SR-VHD or AF-VHD (indicated by gray boxes on the horizontal axis labels). Statistically significant changes in the same direction were observed for 16 genes (yellow boxes). For the 15 remaining genes, statistically significant changes were seen for only one of the VHD groups (green boxes). For no genes were there statistically significant changes in both groups that went in opposite directions. Thus, the SR-VHD group showed substantial changes from control, pointing to important remodeling of cardiac ion-transport genes in VHD and the need to consider VHD effects in the AF group.

To identify the AF-related molecular portrait, AF-VHD patients were compared with SR-VHD patients with 2-class statistical analysis. Among genes that were not significantly altered compared with control in either SR-VHD or AF-VHD alone, 2 genes (Kv4.3 and Cav$\alpha$2$\delta$2) showed a different distribution in SR-VHD compared with AF-VHD patients. Among genes that were significantly up- or downregulated congruently, MinK and Cav$\alpha$2$\delta$1 showed differential expression in SR-VHD versus AF-VHD patients. Among 15 genes that demonstrated significantly altered expression in 1 group only, 8 were differentially distributed in SR-VHD versus AF-VHD patients. In total, 12 genes showed differential expression in AF-VHD versus SR-VHD: the calcium-channel $\beta$-subunit Cav$\alpha$2$\delta$1 (*CACNA2D1*), phospholamban (*PLN*), the chloride-channel subunit *CLCN6*, the voltage-gated potassium channel $\beta$-subunit Kv$\beta$1 (*KCNAB1*), the slow delayed-rectifier $\beta$-subunit MinK (*KCNE1*), the potassium-channel subunit MIRP2 (*KCNE3*), and the 2-pore potassium-channel subunit TWIK-1 (*KCNK1*) were more strongly expressed in AF-VHD patients compared with SR-VHD, whereas the T-type calcium-channel $\alpha$-subunit Cav3.1 (*CACNA1G*), the calcium-channel $\beta$-subunit Cav$\alpha$2$\delta$2 (*CACNA2D2*), inositol-triphosphate receptor 1 (*ITPR1*), the sodium-channel $\beta$-subunit Nav$\beta$2 (*SCN2B*), and the transient-outward potassium-current $\alpha$-subunit Kv4.3 (*KCND3*) were less strongly expressed in AF-VHD than in SR-VHD patients.

## Hierarchical Clustering Analysis Reveals AF-SR Differences

The results of unsupervised 2-way hierarchical clustering analysis are shown in Figure 3. Samples were grouped according to gene expression differences, ordering samples with the most similar expression patterns closest to each other and the most different patterns furthest apart. The hierarchical clustering analysis clearly separated SR-VHD from AF-VHD patients, displaying distinct transcriptional profiles of genes involved in cardiac electrical signaling. All 12 genes identi-

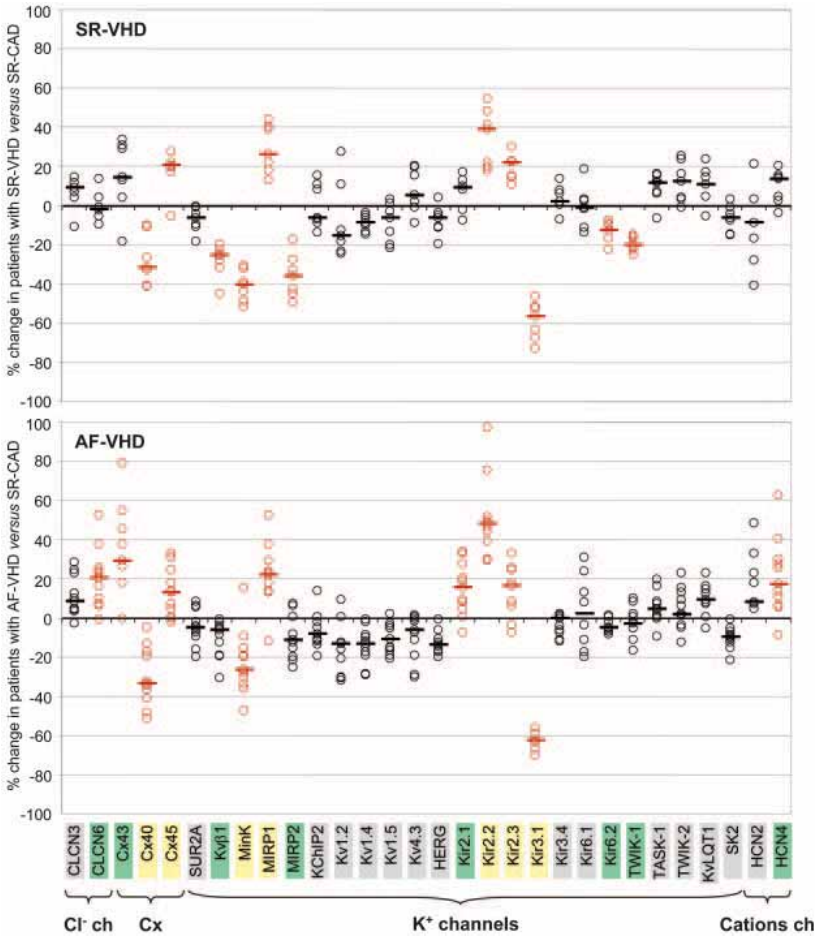



**Figure 2.** Cardiac ion-channel remodeling associated with VHD and AF. Data were obtained for chloride, potassium, and cation channels. Same format as for Figure 1.

fied by 2-class analysis were included in discriminatory gene clusters (indicated at right of figure). Gene groups A and B were more strongly expressed in AF-VHD compared with SR-VHD, whereas the opposite was seen for group C. It should be noted that this hierarchical clustering indicates gene expression levels in SR-VHD relative to the AF-VHD group but not relative to the control group. Thus, some genes in groups A and B may show decreased expression versus the control pool, whereas others in group C may show increased expression (see Figures 1 and 2). Among the genes identified, 6 play potential roles in ionic currents that are modified in models of AF: Kv4.3 ($I_{to}$), Kir2.1 ($I_{K1}$), MinK and MIRP2 ($I_K$), and Cav$\alpha$2$\delta$1 and Cav$\alpha$2$\delta$2 ($I_{Ca}$). Clustering was also attempted in relation to medication or type of VHD, and this analysis showed no detectable effect of these variables. When gene clusters A, B, and C were removed from the analysis, hierarchical clustering did not reveal gene expression differences between AF-VHD and SR-VHD samples (Figure 4), which intermingled rather than being distinctly separated as in Figure 3.

## Expression Assessment by Real-Time RT-PCR and Western Blotting

Microarray data were further confirmed by TaqMan real-time PCR (Figure 5). The expression of the 12 genes showing characteristic expression in AF and clustering on hierarchical analysis was quantified on the RNA samples previously used for microarray experiments. In addition, we validated 2 genes that were not modified by either AF or VHD (SERCA2 and KvLQT1) and 4 genes that are classically altered by AF (NCX1, Cx43, Cx40, and Kir2.1). Transcripts were quantified with fluorescent probes with 100% PCR efficacy. This standardized method permits accurate quantification of transcript expression levels. Statistically significant differences between SR-VHD and SR-CAD groups are shown by asterisks (Figure 5), and differences between AF-VHD and SR-VHD groups are shown by pound symbols (#). Expression differences detected by real-time RT-PCR paralleled those identified by microarray. Of particular note are similar variations in inositoltrisphosphate receptor, phospholamban, CLCN6, Nav$\beta$2, Kv$\beta$1, MiRP2, and Kv4.3. The expression level of Cav3.1 was too low to be reproducibly evaluated by RT-PCR and this is not shown in Figure 5. Finally, the 2 genes that were unchanged by either AF-VHD or SR-VHD showed no change by RT-PCR.

Western-blot experiments (Figure 6) were conducted for ion-channel subunits (Cav1.2, Kir2.1, and Cx40) selected on the basis of evidence for a pathophysiological role in AF.[2] In agreement with gene expression data, Cav1.2 and Cx40 proteins were similarly expressed in SR-VHD and AF-VHD groups. Cav1.2 and Cx40 were significantly downregulated in SR-VHD versus SR-CAD. Kir2.1 was upregulated in the AF-VHD group only. Although changes were quantitatively similar for Cav1.2 and Cx40 mRNA and protein, upregula-

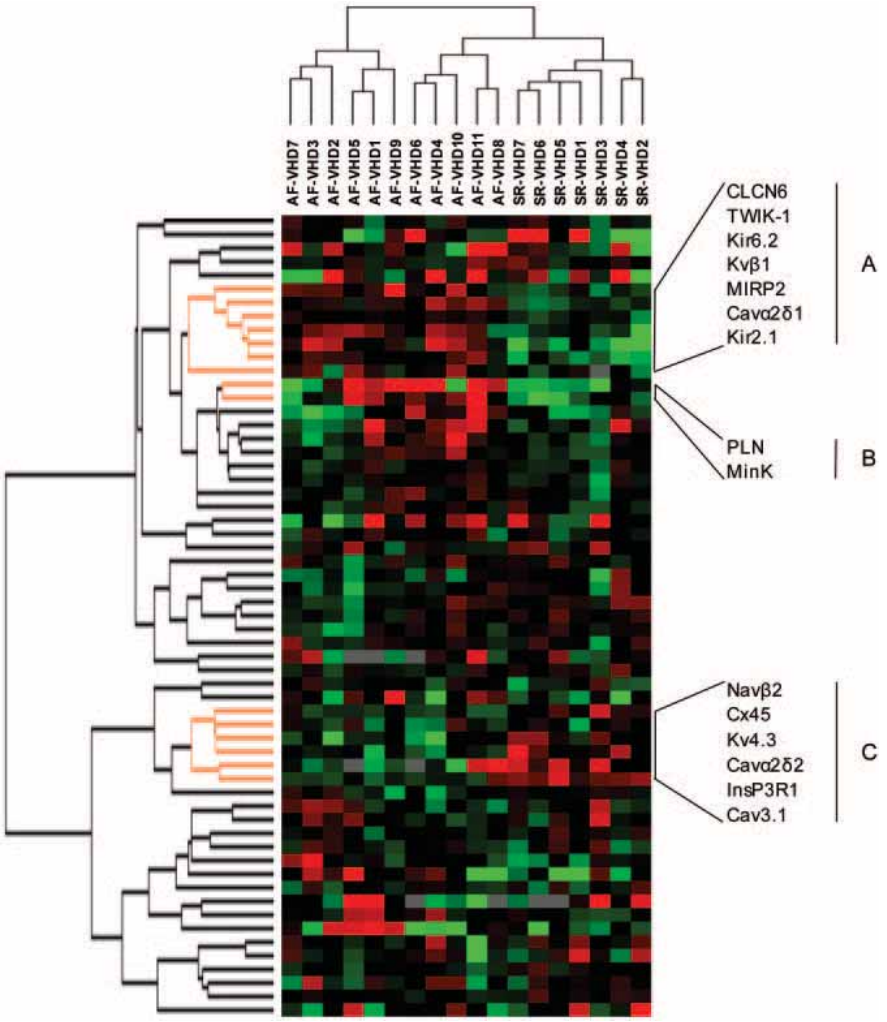



**Figure 3.** Two-way hierarchical clustering. Two-way hierarchical agglomerative clustering applied to 7 SR-VHD patients and 11 AF-VHD patients (horizontally) and to 59 selected genes (vertically). Input consisted of median of replicate expression values for each gene and patient. Each gene is represented by single row of colored boxes, and each patient by single column. Entire gene clustering is shown at left. Three selected clusters containing genes relevant to SR-VHD vs AF-VHD discrimination are shown at right (A, B, C). Each color patch in map represents gene expression level, with continuum of expression levels from dark green (lowest) to bright red (highest). Missing values are coded as gray.

tion of Kir2.1 protein was more pronounced than Kir2.1 mRNA.

## Functional Correlation

$I_{K1}$ recorded with a ramp protocol (Figure 7A and 7B) was significantly larger in AF-VHD than in SR-VHD or SR-CAD groups (Figure 7C). Peak $I_{Ca,L}$ (Figure 7B) was significantly smaller in AF-VHD than in SR-VHD (Figure 7C). $I_{Ca,L}$ was slightly, but not significantly, smaller in SR-VHD than SR-CAD.

## Reversibility of Ion-Subunit Changes

To assess potential changes in ionic remodeling with SR restoration, we used quantitative RT-PCR to study ion-channel subunit expression after SRR. Figure 8 shows mRNA quantification for 13 selected genes (3 of which had been found downregulated, 4 upregulated, and 5 unchanged in the AF-VHD group). For each of the genes altered in the AF-VHD group, values for SRR-VHD were indistinguishable from those of SR-VHD, indicating their reversibility. Genes that did not differ between SR-VHD and AF-VHD were similarly unchanged for SRR-VHD.

# Discussion

Although the link between VHD and AF is well recognized, the electrophysiological profile that distinguishes VHD patients who develop AF from those who do not remains unclear. Furthermore, there are relatively few data in the literature about the changes in ion-channel gene expression caused by VHD. The present study shows substantial remodeling in SR-VHD patients. It also shows that ion-channel subunit profiling discriminates SR-VHD patients from AF-VHD patients, and thus that there is a specific ion-channel subunit expression portrait for VHD patients with AF.

## Ion-Channel Expression Changes in Our Study Compared With the Literature

In accordance with previous studies performed in dogs and humans,[25–27] we found a clear reduction in L-type calcium-channel mRNA expression (Cav1.2 and Cav1.3) associated with a weak but statistically significant decrease in Nav1.5. The Cav1.2 differences were observed in both AF-VHD and SR-VHD patients and were confirmed at the protein level. However, functional data showed that $I_{Ca,L}$ was lower in AF-VHD than SR-VHD, which showed only a small trend to be lower than in SR-CAD. These data support the idea that the mechanism for $I_{Ca,L}$ reduction is complex and could involve altered regulation or other mechanisms distinct from mRNA and protein expression changes. Christ et al[28] recently illustrated this complexity by showing that in AF patients, increased protein phosphatase activity may contribute to

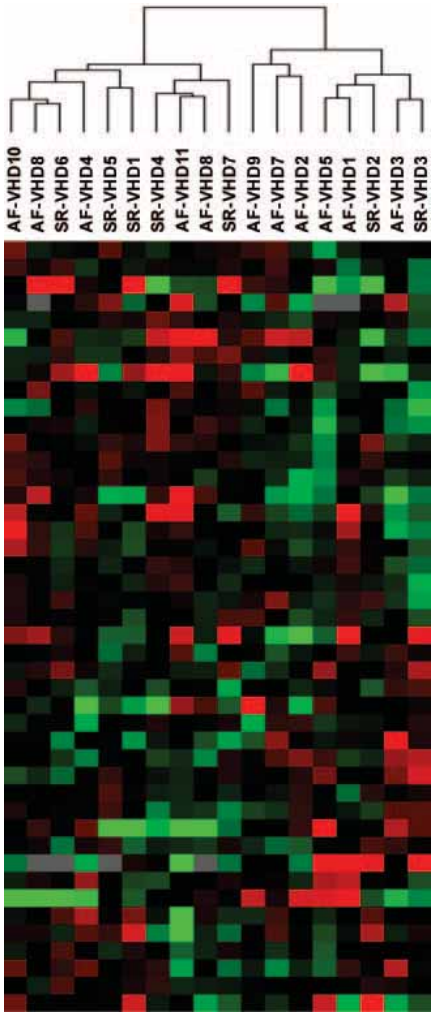



**Figure 4.** Two-way hierarchical clustering applied to 44 selected genes. Removal of gene clusters A, B, and C in Figure 3 eliminates ability to discriminate AF-VHD from SR-VHD patients, illustrating importance of these clusters. Format as in Figure 3.

impaired basal $I_{Ca,L}$. In both SR-VHD and AF-VHD groups, we observed upregulation of calcium-channel $\gamma$-subunits Cav$\gamma$4 and Cav$\gamma$7. The physiological functions of these calcium-channel subunits are complex and their role in the heart is unknown,[29] but they could contribute to determining L-type calcium-channel function in VHD and AF. Cav3.1 was upregulated in SR-VHD but unchanged in AF-VHD. Although there is some evidence of a role for T-type calcium channels in AF-related remodeling,[30] T-type calcium-current density is not altered by atrial tachycardia–induced remodeling.[13]

Our observation of Cx43 upregulation in AF is consistent with observations of Cx43 protein distribution by Elvan et al,[31] and our observation of reduced Cx40 expression in AF agrees with the elegant experiments of Nao et al[32] and Ausma et al.[33] These findings suggest that previously observed connexin-protein changes in AF may be due to transcriptional regulation. However, it should be noted that AF-related connexin expression changes have varied widely across studies, at least in part according to the model used (see Van der Velden and Jongsma[34] for review). Upregulation of Cx45 may compensate for the decreases we observed in Cx40, but the role of Cx45 in the atria is unclear. NCX1 was upregulated in AF-VHD (albeit with pronounced interindividual variability) but not in SR-VHD. Upregulation of NCX1 protein has been reported in cardiac samples from AF patients,[35] and NCX1 upregulation occurs in an animal model of AF related to congestive heart failure.[36]

In agreement with previous findings,[37] we observed downregulation of Kir3.1 subunit expression in AF patients. Several studies have reported increased inward-rectifier current density in cardiomyocytes isolated from atrial tissues from patients with AF.[3,7] The cardiac inward rectifier is carried not only by Kir2.1 channels but also by Kir2.2 and Kir2.3.[38] In our study, Kir2.1 expression was unmodified in SR-VHD but slightly increased in AF-VHD patients, whereas Kir2.2 and Kir2.3 were upregulated in both SR-VHD and AF-VHD. RT-PCR, protein, and patch-clamp data showed a significant upregulation of Kir2.1 in AF-VHD, suggesting that (1)expression changes were slightly underestimated by the microarray technique, consistent with our previous findings,[39] and (2) Kir2.1 is regulated posttranscriptionally. Previous reports have indicated downregulation of Kv4.3 subunit expression in AF patients.[5,37] Our initial microarray analysis did not show statistically significant Kv4.3 changes in either SR-VHD or AF-VHD groups. However, 2-class statistical analysis and hierarchical clustering indicated a significant contribution of Kv4.3 to the discrimination of SR-VHD from AF-VHD groups, with AF-VHD patients showing lower expression. In addition, real-time RT-PCR showed significant decreases in Kv4.3 expression in AF-VHD.

We are not aware of published studies of ion-channel subunit expression changes in VHD. Le Grand et al[12] showed that (1) L-type calcium-current density was decreased in patients with dilated atria, in agreement with our findings (Figure 1), and (2) transient-outward current was decreased, but to a lesser extent.

## Potential Significance of Our Observations

To our knowledge, this is the first study that used gene microarrays to evaluate ion-channel subunit-expression changes associated with AF and VHD. Thijssen et al[40] examined gene-expression changes in a goat model of AF with the differential display technique, but their approach did not address changes in ion-channel subunits. We obtained 2 principal general findings (1) that VHD produces important changes in cardiac ion-channel gene expression and (2) that discrete ion-channel subunit changes differentiate patients with VHD who develop persistent AF from those who do not. A number of the ion-channel and transporter subunits that are differentially affected in AF patients, including calcium-channel $\delta$-subunits, inositoltrisphosphate receptors, CLCN6 chloride-channels, and a variety of potassium-channel $\beta$-subunits, have not previously been reported in AF-related ionic remodeling. Their significance will need to be addressed in further studies. The ion-channel changes differentiating AF-VHD from SR-VHD patients could be a result of AF, which was sustained for at least 1 year before surgery in the AF-VHD group, or could theoretically represent ionic current expression patterns favoring AF development. How-

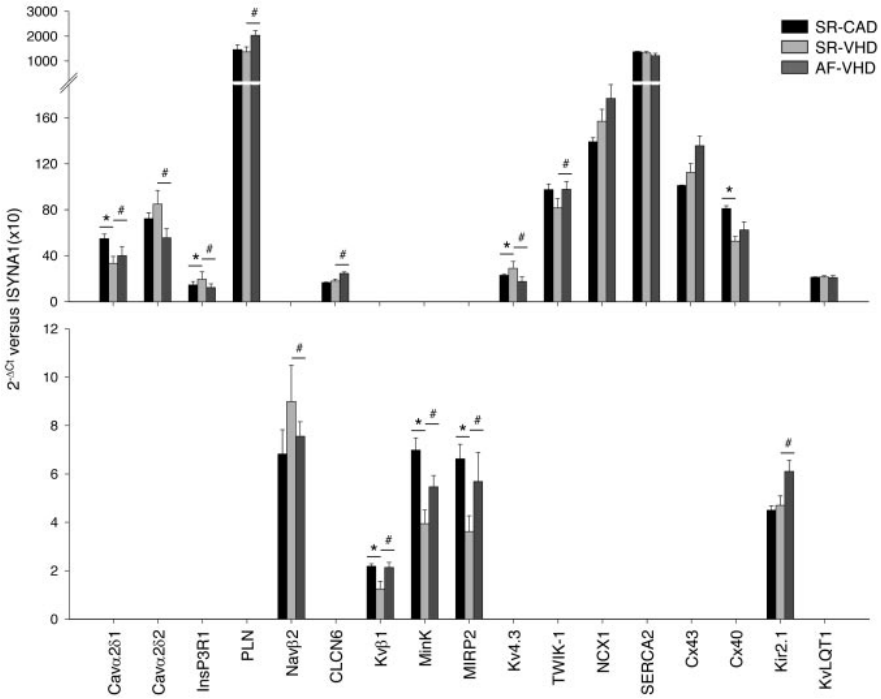



**Figure 5.** Cardiac ion-channel subunit remodeling evaluated by quantitative real-time PCR. Bar graphs show relative quantification (*y* axis) of 17 selected genes (*x* axis). In this figure, as in Figure 8, data are expressed relative to expression of *ISYNA1* (×10) and are mean±SEM from 7 to 11 patients. Top panel shows highly expressed transcripts, whereas bottom panel contains weakly expressed transcripts. Black bars are from SR group, whereas gray bars are from VHD groups. *$P<0.05$ for SR-VHD vs SR-CAD; #$P<0.05$ for AF-VHD vs SR-VHD.

ever, our finding that the SRR-VHD group had a gene expression pattern like that of SR-VHD suggests that the specific AF-VHD profile is more likely a result of AF per se than of predisposing factors.

One important observation in our study is the overall similarity in ion-channel and transporter-expression changes in SR-VHD versus AF-VHD groups (Figures 1 and 2). This highlights the need to obtain a disease-matched control group for studies of gene expression changes associated with AF. Many previous studies of AF-associated remodeling have lacked such control groups, making it difficult to know whether the changes observed in the AF population were due to AF or the underlying heart disease.

## Study Limitations

Overall, changes in the expression of ion-channel genes were small in amplitude, with the exception of changes in the L-type $Ca^{2+}$-channel Cav1.2 and Kir3.1 $K^{+}$-channel transcripts, which were decreased to less than half their control value. However, ion-channel gene expression is finely tuned. A 30% alteration in ion-channel mRNA could have a dramatic effect on cardiomyocyte electrophysiology if it translates into similar alterations at the levels of protein expression and current amplitude.

The AF-specific ionic remodeling described here was obtained in patients with >1 year of sustained arrhythmia. Different durations of AF and paroxysmal forms could produce different alterations. It is well known that the atria are heterogeneous tissues. Therefore, our data obtained in RAA samples may not accurately reflect alterations in the rest of the atria. Furthermore, ion-channel gene-expression changes in thoracic veins like the pulmonary veins,[41] the ligament of Marshall,[42] and the venae cavae[43] may be different from changes in the present study and may be

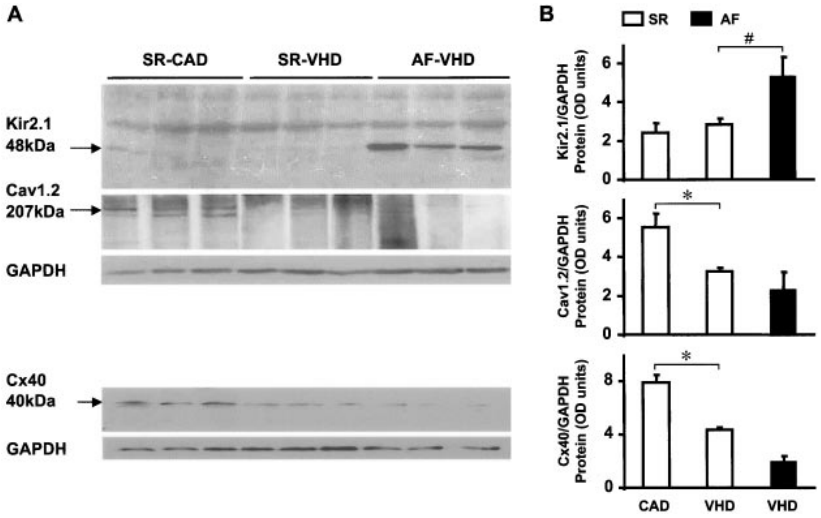



**Figure 6.** Western blot analysis of key channel proteins in SR-CAD, SR-VHD, and AF-VHD groups. A, Whole-tissue membrane proteins probed with anti-Kir2.1, anti-Cav1.2, and anti-Cx40 antibody. Expected molecular masses are indicated. Lower panels show corresponding glyceraldehyde 3-phosphate dehydrogenase (GAPDH) bands to which protein results were normalized. B, Mean±SEM protein expression values vs GAPDH. *$P<0.05$ for SR-VHD vs SR-CAD and #$P<0.05$ for AF-VHD vs SR-VHD, n=6 atrial appendages per group.

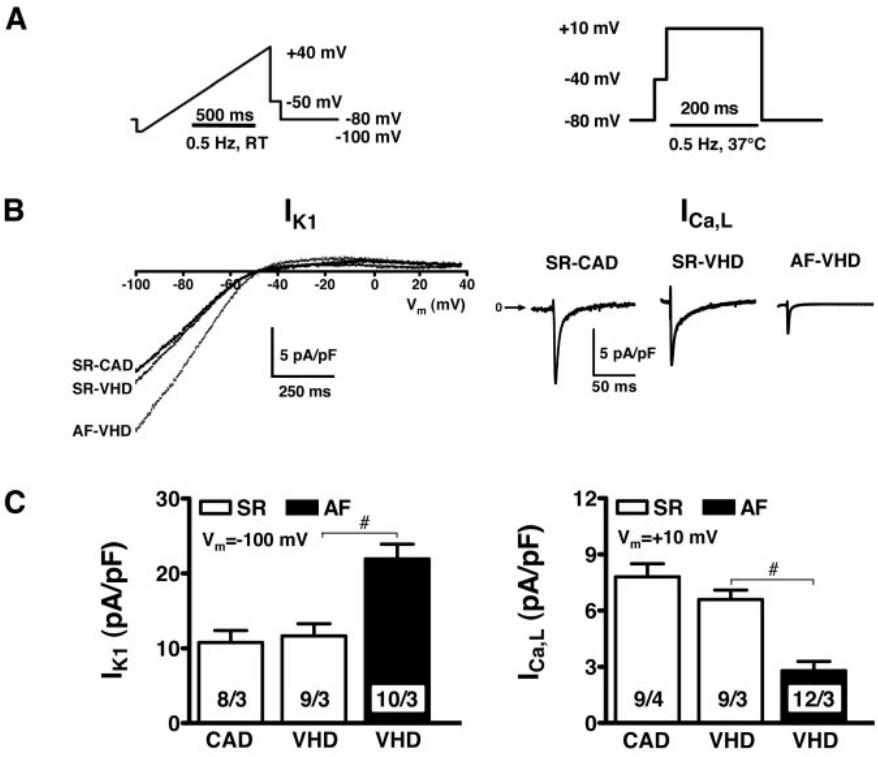



**Figure 7.** Characterization of $I_{K1}$ and $I_{Ca,L}$ in atrial myocytes from patients in SR-CAD, SR-VHD, and AF-VHD. A, Clamp protocols for $I_{K1}$ and $I_{Ca,L}$. B, Original recordings of $I_{K1}$ (only currents during depolarizing ramps are depicted) and $I_{Ca,L}$. C, Current densities of $I_{K1}$ (at 100 mV) and $I_{Ca,L}$ (at +10 mV). Numbers within columns indicate number of myocytes/patients (#$P<0.05$ for AF-VHD vs SR-VHD).

important in determining AF occurrence. The LA may better reflect the effects of VHD on ion-channel remodeling, and the impact of VHD on the right atrium, as shown here, may be an underestimate.

It would be practically impossible to confirm all of the gene expression changes we saw with Western blotting and functional studies. We were able to confirm alterations in Kir2.1, CaV1.2, and Cx40. In addition, our transcriptional data for RYR2, PLN, SERCA2, and CASQ2 agree with Western blot data obtained by Vest et al[44] and El-Armouche et al (unpublished data). However, these data clearly do not bear on the rest of the transcripts studied.

A final limitation is the fact that our patients in AF were taking a variety of cardioactive drugs, which differed to some extent between SR-VHD and AF-VHD patients. An example is the higher incidence of aspirin treatment in the control group and of fluindone in the AF-VHD group. Although there was no obvious relation between specific drug treatments and

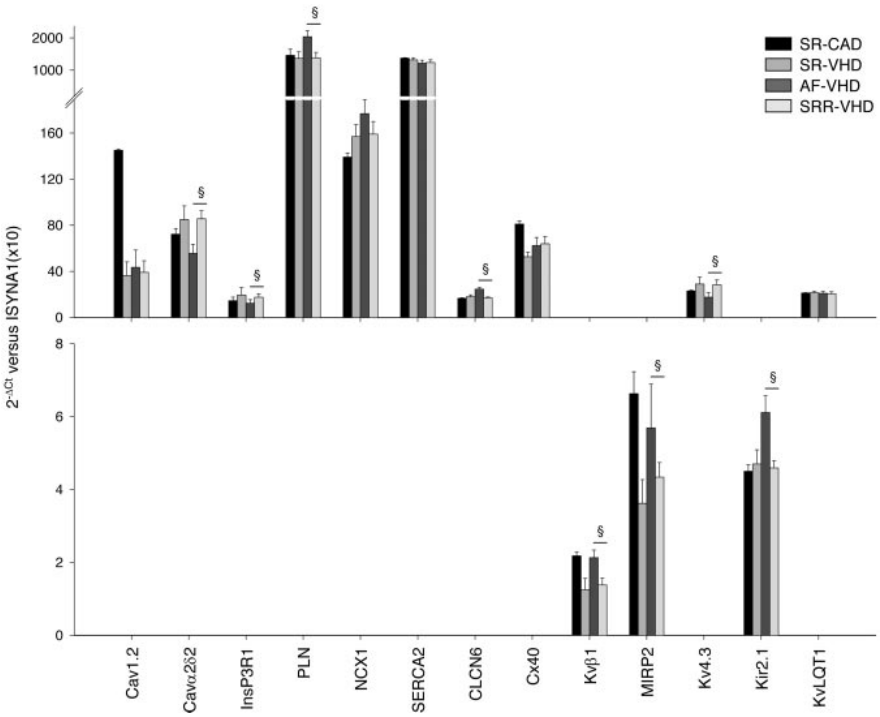



**Figure 8.** Real-time RT-PCR quantification in SRR-VHD patients. As in Figure 5, data are mean±SEM of expression values (n=6 in SRR-VHD group) and are expressed as ratio vs *ISYNA1* (×10). §$P<0.05$ for AF-VHD vs SRR-VHD.

ion-channel expression pattern, as assessed with clustering, we cannot exclude some influence of these inevitable differences in drug therapy between patients in SR and those in AF.

The remodeling changes due to SR-VHD might be expected to parallel those in animal models of congestive heart failure[36] and those associated specifically with AF-VHD to parallel those in models of atrial tachycardia remodeling.[4,27] Although there are many similarities, there are also differences that could be due to species differences, discrepancies in the precise form of heart disease and its duration, and drug therapy effects in the clinical population. In addition, tachycardia-related remodeling is altered by concomitant heart failure, so that tachycardia-induced ionic remodeling in a pathological substrate may differ from what is seen in the normal heart, as in experimental paradigms.[45]

## Acknowledgments

This study was supported by special grants from Ouest genopole, INSERM, CNRS, Ministère français de la Recherche (ACI), Canadian Institutes of Health Research, Quebec Heart and Stroke Foundation, Bundes Ministerium für Bildung und Forschung, and the Deutsche Forschungsgemeinschaft. The authors thank Martine Le Cunff, Isabelle Guisle, Catherine Chevalier, Jean-Marie Heslan, Evelyn Landry, Annegrett Häntzschel, Trautlinde Thurm, and Ulrike Heinrich for their excellent technical assistance.

## References

1. Benjamin EJ, Wolf PA, D'Agostino RB, Silbershatz H, Kannel WB, Levy D. Impact of atrial fibrillation on the risk of death: the Framingham Heart Study. *Circulation*. 1998;98:946–952.
2. Nattel S. New ideas about atrial fibrillation 50 years on. *Nature*. 2002; 415:219–226.
3. Bosch RF, Zeng X, Grammer JB, Popovic K, Mewis C, Kuhlkamp V. Ionic mechanisms of electrical remodeling in human atrial fibrillation. *Cardiovasc Res*. 1999;44:121–131.
4. Yue L, Feng J, Gaspo R, Li GR, Wang Z, Nattel S. Ionic remodeling underlying action potential changes in a canine model of atrial fibrillation. *Circ Res*. 1997;81:512–525.
5. Grammer JB, Bosch RF, Kuhlkamp V, Seipel L. Molecular remodeling of Kv4.3 potassium channels in human atrial fibrillation. *J Cardiovasc Electrophysiol*. 2000;11:626–633.
6. Van Wagoner DR, Pond AL, McCarthy PM, Trimmer JS, Nerbonne JM. Outward $K^+$ current densities and Kv1.5 expression are reduced in chronic human atrial fibrillation. *Circ Res*. 1997;80:772–781.
7. Dobrev D, Graf E, Wettwer E, Himmel HM, Hala O, Doerfel C, Christ T, Schuler S, Ravens U. Molecular basis of downregulation of G-protein-coupled inward rectifying K(+) current (I(K,ACh) in chronic human atrial fibrillation: decrease in GIRK4 mRNA correlates with reduced I(K,ACh) and muscarinic receptor-mediated shortening of action potentials. *Circulation*. 2001;104:2551–2557.
8. Brundel BJ, Henning RH, Kampinga HH, Van Gelder IC, Crijns HJ. Molecular mechanisms of remodeling in human atrial fibrillation. *Cardiovasc Res*. 2002;54:315–324.
9. Van Wagoner DR, Nerbonne JM. Molecular basis of electrical remodeling in atrial fibrillation. *J Mol Cell Cardiol*. 2000;32:1101–1117.
10. Nattel S, Li D. Ionic remodeling in the heart: pathophysiological significance and new therapeutic opportunities for atrial fibrillation. *Circ Res*. 2000;87:440–447.
11. Kannel WB, Wolf PA, Benjamin EJ, Levy D. Prevalence, incidence, prognosis, and predisposing conditions for atrial fibrillation: population-based estimates. *Am J Cardiol*. 1998;82:2N–9N.
12. Le Grand BL, Hatem S, Deroubaix E, Couetil JP, Coraboeuf E. Depressed transient outward and calcium currents in dilated human atria. *Cardiovasc Res*. 1994;28:548–556.
13. Van Wagoner DR, Pond AL, Lamorgese M, Rossie SS, McCarthy PM, Nerbonne JM. Atrial L-type $Ca^{2+}$ currents and human atrial fibrillation. *Circ Res*. 1999;85:428–436.
14. Le Meur N, Lamirault G, Bihouée A, Steenman M, Bedrine-Ferran H, Teusan R, Ramstein G, Leger JJ. A dynamic, web-accessible resource to process raw microarray scan data into consolidated gene expression values: importance of replication. *Nucleic Acids Res*. 2004;32: 5349–5358.
15. Tseng GC, Oh MK, Rohlin L, Liao JC, Wong WH. Issues in cDNA microarray analysis: quality filtering, channel normalization, models of variations and assessment of gene effects. *Nucleic Acids Res*. 2001;29: 2549–2557.
16. Yang YH, Dudoit S, Luu P, Lin DM, Peng V, Ngai J, Speed TP. Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Res*. 2002;30:e15.
17. Iglewicz B, Hoaglin DC. *How to Detect and Handle Outliers*. Milawaukee, Wis: American Society for Quality Control; 1993.
18. Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionising radiation response. *Proc Natl Acad Sci U S A*. 2001;98:5116–5121.
19. Smyth GK, Yang YH, Speed T. Statistical issues in cDNA microarray data analysis. *Methods Mol Biol*. 2003;224:111–136.
20. Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci U S A*. 1998;95:14863–14868.
21. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) method. *Methods*. 2001;25:402–408.
22. Zicha S, Xiao L, Stafford S, Cha TJ, Han W, Varro A, Nattel S. Transmural expression of transient outward potassium current subunits in normal and failing canine and human hearts. *J Physiol*. 2004;561: 735–748.
23. Dobrev D, Wettwer E, Himmel HM, Kortner A, Kuhlisch E, Schuler S, Siffert W, Ravens U. G-protein $\beta_3$-subunit 825T-allele is associated with enhanced human atrial inward rectifier potassium currents. *Circulation*. 2000;102:692–697.
24. Christ T, Boknik P, Wohrl S, Wettwer E, Graf EM, Bosch RF, Knaut M, Schmitz W, Ravens U, Dobrev D. Reduced L-type $Ca^{2+}$ current density in chronic human atrial fibrillation is associated with increased activity of protein phosphatases. *Circulation*. 2004;110:2651–2657.
25. Grammer JB, Zeng X, Bosch RF, Kuhlkamp V. Atrial L-type $Ca^{2+}$-channel, $\beta$-adrenorecptor, and 5-hydroxytryptamine type 4 receptor mRNAs in human atrial fibrillation. *Basic Res Cardiol*. 2001;96:82–90.
26. Lai LP, Su MJ, Lin JL, Lin FY, Tsai CH, Chen YS, Huang SK, Tseng YZ, Lien WP. Down-regulation of L-type calcium channel and sarcoplasmic reticular $Ca^{2+}$-ATPase mRNA in human atrial fibrillation without significant change in the mRNA of ryanodine receptor, calsequestrin and phospholamban: an insight into the mechanism of atrial electrical remodeling. *J Am Coll Cardiol*. 1999;33:1231–1237.
27. Yue L, Melnyk P, Gaspo R, Wang Z, Nattel S. Molecular mechanisms underlying ionic remodeling in a dog model of atrial fibrillation. *Circ Res*. 1999;84:776–784.
28. Christ T, Boknik P, Wohrl S, et al. L-type Ca2+ current downregulation in chronic human atrial fibrillation is associated with increased activity of protein phosphatases. *Circulation*. 2004;110:2651–2657.
29. Kang MG, Campbell KP. $\gamma$-Subunit of voltage-activated calcium channels. *J Biol Chem*. 2003;278:21315–21318.
30. Fareh S, Benardeau A, Thibault B, Nattel S. The T-type $Ca^{2+}$ channel blocker mibefradil prevents the development of a substrate for atrial fibrillation by tachycardia-induced atrial remodeling in dogs. *Circulation*. 1999;100:2191–2197.
31. Elvan A, Huang XD, Pressler ML, Zipes DP. Radiofrequency catheter ablation of the atria eliminates pacing-induced sustained atrial fibrillation and reduces connexin 43 in dogs. *Circulation*. 1997;96:1675–1685.
32. Nao T, Ohkusa T, Hisamatsu Y, Inoue N, Matsumoto T, Yamada J, Shimizu A, Yoshiga Y, Yamagata T, Kobayashi S, Yano M, Hamano K, Matsuzaki M. Comparison of expression of connexin in right atrial myocardium in patients with chronic atrial fibrillation versus those in sinus rhythm. *Am J Cardiol*. 2003;91:678–683.
33. Ausma J, van der Velden HM, Lenders MH, van Ankeren EP, Jongsma HJ, Ramaekers FC, Borgers M, Allessie MA. Reverse structural and gap-junctional remodeling after prolonged atrial fibrillation in the goat. *Circulation*. 2003;107:2051–2058.
34. Van der Velden HM, Jongsma HJ. Cardiac gap junctions and connexins: their role in atrial fibrillation and potential as therapeutic targets. *Cardiovasc Res*. 2002;54:270–279.
35. Schotten U, Greiser M, Benke D, Buerkel K, Ehrenteidt B, Stellbrink C, Vasquez-Jimenez JF, Schoendube F, Hanrath P, Allessie M. Atrial

fibrillation-induced atrial contractile dysfunction: a tachycardiomyopathy of a different sort. *Cardiovasc Res*. 2002;53:192–201.
36. Li D, Melnyk P, Feng J, Wang Z, Petrecca K, Shrier A, Nattel S. Effects of experimental heart failure on atrial cellular and ionic electrophysiology. *Circulation*. 2000;101:2631–2638.
37. Brundel BJ, Van Gelder IC, Henning RH, Tuinenburg AE, Wietses M, Grandjean JG, Wilde AA, Van Gilst WH, Crijns HG. Alterations in potassium channel gene expression in atria of patients with persistent and paroxysmal atrial fibrillation: differential regulation of protein and mRNA levels for K+ channels. *J Am Coll Cardiol*. 2001;37:926–932.
38. Dhamoon AS, Pandit SV, Sarmast F, Parisian KR, Guha P, Li Y, Bagwe S,Taffet SM, Anumonwo JM. Unique Kir2.x properties determine regional and species differences in the cardiac inward rectifier $K^+$ current. *Circ Res*. 2004;94:1332–1339.
39. Le Bouter S, El Harchi A, Marionneau C, Bellocq C, Chambellan A, van Veen T, Boixel C, Gavillet B, Abriel H, Le Quang K, Chevalier JC, Lande G, Leger JJ, Charpentier F, Escande D, Demolombe S. Chronic amiodarone remodels expression of ion channel transcripts in the mouse heart. *Circulation*. 2004;110:3028–3035.
40. Thijssen VL, van der Velden HM, van Ankeren EP, Ausma J, Allessie MA, Borgers M, van Eys GJ, Jongsma JH. Analysis of altered gene expression during sustained atrial fibrillation in the goat. *Cardiovasc Res*. 2002;54:427–437.
41. Nattel S. Basic electrophysiology of the pulmonary veins and their role in atrial fibrillation: precipitators, perpetuators, and perplexers. *J Cardiovasc Electrophysiol*. 2003;14:1372–1375.
42. Hwang C, Wu TJ, Doshi RN, et al. Vein of marshall cannulation for the analysis of electrical activity in patients with focal atrial fibrillation. *Circulation*. 2000;101:1503–1505.
43. Tsai CF, Tai CT, Hsieh MH, Lin WS, Yu WC, Ueng KC, Ding YA, Chang MS, Chen SA. Initiation of atrial fibrillation by ectopic beats originating from the superior vena cava: electrophysiological characteristics and results of radiofrequency ablation. *Circulation*. 2000;102:67–74.
44. Vest JA, Wehrens XH, Reiken SR, Lehnart SE, Dobrev D, Chandra P, Danilo P, Ravens U, Rosen MR, Marks AR. Defective cardiac ryanodine receptor regulation during atrial fibrillation. *Circulation*. 2005;111:2025–2032.
45. Cha TJ, Ehrlich JR, Zhang L, Nattel S. Atrial ionic remodeling induced by atrial tachycardia in the presence of congestive heart failure. *Circulation*. 2004;110:1520–1526.

## CLINICAL PERSPECTIVE

Remodeling of atrial ion-channel expression occurs in atrial fibrillation (AF) and may play an important role in its maintenance. Changes in ion-channel function have potentially significant implications for designing more effective and specific antiarrhythmic drug therapy for AF. In this study, we applied gene microarray technology to examine the expression of a wide range of cardiac ion channel and transporter genes in atrial tissues from patients with no atrial disease undergoing routine coronary-bypass surgery and valvular heart disease (VHD) patients in sinus rhythm (VHD-SR) or with longstanding AF (VHD-AF) undergoing valve-replacement surgery. We found substantial ion-channel gene-expression changes in both VHD-SR and VHD-AF, with roughly two thirds of remodeled genes being common to both; however, there was also a specific pattern of gene-expression change characteristic of VHD-AF, including several ion-channel genes not previously known to be remodeled in AF. Our study shows that VHD substantially affects ion-channel expression, emphasizing the importance of disease-matched controls for any AF-remodeling study, and that there is a pattern of ion-channel gene-expression change specifically related to AF. These findings have potentially important implications for understanding how VHD leads to AF, how AF alters cardiac ion-channel function, and how to design improved ion channel–modulating antiarrhythmic drug targets for AF patients.

Titre : Portraits moléculaires des pathologies cardiaques

Résumé :

Les pathologies cardiaques représentent une cause majeure de morbidité et de mortalité. L'achèvement du séquençage du génome humain a permis le développement de nouvelles approches des maladies comme les études d'expression génique. Au moyen de puces à ADN, les travaux de cette thèse se sont centrés sur l'étude des variations transcriptomales survenant dans l'insuffisance cardiaque et la fibrillation auriculaire chez l'homme. Un profil d'expression génique permettant de prédire le niveau de sévérité clinique des patients en insuffisance cardiaque avancée a été identifié et pourrait servir de base au développement d'un biomarqueur du pronostic. Dans la fibrillation auriculaire chronique, l'analyse des variations d'expression génique a suggéré l'existence d'un phénotype tissulaire pro-thrombotique qui pourrait mieux expliquer le risque thromboembolique majeur de la maladie. Ces résultats ont globalement confirmé l'intérêt de l'étude des variations transcriptomales dans les maladies cardiaques. La poursuite des travaux déjà engagés devrait non seulement permettre une meilleure compréhension des mécanismes moléculaires des pathologies mais également l'identification de biomarqueurs à but diagnostique, pronostique ou d'orientation thérapeutique.

Mots-clés :

Génomique, profils d'expression génique, puces à ADN, insuffisance cardiaque, fibrillation auriculaire.

Title : Molecular portraits of cardiac diseases

Abstract :

Cardiac diseases remains a major cause of mortality and morbidity. With the completion of the sequencing of the human genome, new techniques like gene expression profiling have been developped to study human diseases. Using DNA-microrrays, we analyzed transcriptomal alterations associated with heart failure and atrial fibrillation in humans. We identified a gene expression profile associated with clinical deterioration of advanced heart failure patients which may be used to better define patients' prognosis. Gene expression alteration in chronic atrial fibrillation were suggestive of a hypercoagulable state, a result of potential importance to better understand the pathophysiology of thromboembolic events in atrial fibrillation. Taken together, these results show the great potential of gene expression profiling to study cardiac diseases. Further studies will improve our knowledge of pathological mecanisms occuring in these diseases and will probably provide us with new biomarkers for diagnosis, prognosis and therapy of cardiac diseases.

Key words :

Genomics, Gene expression profiling, DNA microarrays, Heart failure, Atrial fibrillation