

Université de Montréal

**Approches transcriptomiques dans l'étude de la
fibrose kystique**

par

Grégory VOISIN

**Département de Médecine, Université de Montréal
Faculté de Médecine**

**Mémoire présenté à la Faculté des études supérieures
en vue de l'obtention du grade de Maître es sciences (M.Sc.)
en Sciences Biomédicales**

Mars 2010

© Voisin Grégory, 2010

Université de Montréal
Faculté des études supérieures

Ce mémoire intitulé :

**Approches transcriptomiques dans l'étude de la
fibrose kystique**

présenté par :

Grégory VOISIN

a été évalué par un jury composé des personnes suivantes :

Dre Céline Bergeron

Présidente-rapportrice

Dr Yves Berthiaume

Directeur de recherche

Dr Sébastien Lemieux

Codirecteur

Dr Pierre Legendre

Membre du jury

Résumé

Les avancées en biotechnologie ont permis l'identification d'un grand nombre de mécanismes moléculaires, soulignant également la complexité de la régulation génique. Néanmoins, avoir une vision globale de l'homéostasie cellulaire, nous est pour l'instant inaccessible et nous ne sommes en mesure que d'en avoir qu'une vue fractionnée. Étant donné l'avancement des connaissances des dysfonctionnements moléculaires observés dans les maladies génétiques telles que la fibrose kystique, il est encore difficile de produire des thérapies efficaces.

La fibrose kystique est causée par la mutation de gène CFTR (*cystic fibrosis transmembrane conductance regulator*), qui code pour un canal chlorure transmembranaire. La mutation la plus fréquente ($\Delta F508$) induit un repliement incorrect de la protéine et sa rétention dans le réticulum endoplasmique. L'absence de CFTR fonctionnel à la membrane a un impact sur l'homéostasie ionique et sur l'hydratation de la muqueuse respiratoire. Ceci a pour conséquence un défaut dans la clairance mucociliaire, induisant infection chronique et inflammation excessive, deux facteurs fondamentaux de la physiopathologie.

L'inflammation joue un rôle très important dans l'évolution de la maladie et malgré le nombre important d'études sur le sujet, la régulation du processus inflammatoire est encore très mal comprise et la place qu'y occupe le CFTR n'est pas établie.

Toutefois, plusieurs autres facteurs, tels que le stress oxydatif participent à la physiopathologie de la maladie, et considérer leurs impacts est important pour permettre une vision globale des acteurs impliqués.

Dans notre étude, nous exploitons la technologie des puces à ADN, pour évaluer l'état transcriptionnel d'une cellule épithéliale pulmonaire humaine fibro-kystique.

Dans un premier temps, l'analyse de notre expérience identifie 128 gènes inflammatoires sur-exprimés dans les cellules FK par rapport aux cellules non FK où apparaissent plusieurs familles de gènes inflammatoires comme les cytokines ou les calgranulines. L'analyse de la littérature et des annotations suggèrent que la modulation de ces transcrits dépend de la cascade de NF- κ B et/ou des voies de signalisation associées aux interférons (IFN). En outre, leurs modulations pourraient être associées à des modifications épigénétiques de leurs loci chromosomiques.

Dans un second temps, nous étudions l'activité transcriptionnelle d'une cellule épithéliale pulmonaire humaine FK en présence de DMNQ, une molécule cytotoxique. Notre but est d'identifier les processus biologiques perturbés par la mutation du gène CFTR en présence du stress oxydatif.

Fondé sur une analyse canonique de redondance, nous identifions 60 gènes associés à la mort cellulaire et leur variance, observée dans notre expérience, s'explique par un effet conjoint de la mutation et du stress oxydatif. La mesure de l'activité des caspases 3/7, des effecteurs de l'apoptose (la mort cellulaire programmée), montre que les cellules porteuses de la mutation Δ F508, dans des conditions de stress oxydatif, seraient moins apoptotiques que les cellules saines. Nos données transcriptomiques suggèrent que la sous-activité de la cascade des MAPK et la sur-expression des gènes anti-apoptotiques pourraient être impliquées dans le déséquilibre de la balance apoptotique.

Mots clés : Fibrose kystique, Δ F508, puce à ADN, inflammation, apoptose, analyse canonique de redondance.

Abstract

Biotechnical advances have allowed a large number of molecular mechanisms to be identified, and have also underlined the complexity of gene regulation. Nonetheless, we still only have a partial understanding of cellular homeostasis. Even with our current knowledge, molecular dysfunctionality observed in genetic illnesses such as cystic fibrosis remain largely misunderstood, prohibiting us from developing efficient treatments. Cystic fibrosis is caused by a mutation of the CFTR (*cystic fibrosis transmembrane conductance regulator*) gene which codes for a chloric transmembrane channel. The most frequent mutation ($\Delta F508$) causes the unfolding of the protein and its retention in the endoplasmic reticulum. This absence of a functional CFTR to the membrane has an impact on the ionic homeostasis and the hydration of the respiratory mucus. Consequently, the mucociliary clearance is disrupted, which leads to chronic infection and excessive inflammation - two fundamental factors of CF's physiopathology.

Inflammation plays a key role in the evolution of the illness and despite many studies on this subject, the regulation of the inflammatory process remains a mystery and CFTR's role is not well established. Additionally, since the physiopathology cannot be explained by mutation alone, several authors suggest the importance of other factors (genetic and/or environmental) which are factors in the clinical picture of the illness.

In our study, we use microarray technology to evaluate the transcriptional state of an epithelial lung cell issued from a human with cystic fibrosis. Initially, our analysis identifies 128 inflammatory genes which are over-expressed in the cystic fibrosis cells versus non-cystic fibrosis cells, showing several inflammatory gene families such as cytokines or calgranulins. An analysis of the publications and annotations of these transcripts suggests that their modulations depend upon the cascade of NF- κ B and/or signalling pathways

associated with interferons (IFN). In other words, their modulations could be associated with epigenetic modifications of their chromosomal loci. Secondly, we study the transcriptional activity of an epithelial lung cell issued from a human with cystic fibrosis in the presence of DMNQ, a cytotoxic molecule. Our goal is to identify the biological processus which are disturbed by the mutation of the CFTR gene in the presence of oxidative stress.

Based on a canonical redundancy analysis, we identify 60 genes associated with cell death and their modulation in our study explained by the combined effect of mutation and oxidative stress. By measuring the activity of caspase 3/7, an effector of apoptosis (programmed cell death), we see that the cells containing the mutation $\Delta F508$ could present a problem with apoptosis. Our transcriptomic data suggest that decreased activity of the MAPK cascade and over-expression of anti-apoptotic genes would be a factor in apoptotic imbalance.

Keywords: Cystic fibrosis, $\Delta F508$, microarray, inflammation, apoptosis, canonical redundancy analysis

Table des matières

INTRODUCTION

A. Historique de la maladie	1
B. La structure génique du CFTR	3
C. La protéine CFTR	3
D. Fonction de la protéine CFTR.....	5
1. CFTR, transporteur de chlore, d'anion et de molécules organiques	5
2. CFTR, régulateur de canaux ioniques	6
3. D'autres fonctions connues du CFTR.....	7
4. CFTR, alias ABCC7, membre de la famille des ABC.....	7
E. La fibrose kystique	8
1. Facteurs expliquant la physiopathologie	8
a. <i>Les mutations du gène CFTR</i>	<i>8</i>
b. <i>La mutation $\Delta F508$.....</i>	<i>10</i>
c. <i>D'autres facteurs génétiques impliqués.....</i>	<i>10</i>
2. La physiopathologie de la maladie	11
a. <i>L'atteinte pulmonaire</i>	<i>12</i>
b. <i>L'hyperinflammation des cellules FK</i>	<i>15</i>
c. <i>Un facteur aggravant de la physiopathologie : le stress oxydatif.....</i>	<i>17</i>
d. <i>Le rôle de l'apoptose dans l'épithélium pulmonaire.....</i>	<i>17</i>
F. L'approche transcriptomique dans l'étude de la maladie	19
1. La technologie des puces à ADN	20
2. L'approche transcriptomique dans d'autres pathologies	20
3. Les études transcriptomiques dans la fibrose kystique	22
G. Hypothèses.....	24
H. Objectifs	25
A. Approches expérimentales	26

B. Techniques et protocoles expérimentales de l'objectif 1 : Étude de la réponse transcriptomique inflammatoire dans des cellules FK	30
1. Modèles cellulaires	30
2. Conditions de culture des modèles cellulaires	31
3. Mesure de l'activité de NF- κ B	32
4. Expérience de PCR quantitative (qPCR)	34
a. <i>Technique qPCR</i>	34
b. <i>Synthèse de l'ADNc</i>	34
c. <i>Réaction de qPCR</i>	35
d. <i>Analyse des résultats de qPCR</i>	35
5. Protocole de l'objectif 1	36
C. Technique et protocoles expérimentales de l'objectif 2 : Études de la réponse transcriptomique de cellules FK induite par un stress oxydatif	37
1. Modèles cellulaires et conditions de culture	37
2. Choix du modèle de stress oxydatif	37
3. Mesure de l'activité des caspases -3 et -7	37
4. Protocole	39
D. Méthodes de mesures physiopathologiques	40
1. Caractéristiques électrophysiologiques du modèle cellulaire	40
2. La mesure du stress oxydatif	41
E. Expérience de puce à ADN	42
1. Caractéristiques de la plateforme d'Affymetrix® HG-U133 Plus 2.0	42
2. De l'extraction des transcrits au contrôle qualité des données d'expression	44
a. <i>Extraction des ARNm</i>	44
b. <i>Synthèse d'ADNc et hybridation</i>	44
c. <i>Lecture des puces et acquisition des données d'expression</i>	45
d. Contrôle qualité	45
3. Normalisation des données d'expression	51

a. Méthode RMA.....	52
b. Normalisation quantile.....	53
F. Outils bioinformatiques.....	54
G. Détection des gènes différentiellement modulés dans le protocole 1	55
1. Test t modifié.	55
2. Valeur P et correction pour les comparaisons multiples.....	55
H. Détection des gènes différentiellement modulés dans l'objectif 2	57
1. Filtrage des données	57
2. Variance des sondes	60
3. L'ANOVA	62
4. L'analyse canonique de redondance	62
5. La classification des profils d'expression	63
I. L'interprétation biologique de données transcriptomiques.....	64
1. Annotations génomiques	64
2. Analyse ontologique et son approche statistique.....	65
A. Contrôle qualité des données d'expression.....	70
B. Normalisation, normalité et homogénéité de la variance	73
C. Résultats de l'analyse inflammatoire	74
1. Notre analyse statistique isole 9 760 transcripts différentiellement exprimés entre le type CuFi-1 et NuLi-1.....	74
2. 2 760 gènes connus sont significativement modulés dans notre expérience.....	78
3. L'analyse ontologique des 2 760 transcripts d'intérêt identifie 8 fonctions biologiques significativement surreprésentées, associées à des processus inflammatoires	79
4. L'analyse des annotations et des données transcriptomiques révèle qu'une partie des gènes inflammatoires modulés présente des caractéristiques communes.....	80
5. La confirmation de modulation de gènes inflammatoires	83

6. Des activateurs de la voie de signalisation de NF- κ B induisent une réponse différente dans les cellules CuFi-1 et NuLi-1	83
D. Traitement des 2 lignées cellulaires au DMNQ.....	85
E. Résultats de l'analyse de l'impact du stress oxydatif sur le transcriptome d'une cellule FK.....	89
1. Un sous-groupe des transcrits testés présente un intérêt en termes de variance dans notre analyse	89
2. Une analyse ANOVA détectent les effets significatifs de la mutation et de l'oxydant mais manque de puissance pour détecter l'interaction des deux facteurs	90
3. Une analyse de redondance sélectionne 687 transcrits dont la variance est expliquée par l'effet conjoint de la mutation et du stress oxydatif.....	93
4. L'apoptose est un processus significativement surreprésenté dans les transcrits dont la variance est expliquée par la mutation et le stress oxydant.	
97	
5. La classification des profils d'expression des transcrits du secteur D montrent quatre grandes tendances de modulation vis-à-vis des conditions expérimentales.	102
6. De grandes classes fonctionnelles se dégagent des 40 gènes associés à la mort cellulaire et à l'apoptose	105
7. Le DMNQ induit une diminution de l'activité des caspases 3/7 dans la lignée porteuse de la mutation Δ F508 par rapport à la lignée contrôle.....	108
A. Une vision globale du transcriptome	110
1. Apports de l'expérience transcriptomique	110
2. Limites de l'expérience transcriptomique	110
3. Nécessité de la validation expérimentale.....	111
B. Le design exprimental	112
1. Le niveau de réplication	113
2. Les modèles cellulaires.....	113

3. Les conditions expérimentales.....	114
C. Le prétraitement et le filtrage des données d'expression	115
1. Le contrôle qualité.....	115
2. La normalisation.....	116
3. Le filtrage des données.....	117
D. Les développements et limites de nos analyses différentielles	119
1. Aspects statistiques de l'analyse univariée.....	120
2. Caractéristiques de notre analyse multivariée	126
E. Les facteurs impliqués dans l'interprétation biologique	130
1. De la sélection de transcripts à l'information biologique.....	130
2. Les annotations génomiques	132
3. L'analyse ontologique.....	133
4. La réduction de l'information	137
F. L'interprétation biologique de nos résultats	139
1. Caractérisation de l'inflammation de cellule fibro-kystique.....	139
a. <i>Cinq grandes familles inflammatoires</i>	140
b. <i>Deux systèmes de régulation géniques.</i>	146
c. <i>La comparaison avec des expériences similaires.</i>	149
d. <i>Une régulation génique au niveau chromosomique.</i>	151
2. Impact du stress oxydatif sur une cellule fibro-kystique	155
a. <i>Dysfonctionnement de la prolifération et de la mort cellulaire dans notre modèle FK</i>	155
b. <i>Vers un modèle d'étude de l'apoptose</i>	157
c. <i>La balance apoptotique au niveau transcriptomique</i>	159
d. <i>Le contrebalancement anti-apoptotique</i>	162

CONCLUSION

Liste des tableaux

Tableau I:		
	Les différentes expériences de puces à ADN dans le domaine de la fibrose kystique	p 22
Tableau II :		
	Fonctions biologiques inflammatoires surreprésentées dans notre analyse différentielle	p 79
Tableau III :		
	Caractéristiques des transcrits modulés appartenant à des classes de fonctions inflammatoires	p 82
Tableau IV:		
	Résultats des validations de qPCR d'une sélection de gènes inflammatoires d'intérêt	p 83
Tableau V :		
	Résumé des probabilités issues du test ANOVA multivariable à deux facteurs pour chaque groupe sélectionné	p 91
Tableau VI :		
	Résumé des résultats produits par l'analyse canonique de redondance sous forme de variance	p 94
Tableau VII:		
	Résultats du test binomial de proportion	p 102
Tableau VIII :		
	Gènes apoptotiques organisés selon quatre termes GO	p 107
Tableau IX :		
	Modulations observées dans les autres expériences de puces à ADN des 128 gènes identifiés dans notre étude	p 112

Liste des figures

- Figure 1 :**
Les dates clés de la fibrose kystique p 2
- Figure 2 :**
Description du gène CFTR, de son transcript et de sa protéine p 4
- Figure 3 :**
Les différentes classes de mutations du CFTR et leurs effets sur l'adressage p 9
- Figure 4 :**
Matériels et méthodes : de la cellule aux données d'expression p 28
- Figure 5 :**
Matériels et méthodes : des données d'expression à l'interprétation biologique p 29
- Figure 6 :**
Caractéristiques de la puce à ADN utilisée p 43
- Figure 7 :**
Exemple de puces problématiques p 47
- Figure 8:**
Exemple de contrôle qualité basé sur la distribution des intensités des oligonucléotides p 48
- Figure 9:**
Quelques illustrations de l'Inter-Quartile Range (IQR) p 59
- Figure 10:**
Exemple de DAG issu de l'analyse ontologique de nos données transcriptomiques p 69

Figure 11 :	Détermination de la dose de DMNQ à utiliser et son impact sur l'intégrité de la monocouche cellulaire	p 88
Figure 12 :	Résultats du contrôle qualité des 12 hybridations	p 72
Figure 13 :	Normalisation des données d'expression	p 73
Figure 14:	Histogramme de Probabilités issues de l'analyse différentielle CuFi-1/NuLi-1	p 76
Figure 15 :	<i>MA-plot</i> et <i>Volcano-plot</i>	p 77
Figure 16 :	Activité du complexe p50/p65, forme canonique du facteur de transcription NF-kB	p 85
Figure 17:	Histogramme des Inter-Quartile Range	p 90
Figure 18 :	Histogramme des probabilités issues d'analyses ANOVA univariées à deux facteurs réalisées sur chacun des 54 675 transcripts	p 92
Figure 19 :	Répartition des transcripts en fonction de la variance expliquée par les conditions expérimentales	p 96
Figure 20 :	Représentation des analyses ontologiques des transcripts d'intérêt	p 100
Figure 21 :	Description des 40 gènes identifiés dans notre analyse associés au processus de la mort cellulaire	p 101

Figure 22 :

Classification des 687 transcripts sélectionnés et la répartition des gènes associés à la mort cellulaire p 104

Figure 23 :

Mesure de l'activité des caspases 3/7 dans les deux lignées cellulaires p 108

Liste des sigles et abréviations

♀	femelle
♂	mâle
ABC	ATP Binding Cassette
ACP	Analyse en Composantes Principales
ACR	Analyse Canonique de Redondance
ACTB	β-actin
ADN	Acide DésoxyriboNucléique
ADNc	Acide DésoxyriboNucléique complémentaire
ALI	Acute Lung Injury
AMPc	Adénosine MonoPhosphate cyclique
AP-1	Activating protein 1
ATP	Adénosine TriPhosphate
ARNc	Acide RiboNucléique copie
ARNm	Acide RiboNucléique messenger
CFTR	Cystic Fibrosis Transmembrane Regulator
CHUM	Centre Hospitalier de l'Université de Montréal
Cl-	chlore
CLC	Chloride Channel
COPD	Chronic Obstructive Pulmonary Disease
Ct	Cycle threshold
CuFi-1	Cystic Fibrosis, University of Iowa
DAG	DiAcyclic Graph
D-MEM	Dulbeccos Modified Eagles Medium
DMNQ	2,3-dimethoxy-1,4-naphtoquinone
DMSO	diméthylsulfoxyde
DNase	Desoribouclease
DNTB	2-nitrobenzoïque

#dNTP	désoxy Nucléosine Tri-Phosphate
DO	Densité Optique
E	Expected
ENaC	Epithelial amiloride-sensitive sodium (Na) Channel
ERK	Extracellular signal-Regulated Kinase
EXP	Experimental
FBS	Fetal Bovin Serum
FCM	Fuzzy C-Means
FDR	False Discovery Rate
FK.	Fibrose Kystique
GABA	Gamma-AminoButyric Acid
GAPDH	glycéraldehyde-3-phosphate déhydrogénase
GDP	Guanosine DiPhosphate
GMP	Guanosine MonoPhosphate
GO	Gene Ontology
GSH	glutation réduit
GSSG	glutation oxidé
GTP	Guanosine TriPhosphate
GRO	Growth-Related Oncogene
H ₂ SO ₄	acide sulfurique
HDAC	Histone DéACétylase
HG	Human Genome
HIPK2	Homeodomain-Interacting Protein Kinase-2
HRP	Horse Radish Peroxydase
HSP70	Heat Shock Protein 70
HTBE	Human Tracheal Bronchial Epithelial
hTERT	human Telomerase Reverse Transcriptase
IDA	inferred from direct Assay
IEA	Inferred from Electronic Annotation

IFN	Interférons
Ig	Immunoglobuline
IL	Interleukine
IPI	Protein Interaction
IQR	Inter-Quartile Range
IRIC	Institut de Recherche en Immunologie et Cancérologie)
Isc	<i>short-circuit current</i>
IκB	Inhibitor of κB
IκK	IκB Kinase
JNK	Jun N-terminal Kinase
KO	Knock-Out
LIMMA	Linear Models for MicroArray data
LPS	LipoPolySaccharide
MAPK	Mitogen-Activated Protein kinases
MAS5	Affymetrix® Microarray Analysis Software version 5.0
MBL	Mannose-Binding Lectin 2
MM	MisMatch
MyD88	<i>Myeloid Differentiation factor 88</i>
NBD	Nucleotide Binding Domain
NFAT1	Nuclear Factor of Activated T 1
NFκB	Nuclear Factor kappa B
NOS-2	Nitric Oxide Synthase 2
NuLi-1	Normal Lung, University of Iowa
O	Observed
ORCC	Outwardly Rectifying Chlorure Channel
PAMPs	Pathogen-Associated Molecular Patterns
PBS	Phosphate Buffered Saline
PDCD4	Programmed Cell Death 4
pH	potentiel Hydrogène

PM	Perfect Match
qPCR	PCR quantitative
R	Regulatory
RE	Reticulum Endoplasmique
RMA	Robust Multichip Average
ROS	Reactive Oxygen Species
RQ	Relative Quantification
Scc	courant de court-circuit
SOM	Self-Organizing Map
Tg	Transgène
TGF β	transforming Growth Factor beta
TLR4	Toll-Like Receptor 4
TMD	TransMembranaire Domain
UTR	UnTranslated Region
β FGF	Fibroblast Growth Factor-beta
	Δ F508 délétion d'un résidu phénylalanine en position 508 de la chaîne protéique

Liste des unités de mesure

%	pourcent
°C	degré Celsius
μ	micro
cm ²	centimètre carré
g	gramme
kb	kilobase
kg	kilogramme
L	litre
m	mètre
M	molaire
m ²	mètre carré
mg	milligramme
min	minute
ml	millilitre
mm	millimètre
mm ²	millimètre carré
n	nano
p	pico
U	unité
V	volt

A ma fille, à ma femme, à Coco

Remerciements

Bordeaux 1999 : Je décide de partir au Québec pour étudier la bioinformatique:
Merci les services d'immigration

Montréal 2002 : J'intègre le programme du DESS en bioinformatique à l'UQAM : Merci Anne !

Hurley's 2003 : Je rencontre Sarah. Merci la vie.

CHUM-site de l'Hôtel-Dieu 2003 : Je rencontre Yves Berthiaume sur un banc et me fait part de ses projets transcriptomiques. Soleil radieux. Merci ma bonne étoile!

La Dôle (Suisse) 2004 : Discussion philosophe avec le berger sur la nécessité de faire de la science. Merci Fred !... Grand Merci, Ca fait près de 15 ans que nous partageons la science, la musique et la vie! À la vie, À la mort !

IRIC 2009 : Merci à Sébastien qui m'a apporté ce qu'il avait : une vision globale, binaire, rigoureuse, à la fois codée et codante, et qui a participé à mon développement scientifique.

17 avril 2009 : Sarah donne naissance à ma fille, Axelle. Merci la Vie, merci Sarah, merci Axelle.

En arrière plan : André, Émilie, Romane, Cécile, Guillaume (les 2), François, Nono et j'en oublie ! Merci à tous !

Je remercie également ma maman et mon papa, pour leur non-soutien. Je remercie spécialement ma future femme Sarah, qui me fait confiance et qui est présente dans mes élucubrations quotidiennes. Un spécial remerciement à Yves qui depuis longtemps me soutient, me fait confiance. Merci Yves pour votre bonhomie!

Introduction

A. Historique de la maladie

« **Salé!!!** ». Voici comment, pendant des siècles, fut décrite la fibrose kystique (FK). Au moyen-âge, le baiser salé du nouveau-né laissait prévoir son funeste sort. Il faudra attendre le 20^{ième} siècle, pour que la maladie soit décrite de façon médicale. Ainsi en 1936, un pédiatre Suisse, Guido Fanconi, attribue aux symptômes des enfants malades, le nom de «fibrose kystique pancréatique» et en 1938, le médecin Dorothy Hansine décrit pour la première fois cliniquement la maladie.

La figure 1 (page 2) retrace l'historique de la maladie au cours des 70 dernières années.

Les premières avancées réelles dans le domaine de la fibrose kystique furent le clonage du gène responsable et la mise en évidence de la mutation par un groupe de recherche de Toronto au Canada en 1989 (Rommens, Iannuzzi et al. 1989). Ceux-ci présumant de l'importance du CFTR (*Cystic Fibrosis Transmembrane Receptor*) dans le transport du chlore (Cl⁻) ont persévéré dans l'étude de ce gène jusqu'à son séquençage et ont ouvert la porte à d'autres équipes sur l'étude d'autres aspects du CFTR pendant les deux dernières décennies. Peu à peu, les connaissances et hypothèses se sont étayées, dévoilant ainsi les fonctions présumées de ce gène au sein de la cellule, son mécanisme de régulation et sa place dans la physiopathologie.

À ce jour, la fibrose kystique est la maladie autosomique récessive la plus courante dans la population caucasienne où une personne sur 25 est porteuse du gène défectueux. Au Canada, on estime qu'un nouveau-né sur 3600 est atteint par la maladie.

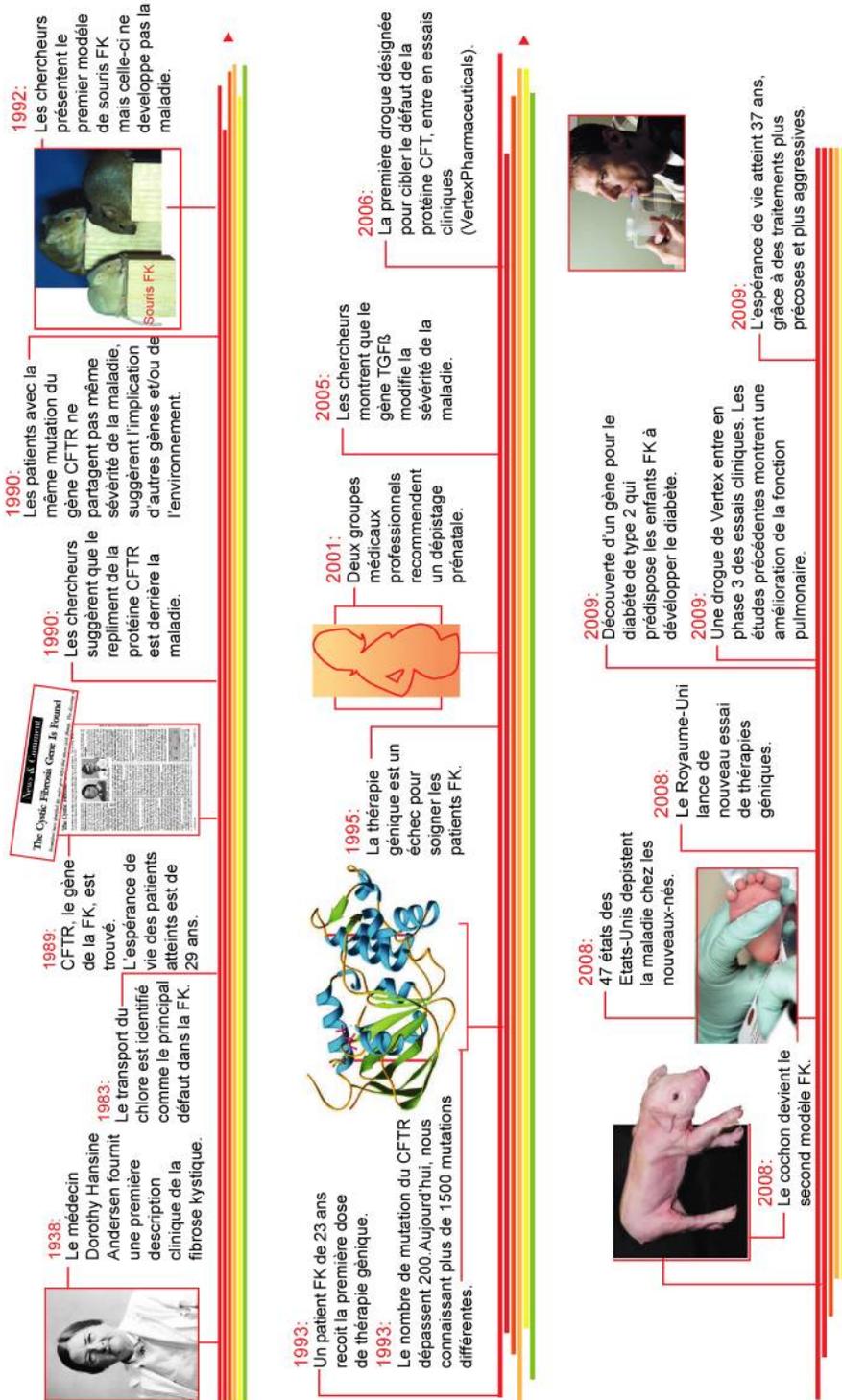


Figure 1 : Les dates clés de la fibrose kystique.

Extrait et adapté de "The Promise of a Cure: 20 Years and Counting. 19 juin 2009. Volume 324. SCIENCE. publié par AAAS.

Bien que 2009 soit le 20^{ième} anniversaire de la découverte du gène responsable de la pathologie et malgré les progrès dans le domaine biomédical, la physiopathologie de la fibrose kystique est encore mal comprise, car la recherche dans le domaine n'a dévoilé que la partie visible de l'affection et il reste encore beaucoup d'efforts pour comprendre pleinement la maladie et aboutir à une guérison totale.

B. La structure génique du CFTR

Le gène CFTR, localisé dans la région q31.2 du bras long du chromosome 7, est codé par une séquence génomique de 250 000 paires de bases (Riordan, Rommens et al. 1989). Il comprend 27 exons, correspondant à un transcript mature de 6,5 kilobases (kb). Chez les mammifères, le nombre moyen d'exon par gène est de 7 et la longueur moyenne des transcripts est de 2,2 kb (*B.Lewin, GENE IV, Editions De Boeck Université*). Ceci présume d'une haute complexité de la régulation spatio-temporelle de son transcript. En effet, plusieurs auteurs soulignent la présence d'éléments régulateurs dans ses parties introniques et leurs rôles restent encore indéterminés (Smith, Barth et al. 1996; Hill, Hong et al. 2006; Ott, Suszko et al. 2009). De la même façon, plusieurs formes épissées ont été observées, leurs fonctions n'ont pas encore été démontrées (Broackes-Carter, Williams et al. 2003). Toutefois, le niveau de synthèse du CFTR est relativement bas dans la plupart des types cellulaires, mais suffisant pour assurer une fonction active de la protéine.

C. La protéine CFTR

Le produit du gène CFTR est une protéine transmembranaire de 1480 acides aminés, dont la structure est présentée à la figure 2 (page 4). De forme symétrique, cette protéine possède cinq domaines : deux domaines de liaison aux nucléotides NBD (*Nucleotide Binding Domain*), deux domaines transmembranaires TMD (*TransMembranaire Domain*) et un domaine de

régulation R (*Regulatory*). Le domaine R possède de nombreux sites de phosphorylation contrôlant l'activité du canal (Chappe, Irvine et al. 2005).

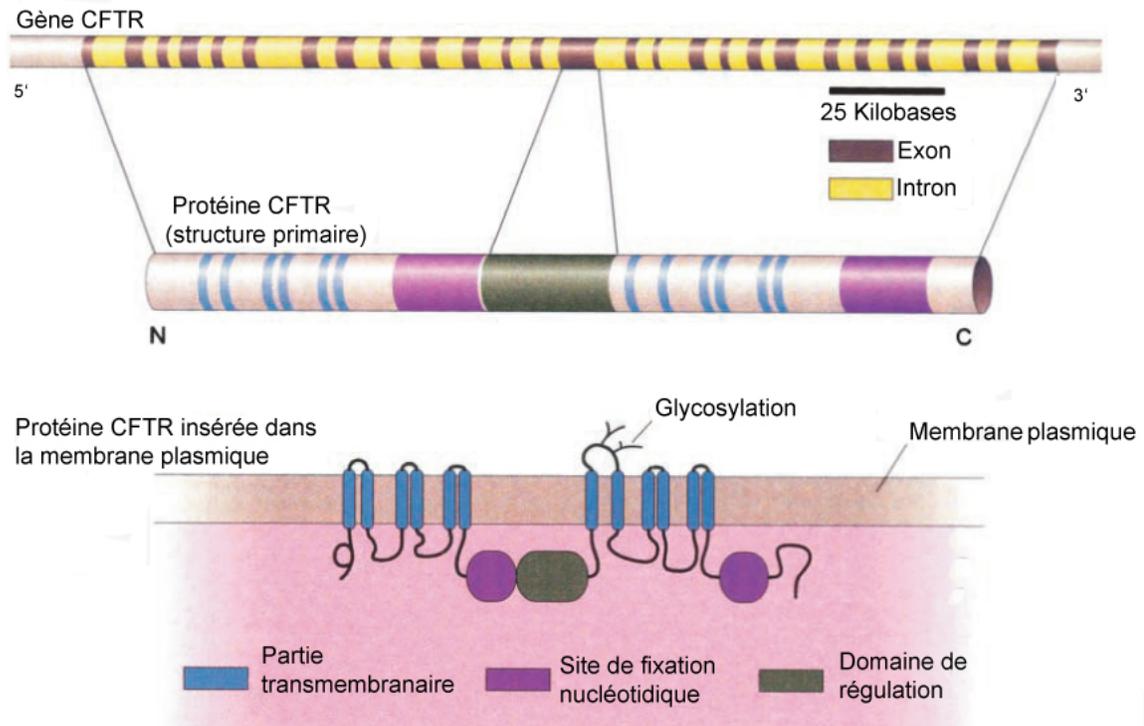


Figure 2: Description du gène CFTR, du transcript et de la protéine

Adapté de Ronald L. Gibson, Jane L. Burns and Bonnie W. Ramsey AJRCCM 2003

Comme toutes les protéines membranaires, le CFTR est synthétisé dans le réticulum endoplasmique (RE), sous la forme de précurseur, que des molécules chaperonnes replient pour lui donner une structure 3D fonctionnelle (forme A) (Gething and Sambrook 1992). Une première étape de maturation permet la glycosylation du précurseur. Toujours immature, cette protéine glycosylée (forme B) est exportée dans l'appareil de Golgi et subit une seconde étape des modifications post-traductionnelles, où les mannoses des chaînes glycosidiques sont remplacés par des glucides plus complexes (forme C).

70 % du CFTR produit restent séquestrés dans le RE en raison d'un défaut de maturation et sont alors rapidement dégradés par le protéosome (Lukacs, Mohamed et al. 1994; Ward, Omura et al. 1995; Ellgaard and Helenius 2001;

Gelman, Kannegaard et al. 2002). Seul 30 % de CFTR synthétisés passent avec succès le contrôle de la qualité de la maturation du RE (Aridor, Bannykh et al. 1995; Schekman and Orci 1996; Wang, Matteson et al. 2004) et sont adressés à la membrane plasmique pour y être intégrés comme transporteur ionique (Prince, Workman et al. 1994).

D. Fonction de la protéine CFTR

Les canaux Cl^- peuvent être divisés en trois grandes classes : (1) les récepteurs canaux, activés par des ligands comme le GABA (*Gamma-AminoButyric Acid*), (2) les canaux CLC (*Chloride Channel*), et (3) le canal CFTR. Ceux-ci sont aussi différents dans leur structure protéique que dans leur fonction et sont connus pour jouer un rôle important dans de nombreux processus physiologiques. Au-delà du transport de chlore, ils sont impliqués dans la régulation du potentiel de membrane, du volume et du pH intracellulaire, dans le transport transépithélial, ainsi que dans la dégradation des protéines et possiblement dans la croissance cellulaire (Suzuki, Morita et al. 2006). De plus, les impacts physiologiques importants observés dans des maladies où sont impliqués les canaux Cl^- , comme la fibrose kystique et la myotonie de Becker, montrent l'importance fonctionnelle de ces canaux pour une bonne homéostasie cellulaire.

1. CFTR, transporteur de chlore, d'anion et de molécules organiques

Les caractéristiques du canal CFTR ont été largement étudiées et ont permis de détailler sa topologie (Rosenberg, Callaghan et al. 2005), sa composition en domaine, son système d'activation à AMPc (adénosine monophosphate cyclique) (Cheng, Rich et al. 1991), ses propriétés biophysiques (Riordan, Rommens et al. 1989; Lee, Wigley et al. 1999; Kunzelmann, Schreiber et al.

2001), ainsi que sa perméabilité (Tabcharani, Rommens et al. 1993; Linsdell, Tabcharani et al. 1997).

La principale fonction connue de la protéine CFTR est le transport du chlore (Anderson, Rich et al. 1991) au niveau de la membrane apicale des cellules épithéliales, des glandes sous-muqueuses des voies aériennes, du pancréas, des intestins, du rein et du foie.

Il a été également démontré que le canal CFTR est perméable à d'autres molécules anioniques selon la perméabilité croissante suivante : $\text{NO}_3^- > \text{Cl}^- > \text{HCO}_3^- > \text{HCOO}^- > \text{CH}_3\text{COO}^-$ (Tabcharani, Rommens et al. 1993; Linsdell, Tabcharani et al. 1997). D'autre part, son implication dans le transport du glutathion réduit a été démontrée (Linsdell and Hanrahan 1998), et ce peptide est bien connu pour sa participation dans la protection cellulaire contre les dommages causés par les espèces réactives de l'oxygène (Roum, Buhl et al. 1993). Plusieurs études ont montré l'influence de la mutation du CFTR sur le transport transmembranaire et transépithélial du glutathion, ainsi qu'une diminution de la concentration en GSH (glutathion) du côté apical de cellules épithéliales pulmonaires (Gao, Kim et al. 1999; Kogan, Ramjeesingh et al. 2003). Ces éléments suggèrent un rôle du CFTR dans le flux du glutathion et par conséquent celui-ci pourrait avoir une place importante dans la gestion du stress oxydatif cellulaire.

2. CFTR, régulateur de canaux ioniques

Les dysfonctionnements épithéliaux observés dans la fibrose kystique ne pouvant pas être expliqués uniquement par la perte d'activité du canal CFTR, les chercheurs ont émis l'hypothèse de l'existence d'une fonction régulatrice de ce canal (Short, Trotter et al. 1998).

Il a été démontré que le canal chlore ORCC (*Outwardly Rectifying Chlorure Channel*) ne fonctionnait qu'en présence du canal CFTR et que ce dernier

assure une activité maximale au canal ORCC (Gabriel, Clarke et al. 1993; Schwiebert, Flotte et al. 1994; Schwiebert, Morales et al. 1998). Le CFTR intervient également dans la régulation d'activité du canal ENaC (*Epithelial amiloride-sensitive sodium Channel*) en exerçant un contrôle négatif sur lui (Boucher, Chinet et al. 1991; Mall, Bleich et al. 1998; Kunzelmann, Schreiber et al. 2000; Konstas, Koch et al. 2003). Néanmoins, l'interaction entre les deux canaux est encore mal comprise et nécessite plus d'investigations.

D'autres canaux semblent également sous l'influence du CFTR, comme les canaux potassiques (Loussouarn, Demolombe et al. 1996) et les aquaporines (Schreiber, Greger et al. 1997).

3. D'autres fonctions connues du CFTR

Il a été démontré que la protéine CFTR joue un rôle dans le mécanisme d'endocytose-exocytose (Bradbury, Jilling et al. 1992), influençant le transport vésiculaire et la fusion des endosomes entre eux (Bowers, Emans et al. 1996). Le CFTR occupe également une place importante dans la réponse immunitaire et le processus d'inflammation/infection. Son implication dans l'activation de la réponse immunitaire suite à l'infection à *P.aeruginosa* (*Pseudomonas aeruginosa*) a été démontrée par plusieurs équipes (Schroeder, Lee et al. 2002; Cannon, Kowalski et al. 2003; Kowalski and Pier 2004).

Par ailleurs, il a été démontré que la déficience d'expression de CFTR a un impact sur l'homéostasie du calcium intracellulaire et que le canal Cl⁻ serait un régulateur de la concentration en calcium intracellulaire (Walsh, Harvey et al. 2000).

4. CFTR, alias ABCC7, membre de la famille des ABC

La famille des ABC (*ATP Binding Cassette*) est une famille de 48 transporteurs membranaires d'une grande variété de substrats (acides aminés, lipides, ions,

sucres, drogues). Cette famille se caractérise par la présence de nombreux domaines NBD au sein de leurs protéines (Holland and Holland 2005). Bien que peu d'entre elles aient été analysées, elles sont considérées comme des transporteurs de molécules contre un gradient de concentration stimulées par l'ATP (adénosine triphosphate).

Plusieurs transporteurs ABC sont impliqués dans des pathologies humaines : la fibrose kystique, l'adrénoleucodystrophie, la maladie de Tangier, le syndrome de Dubin-Johnson. Cela suggère que cette famille de transporteurs, dont est membre le CFTR, occupe une place centrale dans la physiologie cellulaire normale.

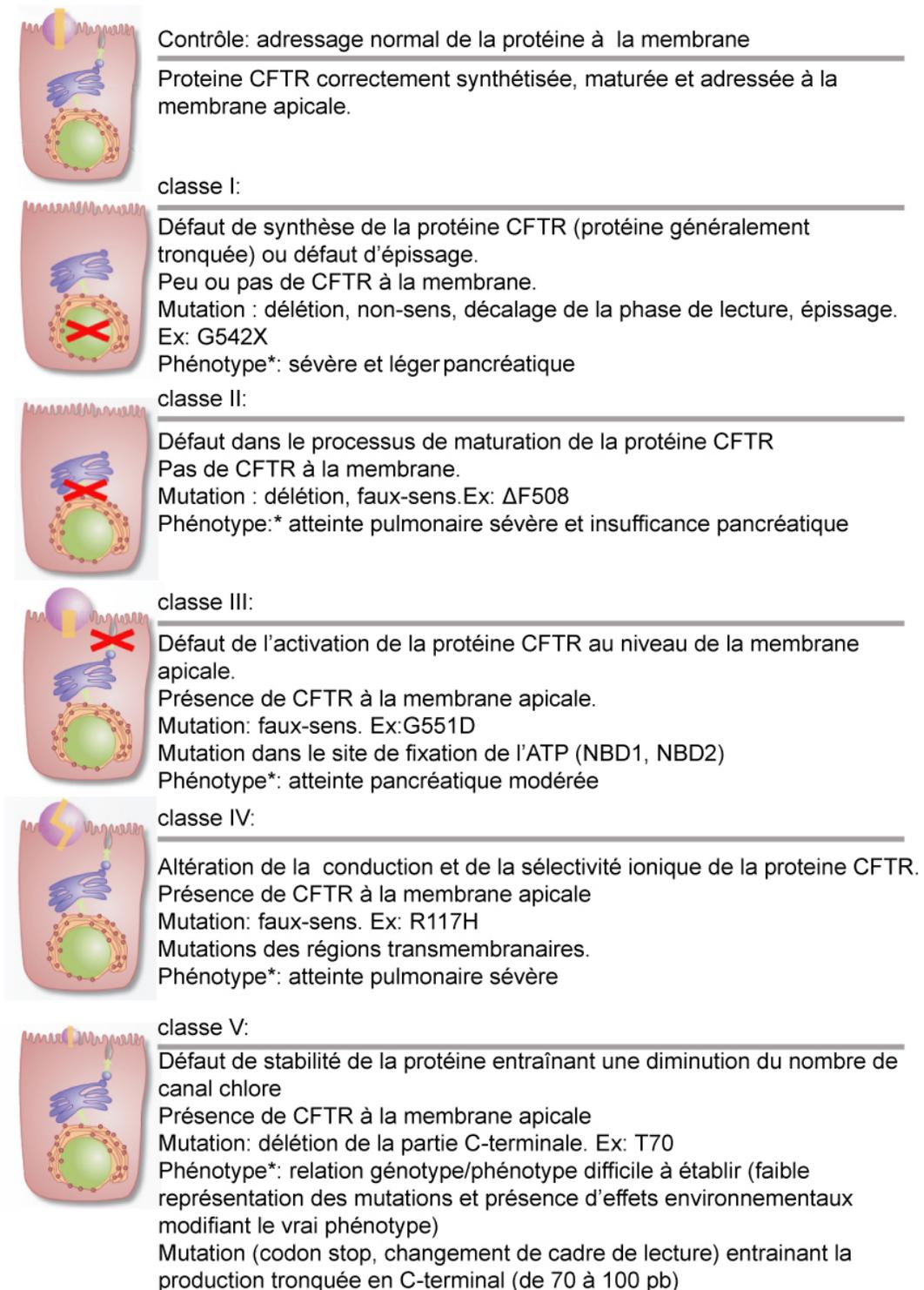
E. La fibrose kystique

1. Facteurs expliquant la physiopathologie

a. Les mutations du gène CFTR

Le défaut majeur qui engendre la fibrose kystique est un facteur génétique. Actuellement, plus de 1600 mutations du gène CFTR sont référencées et n'ont pas la même fréquence d'apparition. Environ 50 d'entre elles sont communément rencontrées chez les caucasiens (Estivill, Bancells et al. 1997), tandis que d'autres sont extrêmement rares. De plus, ces mutations sont très dépendantes de groupes ethniques. Par exemple, la mutation non-sens W1282X est surtout présente chez les Juifs Ashkénazes.

Anciennement regroupées en six classes de mutations, elles le sont désormais en cinq groupes définies par l'impact de l'anomalie génétique sur la protéine (Welsh and Smith 1993; Haardt, Benharouga et al. 1999). La figure 3 (page 9) résume les caractéristiques des différentes classes de mutations.



*à chaque classe est associé un phénotype, lequel peut-être variable selon les patients FK et dépendant de l'environnement.

Figure 3 : Les différentes classes de mutations du CFTR et leurs effets sur son adressage.

b. La mutation $\Delta F508$

La délétion d'un codon provoquant la perte d'une phénylalanine dans la protéine CFTR en position 508 est l'anomalie génétique la plus fréquente dans la population caucasienne (70 % des cas) et suite à une vague d'immigration venant d'Europe, elle est devenue prépondérante en Amérique du Nord et en Australie (*The Cystic Fibrosis Genetic Analysis Consortium*. 1994). Cette mutation, communément désignée $\Delta F508$, empêche une maturation efficace du précurseur en raison d'un défaut de repliement de la protéine CFTR (mutation de classe II)

Parce que cette mutation se situe dans le domaine NBD1, les chercheurs ont longtemps émis l'hypothèse que le repliement local inadéquat du domaine NBD1 entraînait une mauvaise structure 3D de la protéine CFTR (Denning, Anderson et al. 1992; Thomas, Ko et al. 1992). Des travaux de 2005, utilisant la technique du dichroïsme circulaire, ont montré que la mutation $\Delta F508$ affecte très peu la topologie de la région NBD1 (Lewis, Zhao et al. 2005). De plus, il a été montré que le mauvais repliement se situe en réalité dans la partie C terminale du CFTR comprenant la partie NBD2 et également que la chaîne latérale de la phénylalanine en position 508 est un élément central dans les interactions intra-domaines et en particulier entre NBD1 et les domaines transmembranaires (Du, Sharma et al. 2005).

En définitive, ce mauvais repliement affecte le transport de la protéine au sein de la cellule puisqu'elle est séquestrée dans le RE. Toutefois, très peu de CFTR muté parvient à atteindre la membrane plasmique.

c. D'autres facteurs génétiques impliqués

L'existence d'autant de génotypes explique en partie les différents phénotypes observés chez les patients, mais la relation de cause à effet reste difficile à

établir. Fondé sur la mutation $\Delta F508$ et son impact sur la fonction pulmonaire, plusieurs équipes ont étudié la corrélation entre les deux sans aboutir à un consensus (Liechti-Gallati, Bonsall et al. 1992; Gan, Geus et al. 1995; Hubert, Bienvenu et al. 1996). Toutefois, les travaux d'une équipe québécoise ont montré une forte corrélation entre la concentration de chlore dans la sueur de deux groupes de malades et leurs mutations de classe IV et V (De Braekeleer, Allard et al. 1998).

D'une manière plus générale, les caractéristiques du phénotype FK pour certains organes comme le pancréas, sont déterminées essentiellement par la mutation. Celles-ci sont peu ou pas influencées par d'autres facteurs génétiques et/ou environnementaux, tandis que le phénotype pulmonaire observé indique l'implication d'autres facteurs (Cutting 2005).

En 1996, Rozmahel a montré l'importance de l'environnement génétique dans la physiopathologie de la maladie chez des souris $CFTR^{-/-}$ croisées avec des souris aux fonds génétiques variés (Rozmahel, Wilschanski et al. 1996). Ainsi, d'autres gènes que le $CFTR$ participeraient à la physiopathologie de la fibrose kystique. Plusieurs études ont mené à l'identification de deux d'entre eux. Le gène de MBL (*mannose-binding lectin 2*) qui code pour une protéine de la réponse immune innée dont la déficience serait significativement associée à une infection précoce chez les patients FK et le gène $TGFB1$ (*transforming growth factor beta 1*) qui est également associé au phénotype pulmonaire FK.

2. La physiopathologie de la maladie

Comme décrit plus tôt, la fibrose kystique est une maladie aux causes génétiques multifactorielles et ceci explique un tableau clinique complexe et propre à chaque patient. Ainsi, chez des patients partageant la même mutation, la physiopathologie observée peut être très variable.

Les organes touchés sont principalement les poumons, le pancréas, le foie, les intestins et le système reproducteur. L'atteinte pulmonaire est la première cause de mortalité et conditionne la qualité de vie du patient. L'atteinte pancréatique, quant à elle, affecte environ 80 % des malades et mène également à la destruction de l'organe. À des degrés variables de sévérité, d'autres dommages moins fréquents apparaissent au niveau du foie, les intestins, les articulations et viennent compléter le tableau clinique. De plus, la stérilité des hommes est observée en raison de l'absence bilatérale congénitale de vas deferens provoquant l'azoospermie. La stérilité est rarement observée chez la femme.

Au-delà des altérations organiques inhérentes à la mutation du gène CFTR, cette dernière influence considérablement les processus biomoléculaires, touchant l'intégrité fonctionnelle de la cellule. Ces mécanismes perturbés participent de façon importante à la physiopathologie de la maladie et sont au centre des recherches entreprises, car de leurs compréhensions découleront des voies thérapeutiques efficaces. Des mécanismes moléculaires tels que la réponse immune innée, l'inflammation ou le stress oxydatif, sont considérablement perturbés chez le patient FK, et sont le sujet de plusieurs études. En outre, de nombreux gènes ont été identifiés comme ayant un rôle important dans la maladie. Bien que plusieurs constatations aient été faites sur leurs dérégulations, il est encore difficile de comprendre l'implication totale de ces mécanismes dans le développement de la physiopathologie.

a. L'atteinte pulmonaire

Les poumons sont les organes emblématiques de la maladie, car leurs dysfonctionnements sont fatales et conditionnent entièrement l'évolution de la maladie. Chaque inspiration remplit nos poumons de 0.8 litre d'air par minute contenant plus de 2000 éléments extérieurs potentiellement dangereux pour

notre organisme. Face à ces agressions, les poumons représentent une barrière de protection efficace. Les lignes de défense de la muqueuse pulmonaire sont composées de différents mécanismes comme la clairance muco-ciliaire, la régulation ionique et hydrique, ainsi que la sécrétion de nombreuses molécules de reconnaissance et de défense. Celles-ci permettent la mise en place d'une réponse immunitaire innée appropriée. Puisque la déficience du CFTR altère le transport ionique, la régulation du volume périciliaire est affectée et modifie de ce fait la clairance d'un mucus rendu plus visqueux. Ce mécanisme de défense étant perturbé, les poumons deviennent dès lors plus sensibles aux agressions et sujet aux infections chroniques à différents germes pathogènes comme *Pseudomonas.aeruginosa*, *Haemophilus influenzae* et *Staphylococcus aureus* (Lyczak, Cannon et al. 2002).

Une étude récente (Divangahi, Balghi et al. 2009) a montré que la déficience du CFTR aurait un rôle intrinsèque dans la faiblesse des muscles respiratoires alors que pendant longtemps les chercheurs l'ont associé au manque d'exercice physique et à la malnutrition (Pinet, Cassart et al. 2003). Ceci souligne que de nombreux aspects du CFTR sont encore à découvrir.

Le défaut physique de la première ligne de défense ne serait pas la seule cause de ces infections répétées (Chmiel and Davis 2003). En effet, il apparaît que les défenses moléculaires immunes, représentant la seconde ligne de défense pulmonaires, sont considérablement perturbées et ne répondent plus correctement aux agressions répétées que subit l'organe. Les composantes moléculaires anti-bactériennes, anti-protéolytiques que l'on retrouve normalement dans le liquide de surface pulmonaire semblent être diminuées de façon significative chez les patients FK. Par exemple, NOS-2 (*nitric oxide synthase 2*), un composant considéré comme important dans la défense innée des cellules épithéliales pulmonaires (Zheng, De et al. 2003), serait moins présent dans l'épithélium pulmonaire FK par rapport au patient sain (Kelley and

Drumm 1998). D'autres acteurs comme les protéines du surfactant, dont les rôles anti-inflammatoires et anti-oxydants sont moins connus que leur rôle de réduction de la tension de la surface pulmonaire (Meyer and Zimmerman 2002), sont également en concentrations pulmonaires moins importantes chez les patients FK que chez les patients contrôles.

La déficience de la réponse immune innée se traduit également au niveau cellulaire par un comportement inadéquat des acteurs de cette réponse. Plusieurs études ont montré le comportement inapproprié de quelques types cellulaires inflammatoires chez les patients FK. L'équipe de Vandivier a souligné le rôle déficient des macrophages dans la phagocytose des cellules apoptotiques, d'autres ont mis en évidence l'accumulation de neutrophiles dans les voies aériennes, des différences dans leur fonction phagocytaire, leur dégranulation excessive, ainsi que dans l'apoptose de ces cellules (Koller, Urbanek et al. 1995; Watt, Courtney et al. 2005; Alexis, Muhlebach et al. 2006).

Les cellules épithéliales ont leur rôle à jouer dans la défense immunitaire et interviennent de façon majeure dans la reconnaissance des pathogènes. Effectivement, ces cellules présentent à leurs surfaces des récepteurs *toll-like* capables de reconnaître les PAMPs (*Pathogen-Associated Molecular Patterns*) à la surface des pathogènes. Ainsi stimulés, ces récepteurs activent les voies de signalisation NF- κ B et AP-1 desquelles résultent, entre autres, la synthèse de facteurs pro-inflammatoires.

Plusieurs équipes se sont intéressées à l'expression de ces récepteurs dans des lignées de cellules épithéliales pulmonaires (Muir, Soong et al. 2004; Greene and McElvaney 2005; Guillot, Le Goffic et al. 2005) sans démontrer de défaut majeur de leur expression chez des cellules FK, mais soulignant l'importance que pourrait avoir le TLR4 (*toll-like receptor 4*) dans la reconnaissance de *P. aeruginosa*, le pathogène le plus observé chez les

patients FK (West and Brunskill 2002) et dont l'infection est associée au déclin de la fonction respiratoire.

b. L'hyperinflammation des cellules FK

Pendant longtemps, il a été logique de penser que l'origine de l'hyperinflammation observée chez les patients FK était d'ordre infectieux. Cependant, certaines constatations laissent penser que l'inflammation pourrait précéder l'infection (Muhlebach, Stewart et al. 1999; Tirouvanziam, de Bentzmann et al. 2000). En effet, la prédominance de neutrophiles (et son inflammation associée) dans les poumons est observée très tôt chez le nouveau-né. Bien plus tard, la susceptibilité aux infections chroniques apparaît. Toutefois, l'hypothèse que l'inflammation précède l'infection reste très controversée dans la communauté scientifique.

L'hyperinflammation chez le patient FK se caractérise par une accumulation constante de neutrophile dans les poumons, à laquelle est associée une importante activité protéolytique et à des taux élevés de chemokines, tels que l'interleukine 8 (IL8) (Sagel, Chmiel et al. 2007).

En 2000, l'équipe de Tirouvanziam suggère que la sécrétion importante d'IL8 pourrait être indépendante d'une infection et serait intrinsèque à la mutation FK. De plus, ils ont montré que l'accumulation de neutrophiles dans les poumons, via leurs activités protéolytiques mènent à la destruction des tissus (Tirouvanziam, de Bentzmann et al. 2000).

A ce jour, de nombreuses études sur le sujet s'emploient à dévoiler les dysfonctionnements de la réponse inflammatoire et à découvrir la place du CFTR dans cette régulation. Désormais, plusieurs faits sont acquis car ils sont soutenus par plusieurs équipes :

- (1) **L'augmentation d'activité des voies inflammatoires des cellules FK par rapport au contrôle.** La voie NF- κ B est la voie de signalisation

découverte comme étant activée chez les cellules FK (DiMango, Ratner et al. 1998; Tabary, Escotte et al. 1999) et d'autre part, nous notons une déficience dans les molécules régulatrices de cette voie (IL10, NOS) (Chmiel, Konstan et al. 1999). Une autre étude utilisant le CFTR_{inh-172} sur des cellules épithéliales pulmonaires a mis en évidence le même phénotype inflammatoire observé chez des cellules FK (Perez, Issler et al. 2007).

D'autres voies de signalisation sont décrites comme activées dans les cellules pulmonaires FK comme la cascade des MAPK (Mitogen-Activated Protein kinases) et la voie AP-1, néanmoins leur états semblent être également dépendant du stress oxydatif de la cellule FK (Sadikot, Christman et al. 2004; Raia, Maiuri et al. 2005; Verhaeghe, Remouchamps et al. 2007; Boncoeur, Criq et al. 2008).

(2) Résultant de l'augmentation d'activité de NF- κ B, l'augmentation des niveaux d'expression de molécules inflammatoires est constatée dans plusieurs types cellulaires que ce soit en présence et en absence de stimulation bactérienne. Les modulations des molécules pro-inflammatoires IL8, IL6, IL1 β , sont largement décrits dans la littérature dans les cellules épithéliales FK (Sagel, Chmiel et al. 2007). D'autres molécules inflammatoires sont sur-exprimées dans les cellules FK, mais leurs rôles dans l'inflammation et la maladie sont encore mal définis.

(3) L'augmentation de l'expression de molécules d'adhésion à la surface des cellules épithéliales pulmonaires a été également montrée (Tabary, Corvol et al. 2006). Ces molécules de surface contribueraient au processus inflammatoire en favorisant l'adhérence et l'accumulation de neutrophiles dans les voies aériennes.

c. Un facteur aggravant de la physiopathologie : le stress oxydatif

La diminution des défenses anti-oxydantes est un signe caractéristique de la maladie et a une conséquence importante sur le stress oxydatif. Le glutathion est le composant le plus important du mécanisme de défense anti-oxydante dans les poumons et a une fonction détoxifiante pour un grand nombre de molécules toxiques. Étant donné l'exposition des poumons à l'environnement extérieur, il n'est pas étonnant que la concentration de glutathion dans cet organe soit plus importante que dans n'importe quel autre (Townsend, Tew et al. 2003).

La perméabilité du CFTR au glutathion a déjà été démontré (Gao, Kim et al. 1999), ainsi que la diminution de sa concentration dans les poumons de patients FK comparée à des patients contrôles (Roum, Buhl et al. 1993). De plus, sa déficience serait associée avec l'activation de NF- κ B (Townsend, Tew et al. 2003).

La physiopathologie de la fibrose kystique est très complexe et par conséquent difficile à cerner, autant au niveau des signes cliniques qu'au niveau moléculaire. Pour ces raisons, il est impossible de dresser un tableau clinique précis de la maladie.

Les défauts fonctionnels de la réponse immune innée, de l'hyperinflammation et du stress oxydatif sont des éléments invariants dans la physiopathologie et il est nécessaire de bien appréhender ces mécanismes dans leur ensemble pour développer des thérapies futures.

d. Le rôle de l'apoptose dans l'épithélium pulmonaire

Face aux agressions diverses, l'épithélium bronchique est une barrière physique efficace. Son intégrité structurelle est conservée par l'aptitude que possède l'épithélium à se reconstituer en éliminant des cellules épithéliales

indésirables ou endommagées, laissant ainsi la place à des cellules normales. L'efferocytose est le processus de clairance des cellules présentant des caractéristiques apoptotiques. Celles-ci sont reconnues par les phagocytes (ou d'autres types cellulaires). Ce processus maintient ainsi l'homéostasie cellulaire et sa déficience entraîne des pathologies pulmonaires chroniques et inflammatoires.

Le processus de l'apoptose est un mécanisme structuré, caractérisé par la condensation nucléaire, la fragmentation de l'ADN, le bourgeonnement de la membrane, l'activation des caspases et l'expression membranaire de marqueurs apoptotiques telles que la phosphatidylserine (Fadok, Voelker et al. 1992; Fadok, de Cathelineau et al. 2001), CD14 (Schagat, Wofford et al. 2001) et CD36 (Savill, Hogg et al. 1992). Contrairement à la nécrose, le mécanisme apoptotique est anti-inflammatoire, car inhibe la production de molécules pro-inflammatoires et stimule l'environnement à produire des molécules anti-inflammatoires.

L'apoptose peut-être initiée par deux voies distinctes. La voie extrinsèque implique des récepteurs membranaires et est dépendante de la caspase 8. La voie intrinsèque est induite par la mitochondrie et est dépendante de la caspase 9. Ces deux voies, initiatrices de l'apoptose, conduisent à la phase d'exécution de l'apoptose par l'activation des caspases exécutrices (caspases 3, 6, 7) qui initie finalement la phase de dégradation protéolytique et nucléaire qui est irréversible et se traduit par des multiples changements cellulaires.

Le défaut de l'apoptose dans des cellules FK a été suggéré par plusieurs équipes sans que la cause ou le dysfonctionnement soient clairement établies. Gottlieb montra que le rôle du pH intracellulaire inhibe le processus de mort cellulaire programmée dans des cellules épithéliales $\Delta F508$ (Gottlieb et Dosanjh 1996). Maiuri en 1997 rapporte une augmentation de la fragmentation de l'ADN, et de l'expression de Fas et FasL, médiateurs du signal apoptotique

dans des cellules de l'épithélium bronchique de patients FK (Maiuri, Raia et al. 1997). Durieu confirme le haut niveau d'expression de FasL dans des cellules de l'épithélium bronchique humain de patient FK, (Durieu, Amsellem et al. 1999). Jungas est le premier à s'intéresser au rôle du stress oxydatif sur le processus de l'apoptose dans le contexte de la fibrose kystique et met en évidence une résistance à l'apoptose induite par l'hyperoxie. (Jungas, Motta et al. 2002). Plus récemment, le Dre Boncoeur montre la dépendance de l'activité des caspases 3/7 au stress oxydatif dans des cellules FK. (Boncoeur, Tabary et al. 2006; Boncoeur, Roque et al. 2008). Finalement, tout récemment, les études menées par Vandivier permettent de constater une augmentation de cellules apoptotiques dans les crachats et les tissus de patients fibro-kystiques (Vandivier, Fadok et al. 2002). Il démontre également l'inhibition de l'efferocytose. L'ensemble des travaux recents souligne l'importance du stress oxydatif dans le devenir apoptotique de cellules FK.

F. L'approche transcriptomique dans l'étude de la maladie

Les 20 ans de recherche sur la maladie se sont naturellement focalisés sur l'étude du CFTR et sur les conséquences de son dysfonctionnement. Néanmoins, la fibrose kystique est une maladie complexe et sa sévérité et son phénotype ne peuvent être expliqués uniquement par la déficience du gène CFTR. Ainsi, d'autres acteurs biologiques, comme les gènes modificateurs participent à la physiopathologie et il est encore laborieux de déterminer pleinement l'implication du CFTR et le rôle de l'environnement génique dans le développement de la maladie.

L'approche transcriptomique dans l'analyse de la fibrose kystique permet une étude plus globale des perturbations biologiques au sein d'une cellule FK et juge de l'ensemble des modulations présentes dans la cellule. L'identification des transcrits modulés et des voies métaboliques perturbées conduit à des

hypothèses plus fines, améliorant ainsi notre compréhension de la physiopathologie. Comprendre la place qu'occupe le CFTR dans l'homéostasie cellulaire et considérer d'autres acteurs moléculaires d'intérêt, permettront d'envisager des thérapies nouvelles, plus efficaces, lesquelles pourraient cibler de manière plus spécifique des molécules clés voire agir directement sur les effets primaires de la déficience du CFTR.

1. La technologie des puces à ADN

La technologie, relativement récente, des puces à ADN, permet l'analyse globale des mécanismes moléculaires intracellulaires. Effectivement, cette approche évalue en une seule expérience, au sein d'un échantillon biologique, le profil d'expression de l'ensemble des gènes. L'étude à large échelle de l'expression des gènes reflète l'activité transcriptionnelle de la cellule à un moment t dans des conditions expérimentales données. De cette activité, les modulations de certains processus biologiques nous apparaissent plus clairement et des conclusions biologiques d'ordre plus général peuvent s'appliquer.

L'équipe de Brown est la première en 1995, à démontrer que l'analyse de l'ensemble du transcriptome permet d'obtenir des signatures moléculaires caractéristiques d'un état cellulaire (Schena, Shalon et al. 1995). Pour cette raison, la technologie des puces à ADN est largement utilisée pour un usage physiopathologique dans le but d'identifier des gènes modulés dans un contexte pathologique, pour un usage diagnostique dans l'analyse de la modulation de marqueurs connus d'une pathologie, ou encore pour un usage pronostic pour la prédiction de l'évolution de la maladie.

2. L'approche transcriptomique dans d'autres pathologies

C'est en cancérologie que la technologie des puces à ADN a permis des avancées remarquables. La classification moléculaire de tumeurs notamment

pour les lymphomes et leucémies, a mis en avant des sous-types particuliers de tumeurs indétectables par les méthodes cliniques ou histo-pathologiques classiques (Golub, Slonim et al. 1999; Alizadeh, Eisen et al. 2000; Bertucci, Houlgatte et al. 2000; Klein, Tu et al. 2001; Rosenwald, Alizadeh et al. 2001; Hedenfalk, Ringner et al. 2002). De plus, ces classifications de profil d'expression sont devenues une aide précieuse pour le pronostic évolutif ainsi que la réponse aux thérapies. D'une façon similaire, l'usage des puces à ADN s'avère indispensable dans l'investigation des pathologies inflammatoires, d'une part parce que leur étiologie est souvent multifactorielle (génétiques et environnementaux) et complexe, et d'autre part parce que leurs physiopathologie est mal comprise. Ainsi, l'analyse transcriptomique de maladies inflammatoires permet de déterminer la part de la génétique et de l'environnement dans le développement de la maladie et de plus autorise des associations hypothétiques entre la physiopathologie et les modulations moléculaires. Pour ces raisons, la technologie des puces à ADN est adéquate pour nous aider à détailler et comprendre les mécanismes moléculaires inflammatoires impliqués.

Le premier usage de cette technologie dans l'étude de l'inflammation date de 2000 (Dieckgraefe, Stenson et al. 2000) et a aidé à mettre en évidence l'implication d'un groupe de gènes exprimés par la muqueuse digestive dans la physiopathologie des colites ulcéraives. De la même façon, deux études sur le lupus érythémateux disséminé montre ainsi le rôle de l'interféron dans la physiopathologie de la maladie (Baechler, Batliwalla et al. 2003; Bennett, Palucka et al. 2003).

3. Les études transcriptomiques dans la fibrose kystique

Plusieurs équipes de recherche ont utilisé l'approche transcriptomique dans l'étude de la fibrose kystique. Celles-ci sont présentées dans le tableau I (page 22).

Référence	Modèle FK (tissu)	Puce à ADN	Conditions
Modèle murin			
Xu, Clark et al. 2003	souris transgénique CFTR KO (tissu pulmonaire entier)	U74AV2	4 CF / 4 non CF
Kaur, Norkina et al. 2004	souris transgénique CFTR KO (tissu pancréatique entier)	U74AV2	3 CF / 3 non CF
Haston, Cory et al. 2006	souris transgénique CFTR KO (tissu pulmonaire gauche entier)	MOE430 2.0	3 CF / 3 non CF
Modèle humain			
Zabner, Scheetz et al. 2005	cellules primaires humaines trachéales et bronchiques	HGU133 A	10 CF / 10 non CF
Wright, Merlo et al. 2006	cellules primaires humaines nasales	HGU133 (A et B)	12 CF / 12 non CF
Gambardella, Biancolella et al. 2006	2 lignées cellulaires humaines épithéliales bronchiques: CuFi-1 ($\Delta F508/\Delta F508$) et CuFi-3 ($\Delta F508/R553X$)	puce maison (144 gènes)	3 CF / 3 non CF
Verhaeghe, Remouchamps et al. 2007	lignée humaine trachéale fœtale CFT-2 ($\Delta F508/\Delta F508$)	HGU133 Plus 2.0	3 CF / 3 non CF
Adib-Conquy, Pedron et al. 2008	neutrophile	puce maison (1050 gènes)	5 CF / 3 non CF

Tableau I: Les différentes expériences de puces à ADN dans le domaine d'étude de la fibrose kystique

La première étude entreprise fut celle du Dr Jeffrey A. Whitsett en 2003 (Xu, Mayhugh et al. 2003). Ce dernier présente une analyse transcriptomique de modèle de souris transgénique présentant les génotypes $CFTR^{\Delta F508}$, $CFTR^{tg}$ (modèle transgénique sur-exprimant la protéine CFTR), et $cftr^{-/-}$. Leurs conclusions présentent une modulation de transcripts liée à la transcription, à l'inflammation, au transport intracellulaire, au signal de transduction et au transport ionique.

En 2005, Le Dr Joseph Zabner (Zabner, Scheetz et al. 2005) présente la première étude transcriptomique sur des cellules primaires FK humaines et

identifie 22 gènes modulés entre la condition contrôle et la condition mutante $\Delta F508$, lesquelles sont pauvrement annotés et ne permet pas d'interprétation biologique.

Jerry M. Wright réalise une expérience similaire à celle du Dr Zabner, en 2006, en considérant dans son design expérimental le degré de sévérité de la maladie (modéré et sévère) (Wright, Merlo et al. 2006). Ces conclusions suggèrent que les différences dans les phénotypes sont associées avec des différences dans l'expression des gènes notamment dans les fonctions associées aux défenses pulmonaires, de l'ubiquitination des protéines, de l'activité oxydoréductase mitochondriale.

Dans la continuité de ses travaux de 2003, le Dr Jeffrey A. Whitsett, utilisant le même modèle (Xu, Liu et al. 2006), cible son analyse sur l'aspect multifactoriel de la maladie et présente les relations de la balance d'oxydoréduction, de l'inflammation, et de l'apoptose. De plus, il complète ses travaux avec une analyse plus complexe des promoteurs des gènes modulés.

La même année, les travaux de Guilbault sont présentés (Guilbault, Novak et al. 2006). Ils utilisent également un modèle murin KO-CFTR. Dans cette étude, elle caractérise la modulation de nombreux gènes pro-inflammatoires, ou appartenant aux signaux de transduction et au remodelage de la matrice cellulaire. Dans un modèle analogue, Dr Haston souligne la modulation de gènes associés à la réponse inflammatoire et à l'apoptose (Haston, Cory et al. 2006).

Parce que la maladie a un impact sur d'autres organes que les poumons, des études se sont intéressées à la modulation du transcriptome dans différents types tissulaires. En 2004, deux études portent leurs intérêts sur le profil d'expression de cellules extraites du pancréas (Kaur, Norkina et al. 2004) et du petit intestin chez la souris (Norkina, Kaur et al. 2004). Tous les deux décrivent la modulation des facteurs pro-inflammatoires.

En 2006, basée sur un modèle cellulaire identique au notre, une étude exploite une puce à ADN « maison » de 144 gènes sélectionnés pour leurs interactions avec le CFTR ou leurs modulations en présence de la forme mutée de CFTR (Gambardella, Biancolella et al. 2006). Cette étude évalue leurs profils d'expression en présence de deux génotypes FK différents et montrent que l'absence de la protéine CFTR (génotype de classe I, $\Delta F508 / G553X$) a un impact plus important sur l'expression des gènes que la présence d'une protéine CFTR non fonctionnelle (génotype de classe II, $\Delta F508 / \Delta F508$).

En 2007, le Dr Adib-Conquy présente l'analyse transcriptomique issue de neutrophiles de patients FK porteurs de deux différentes classes de mutations (Adib-Conquy, Pedron et al. 2008). Basé sur une puce « maison », il évalue le niveau d'expression d'une sélection de 1050 gènes d'intérêt dans la fibrose kystique. Son équipe conclut à un état d'activation des neutrophiles chez les patients fibro-kystiques.

La même année, l'équipe de Verhaeghe, s'appuie sur deux lignées humaines de cellules trachéales FK, pour identifier les mécanismes intrinsèques de l'inflammation présente chez les cellules mutantes (Verhaeghe, Remouchamps et al. 2007). Après avoir validé l'activation de NF- κ B et d'AP-1 dans leur modèle, ils mettent en évidence l'implication de la phosphorylation d'ERK, une composante de la voie des MAPK dans le mécanisme inflammatoire.

G. Hypothèses

Les résultats présentés dans ce mémoire sont orientés par 2 hypothèses :

- (1) la mutation du gène CFTR a un large effet sur la modulation de transcrits pro-inflammatoires dans les cellules épithéliales pulmonaires humaines.
- (2) le stress oxydatif perturbe des voies métaboliques de cellules FK, lesquelles pourraient expliquer la physiopathologie de la maladie. L'une

des voies perturbée pourrait être l'apoptose, la mort cellulaire programmée.

H. Objectifs

Notre expérience de transcriptomique a pour objectif de déterminer les modulations du transcriptome dans une cellule épithéliale pulmonaire déficiente en CFTR. Toutefois, pour atteindre ces objectifs, nous exploitons deux protocoles différents autant dans leur finalité que dans leur méthodologie d'analyse.

Le premier protocole exploite une méthodologie d'analyse largement utilisée dans la communauté exploitant les puces à ADN. Celle-ci se focalise volontairement sur la population de transcripts identifiés comme participant aux processus inflammatoires et modulés dans notre expérience transcriptomique. Cette approche nous conduit à isoler les molécules inflammatoires qui reflètent l'état inflammatoire induit par la mutation.

Le second protocole, quant à lui, est basé sur une approche moins conventionnelle dans le domaine de la transcriptomique, et nous aide à isoler les mécanismes modulés dans une cellule FK en présence de stress oxydatif. La méthodologie développée dans ce protocole a pour but de pallier aux défauts des outils statistiques couramment utilisés dans les analyses transcriptomiques.

Les deux angles d'étude de notre expérience nous conduisent à étudier deux aspects fondamentaux de la maladie : l'inflammation et l'impact du stress oxydatif sur une cellule fibro-kystique. Ils devraient conduire à affiner les hypothèses de recherche et à améliorer les connaissances dans le domaine.

Matériel et méthodes

A. Approches expérimentales

Pour explorer la dérégulation des fonctions biologiques dans une cellule épithéliale pulmonaire porteuse de la mutation du gène CFTR, nous réalisons une expérience de puce à ADN sur deux lignées cellulaires épithéliales bronchiques humaines en présence et en absence d'un stress oxydatif. Ce design expérimental à deux facteurs (mutation, stress oxydatif) à deux niveaux (absence et présence) nous permet de comparer le transcriptome des deux lignées en condition de stress. Basé sur la technologie d'Affymetrix® (figure 4, page 28), nous hybridons les transcripts isolés des quatre conditions expérimentales sur la puce HG-U133 Plus 2.0. Cette plateforme d'analyse transcriptomique permet l'évaluation du niveau d'expression absolu d'environ 38 000 transcripts dans chacune des conditions de notre expérience.

Dans une première approche, nous étudions la perturbation de la réponse inflammatoire présente dans notre modèle cellulaire FK. Un résumé du protocole d'analyse est présenté à la figure 5 partie 8 (page 29). Dans ce protocole nous exploitons uniquement le facteur mutation de notre expérience. Orienté par une analyse statistique et ontologique, nous focalisons nos observations sur un ensemble de gènes inflammatoires dont les niveaux d'expression sont significativement modulés entre les deux types cellulaires. L'étude de leurs annotations génomiques et une analyse comparative avec d'autres expériences de puce à ADN permettent une caractérisation plus précise de l'inflammation induite dans le contexte de la fibrose kystique.

Dans un second temps, nous nous intéressons à l'impact du stress sur le transcriptome de la cellule épithéliale porteuse de la mutation $\Delta F508$. Ce protocole se différencie de notre première approche par deux aspects : (1) nous exploitons les deux facteurs de notre design expérimental (la mutation et le stress oxydatif) et (2) nous développons une méthodologie d'analyse basée sur la variance des données d'expression et qui tente de pallier aux défauts des tests statistiques traditionnellement utilisés dans l'analyse transcriptomique.

Figure 4: Matériels et méthodes : de la cellule aux données d'expression

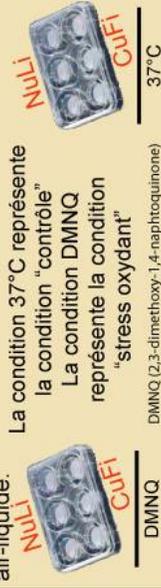
2. Mesure de la résistance



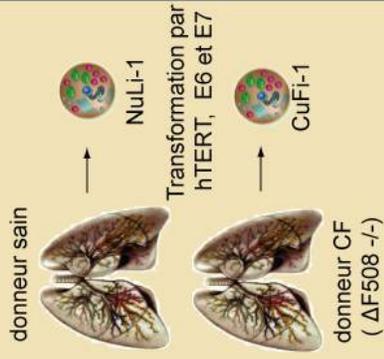
Pour les deux types cellulaires, la mesure de la résistance est réalisée par les techniques de l'Ussing ou par EVOM dans le but de vérifier l'activité ionique liée au canal CFTR.

3. Conditions de culture

Chaque type cellulaire est mis en culture dans le milieu BEGM contenant suppléments nutritifs et antibiotiques dans deux conditions expérimentales différents en triplicatas. Après un mois environ, une monocouche cellulaire est formée et les cellules sont en interface air-liquide.



1. Lignées cellulaires



La transformation des cellules épithéliales bronchiques est réalisée par le Dr J. Zabner (Zabner J et al, Am J Physiol Lung Cell Mol Physiol, 2003).

4. Extraction ARNm - Synthèse ADNc

Pour chaque condition expérimentale, les ARNm sont extraits afin de mesurer le niveau de transcription des gènes des deux types cellulaires.



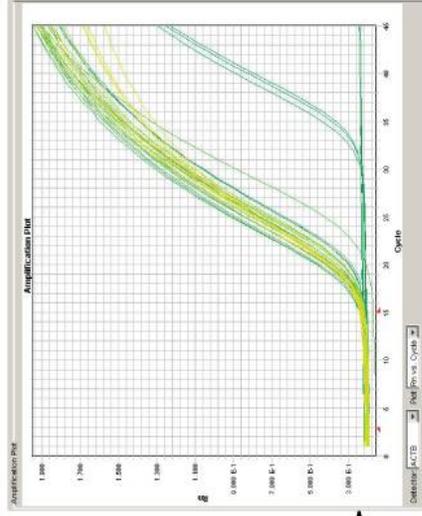
5. Hybridation



Cette partie décrit les étapes techniques de l'expérience de puce à ADN.

5bis. qPCR

Basé sur les résultats issus de l'expérience de puce à ADN, une sélection de gènes est faite et leurs modulations sont confirmées par qPCR.



6. Acquisition

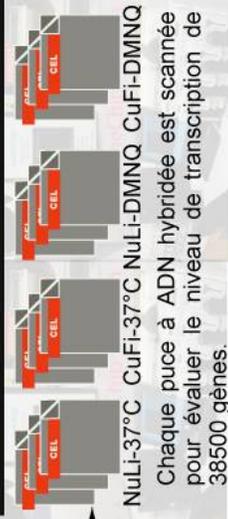
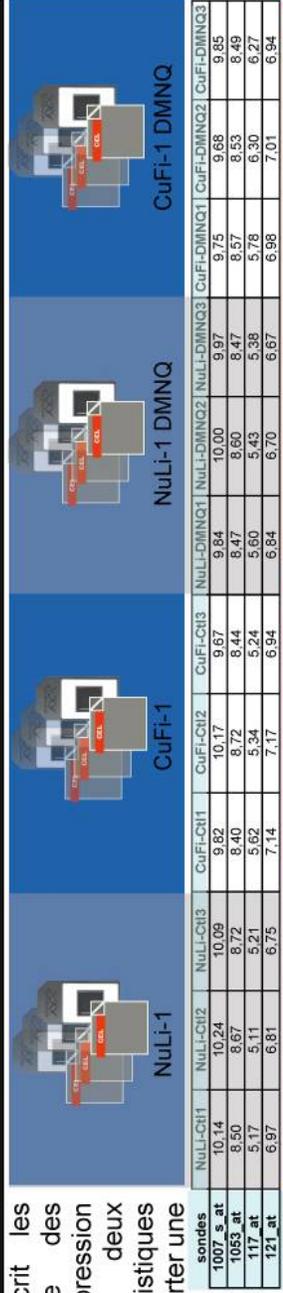


Figure 5: Matériels et méthodes : des données d'expression à l'interprétation biologique

Cette partie décrit les étapes d'analyse des données d'expression ainsi que les deux approches statistiques utilisées pour apporter une interprétation biologique à notre expérience.

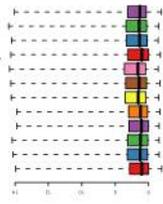


6. Contrôle Qualité

1. Inspection visuelle des hybridations
2. Distribution des intensités
3. Critères du contrôle qualité d'Affymetrix:
 - bruit de fond
 - facteur de réduction
 - pourcentage de present ratio 3'/5'
 - contrôles "Spike-In"

7. Normalisation

Les données d'expression sont normalisées par RMA.



8. Première approche

Quelles sont les caractéristiques moléculaires de la réponse inflammatoire d'une cellule épithéliale pulmonaire porteuse de la mutation ΔF508?

Analyse différentielle

Basé sur les conditions expérimentales NuLi-1 (non FK) et CuFi-1 (FK), nous isolons les transcrits significativement modulés entre les deux conditions (P-Value < 0.05) à l'aide d'un test t modifié.

Analyse ontologique

Une analyse ontologique permet de connaître les fonctions biologiques significativement surreprésentées liées aux processus inflammatoires et immunes.

Annotation

Après un contrôle des annotations génomiques, nous ne retenons que les transcrits présentant une consistance dans les annotations d'Affymetrix, Ensembl et Resnet.

Analyse qPCR

Les modulations des transcrits significativement modulés entre les deux conditions expérimentales (NuLi-1 et CuFi-1) sont confirmées par qPCR.

9. Seconde approche

Quelle sont les fonctions biologiques modulées par la mutation CFTR en présence du DMNQ, inducteur du stress oxydatif ?

Filtrage des données

Considérant les quatre conditions expérimentales, l'IQR nous sert à sélectionner les transcrits avec des variances d'intérêt. De cette façon, les transcrits avec très peu de variance sont écartés de l'analyse.

Analyse de redondance

L'analyse de redondance isole les transcrits dont la variance est attribuable aux effets des deux facteurs, reflétant une interaction potentielle.

Classification

L'ensemble des transcrits d'intérêt est soumis à un algorithme de classification non supervisé lequel distingue les tendances dans la modulation.

Analyse ontologique

Pour chaque groupe issu de la classification, l'analyse ontologique met en avant les processus biologiques surreprésentés.

B. Techniques et protocoles expérimentales de l'objectif 1 : Étude de la réponse transcriptomique inflammatoire dans des cellules FK

1. Modèles cellulaires

Pour étudier le transcriptome de cellules FK, nous utilisons deux lignées cellulaires, don du Dr J. Zabner de l'Université d'Iowa.

La lignée cellulaire contrôle est désignée « NuLi-1 » (*Normal Lung, University of Iowa*) et a été développée à partir de cellules épithéliales bronchiques humaines de patient sain. La lignée cellulaire FK, désignée « CuFi-1 » (*Cystic Fibrosis, University of Iowa*), a été développée à partir de cellules épithéliales bronchiques humaines porteuses de la mutation homozygote $\Delta F508$ (Zabner, Karp et al. 2003). Les deux types cellulaires ont été transformés par la hTERT (*human Telomerase Reverse Transcriptase*), une sous-unité catalytique de la télomérase humaine et par deux oncogènes (E6 et E7) du papillomavirus de type 16 (figure 4, partie 1, page 28).

Les deux lignées présentent 47 chromosomes et le type NuLi-1 possède les chromosomes sexuels ♂ (XY) tandis que les cellules CuFi-1 possède les chromosomes sexuels ♀ (XX).

Dans notre expérience, nous utilisons des cellules essentiellement au passage 14 et au passage 16. Le passage est le procédé de culture cellulaire permettant la production d'un grand nombre de cellules à partir de cellules préexistantes. Puisqu'un nombre de passages trop élevé peut affecter des fonctions cellulaires, nous avons choisi un intervalle de passage où les cellules CuFi-1 et NuLi-1 expriment le CFTR.

2. Conditions de culture des modèles cellulaires

La mise en culture des deux lignées nous permet d'obtenir une monocouche confluente de cellules et dans des conditions de culture interface air-liquide (ALI), celles-ci forment un épithélium polarisé présentant une résistance trans-épithéliale. De cette façon, nous nous rapprochons au mieux des conditions *in vivo* pour comparer les transcriptomes de cellules épithéliales pulmonaires normales et porteuses de la mutation $\Delta F508$ (figure 4, partie 3, page 28).

Initialement les cellules sont décongelées en boîte de Pétri préalablement traitées au collagène de type VI (Sigma, Saint Louis, MO) pour un minimum de 18 heures à température ambiante et rincées deux fois au PBS (*Phosphate Buffered Saline*). Les cellules sont mises en culture dans le milieu *Clonetics™ Bronchial Epithelial Cell Growth Media* (Cambrex, Charles City, IA) incluant des suppléments et les antibiotiques suivants : pénicilline-streptomycine (50 $\mu\text{g/ml}$), gentamycine (50 $\mu\text{g/ml}$), (fluconazole à 2 $\mu\text{g/ml}$) (ANNEXE 1 : Composition des suppléments et antibiotiques du milieu de culture). Le milieu des cellules est changé tous les jours.

Au jour 8, lorsque 80 à 100 % des cellules sont confluentes, elles sont trypsinisées (trypsine 0.25 % EDTA 1mM, Gibco®), (Invitrogen Carlsbad, CA), puis passées sur des supports semi-perméables, filtres de 0,4 μm de diamètre (Corning, One Riverfront Plaza, NY), le fond recouvert de collagène de type VI, celui-ci agissant comme une matrice qui facilite l'adhésion et la différenciation cellulaire. Ensuite, elles sont mises en culture dans un milieu de base composé de D-MEM (*Dulbeccos Modified Eagles Medium*) + F12 (50 % D-MEM, Gibco®, et 50 % de F-12, Gibco®) additionné des antibiotiques suivants: pénicilline-streptomycine à 50 $\mu\text{g/ml}$, gentamycine à 50 $\mu\text{g/ml}$, fluconazole à 2 $\mu\text{g/ml}$ et 1.25 $\mu\text{g/ml}$ amphotéricine

(ANNEXE 1 : Composition des suppléments et antibiotiques du milieu de culture). Pour la première journée de mise en culture, au milieu de base, on ajoute du FBS (*Fetal Bovin Serum*, Gibco®) à 5 %.

Au jour 9, le milieu est changé pour du milieu de base additionné de Ultrosor G à 2 % (Pall Corporation, Port Washington, NY).

Au jour 12, le milieu de culture du côté apical est retiré et le milieu du côté basolatéral est changé.

Entre les jours 12 et 28, le milieu du côté basolatéral est changé tous les 2 jours et le milieu qui remonte du côté apical est aspiré.

Entre les jours 28 et 30, lorsque le milieu en basolatéral ne remonte plus du côté apical, les cellules sont prêtes à être utilisées. Elles sont alors dans des conditions ALI.

3. Mesure de l'activité de NF- κ B

NF- κ B est un facteur de transcription important dans de nombreux processus biologiques, dont la réponse inflammatoire. Le suivi de son activité est important pour l'étude des réponses cellulaires.

Les cellules ont été cultivées dans les mêmes conditions que celles de notre expérience de puce à ADN. Toutefois, le nombre de passages est différent puisqu'elles étaient au passage 16 pour ce dosage protéique. Réalisée en triplicata biologique, l'activité de base de NF- κ B a été mesurée pour les cellules NuLi-1 et CuFi-1 en absence et en présence de LPS (lipopolysaccharide) (1 μ g/ml) ou de TNF (10 ng/ml). La mesure de l'activité de p65, déterminée par la translocation nucléaire de la sous-unité, a été déterminée après 30 minutes de stimulation au LPS et TNF sur 2.5 μ g de protéine nucléaire.

L'expression des formes actives de p65 et de p50 a été évaluée à l'aide du kit TransAM™ *Flexi NF-κB p50 / NF-κB p65*, d'Active Motif® (Carlsbad, CA, USA). Ce kit contient une plaque de 96 puits dans lesquels sont fixés des oligonucléotides contenant la séquence consensus 5' GGGACTTCC 3' qui reconnaît spécifiquement la forme active de p65. Ainsi, si p65 est présente dans l'extrait de protéine nucléaire, la protéine peut se lier à sa séquence consensus. Le complexe p65/ADN peut alors être détecté immunologiquement par un anticorps dont la présence est quantifiable par une réaction chromogénique.

L'extraction nucléaire est réalisée avec le kit *TransAM™ nuclear extract Kit*, d'Active Motif®, selon les recommandations du fabricant. Après avoir déterminé la concentration de protéine nucléaire de chaque essai, nous avons déposé dans chaque puit un volume correspondant à 2.5 µg de protéine.

La réaction antigénique utilise un anticorps primaire (ou anti p50 NF-κB IgG de lapin dirigé contre p65 ou p50 humain). La plaque a été incubée 1 heure à température ambiante avec un anticorps primaire anti p65 NF-κB, puis lavée à trois reprises.

L'anticorps secondaire (anticorps de chèvre dirigé contre les IgG de lapin), conjugué à l'enzyme HRP (*Horse Radish Peroxydase*), a été incubé dans les mêmes conditions que l'anticorps primaire. Les puits ont été rincés quatre fois avant d'être incubés avec la solution de développement (3,3',5,5' tetramethylbenzidine dans DMSO 1 %).

Après une incubation de 4 minutes à température ambiante, la réaction fut arrêtée avec du H₂SO₄ à 0,5 M. Par la suite la densité optique de chaque puit est lue au lecteur de plaques à 450 nm pour le dosage de p65 et 595 nm pour le dosage de p50.

4. Expérience de PCR quantitative (qPCR)

a. *Technique qPCR*

Les quantifications issues d'une expérience de puce à ADN sont discutables pour plusieurs raisons: la spécificité des sondes, la transformation et la normalisation des données d'expression, les hybridations croisées (Murphy 2002). Par conséquent, il est nécessaire de valider les modulations observées par une autre méthode d'évaluation, comme la PCR quantitative, qui permet de quantifier de façon relative les transcripts. La quantification relative décrit les changements dans l'expression du gène considéré par rapport à l'expression d'un gène contrôle, appelé contrôle endogène (ACTB, GAPDH,...), dont le niveau d'expression reste relativement constant quelles que soient les conditions expérimentales. À cette fin, nous utilisons le système de qPCR développé par *Applied Biosystems* (Applied Biosystems, Foster City, CA) : *Taqman® Gene Expression Assay*.

b. *Synthèse de l'ADNc*

La synthèse de l'ADNc (ADN complémentaire) est réalisée au Centre de Recherche du CHUM (Centre Hospitalier de l'Université de Montréal), site de l'Hôtel-Dieu, dans le laboratoire du Dr Yves Berthiaume.

Après l'extraction des ARNm (ARN messenger) totaux, ceux-ci sont utilisés pour synthétiser l'ADNc. Dans un premier temps, les ARNm totaux sont traités avec un volume de DNase 1 (Sigma), pour éliminer l'ADN résiduel. Dans un second temps, les mRNAs totaux sont rétro-transcrits par l'ajout de 1.1 volume de dNTP à 10 mM (Invitrogen), de 1.1 volume de *random primer* à 200 ng/ul (Invitrogen), de 1.1 volume de MMLV *reverse transcriptase* (Invitrogen). La retro transcription est réalisée en trois étapes : 10 minutes à 25°C, 60 minutes à 37°C et 15 minutes à 70°C (figure 4, partie 4, page 28)

c. Réaction de qPCR

La quantification des transcrits par qPCR est réalisée à l'IRIC (Institut de Recherche en Immunologie et Cancérologie), dans les laboratoires de la plateforme de génomique fonctionnelle. La liste des transcrits testés est présentée en annexe 11

Toutes les réactions sont réalisées sur des plaques de 384 puits.

Chaque réaction est réalisée dans un volume final de 10.05 μ l contenant 80 ng d'ADNc pour un volume de 4 μ l, 5 μ l de *Fast Universal Master Mix* qPCR (Applied Biosystems), 1 μ l de sonde universelle Taqman (20X) et de 0.05 μ l d'amorces spécifiques (concentration équimolaire de 100 μ M).

Les contrôles endogènes utilisés sont GAPDH (glyceraldehyde-3-phosphate déhydrogénase) et ACTB (β -actin). (figure 4, partie 5 bis, page 28)

d. Analyse des résultats de qPCR

La quantité relative du transcript testé est déterminée par la méthode du $\Delta\Delta$ Ct (Livak et Schmittgen 2001). La valeur Ct (*cycle threshold*) est définie par le nombre de cycles d'amplification de l'ADNc nécessaire pour que le niveau signal de la fluorescente dépasse un seuil limite. Les valeurs de Ct sont inversement proportionnelles à la quantité d'acide nucléique dans l'échantillon : plus le Ct est bas, plus la quantité d'ADNc est élevée.

La valeur Ct du transcript est normalisée par le Ct d'un gène contrôle endogène (GAPDH ou ACTB). Ainsi, le Δ Ct = Ct_{target} - Ct_{GAPDH}. Ce Δ Ct est comparé arbitrairement avec le Ct d'un gène calibrateur : $\Delta\Delta$ Ct = Δ Ct_{sample} - Δ Ct_{calibrator}. L'expression relative est calculée par le logiciel SDS2.2.2 d'Applied Biosystems selon la formule $RQ = 2^{-\Delta\Delta Ct}$ (figure 4, partie 5bis, page 28).

5. Protocole de l'objectif 1

L'influence de la mutation du gène CFTR sur le transcriptome d'une cellule épithéliale bronchique peut-être déterminée en comparant le niveau d'expression de l'ensemble des transcrits de deux lignées cellulaires humaines. Le type contrôle, NuLi-1, est issu d'un patient sain et le type FK, CuFi-1, est issu d'un patient porteur homozygote de la mutation $\Delta F508$.

Les deux lignées cellulaires sont utilisées au passage 14 et sont mises en culture dans le milieu BEGM avec des antibiotiques jusqu'à formation d'une monocouche cellulaire. Ensuite celles-ci sont mises en culture dans le milieu DMEN+F12, jusqu'à qu'elles soient en condition ALI, reproduisant *in vitro* les conditions physiologiques d'un épithélium pulmonaire.

Les ARNm polyadénylés des deux conditions expérimentales sont extraits (en triplicata biologiques) pour être hybridés à la puce HG-U133 Plus 2.0 d'Affymetrix®.

Après l'acquisition des données d'expression et leur normalisation, une analyse statistique univariée optimisée exploite les modèles linéaires (Smyth 2004) et isole les transcrits avec des modulations significativement différentes entre les conditions NuLi-1 et CuFi-1.

Par la suite, l'analyse ontologique des transcrits met en évidence les fonctions biologiques surreprésentées, parmi lesquelles apparaissent celles liées aux mécanismes inflammatoires. De plus, nous validons par qPCR, les niveaux d'expression de quelques transcrits d'intérêt qui sont différentiellement exprimés dans notre analyse et connus dans la littérature FK.

L'ensemble du protocole 1 est présenté dans la figure 5 partie 8 (page 29).

C. Technique et protocoles expérimentales de l'objectif 2 : Études de la réponse transcriptomique de cellules FK induite par un stress oxydatif

1. Modèles cellulaires et conditions de culture

Les modèles cellulaires et les conditions de culture utilisés dans le second protocole sont identiques à ceux employés dans le premier protocole. Néanmoins, 24 heures avant l'extraction des ARNm pour l'hybridation sur les puces à ADN, 15 µl de DMNQ (2,3-dimethoxy-1,4-naphtoquinone) à une concentration de 15 µM sont ajoutés du côté apical de la couche monocellulaire pour induire un stress oxydatif.

2. Choix du modèle de stress oxydatif

Pour induire le stress oxydatif dans notre expérience, nous choisissons le DMNQ suite à une étude dose-réponse, dans notre laboratoire, de plusieurs molécules cytotoxiques. La dose et la durée d'exposition auxquelles les cellules doivent être exposées sont déterminées pour induire le stress tout en assurant l'intégrité de la monocouche cellulaire. Les durées d'expositions testées sont de 6 h et de 24 h, et les doses testées sont de 10 µM, 15 µM et 20 µM.

3. Mesure de l'activité des caspases -3 et -7

Pour évaluer l'apoptose dans nos modèles cellulaires, nous mesurons l'activité de caspases impliqués dans la cascade apoptotique. Les caspases (*cysteine aspartic acid-specific protease*) sont des molécules importantes dans les voies de l'apoptose chez les mammifères (Thornberry et Lazebnik 1998). Elles sont classées selon la taille de leurs pro-domaines (LeBlanc 2003), lesquels contiennent des motifs essentiels à l'activité caspasique. Les

caspases 3, 7, et 6 possédant un prodomaine court (< 30 acides aminés) sont des caspases effectrices et leurs activations conduisent au clivage de nombreuses protéines cellulaires, aboutissant aux étapes terminales et ultimes de l'apoptose.

Le dosage de l'activité des caspases -3 et -7 (3/7) autorise l'interprétation de l'activité apoptotique d'une cellule (Liu, Li et al. 2004; Ren, Wagner et al. 2004). À partir d'une préparation cellulaire, le kit de détection *Caspase-Glo® 3/7 Assay* de Promega (Madison, WI, USA) permet une mesure de l'activité des caspases 3/7 sans distinguer leurs activités respectives. Ce réactif contient un substrat des caspases 3/7, lequel contient le térapeptide DEVD (Asp-Glu-Val-Asp). Ce substrat clivé relâche l'aminoluciférine, un substrat de la luciférase, produisant de la luminescence.

Nous avons cultivé chaque type de cellule dans les conditions expérimentales similaires à celles de notre expérience de puce à ADN. Cette expérience est réalisée en triplicata biologique, réunissant les conditions suivantes : NuLi-1 (en absence et présence de DMNQ), CuFi-1 (en absence et présence de DMNQ).

Pour évaluer l'impact du stress oxydatif, nous avons utilisé le même modèle que celui de l'expérience de puce à ADN: un traitement au DMNQ à une concentration de 15 μ M pendant une période de 24 heures.

Pour chaque condition, les filtres sont rincés avec du PBS du côté apical et basolatéral. Ensuite, un volume de réactif suffisant pour recouvrir la surface du filtre (200 μ l à 5 mU) est rajouté du côté apical. En accord avec les recommandations du constructeur, nous mesurons la luminescence dégagée par la luciférase après 30 minutes de traitement dans l'obscurité.

Afin de normaliser les luminescences mesurées, nous les rapportons à la quantité de protéine présente dans chaque échantillon testé.

4. Protocole

Notre second protocole a pour objectif de rendre compte de l'implication du stress oxydatif sur le transcriptome d'une cellule FK. Pour cela, nous exploitons les deux conditions de notre expérience de puce à ADN : la mutation et le stress oxydatif. Ainsi, pour chaque transcript, nous évaluons le niveau d'expression dans les quatre conditions expérimentales (NuLi-1 et CuFi-1 en présence et en absence du DMNQ) et sélectionnons ceux présentant un profil d'expression influencé par les deux facteurs (la mutation $\Delta F508$ et le traitement par l'oxydant) reflétant une interaction potentielle entre les deux conditions.

Tout d'abord, nous écartons de notre analyse l'ensemble des transcripts présentant des niveaux d'expression très peu variables au cours de l'expérience. Ensuite, sur les transcripts présentant une modulation suffisante, nous appliquons une méthode d'analyse non paramétrique basée sur la variance de leur expression. Cette méthode quantifie l'impact de chaque facteur sur le profil d'expression des transcripts en décomposant leur variance. Cette approche nous conduit à ne s'intéresser qu'aux transcripts dont la modulation observée peut être expliquée par une combinaison des deux facteurs. Pour raffiner notre analyse, les transcripts d'intérêt sont séparés en quatre groupes par un algorithme de classification non-supervisé. Chaque groupe rassemble des transcripts avec des profils d'expression similaires. Cette classification distingue les grandes tendances de modulation dans nos données. Finalement une analyse ontologique est réalisée pour chaque groupe soulignant ainsi les fonctions biologiques surreprésentées lesquelles sont représentatives de l'impact du stress oxydatif sur le transcriptome d'une cellule FK.

A travers cette méthodologie, nous sommes capables d'isoler les fonctions biologiques modulées à la hausse ou à la baisse dans notre expérience et produire des hypothèses expliquant les perturbations engendrées par le stress oxydatif.

L'ensemble du protocole 2 est présenté dans la figure 5 partie 9 (page 29).

D. Méthodes de mesures physiopathologiques

1. Caractéristiques électrophysiologiques du modèle cellulaire

La membrane cellulaire est un système complexe qui répond à des stimulations externes physiques et chimiques. Pour comprendre les mécanismes en partie mis en place par de tels stimuli, une approche consiste à étudier la diffusion et le transport de molécules biologiques (et/ou pharmacologiques). Dans le cadre d'étude de la fibrose kystique où le défaut d'un transporteur ionique est impliqué, caractériser le transport des ions de part et d'autre de la membrane plasmique est nécessaire pour juger de l'incidence de la perte de fonction du transporteur ionique CFTR.

Pour caractériser le transport ionique de nos deux types cellulaires sous différentes conditions, nous utilisons la technique de la chambre d'Ussing, développée par l'équipe d'Ussing en 1958 (Ussing et Zerahn 1951). Cette technique permet de mesurer les paramètres électriques de la monocouche (le courant de court-circuit et le potentiel trans-épithélial).

La chambre d'Ussing est composée de deux chambres, apicale et basolatérale, séparées par le filtre portant la monocouche cellulaire polarisée (ou un fragment tissulaire excisé) dont on souhaite déterminer le mécanisme de transport ionique. Les deux chambres baignent dans un

volume identique de composition précise limitant du fait les biais possibles qui pourraient être engendrés par un gradient ionique.

Dans de telles conditions, lorsque le voltage est maintenu à zéro, il est possible de mesurer le courant court-circuit qui correspond au courant injecté par l'appareil pour maintenir le potentiel à zéro. Ce courant, que l'on appelle le courant de court-circuit (I_{sc} pour *short-circuit current*), correspond au courant nécessaire pour maintenir la différence de potentiel entre le niveau apical et basolatéral, ce qui correspond à la quantité de transport actif des ions d'une cellule.

Dans l'épithélium respiratoire, le courant de court-circuit résulte principalement de l'absorption transépithéliale de Na^+ généré par l'activité du canal ENaC du côté apical et de la pompe Na^+/K^+ -ATPase du côté basolatéral, ainsi que de la sécrétion de Cl^- du compartiment basolatéral vers le compartiment apical. Après inhibition de l'absorption de sodium par l'amiloride et stimulation du canal CFTR, cette technique permet de mesurer l'activité membranaire de ce dernier.

2. La mesure du stress oxydatif

La mesure du stress oxydatif est réalisée par une mesure du GSH (glutathion) intracellulaire total. Pour cela, la forme oxydée du glutathion (GSSG) est réduite par la glutathion réductase en présence du NADPH pour produire du GSH. Ensuite le glutathion réduit est mesuré par une réaction colorimétrique où le DNTB (2-nitrobenzoïque) réagit spécifiquement avec les groupements thiol, donne naissance à un composé coloré qui absorbe à 412 nm (Mourad, Min et al. 2000). Cette méthode mesure à la fois le GSH produit à partir de GSSG et le GSH cellulaire. De plus, afin de vérifier l'intégrité de la monocouche cellulaire, le courant total est mesuré par la méthode d'Ussing.

E. Expérience de puce à ADN

1. Caractéristiques de la plateforme d’Affymetrix® HG-U133 Plus 2.0

L’expérience de puce à ADN est réalisée par le service de biopuces du Centre de Recherche du CHUM.

Pour notre étude, nous utilisons la technologie des puces d’expression à oligonucléotides d’Affymetrix® (Affymetrix®, Santa Clara, CA). Les caractéristiques importantes de la plateforme d’Affymetrix® HG-U133 Plus 2.0 sont résumées dans la figure 6 (page 43)

Les biopuces d’Affymetrix® sont composées d’une plaque de silicate de 50 µm par 50 µm, sur laquelle sont fixées par photolithographie 1 352 569 oligonucléotides de 25 bases qui agissent comme des sondes. Selon la terminologie du constructeur, ces sondes sont de deux types: PM (*Perfect Match*) et MM (*MisMatch*). Les sondes PM hybrident de façon presque spécifique une région de la partie 3’UTR d’un transcript. Chaque sonde PM est associée à une sonde MM, de même composition nucléotidique à l’exception de la 13^{ième} base : elle forme une paire de sondes. Étant donné la longueur des sondes PM, celles-ci sont susceptibles d’hybrider d’autres transcripts. Ces hybridations, appelées « hybridations croisées », sont indésirables et correspondent au bruit de fond spécifique à la paire de sonde. Ce bruit de fond est calculé à partir de la moyenne de log ratio des valeurs des PM et des MM et est utile pour pondérer le signal PM. De plus, pour tester la spécificité de l’hybridation de chaque transcript, 14 paires de sondes hybrident le même transcript dans la région 3’UTR ne dépassant pas 500 bases, et forment un jeu de sondes, identifié par une nomenclature propre à Affymetrix®. La répartition des jeux de sondes hybridant le même

transcript sur la puce à ADN est aléatoire et non contiguës, tandis que le couple de sondes PM/MM est physiquement situé l'un à côté de l'autre.

Dans notre expérience, nous utilisons la plateforme pangénomique humaine d'Affymetrix® nommée « *Human Genome HG-U133 Plus 2.0* » composée de 1.3 millions d'oligonucléotides, représentant 54 675 *probesets*, terme propre à la technologie d'Affymetrix®. **Dans ce mémoire, nous utilisons le mot « sonde » à la place du terme *probeset*, et l'ensemble des analyses présentées s'intéressent à la modulation de 54 675 sondes dans nos conditions expérimentales reflétant le niveau de transcription d'environ 38 500 transcripts, et par conséquent l'activité de 38 500 gènes.**

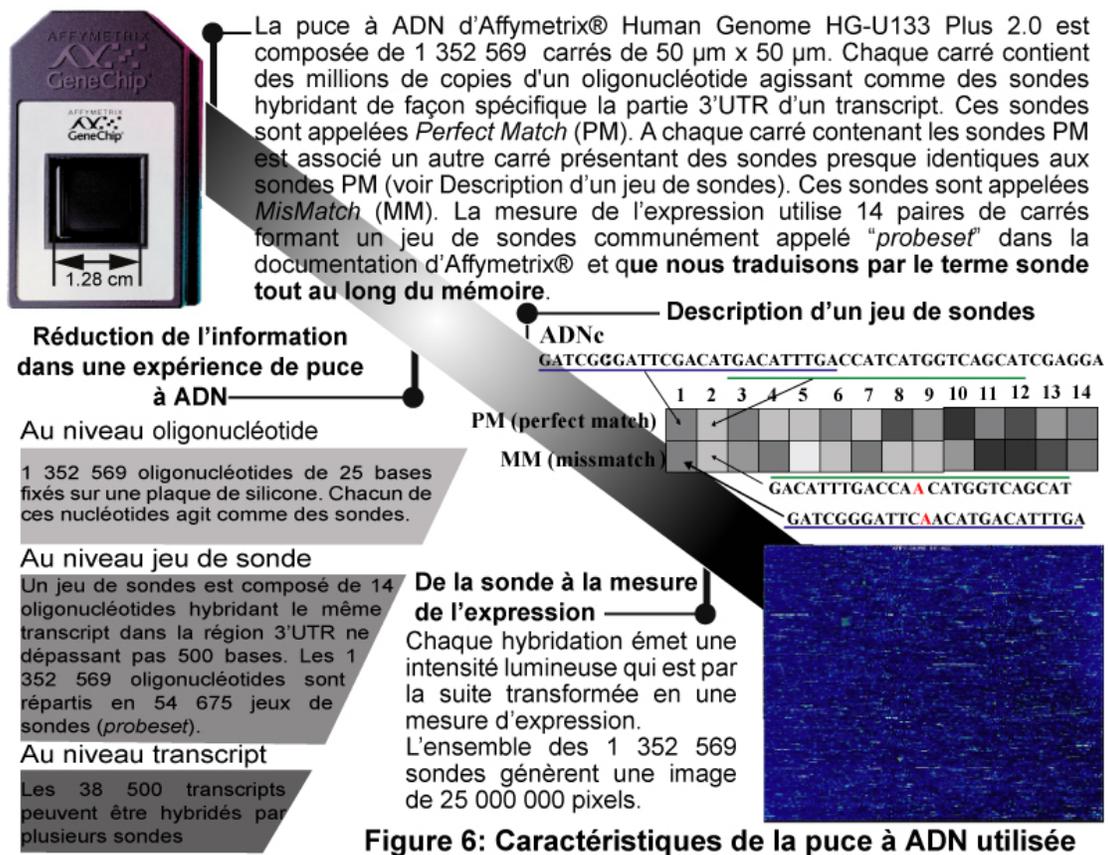


Figure 6: Caractéristiques de la puce à ADN utilisée

2. De l'extraction des transcrits au contrôle qualité des données d'expression

a. *Extraction des ARNm*

Pour chaque réplicat biologique de chaque condition expérimentale, les ARNm totaux sont isolés en utilisant le TriZOL (Invitrogen) en accord avec les recommandations du fabricant. Les ratios d'absorbance 260nm/280nm compris entre 1.8 et 2.2 jugent d'une bonne qualité de l'extraction. Ensuite, les ARNm totaux sont purifiés avec le kit *Rneasy Mini kit* (Qiagen, Valencia, CA) selon les recommandations du fabricant. La qualité finale des ARNm totaux purifiés est vérifiée par le système d'Agilent 2100 Bioanalyzer (Agilent, Palo Alto, CA) (figure 4 partie 4 page 28).

b. *Synthèse d'ADNc et hybridation.*

À partir des ARNm totaux, une transcription inverse synthétise les ADNc totaux suivant les recommandations du fabricant. Les ADNc totaux permettent de synthétiser des ARNc totaux utilisant comme amorce un oligonucléotide T7-(dT) 24, marqué à la biotine avec les produits fournis par Affymetrix®. Les ARNc totaux sont ensuite fragmentés et hybridés sur la plateforme HG-U133 Plus 2.0 pendant une période de 16 heures, à 45°C, à 60 tours par minute dans le four d'hybridation 640 d'Affymetrix®. La plateforme est ensuite lavée, fixée par la streptavidine-phycoerythine, puis par un anticorps biotiné antistreptavidine. Ce processus de trois étapes est réalisé dans l'Affymetrix® Fluidic Station 450 (figure 4 partie 4 page 28).

c. Lecture des puces et acquisition des données d'expression

La lecture des puces est réalisée par le Scanner d’Affymetrix® 3000 7G et les données d’expression de la puce à ADN sont analysées par la suite logicielle d’Affymetrix® MAS5 (*Affymetrix® Microarray Analysis Software version 5.0*). À partir de l’image de la puce scannée, le logiciel calcule les intensités brutes pour chaque sonde et crée un fichier de signature « .CEL » pour chaque puce scannée. Ce fichier sert de base à l’analyse bioinformatique de puce à expression d’Affymetrix®.

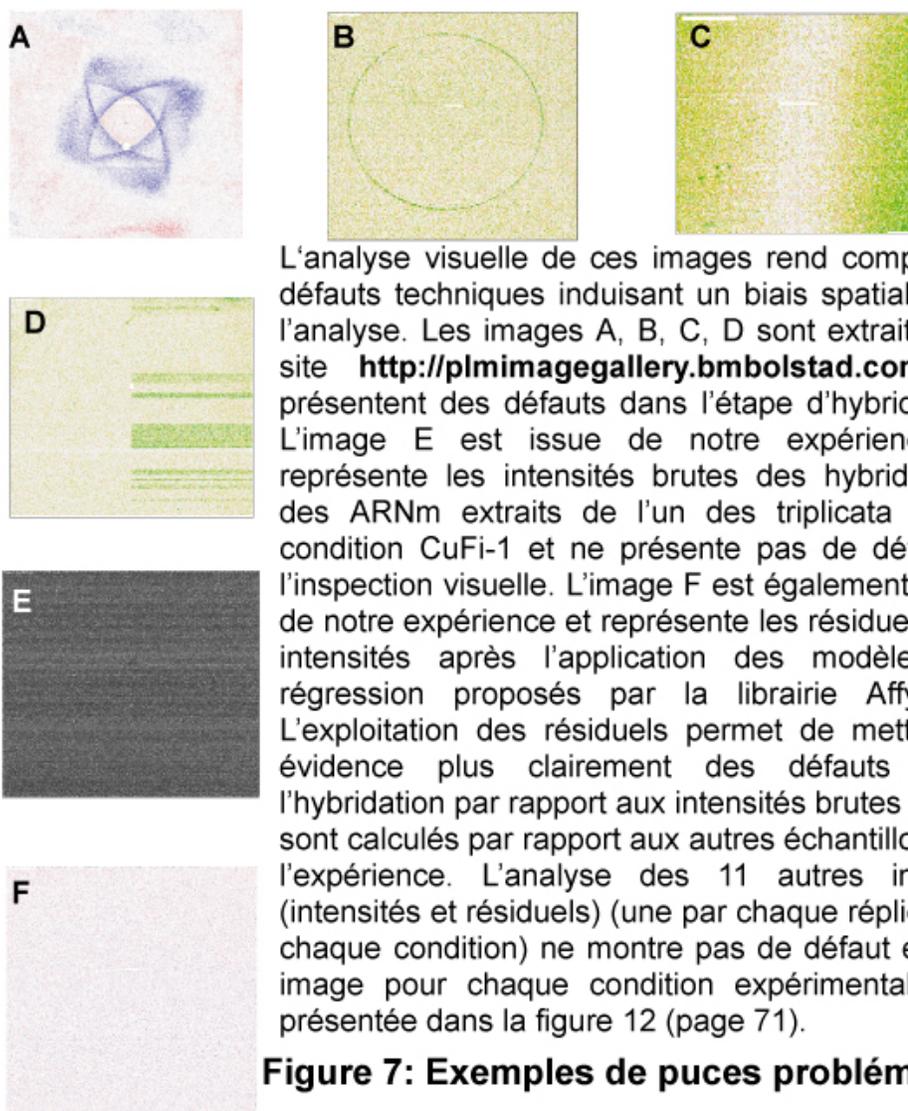
d. Contrôle qualité

Pour assurer une bonne analyse des données d’expression, il est indispensable de juger de la qualité de notre expérience et d’écarter toute puce hybridée présentant des défauts d’origine technique. Par conséquent, notre analyse exploite uniquement des données respectant certains critères de qualité.

Le contrôle qualité présente deux niveaux d’examen. Le premier niveau permet un contrôle global jugeant du comportement des intensités des 1.3 millions d’oligonucléotides dans chaque puce à ADN de l’expérience. Cette étape comprend : (1) l’analyse visuelle des fichiers images et (2) l’étude de la distribution des intensités pour chaque puce. Le second niveau du contrôle qualité évalue le comportement des intensités d’oligonucléotides par rapport à des critères définis par Affymetrix® (*Affymetrix® technical report*) : le bruit de fond moyen, le facteur de réduction, le pourcentage de présent, ainsi que le ratio 3’/5’, les contrôles étalons.

Premier niveau : étude des intensités des oligonucléotides.

L'étude de l'intensité des hybridations des oligonucléotides nous permet d'évaluer la qualité de chaque puce à ADN composant l'expérience. Pour cela, à partir du fichier .CEL contenant les intensités, nous générons un fichier image représentant les intensités en \log_2 et repérons ainsi d'éventuels défauts dans l'hybridation. La figure 7 (page 47) montre quelques exemples d'images générées à partir de données d'expression.

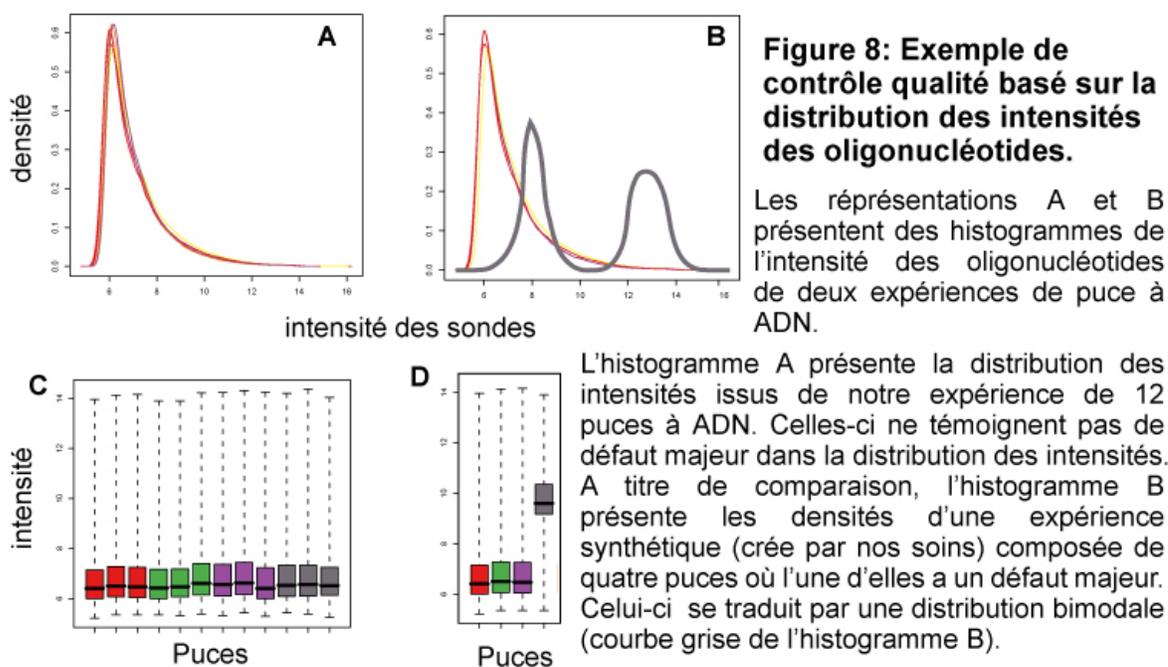


L'analyse visuelle de ces images rend compte de défauts techniques induisant un biais spatial dans l'analyse. Les images A, B, C, D sont extraites du site <http://plmimagegallery.bmbolstad.com/> et présentent des défauts dans l'étape d'hybridation. L'image E est issue de notre expérience et représente les intensités brutes des hybridations des ARNm extraits de l'un des triplicatas de la condition CuFi-1 et ne présente pas de défaut à l'inspection visuelle. L'image F est également issue de notre expérience et représente les résiduels des intensités après l'application des modèles de régression proposés par la librairie AffyPLM. L'exploitation des résiduels permet de mettre en évidence plus clairement des défauts dans l'hybridation par rapport aux intensités brutes car ils sont calculés par rapport aux autres échantillons de l'expérience. L'analyse des 11 autres images (intensités et résiduels) (une par chaque réplicat de chaque condition) ne montre pas de défaut et une image pour chaque condition expérimentale est présentée dans la figure 12 (page 71).

Figure 7: Exemples de puces problématiques

De la même façon, l'étude de la distribution des intensités des hybridations des oligonucléotides dans chacune des puces à ADN se révèle indispensable pour repérer la ou les puces présentant des comportements différents des autres. Nous nous attendons à avoir des distributions similaires dans les triplicatas ainsi qu'entre les différentes conditions expérimentales puisque notre expérience n'utilise qu'un type cellulaire où une faible quantité de gènes sont

modulés. Un histogramme des intensités pour chaque puce ou une représentation de « boîtes à moustaches » permet de mettre en évidence, par exemple, une distribution bimodale indiquant un artefact spatial. La figure 8 (page 48) présente la distribution des figures des intensités des oligonucléotides.



Les boîtes à moustaches dans les représentations C et D, sont une autre visualisation des distributions des oligonucléotides de la figure A et de la figure B. En C, la répartition des intensités est constante dans l'ensemble des 12 puces. En D, nous retrouvons le défaut de l'expérience synthétique sous la forme de la boîte à moustaches grise avec une médiane plus élevée que les trois autres puces à ADN, en raison de la distribution bimodale.

Second niveau : l'analyse des paramètres du contrôle qualité d'Affymetrix®.

Le second niveau de contrôle qualité exploite des paramètres fournis par Affymetrix® et permet de rendre compte de la qualité de l'expérience à travers cinq contrôles :

(1) Le bruit de fond : Les deux principaux facteurs qui génèrent du bruit dans une expérience de puce à ADN sont : (1) la variabilité de la mesure lors de l'étude de lecture de la puce hybridée et (2) la qualité des échantillons des ARNm analysés. Pour permettre une comparaison des intensités entre les puces, il est important de vérifier que le bruit de fond pour chaque puce soit similaire. Affymetrix® recommande que celui-ci se situe entre les valeurs critiques 20 et 100. Des écarts élevés dans ces valeurs pourraient indiquer un problème au niveau de l'hybridation et biaiser les intensités. Le bruit de fond moyen est estimé sur la moyenne de 16 valeurs de bruit de fond. Pour le calculer, la grille de 1.3 millions pixels est divisée en 16 régions rectangulaires égales et pour chaque région, 2 % des intensités les plus basses sont utilisées pour calculer le bruit de fond. L'intensité de chaque oligonucléotide est ainsi ajustée en fonction de la valeur moyenne du bruit de fond.

(2) Le facteur de réduction : Pour permettre la comparaison des intensités entre les puces, une transformation des données est nécessaire. Pour cela, l'intensité des oligonucléotides de chaque puce est transformée par un facteur de réduction, qui permet à toutes les puces de l'expérience d'avoir une moyenne des intensités identique. Affymetrix® ne recommande pas d'utiliser une valeur critique pour juger de la bonne ou mauvaise qualité de l'expérience, mais suggère que les facteurs de réduction des puces constituant une même expérience ne soient pas plus grands qu'un facteur 3. Cette analyse peut mettre en évidence d'éventuels problèmes lors de l'étape d'amplification des ARNm.

(3) Le pourcentage de présents : Ce paramètre est défini comme étant le pourcentage de sondes détectés comme « PRESENT », sur une puce à ADN hybridée, en accord avec l'algorithme de détection d'Affymetrix®. En effet, cet algorithme attribue, à chacune des 54 675 sondes, une mesure qualitative de l'expression. Basée sur le test de rang de Wilcoxon des valeurs de PM et de MM, cette mesure qualitative est calculée et permet de déclarer une sonde « PRESENT » si le gène est exprimé, « ABSENT » s'il n'est pas exprimé et « MARGINAL » en cas de doute.

Ainsi pour chaque puce à ADN composant une expérience, MAS5 calcule le pourcentage de « PRESENT » qui devrait être similaire dans un groupe de répétition. Toutefois, ce pourcentage de « PRESENT » varie également considérablement selon les types cellulaires et les conditions expérimentales, mais devrait être compris, dans les meilleurs cas, entre 40 et 50 pour cent (recommandation d'Affymetrix®).

Le contrôle du pourcentage de transcripts déclarés « PRESENT » permet de juger de la reproductibilité de la détection de l'hybridation dans un même groupe de réplication et, de plus, des pourcentages extrêmement faibles pourraient révéler un problème dans la procédure d'amplification et d'hybridation.

(4) Le ratio 3'/5' : La réaction de marquage des transcripts est une étape importante car elle conditionne leur détection. Ainsi, le ratio 3'/5' pour deux contrôles endogènes (GAPDH et ACTB) nous informe sur la performance de cette réaction et indique s'il existe un défaut dans le marquage. Chaque contrôle endogène est représenté par 3 sondes : une pour la partie 5' du gène, une pour

le milieu du gène, une pour la partie 3' du gène. Affymetrix® suggère que ce ratio ne soit pas supérieur à 3.

(5) Les contrôles étalons : Pour apprécier la sensibilité de la détection de l'hybridation par le scanner, 4 contrôles (BioB, BioC, BioD, Cre), sont rajoutés aux échantillons en différentes concentrations (1.25, 5, 25, 100 pM). Les trois copies d'ARNm par cellule du gène BioB (1.25pM) aux 200 copies par cellule du gène Cre (100 pM) permettent de fournir le seuil limite de détection de l'hybridation. Ainsi, l'algorithme de détection d'Affymetrix® devra toujours détecter ces gènes comme étant « PRESENT ». Dans le cas contraire, cela pourra révéler un problème dans l'hybridation.

3. Normalisation des données d'expression

La normalisation réduit l'impact des effets systématiques (non aléatoires) présents dans des données d'expression. Ces effets ne sont pas liés à des variations biologiques, mais à des variations techniques cumulées à chaque étape de la préparation de l'expérience. Cette transformation rend ainsi possible la comparaison des données d'expression entre plusieurs puces d'une même expérience.

La normalisation repose sur le postulat que pour les échantillons d'un même tissu testé dans la même expérience, les intensités devraient être constantes d'une puce à l'autre.

Cette dernière décennie, plusieurs stratégies de normalisation ont été développées. Irizarry en 2003, ont mis à l'épreuve ces algorithmes de normalisation sur une expérience de type Affymetrix® utilisant des données d'expression synthétique (Irizarry, Hobbs et al. 2003). Ils ont conclu que l'algorithme RMA (*Robust Multichip Average*) était le plus performant en

termes de précision, de spécificité et de sensibilité. Pour ces raisons, nous choisissons d'utiliser la méthode RMA pour normaliser nos données d'expression.

a. Méthode RMA

Cette méthode transforme les données d'expression issues de fichier .CEL. Elle estime le bruit de fond, réduit la taille des données numériques, normalise les données et les transforme en \log_2 .

Son algorithme comprend quatre étapes :

(1) Estimation/correction du bruit de fond : pour une sonde donnée, son intensité peut être décomposée en une intensité intrinsèque et une intensité liée au bruit de fond. La distribution des valeurs \log de l'intensité de l'ensemble des oligonucléotides composant la sonde permet d'établir l'intensité du bruit de fond et de corriger la valeur de l'intensité intrinsèque.

(2) La « summarization » : cette étape convertit les intensités de 1.3 millions de sondes en mesures d'expression de 54 675 sondes. Le calcul de ces mesures d'expression n'utilise que les valeurs du PM et n'exploite pas les valeurs de MM, qui correspondent au bruit de fond spécifique à la sonde elle-même. Irizarry justifie le fait de ne pas exploiter les signaux des sondes MM, car ceux-ci ajoutent plus de bruit au signal et n'apportent pas une meilleure qualité à l'estimation des intensités.

(3) La normalisation quantile : (décrite dans le paragraphe suivant).

(4) Transformation des données : transformées en logarithme de base 2.

b. Normalisation quantile

L'étape de normalisation permet d'imposer une même distribution empirique des intensités pour chaque puce à ADN d'une expérience.

Voici son algorithme comme le décrit l'auteur, B.M Bolstadt, dans sa thèse (Bolstad 1998) :

Soit M une matrice de données d'expression de dimension p par n , représentant l'expérience,
 p étant le nombre de sondes,
 n étant le nombre de puce à ADN de l'expérimentation
(chaque colonne de la matrice représente les données issues d'une puce).
Trier en ordre croissant chaque colonne de M séparément.
 $M \rightarrow Mt$ (Mt étant la matrice M triée)
Pour chaque ligne de Mt
Calculer la valeur moyenne μ .
Remplacer tous les valeurs de la ligne par μ . $Mt \rightarrow M\mu$
($M\mu$ étant la matrice dont les valeurs de chaque ligne sont remplacées par la moyenne correspondante)
Réarranger chaque colonne de $M\mu$ dans le même ordonnancement de chaque colonne de M .

F. Outils bioinformatiques

Dans les années 2000, les avancées des biotechnologies (séquençage, plateforme d'expression transcriptomique,...) ont permis de générer des quantités importantes de données de nature biologique. Pour développer des stratégies d'analyse adéquates, le projet BioConductor associa le savoir-faire des statisticiens, mathématiciens, biologistes et développeurs pour créer BioConductor, un ensemble d'outils informatiques, algorithmiques et logiciels adaptés à des problèmes bioinformatiques intégrant les méta-données (séquences génomiques, données d'expression, interaction moléculaires,...) (Gentleman, Carey et al. 2004). Chaque application de ce projet se présente sous la forme d'une librairie, comprenant la mise en oeuvre des algorithmes sous la forme de fonctions orientées-objet.

Ce logiciel libre d'accès est basé sur le langage de programmation statistique R, mais contient également des fonctions en langage C.

De plus, de nombreuses librairies, ne faisant pas partie du projet d'origine ont été développées pour répondre à des problématiques orientées sur l'analyse de données d'expression transcriptomique.

Pour ces raisons, R et BioConductor sont devenus des logiciels indispensables pour l'analyse des données d'expression de puce à ADN, mais également dans le domaine de la spectrométrie de masse, de la gestion des annotations génomiques, ainsi que d'autres domaines comme l'écologie, l'astronomie et l'économétrie.

L'intégralité des analyses présentées dans ce mémoire est codée en langage R et l'annexe 2 présente les librairies utilisées. Les annexes 3 à 10 présentent les différentes parties de code R utilisées pour produire nos résultats.

G. Détection des gènes différentiellement modulés dans le protocole 1

Dans notre premier protocole, l'étude de la modulation des gènes dans les conditions expérimentales NuLi-1 et CuFi-1 est réalisée en triplicat biologique sur des cellules épithéliales pulmonaires en culture pour chacune des conditions. L'usage d'un test statistique adéquat nous permet d'identifier des gènes différentiellement exprimés entre les deux conditions.

1. Test t modifié.

Pour analyser des données d'expression regroupées en deux conditions expérimentales CuFi-1 et NuLi-1 en triplicata biologiques, nous utilisons une version modifiée du test t de Student où la variance est estimée grâce à une méthode basée sur le théorème de Bayes (Baldi et Long 2001) combinée à une approche linéaire (Smyth 2004). Cette méthode statistique est mise en oeuvre dans la librairie LIMMA (*Linear Models for MicroArray data*) (R. Gentleman 2005).

2. Valeur P et correction pour les comparaisons multiples.

Pour décrire la signification d'une modulation observée d'un transcript donné entre deux conditions expérimentales, la statistique issue d'un test peut être convertie en utilisant sa distribution de référence. Ainsi la distribution de Student à n_1+n_2-2 degrés de liberté (n_1 et n_2 étant le nombre de répétitions dans chaque condition) permet de transformer la variable t calculée en une valeur P qui représente la probabilité d'observer de telles données si l'hypothèse H_0 (il n'y a pas de modulation entre les deux conditions expérimentales) est vraie. En d'autres mots, la valeur P représente la probabilité que la différence de moyenne observée pour une sonde donnée

soit due au hasard. Nous choisissons notre seuil de signification à 0.05, donc toutes les sondes présentant des valeurs P inférieures à ce seuil sont jugées significativement modulées entre les deux conditions.

Les comparaisons multiples sont caractéristiques de l'analyse de puce, car nous testons l'hypothèse H_0 (hypothèse nulle) pour chaque sonde de notre expérience, laquelle est composée de 54 675 sondes. Chaque test s'accompagne d'une prise de risque (niveau de confiance, seuil de signification), α , bien souvent associée à la valeur 0.05. Concrètement, cela signifie que chaque fois que l'on teste H_0 , nous avons 5 % de chance de rejeter l'hypothèse nulle alors qu'il faudrait l'accepter (Pounds 2006). Puisqu'une expérience de puce est composée d'un nombre important de sondes (dans notre cas, 54 675), le test statistique est répété autant de fois. Par conséquent, il est nécessaire d'estimer la quantité de faux-positifs générée par ces tests consécutifs, car elle permet de contrôler le nombre de faux positifs.

Dans notre analyse, nous utilisons la méthode développée par Benjamin et Hochberg en 1995 (Benjamini 1995) : le *False Discovery Rate*.

Le *False Discovery Rate* (FDR) permet une estimation du taux de faux-positifs dans une sélection de sondes. Cette méthode permet de calculer pour chaque sonde une valeur P ajustée. Il est important de bien distinguer la valeur P et la valeur P ajustée. La valeur P est la probabilité des données sous l'hypothèse H_0 pour un seul test. Elle est associée à la qualification de faux positif. Tandis la valeur P ajustée est le seuil de faux-positifs toléré dans une sélection de sondes d'intérêt. Elle quantifie le nombre des faux-positifs espéré dans une liste et n'informe en rien sur la répartition des faux-positifs dans cette liste.

L'algorithme de calcul du FDR peut se résumer de la façon suivante :

l , une liste de n sondes présentes sur la puce à ADN
 pv , la valeur P individuelle
 fdr , le niveau espéré de faux-positifs.
Pour chaque sonde n_i de l ,
 Calculer pv_i .
Former la liste $l.ordonnée$, classer les sondes de l dans
l'ordre croissant de leur pv_i ,
A partir de $l.ordonnée$, extraire une liste $l.extrait$ de
2 à n sondes.
Pour chaque liste $l.extrait$ formée de taille m
 Calculer le $fdr = (pv_m * n) / m$

Ainsi pour une liste « *l.extrait* » de 100 sondes, présentant une valeur P inférieure à 0.00001 implique que cette liste a un FDR de $(0.00001 * 54\ 675) / 100$, ceci correspondant à un taux de faux-positifs de 0.0054.

H. Détection des gènes différentiellement modulés dans l'objectif 2

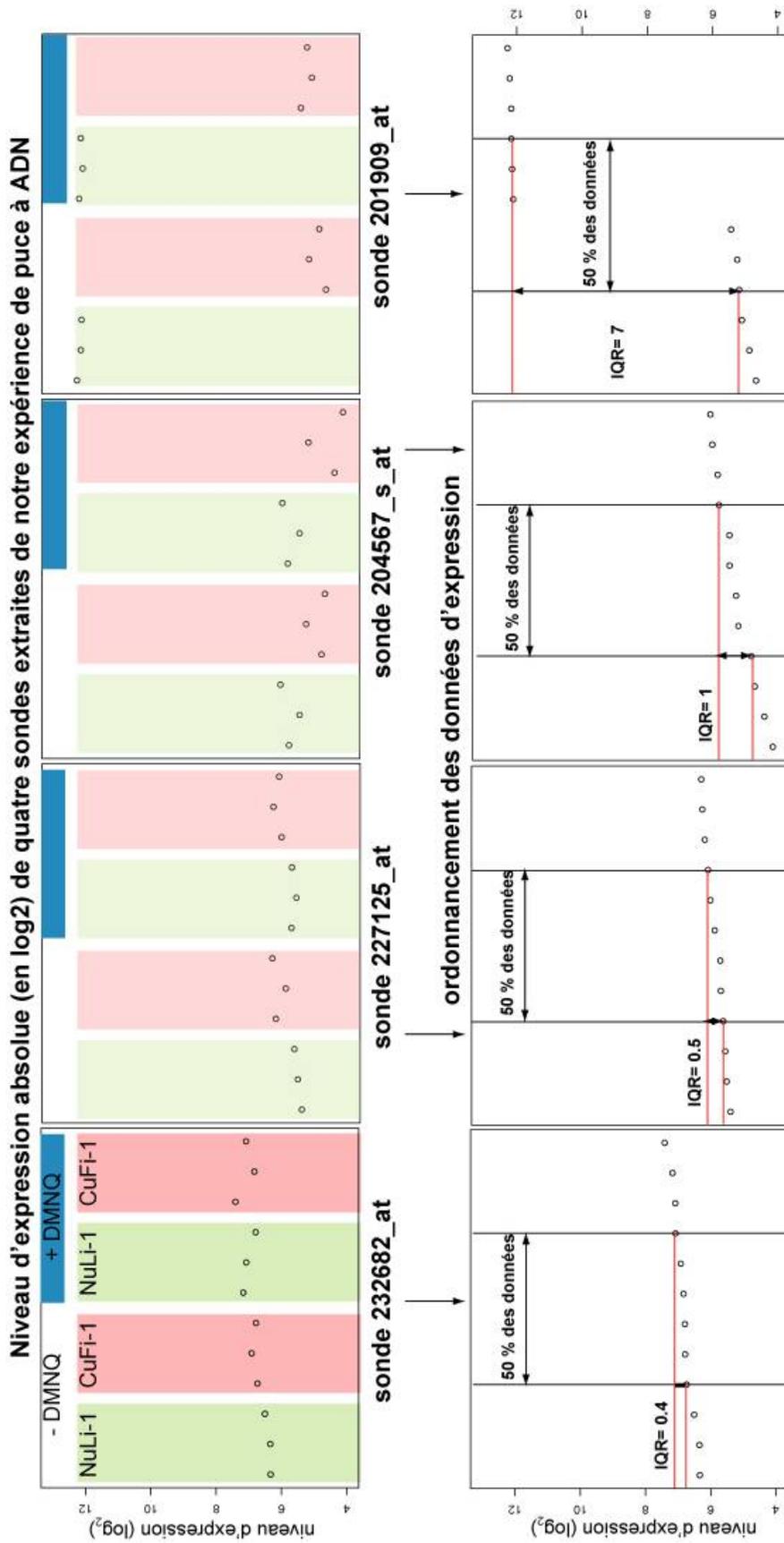
Dans ce protocole nous exploitons les deux facteurs (mutation, oxydant) à deux niveaux (absence/présence) de notre expérience. L'analyse de la variance des données d'expression de chaque sonde, nous permet d'interpréter la contribution des facteurs sur la variance observée de chacune d'elles.

1. Filtrage des données

L'analyse de variance utilisée dans notre approche exploite le calcul matriciel dont le temps de calcul peut être important, car il est fonction de la quantité de données soumises. Sur l'ensemble des 54 675 sondes présentes sur la plateforme HG-U133 Plus 2.0, très peu sont modulées et

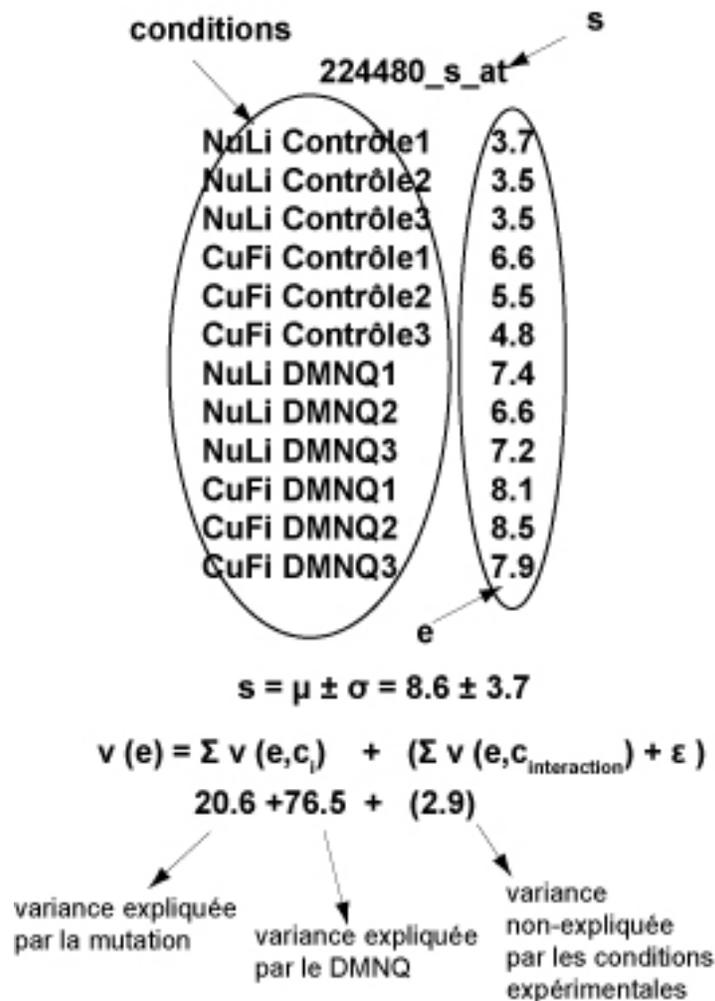
présentent une variance très faible. Par conséquent, les retirer de notre analyse n'affecte pas nos résultats et seules les sondes ayant une variance supérieure à un seuil fixé seront analysées. Pour filtrer nos données sur la base de leur variance, l'IQR (*Inter-Quartile Range*) est utilisé. L'IQR est une mesure de dispersion statistique autour d'une valeur centrale, la médiane. Cette mesure correspond à la différence entre le troisième et le premier quartile des données d'expression. En d'autres mots, l'IQR indique l'écart de la mesure de 50 % des données autour de la médiane et présente l'avantage de ne pas être affecté par des valeurs extrêmes. La figure 9 (page 59) illustre le calcul de l'IQR pour quatre sondes représentatives extraites de notre expérience.

Pour pallier aux biais liés à la dépendance du seuil choisi lors de notre analyse, l'interprétation des résultats est réalisée à partir de plusieurs seuils de filtrage dans un intervalle d'IQR compris entre 0.5 et 1.



2. Variance des sondes

Par définition, la variance représente la variation des mesures d'expression autour de la moyenne et fournit une valeur numérique, qui est peu informative par elle-même sur la source de la variance. Dans le cadre d'une analyse multivariable (pour les designs expérimentaux à plusieurs facteurs), cette variance peut être décomposée et la contribution de chaque facteur à la variance observée peut être estimée.



Considérons une sonde s , et e , son vecteur des mesures d'expression dans différentes conditions expérimentales c .

On peut modéliser simplement la mesure de l'expression de s

$$s = \mu \pm \sigma$$

avec μ , la moyenne des mesures d'expression de e et σ , l'écart-type des mesures d'expression de e .

La variance étant le carré de l'écart-type, on peut aussi écrire

$$s = \mu \pm v(e)$$

avec $v(e)$, la variance des mesures d'expression e .

Selon i , le nombre de conditions expérimentales c composant le vecteur e , nous pouvons décomposer la variance en plusieurs parties, chacune induite par chacun des facteurs composant c_i et les interactions possibles entre chaque facteurs ($C_{interaction}$).

$$v(e) = \sum v(e, c_i) + \sum v(e, C_{interaction}) + \varepsilon$$

$v(e)$, la variance totale du vecteur d'expression e .

$v(e, c_i)$, la variance présente dans le vecteur d'expression liée à la condition i .

$v(e, C_{interaction})$, la variance présente dans le vecteur d'expression due à l'interaction entre les différents facteurs expérimentaux.

ε , la variance non expliquée par le modèle.

Pour analyser la variance présente dans nos données d'expression, deux outils statistiques sont utilisés :

(1) le **test ANOVA multivariable** nous permet de statuer sur le changement significatif ou non de la moyenne entre les différentes conditions expérimentales.

(2) **l'analyse canonique de redondance**, méthode d'ordination en espace réduit sous contrainte modélisant les relations entre les variables réponse (expression des gènes) et les facteurs de l'expérience (la mutation et l'oxydant).

3. L'ANOVA

L'analyse de variance (ANOVA) (Fisher 1954) est un outil statistique approprié pour étudier la différence de moyenne de plusieurs groupes expérimentaux lorsqu'il y a plusieurs observations (répétition) au sein de chaque groupe. Le rapport des variances inter-groupe et intra-groupe détermine la valeur de la statistique F . Pour tester l'hypothèse H_0 (les différences de moyenne observées sont purement aléatoires), la statistique F est comparée à une distribution adéquate afin d'estimer la signification des différences observées entre les moyennes des groupes.

L'avantage de l'ANOVA est qu'elle permet de comparer plusieurs groupes en une seule statistique et de se prononcer sur la signification de l'interaction entre facteurs étudiés si l'analyse comporte des facteurs croisés et le design expérimental utilise des répétitions. Dans notre analyse, nous utiliserons des tests par permutations pour le calcul de l'analyse au lieu de nous référer à la distribution de F . Le test anova multivariable sera réalisé à l'aide des fonctions « rda » et « anova.cca » de la librairie `vegan` (Oksanen et al 2009) en langage R.

4. L'analyse canonique de redondance

L'analyse canonique de redondance (ACR) est une généralisation de la régression multiple pour des données-réponse multivariées. Nous utiliserons l'ACR dans le contexte particulier où le tableau des variables explicatives contient les facteurs expérimentaux ainsi que l'interaction. Ce type d'application, déjà connu pour la régression multiple, a été proposé pour l'ACR par Legendre et Anderson (Legendre 1999).

Une analyse canonique transforme une information sous une forme plus simple et plus compréhensible sans perte de généralité. L'analyse de

redondance (Rao 1964) est une analyse canonique simultanée de deux matrices de données, une matrice Y contenant les variables-réponses (c'est-à-dire les niveaux d'expression des sondes) et une matrice X décrivant les variables explicatives (c'est-à-dire les facteurs expérimentaux et, si on le désire, leurs interactions). À partir de ces régressions multiples, est formée une matrice Z , contenant les valeurs d'expression ajustées, résultant de la régression. Le test de signification se fait par le calcul d'une statistique $R^2 = (\text{variance totale de } Z) / (\text{variance totale de } Y)$, puis par transformation du R^2 en statistique F laquelle est testée par permutation. Si le résultat du test est significatif, on peut faire subir une analyse en composante principale à la matrice Z , d'où résultent des vecteurs propres et valeurs propres qui permettent de calculer des axes d'ordination des données en quelques dimensions, produisant un graphique informatif des relations entre les variables-réponse (l'expression des gènes) et les facteurs expérimentaux. Cette méthode est décrite en détail dans Legendre & Legendre (Legendre et Legendre 1998). Grâce à l'analyse de redondance, les données d'expression sont transformées en deux axes d'ordination donnant la position de chaque sonde sur deux axes correspondant aux deux directions de plus grande variance de la relation entre le tableau-réponse et les facteurs. Nous utiliserons ces deux variables comme l'abscisse et l'ordonnée d'un graphique d'ordination. La fonction d'analyse de redondance utilisée pour les graphiques fait partie de la librairie « `rdaTest` » (Legendre 2009) du langage statistique R.

5. La classification des profils d'expression

Exploitant l'hypothèse que des groupes participant à la même fonction biologique partageraient des profils d'expression similaires, nous groupons les

sondes sélectionnées en fonction de leurs profils d'expression basée sur les valeurs standardisées des 12 données d'hybridation.

Pour produire cette classification, nous choisissons une méthode de groupement non supervisé (produisant des regroupements à priori, sans connaissance au préalable des groupes) appelée *fuzzy c-means*.

Le *fuzzy c-means* (FCM) est un algorithme de classification non supervisé flou (Dunn 1973). Issu d'un algorithme plus simple, le *c-means*, il intègre la notion d'ensemble flou dans la définition des classes. Cela signifie, dans notre cas, que chaque sonde pourra appartenir à une classe avec un certain degré d'appartenance et ainsi être associée à plusieurs classes (Bezdek 1981).

L'algorithme FCM est disponible dans la librairie « Mfuzz » (Futschik 2009).

I. L'interprétation biologique de données transcriptomiques

Après avoir interprété statistiquement nos données d'expression et sélectionné une liste de transcripts d'intérêt, il est important de leur donner une interprétation biologique laquelle pourra être reliée à la physiopathologie de la maladie. Pour cela, après avoir identifié les sondes modulées par des noms de gène, nous les associons à des fonctions biologiques à l'aide d'annotations génomiques de différentes sources.

1. Annotations génomiques

Les 54 675 sondes présentes sur la puce HG-U133 Plus 2.0 sont associées à une grande quantité d'annotations génomiques (nom du gène, position chromosomique, voie métabolique, fonction biologique, identifiant divers...). Étant donné que plusieurs sondes peuvent hybrider le même transcript, plusieurs identifiants présentent les mêmes annotations génomiques. Par exemple, le transcript codant pour la thrombospondine 1 (THBS1) est hybridé

par sept sondes, identifiées 201107_s_at, 201108_s_at, 201109_s_at, 201110_s_at, 215775_at, 235086_at, 239336_at. Dans notre analyse, l'ensemble des résultats sont présentés en utilisant uniquement le nom des gènes et dans le cas de sonde multiples, nous utilisons la sonde qui présente la signification la plus élevée.

Toutefois, afin d'optimiser notre interprétation biologique, il est nécessaire d'avoir le contrôle sur la qualité des annotations en croisant différentes sources de données. Notre source principale de données est la base d'annotation d'Affymetrix®, NetAffx (Liu, Loraine et al. 2003) qui rassemble 61 annotations génomiques pour chaque sonde et utilise l'identifiant Unigene comme référence. Bien que les services d'Affymetrix® s'emploient à entretenir le contenu de NetAffx, certaines discordances entre les bases de données apparaissent. Ainsi pour un même identifiant, le nom du gène et sa fonction biologique peuvent varier (Perez-Iratxeta et Andrade 2005; Stalteri et Harrison 2007). Pour cette raison, nous soumettons les annotations d'intérêt fournies par Affymetrix® à un contrôle en utilisant la librairie biomaRt de Bioconductor (Nie, Neerinx et al. 2009), qui autorise un accès direct à la base de données d'Ensembl (www.ensembl.org). Ainsi, pour chaque transcript sélectionné dans nos analyses, nous confirmons ses annotations d'Affymetrix® par celles d'Ensembl et validons son existence protéique en interrogeant la base de données Resnet (www.arianegenomics.com).

2. Analyse ontologique et son approche statistique

Le *Gene Ontology* (GO) Consortium (www.geneontology.org) a développé un vocabulaire structuré, contrôlé, repartit en trois ontologies correspondant aux processus biologiques, aux fonctions moléculaires et aux localisations cellulaires pour décrire le rôle des gènes et produits de gènes de l'ensemble des eucaryotes.

Chaque gène est défini par un processus biologique qui reflète l'objectif biologique auquel il participe. Ce processus biologique peut être décomposé en une ou plusieurs fonctions moléculaires, décrivant l'action biomoléculaire du produit de l'expression du gène. Finalement, la localisation cellulaire est l'endroit où la forme active du gène agit. Ainsi un gène peut être associé à un ou plusieurs termes GO. Par exemple, la thrombospondine 1 est associée, entre autres, au processus de fixation des ions calciques, processus défini par le GO Consortium comme étant le terme GO « 0005509 » (*calcium ion binding*), et également au terme GO « 0002544 » (*chronic inflammatory response*).

Les trois ontologies GO sont structurées sous forme d'un graphe orienté acyclique (DAG). Chaque nœud de ce graphe est un terme GO, et peut être un « enfant » d'un nœud et être « parent » de plusieurs nœuds. En effet, la majorité des gènes sont pléiotropes, c'est-à-dire qu'un gène peut avoir plusieurs produits et les produits d'un gène possèdent une ou plusieurs fonctions biochimiques. Sachant que le terme « enfant » est toujours plus spécifique que le ou les termes « parents », plus on descend dans l'arborescence du graphe, plus les GO termes sont précis. Un exemple de DAG (et son utilité) est présenté dans la figure 10 (page 69).

À partir d'une liste de sondes d'intérêt, annotées par des GO termes, il est intéressant de connaître la ou les fonctions biologiques, moléculaires, compartimentations cellulaires, qui apparaissent plus qu'attendu dans cette liste. L'analyse ontologique consiste à une comparaison de la distribution de GO termes présents dans une liste sélectionnée de sondes avec les GO termes de la liste d'origine (sonde sélectionnées et non sélectionnées). De cette façon, nous mettons en évidence des fonctions biologiques, moléculaires et des localisations sur- ou sous-représentées dans notre sélection et ainsi

nous transformons une liste de sondes difficilement interprétable d'un point de vue biologique en information biologique d'intérêt.

Les outils d'analyse ontologique sont implémentés dans la librairie GOstats (Gentleman. 2007).

L'analyse ontologique s'accompagne du calcul d'une statistique qui permet de décider de la pertinence de la surreprésentation d'un terme GO dans une liste d'intérêt.

Nous présentons cette approche statistique par analogie au problème du tirage de X balles dans une urne contenant N balles de deux couleurs : M balles rouges et $N-M$ balles vertes (n sonde au total dans l'analyse, soit m sonde associées à la fonction biologique r et $n-m$ sonde non associées à la fonction biologique r)

Si nous choisissons aléatoirement K balles dans l'urne, X seront de couleur rouge et $K-X$ seront de couleur verte (sélectionnons k sonde dans notre analyse sur la base de leur modulation, x sonde appartiennent à la fonction biologique r et $k-x$ sonde appartiennent à la fonction biologique r). Le problème se résume en une question simple : quelle est la probabilité de sélectionner k balles dans l'urne et que x balles soit rouge ? Par analogie, quelle est la probabilité de sélectionner x sondes associées à la fonction biologique r , connaissant n, m, k ?

Il est connu que cette probabilité suit une distribution hypergéométrique ayant pour paramètre n, m et k . Dès lors, nous pouvons calculer la probabilité d'avoir x ou moins sonde sélectionné appartenant à la catégorie r .

Une méthode alternative basée sur le test χ^2 (chi carré) permet également d'évaluer l'égalité des proportions d'un terme GO : le test de Pearson ou du χ^2 . A partir d'une table de contingence, nous pouvons calculer O , la fréquence Observées et E la fréquence attendue (E pour *Expected*). Cette table de

contingence de 2 x 2, dénombre le nombre total de sonde de l'expérience, divisée en deux groupes (celles qui sont impliquées dans la fonction biologique f et celles qui ne sont pas impliquées), et le nombre de sonde sélectionnée, divisée également en deux groupes (celles impliquées dans la fonction biologique f et celles qui ne sont pas impliquées).

A partir de ces quatre quantités transformées en fréquence, nous calculons la

statistique $\chi^2 = \frac{\sum(O - E)^2}{O}$, qui, comparée à la distribution de Fisher à un degré

de liberté, nous indique si l'égalité des proportions d'un GO terme donné. Dans le cas d'inégalité dans les proportions, la Probabilité, issue de la statistique, nous informe jusqu'à quel point l'inégalité est significative.

Dans le cas où les fréquences espérées sont petites (moins de 4), le résultat du test de Pearson peut-être biaisée car ces petites valeurs augmentent artificiellement la valeur du test statistique. Cette valeur ne suit pas la distribution de Fisher. Pour pallier à ce problème, nous pouvons utiliser le test exact de Fisher qui est basé sur le dénombrement de chacune des quantités présentes dans la table de contingence.

L'ensemble des probabilités calculées dans les analyses ontologiques produit dans nos travaux sont fournis par une distribution hypergéométrique de la fonction `hyperGTest()` de la librairie `Gostats` (Gentleman. 2007).

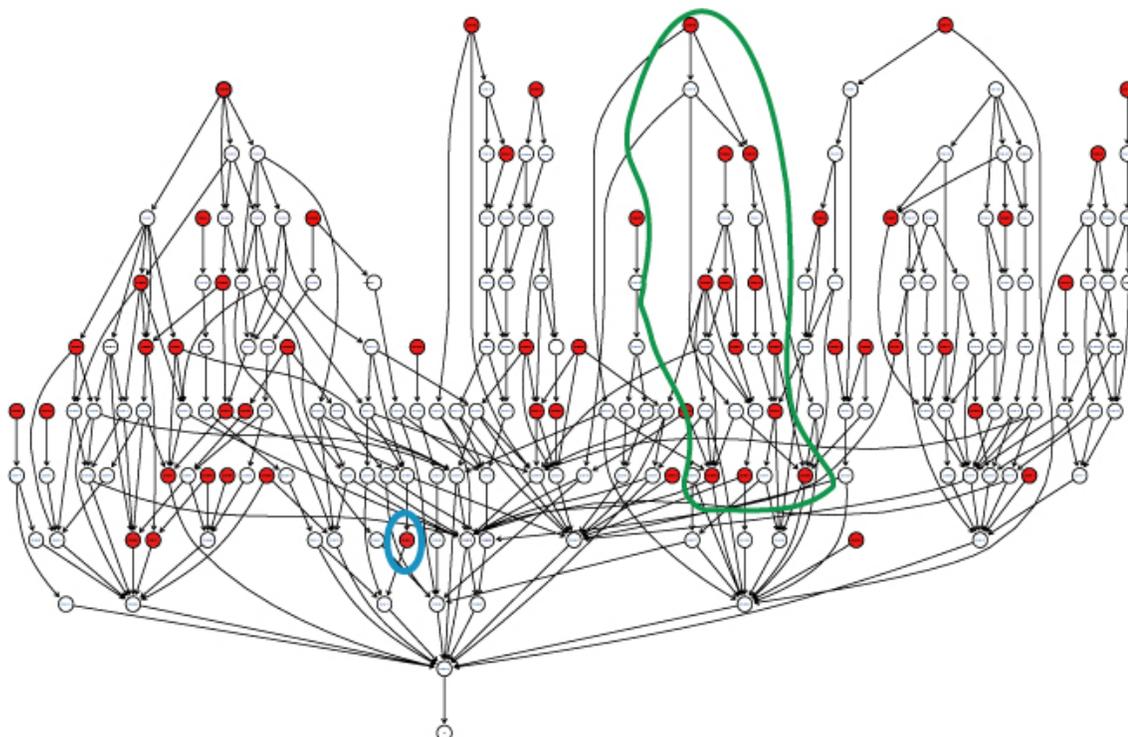


Figure 10 : Exemple de DAG issu de l'analyse ontologique de nos données transcriptomiques.

Ce graphe est réalisé à partir d'un ensemble de transcripts différentiellement exprimés dans notre analyse. L'analyse ontologique identifie 63 processus biologiques surreprésentés dans l'ensemble de transcripts d'intérêt. Ceux-ci sont identifiés par des nœuds rouges du graphe. Les autres nœuds (en blanc) sont des processus (non surreprésentés) qui relient les nœuds rouges.

L'intérêt de cette présentation est de se rendre compte des relations qui existent entre les processus et permet de souligner la surreprésentation des certaines branches du graphe qui peut retenir notre attention. En d'autres mots, porter son attention sur un groupe de fonctions surreprésentées et reliées dans le graphe (l'exemple est entouré en vert) pourrait être plus intéressant que de porter son attention sur un nœud rouge isolé (l'exemple est entouré en bleu). Toutefois, à notre connaissance, cette considération subjective n'est pas intégrée dans les logiciels d'analyse, qui se restreignent souvent à une liste de fonctions surreprésentées.

Résultats

A. Contrôle qualité des données d'expression

Le contrôle qualité nous montre que les données d'expression issues de l'expérience de puce à ADN à quatre conditions expérimentales sont de bonne qualité et pleinement exploitable. L'ensemble des graphiques et tableau de ce contrôle qualité est présenté à la figure 12 (page 72)

La partie A présente le résultat de l'hybridation de l'ensemble des transcripts polyadénylés sur la plateforme d'Affymetrix®. Sous la forme d'une image, nous présentons uniquement l'hybridation pour un représentant de chaque condition expérimentale car l'analyse visuelle de ces fichiers images n'a pas révélé d'artefacts liés à l'hybridation. Nous observons une répartition uniforme des intensités sur l'ensemble de la puce à ADN, indiquant l'absence de biais spatial. À titre de comparaison, des exemples de ce type de biais sont présentés à la figure 7 (page 47).

La partie B décrit la distribution de l'intensité des sondes dans chaque puce pour chacune des conditions expérimentales. Le graphique de gauche présente cette distribution sous la forme de *boxplot* et celui de droite sous la forme d'un histogramme de fréquence. Ces deux représentations nous montrent une distribution similaire à travers l'ensemble de l'expérience et également à travers les groupes de répétition. Ces résultats suggèrent qu'il n'y a pas de défaut dans la distribution des intensités des sondes.

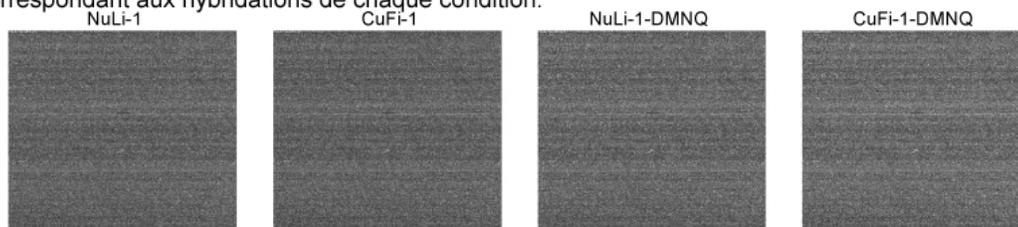
La partie C présente les régressions linéaires de la moyenne des intensités pour l'ensemble des sondes ordonnées selon leur positionnement physique sur le transcript. La linéarité de ces régressions souligne un comportement similaire de l'ensemble des sondes à travers l'expérience. Celles-ci témoignent de la bonne qualité des ARNm et d'une minime dégradation de leur partie 3' (correspondant aux régions d'hybridation).

La partie D présente les intensités des contrôles étalons au travers de chacune des puces. Leurs niveaux d'expression aux concentrations fixées sont

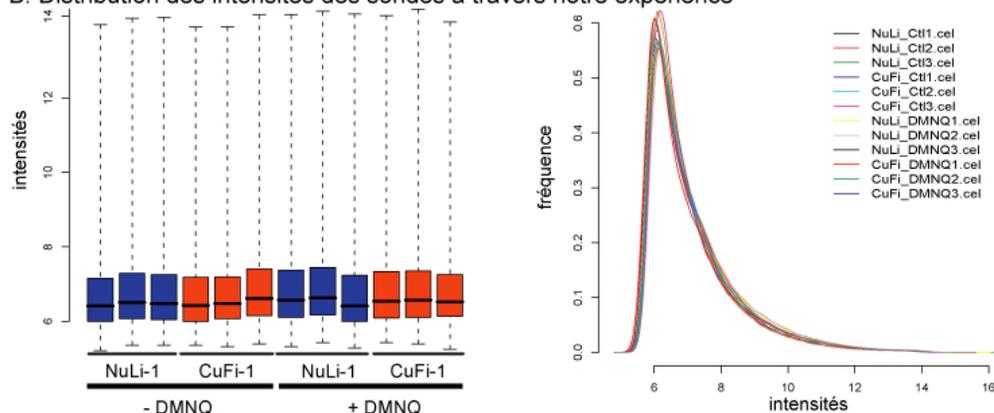
similaires dans l'ensemble des puces à ADN. Ainsi, une tendance linéaire est observée pour l'ensemble de nos échantillons indiquant la bonne performance dans l'étape d'hybridation.

Le tableau de la partie E présente les résultats pour chacune des puces des paramètres du contrôle qualité d'Affymetrix®. Ceux-ci sont en accord avec les recommandations du constructeur. Le bruit de fond et le pourcentage de « PRESENT » sont constants dans l'ensemble de l'expérience, les facteurs de réduction ne présentent pas d'écart important entre eux, les ratios 3'/5' de la GAPDH et ACTB sont inférieurs à 3. L'ensemble de ces paramètres montrent l'absence de biais dans notre expérience.

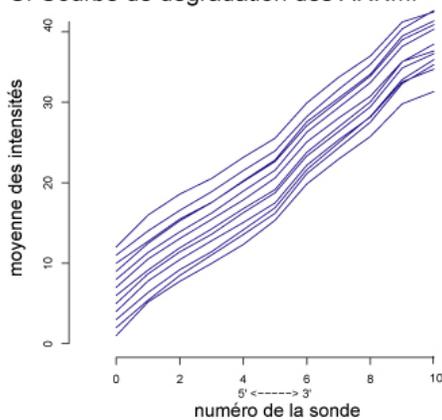
A. Inspection visuelle . Nous présentons une image pour chacune des conditions expérimentales correspondant aux hybridations de chaque condition.



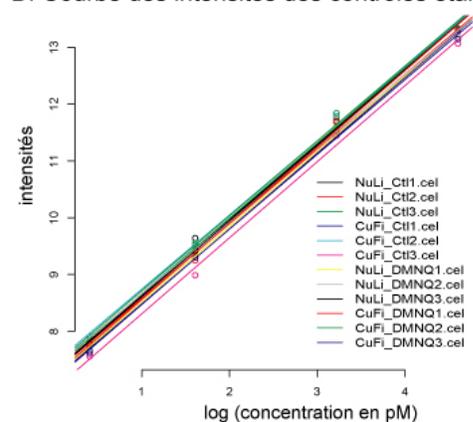
B. Distribution des intensités des sondes à travers notre expérience



C. Courbe de dégradation des ARNm.



D. Courbe des intensités des contrôles étalons.



E. Critères de contrôle qualité d'Affymetrix.

puce	bruit de fond	facteur de réduction	pourcentage de « PRESENT »	ratio 3'/5' GAPDH	ratio 3'/5' B-actine
NuLi_Ct11	44,07	1,51	41,09	1,06	1,44
NuLi_Ct12	48,16	1,19	42,73	1,04	1,42
NuLi_Ct13	47,4	1,22	42,51	1,09	1,45
CuFi_Ct11	46,11	1,49	40,79	1,09	1,39
CuFi_Ct12	48,19	1,58	40,4	1,05	1,52
CuFi_Ct13	50,13	1,02	43,73	1,09	1,3
NuLi_DMNQ1	47,63	1,16	42,2	1,01	1,29
NuLi_DMNQ2	49,98	0,99	43,31	1,14	1,4
NuLi_DMNQ3	45,77	1,17	42,8	1,07	1,46
CuFi_DMNQ1	48,31	1,14	42,19	1,1	1,38
CuFi_DMNQ2	48,95	1,14	42,4	1,1	1,54
CuFi_DMNQ3	49,64	1,39	40,84	1,02	1,53

Figure 12: Résultat du contrôle qualité des 12 hybridations.

B. Normalisation, normalité et homogénéité de la variance

Les données d'expression issues des signaux d'hybridation nécessitent un traitement pour permettre une analyse statistique adéquate. La figure 13 (page 73) présente l'effet de l'étape de la normalisation sur la distribution des données d'expression. La partie A présente la variabilité des données d'expression sous forme de *boxplot* pour les triplicatas de chaque condition expérimentale. La normalisation basée sur une transformation quantile rend identique les distributions des données d'expression présentées dans la partie B.

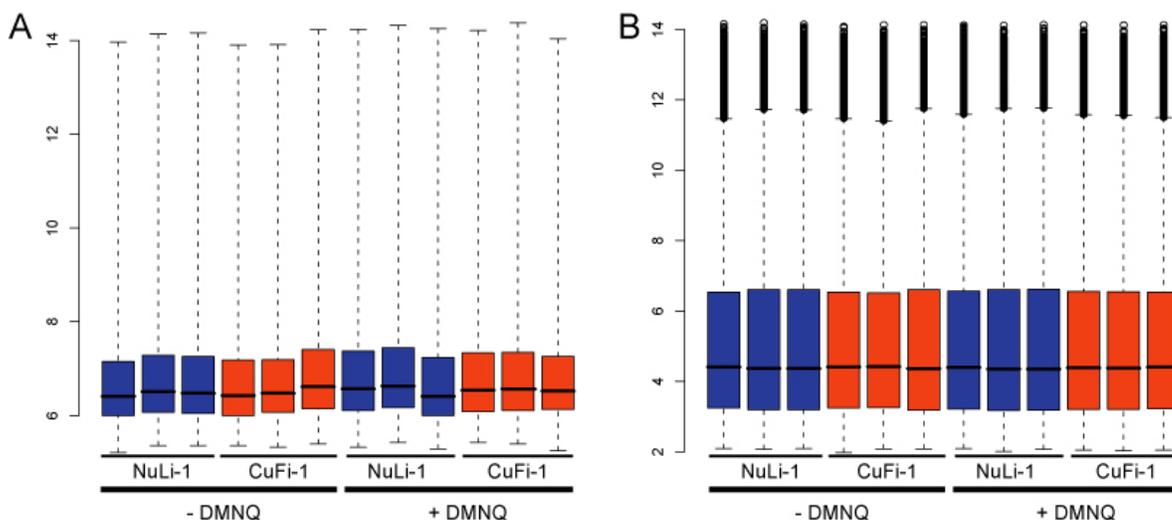


Figure 13: Normalisation des données d'expression. Représentation sous la forme de boxplot des intensités des données d'expression avant normalisation (A) et après normalisation (B).

L'étape de normalisation permet une transformation globale des données. Toutefois, il est important, pour chaque sonde considérée dans notre expérience, d'estimer sa normalité et l'homogénéité de sa variance. Ces deux critères sont des suppositions nécessaires à la bonne application de méthodes paramétriques telles que le test t et l'ANOVA, tests couramment utilisés dans l'analyse de données transcriptomiques. Ainsi, nous testons la normalité par le test de Shapiro et l'homogénéité de la variance par le test de Kligner-Killwen. Ces analyses montrent que 50629 transcripts (92.6 % des transcripts totaux

testés) ont des niveaux d'expression distribués normalement et 50776 transcripts (93.7 %) ont une variance consistante. Toutefois, 1.8 % d'entre eux échouent aux deux tests en ne présentant pas une distribution normale de leurs données et ont une variance hétérogène entre les quatre groupes expérimentaux.

Le test de normalité des données d'expression et le test d'homogénéité de la variance entre les conditions sont deux critères importants qui permettent de se prononcer sur le choix de la mesure à considérer pour filtrer ou sélectionner les transcripts. De plus, 287 transcripts présentent des valeurs extrêmes (une valeur extrême est définie comme ayant une valeur d'expression 1.5 fois la valeur de l'IQR), caractéristique qui pourrait biaiser l'interprétation d'une mesure paramétrique telle que la variance ou la moyenne.

Globalement, une quantité non négligeable de transcripts testés ne respecte pas les suppositions nécessaires pour l'emploi de tests paramétriques. Ceci justifie l'utilisation de test non paramétrique et également la prudence dans l'interprétation des résultats produits par des méthodes paramétriques.

C. Résultats de l'analyse inflammatoire

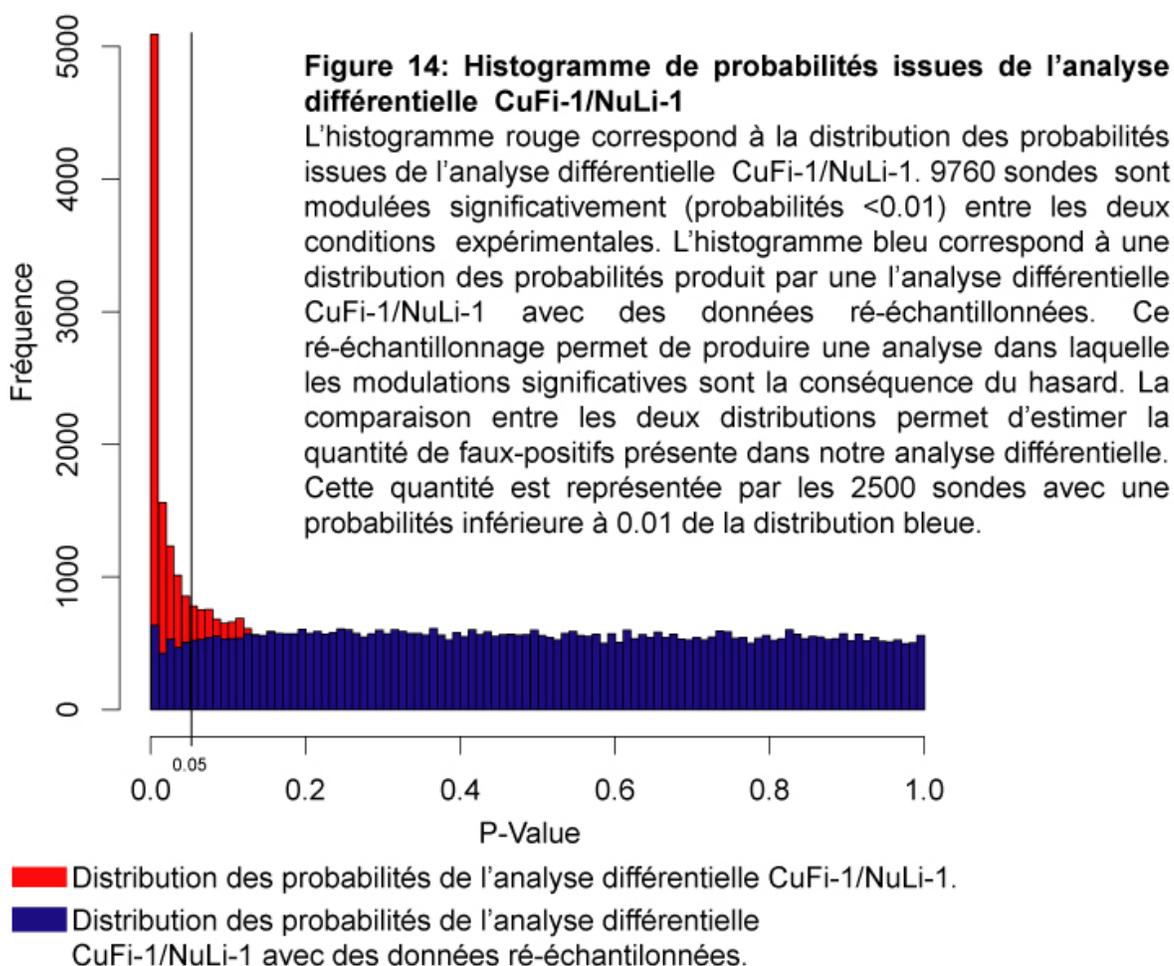
1. Notre analyse statistique isole 9 760 transcripts différentiellement exprimés entre le type CuFi-1 et NuLi-1

A partir des 54 675 sondes présentes sur la puce HG-U133 Plus 2.0 et en nous basant sur un test t modifié intégrant un modèle linéaire (Smyth 2005), nous sélectionnons 9 760 sondes (17.8 % de la totalité) avec une expression différentielle significative entre les lignées CuFi-1 et NuLi-1 (probabilité < 0.05).

Pour estimer la proportion de sondes faussement identifiées comme étant différentiellement exprimés dans notre expérience, nous produisons une analyse différentielle sur nos données d'expression ré-échantillonnées. Pour être plus explicite, ce ré-échantillonnage de nos données élimine de notre

expérience les expressions différentielles associées à nos conditions expérimentales. Par conséquent, toute expression différentielle détectée est considérée être le fruit du hasard et cette méthode vise à quantifier celles-ci. De cette façon, nous estimons le nombre de sondes faux-positives présentes dans nos données d'expression.

Cette méthode de ré-échantillonnage est illustrée à la figure 14 (page 76). La comparaison des histogrammes des Probabilités issues des analyses différentielles sur les données d'origine et les données ré-échantillonnées permet d'estimer à environ 2 500 sondes faussement positives présents dans la sélection. Ceci est confirmé par l'estimation de 27 % de faux-positifs produit par LIMMA.



Filtrées en fonction de leur ratio du niveau d'expression entre les deux types cellulaires, nous identifions 4 397 sondes sur-modulés et 5 363 sondes sous-modulées dans la condition CuFi-1 par rapport à notre condition contrôle. Ces résultats sont présentés à la figure 15 (page 77), dans des représentations communément utilisées pour visualiser les données d'expression transcriptomiques : le *MA-plot* et le *volcanot-plot*.

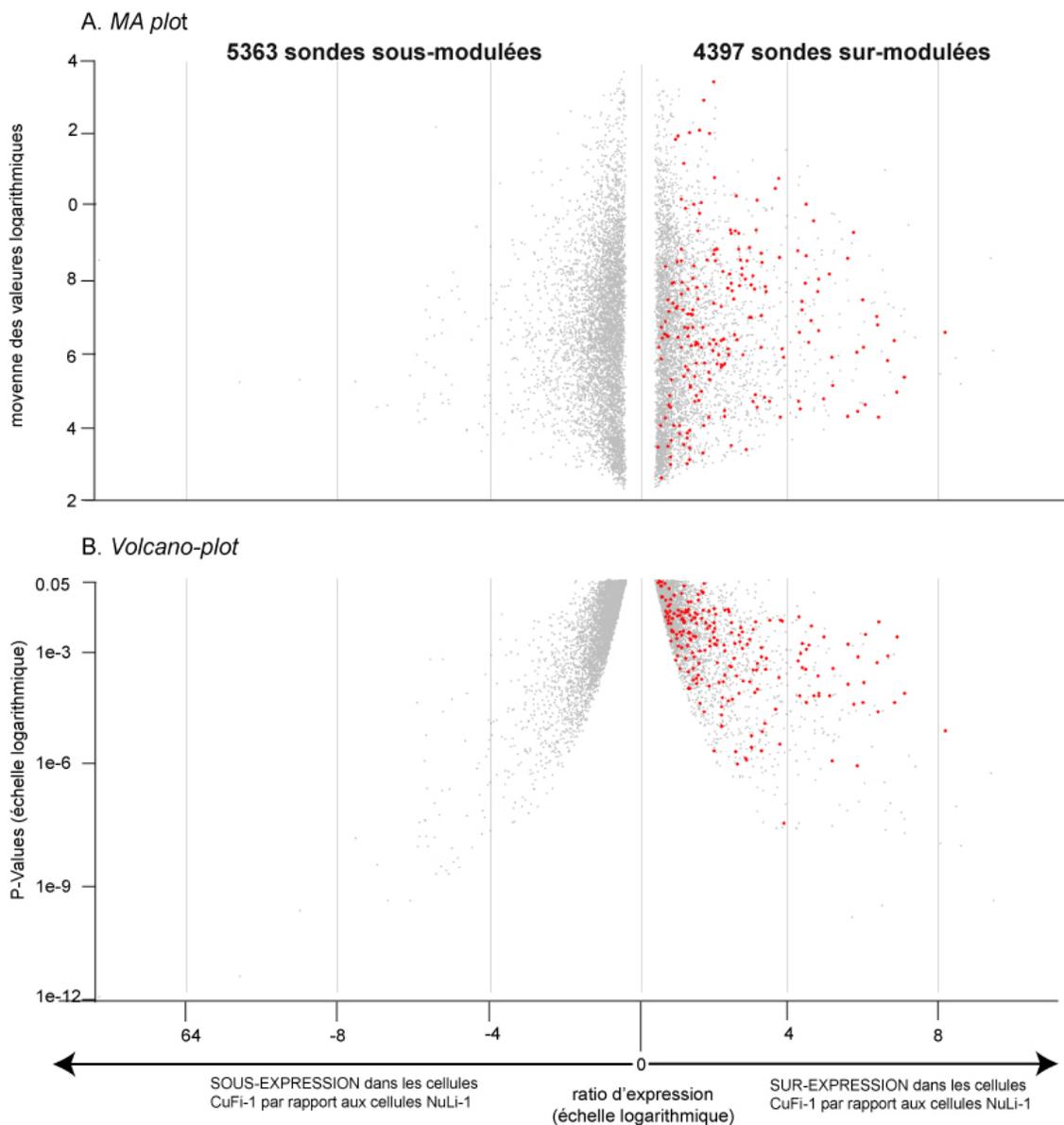


Figure 14: *MA-plot* et *Volcano-plot*

Les graphiques A et B, représentant respectivement le *MA-plot* et le *volcano-plot* sont deux représentations visuelles communément utilisées dans l'analyse d'une expérience de puce à ADN. Les points gris des graphiques représentent l'ensemble des transcrits significativement modulés dans notre expérience (probabilité < 0.05), et les points rouges représentent les transcrits inflammatoires significativement modulés entre les deux conditions expérimentales et associés aux 8 fonctions biologiques inflammatoires identifiées par l'analyse ontologique.

La figure 14.A présente les ratios des valeurs d'expression des transcrits dans les deux conditions ($M = \log_2 \text{condition1} - \log_2 \text{condition2}$) versus la moyenne de leurs valeurs ($A = \frac{1}{2} \times (\log_2 \text{condition1} + \log_2 \text{condition2})$).

La figure 14.B présente les ratios des valeurs d'expression versus les P-Values calculées par un test-t modifié.

2. 2 760 gènes connus sont significativement modulés dans notre expérience

Trois étapes nous conduisent à réduire les 9 760 sondes modulées dans notre expérience en 2 760 gènes connus.

(1) Dans le cas de transcript hybridant plusieurs sondes spécifiques présent sur la plateforme HG-U133 Plus 2.0, nous pouvons éliminer les sondes multiples (c'est-à-dire les sondes hybridant le même transcript) et arbitrairement nous ne conservons que les résultats de celle qui est la plus significative.

(2) Parce que les souches CuFi-1 et NuLi-1 ont des caryotypes différentes et en particulier en ce qui concerne les chromosomes sexuels, nous retirons de notre analyse 1467 sondes associées à ces chromosomes et dont la modulation différentielle ne serait pas une conséquence de la mutation.

(3) Finalement, nous souhaitons éviter également d'intégrer dans notre analyse des biais liés aux erreurs d'annotation et nous nous assurons que les transcripts d'intérêt ont une activité protéique. Pour cela, nous validons les annotations auprès de trois sources génomiques différentes : les annotations d'Affymetrix sont validées par les annotations issues de la base de données Ensembl pour le génome humain (www.ensembl.org), qui sont elles-mêmes validées au niveau protéique par la base de données Resnet (www.ariadegenomics.com).

Ces trois filtres permettent ainsi de sélectionner 2 760 protéines annotées dont les transcripts sont significativement modulés dans notre expérience.

3. L'analyse ontologique des 2 760 transcrits d'intérêt identifie 8 fonctions biologiques significativement surreprésentées, associées à des processus inflammatoires

Pour déterminer les processus biologiques significativement affectés par la mutation dans notre expérience, nous réalisons une analyse ontologique. Notre but étant essentiellement de caractériser la réponse inflammatoire dans les cellules FK par rapport aux cellules contrôles, nous centrons notre intérêt sur les 8 fonctions biologiques associées aux processus inflammatoires surreprésentés. Les caractéristiques de ces fonctions dans notre analyse sont présentées dans le tableau II (page 79).

Identifiant GO	Terme	Définition	P-Value	Espéré	Observé	Protéine
GO:0006955	immune response	Any immune system process that functions in the calibrated response of an organism to a potential internal or invasive threat	0.00007	36	60	59
GO:0006954	inflammatory response	The immediate defensive reaction (by vertebrate tissue) to infection or injury caused by chemical or physical agents. The process is characterized by local vasodilation, extravasation of plasma into intercellular spaces and accumulation of white blood cells and macrophages	0.00005	22	41	38
GO:0006935	chemotaxis	The directed movement of a motile cell or organism, or the directed growth of a cell guided by a specific chemical concentration gradient. Movement may be towards a higher concentration (positive chemotaxis) or towards a lower concentration (negative chemotaxis)	0.00093	14	27	25
GO:0001816	cytokine production	The appearance of a cytokine due to biosynthesis or secretion following a cellular stimulus, resulting in an increase in its intracellular or extracellular levels	0.00099	14	26	17
GO:0009615	response to virus	A change in state or activity of a cell or an organism (in terms of movement, secretion, enzyme production, gene expression, etc.) as a result of a stimulus from a virus	0.00001	10	25	23
GO:0050776	regulation of immune response	Any process that modulates the frequency, rate or extent of the immune response, the immunological reaction of an organism to an immunogenic stimulus	0.00898	14	23	18
GO:0002684	positive regulation of immune system process	Any process that activates or increases the frequency, rate, or extent of an immune system process	0.01430	14	23	18
GO:0045087	innate immune response	Innate immune responses are defense responses mediated by germline encoded components that directly recognize components of potential pathogens	0.01075	13	22	18

Tableau II: Fonctions biologiques inflammatoires surreprésentées dans notre analyse différentielle

L'analyse ontologique des gènes différentiellement significativement surexprimés (P-Value <0.05) entre la condition CuFi-1 et la condition NuLi-1 identifie 57 fonctions biologiques significativement surreprésentées (P-Value <0.05) et composées au minimum de 20 gènes. La table présente uniquement les 8 fonctions biologiques surreprésentées, associées à l'inflammation. Les 3 premières colonnes (GO identifiant, Termes, Définition) correspondant aux informations de ces fonctions fournies par le « Gene Ontology consortium » (www.geneontology.org). La colonne « Espéré » représente la quantité de gènes espérés (ou attendus), associés à un GO terme donné (utilisant comme référence l'ensemble des gènes présents sur la puce à ADN, HGU-133 Plus 2.0). La colonne « Observé », représente la quantité de gènes observés dans notre analyse différentielle, associées à un GO terme donné. Le rapport entre ces 2 quantités nous informe de l'enrichissement d'une fonction biologique dans l'expérience. La colonne « Protéine », indique la quantité des gènes observés qui ont un identifiant protéique dans la base de données Resnet (www.ariadnegenomics.com).

Les 8 classes inflammatoires sont composées de 128 gènes annotés et connus. En raison du rôle pléiotrope des gènes, ceux-ci peuvent être associés à plusieurs fonctions biologiques. Ainsi la plupart des gènes identifiés

peuvent appartenir à plusieurs classes inflammatoires. L'annexe 10 détaille l'appartenance aux classes inflammatoires des 128 gènes ainsi que certaines caractéristiques génomiques et leur modulation au sein de notre expérience.

4. L'analyse des annotations et des données transcriptomiques révèle qu'une partie des gènes inflammatoires modulés présente des caractéristiques communes

Parmi les transcripts inflammatoires sur-modulés dans la condition CuFi-1 par rapport au NuLi-1, nous notons la présence de plusieurs familles inflammatoires. Le tableau III (page 82) présente cinq familles d'intérêts ayant au moins quatre transcripts modulés dans notre expérience. Les deux tiers de ces transcripts présentent une modulation supérieure à deux, c'est-à-dire que la quantité de transcripts mesurée dans les CuFi-1 est au moins deux fois supérieure à celle mesurée dans les NuLi-1. Les transcripts les plus modulés (avec des ratios d'expression supérieurs à 10) sont CXCL10, CXCL3, IL1R2, IL6.

L'organisation des transcripts en fonction de leur appartenance à différentes familles de gènes montre que ceux-ci peuvent être associés à un locus chromosomique donné. Ainsi les transcripts modulés appartenant à la famille des chemokines (CXC), des interleukines (IL), des *Guanylate Binding Protein* (GBP), des 2',5' Oligoadénylate Synthétase (2'5'OAS), des calgranulines (S100) sont majoritairement respectivement codés par le chromosome 4 (locus q21), le chromosome 2 (locus q12, q14), le chromosome 1 (locus q22.2), le chromosome 12 (locus q24.2), le chromosome 1 (locus q22.2), le chromosome 1 (locus q22.2), le chromosome 1 (locus q21.2). Le récepteur CXCR4 appartenant à la famille des chemokines est associé au locus du groupe des interleukines.

La comparaison avec des expériences de puce à ADN étudiant, d'une part, la réponse transcriptomique induite par NF- κ B (Hinata, Gervin et al. 2003) et d'autre part la réponse transcriptomique induite par la voie de signalisation dépendante des IFN (Der, Zhou et al. 1998; Browne, Wing et al. 2001; Radaeva, Jaruga et al. 2002; Sana, Janatpour et al. 2005) montre que 22(/128) transcripts de notre sélection sont sensibles à l'activité de NF- κ B et 32(/128) transcripts sont sensibles à l'activité des IFN (voir ANNEXE 10). Le tableau III (page 82) montre que l'ensemble des chemokines (à l'exception de CXCR4) ont une expression dépendante du système NF-KB. De manière similaire, les membres composant la famille des 2'5'OAS sont exclusivement modulés par les systèmes des interférons. Quant à IL6, IL8, IL1 β , considérées comme des molécules clés de la réponse inflammatoire semblent être modulées par les 2 systèmes de régulation.

Tableau III : caractéristiques des transcrits modulés appartenant à des classes de fonctions inflammatoires
 Les annotations (locus, gène symbol, description) sont fournies par NCBI (<http://www.ncbi.nlm.nih.gov/>). Le ratio d'expression de CuFi-1 et Nuli-1 et la P-Value issue du test t modifié sont calculées par LIMMA. La colonne « sonde d'Affymetrix Plus 2.0 » identifie les sondes d'Affymetrix ayant servies pour l'analyse des transcrits.

CHEMOKINES (CXC)	LOCUS	GENE SYMBOL	RATIO P-VALUE	DESCRIPTION	SONDE (HG-U133 Plus 2.0)
	2q21	CXCR4	1.65 0.0076	receptor 4	209201_x_at, 211919_s_at
	4q12-q13	CXCL5	9.85 0.0006	ligand 5	214974_x_at, 215101_s_at
	4q12-q13	PPBP	3.16 0.0016	pro-platelet basic protein (ligand 7)	214146_s_at
	4q12-q21	PF4	1.88 0.0062	platelet factor 4 (ligand 4)	206390_x_at
	4q21	CXCL1	5.74 0.0001	ligand 1	204470_at
	4q21	CXCL10	8.91 0.0004	ligand 10	204533_at
	4q21	CXCL2	4.46 0.0003	ligand 2	209774_x_at
	4q21	CXCL3	10.45 <0.0001	ligand 3	207850_at
	4q21	CXCL6	5.94 0.0003	ligand 6	206336_at
	4q21.2	CXCL11	10.70 0.0018	ligand 11	211122_s_at
INTERLEUKINES	2q12	IL18R1	1.88 0.0083	interleukin 18 receptor 1	206618_at
	2q12	IL1RL1	4.84 0.0034	interleukin 1 receptor-like 1	207526_s_at, 234066_at, 242809_at
	2q12-q21	IL1F9	3.76 <0.0001	interleukin 1 family, member 9	220322_at
	2q12-q22	IL1R2	11.46 0.0001	interleukin 1 receptor, type II	205403_at, 211372_s_at
	2q14	IL1A	2.87 0.0001	interleukin 1, alpha	210118_s_at
	2q14	IL1B	2.46 <0.0001	interleukin 1, beta	205067_at, 39402_at
	2q14	IL1F5	1.54 0.0010	interleukin 1 family, member 5	222223_s_at
	2q14.2	IL1RN	2.48 0.0001	interleukin 1 receptor antagonist	212657_s_at, 212659_s_at, 216243_s_at, 216244_at
	3q28	IL1TRAP	1.50 0.0361	interleukin 1 receptor accessory protein	205227_at, 210233_at
	4q13-q21	IL8	4.95 0.0001	interleukin 8	202859_x_at, 211506_s_at
	7p21	IL6	16.72 <0.0001	interleukin 6	205207_at
CALGRANULINES	1q21	S100A12	7.40 <0.0001	S100 calcium binding protein A12	205863_at
	1q21	S100A7	1.73 0.0008	S100 calcium binding protein A7	205916_at
	1q21	S100A8	1.98 <0.0001	S100 calcium binding protein A8	202917_s_at
	1q21	S100A9	1.80 <0.0001	S100 calcium binding protein A9	203535_at
GUANYLATE BINDING PROTEIN	1p22.2	GBP1	2.93 0.0004	guanylate binding protein 1	202270_at, 231577_s_at
	1p22.2	GBP3	1.64 0.0006	guanylate binding protein 3	223434_at
	1p22.2	GBP5	1.58 0.0016	guanylate binding protein 5	229625_at, 238681_at
	1p22.2	GBP6	2.84 0.0002	guanylate binding protein 6	1559607_s_at
2'-5'-OLIGOADENYLATE SYNTHETASE	12q24.1	OAS1	3.63 <0.0001	2'-5'-oligoadenylate synthetase 1	202869_at, 205552_s_at
	12q24.2	OAS2	4.60 0.0003	2'-5'-oligoadenylate synthetase 2	204972_at, 228607_at
	12q24.2	OAS3	2.74 0.0001	2'-5'-oligoadenylate synthetase 3	218400_at, 218400_at
	12q24.2	OASL	4.64 0.0008	2'-5'-oligoadenylate synthetase-like	205660_at, 210797_s_at

LEGENDES DES FONCTIONS

- INFLAMMATOIRES
- immune response
- inflammatory response
- chemotaxis
- cytokine production
- response to virus
- regulation of immune response
- positive regulation of immune system process
- innate immune response
- dépendant NF- κ B
- dépendant interférons

5. La confirmation de modulation de gènes inflammatoires

Gene Symbol	puce à ADN		qPCR	
	Ratio	P-Value	Ratio	P-Value
IL6	16.72	0.001	5.21	0.0099
IL8	4.95	0.004	3.80	0.015
IL1B	2.46	< 0.0001	4.87	0.011
CXCL10	8.91	0.014	7.27	0.000
CXCL11	10.70	0.037	6.65	0.000
STAT1	2.50	0.031	2.06	0.010
MYD88	1.52	0.016	1.22	0.716
TNFAIP6	7.49	0.018	58.32	0.000
S100A12	7.40	< 0.0001	2.44	0.002
C1S	4.727	0.027	24.49	<0.0001

Tableau IV: Résultats des validations de qPCR d'une sélection de gènes inflammatoires d'intérêt.

Afin de valider les modulations observées d'un ensemble de gènes inflammatoires dans notre expérience, nous testons leur sur-expression dans le type CuFi-1 versus NuLi-1 par la méthode de qPCR. Ces résultats sont présentés dans le tableau IV (page 83). Ces transcripts ont déjà été rapportés modulés dans d'autres expériences de puce à ADN. Les amorces utilisées pour la quantification par PCR des transcripts sont présentées en annexe 11.

6. Des activateurs de la voie de signalisation de NF-κB induisent une réponse différente dans les cellules CuFi-1 et NuLi-

1

Parce qu'il a été démontré par plusieurs études que la réponse inflammatoire chez les cellules FK pouvait être dépendante du complexe NF-κB, nous évaluons l'activité de ces sous-unités, p65 et p50, en mesurant leur translocation nucléaire dans les deux types cellulaires. Les résultats de cette expérience, présentés à la figure 16 (page 85), montrent qu'il n'y a pas de différence dans la translocation des sous-unités p50 (figure 16.B) et p65 (figure

16.B) entre les deux types cellulaires. Toutefois en présence d'activateur de la voie de signalisation de NF- κ B (LPS à une concentration de 1 μ g/ml et TNF à une concentration de 10 ng/ml), nous observons une augmentation significative de la translocation de p65 dans le type cellulaire CuFi-1 par rapport au type NuLi-1. Suite à la stimulation au TNF et au LPS, nous ne notons pas de différence dans la translocation de p50 entre les deux types cellulaires.

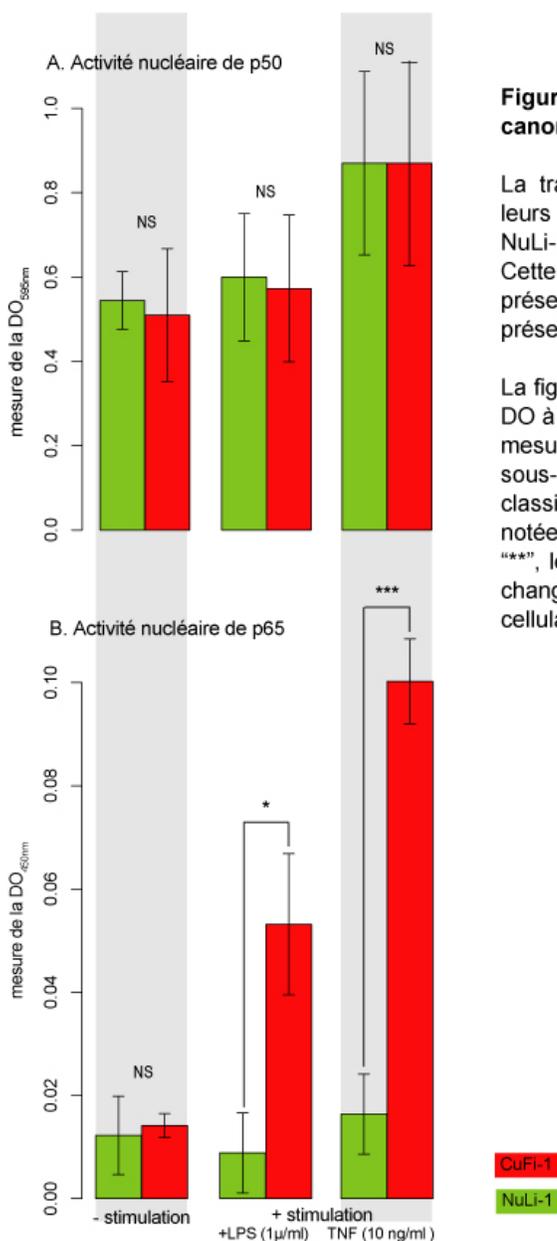


Figure 16: Activité du complexe p50/p65, forme canonique du facteur de transcription NF- κ B.

La translocation nucléaire des 2 sous-unités, corrélée à leurs activités, est mesurée dans les cellules CuFi-1 et NuLi-1 (passage 16) dans 2.5 μ g de protéine nucléaire. Cette mesure est réalisée en absence de stimulation et en présence de LPS à une concentration de 1 μ g/ml et en présence de TNF à une concentration de 10 ng/ml.

La figure A et B représente respectivement la mesure de la DO à 595 nm, reflétant l'activité de la sous-unité (p50), et la mesure de la DO à 450 nm, qui évolue l'activité de la sous-unité (p65). Les probabilités sont calculées avec le classique test t. Les probabilités inférieures à 0.001 sont notées "****", les probabilités inférieures à 0.01 sont notées "***", les probabilités inférieures à 0.05 sont notées "**". Les changements d'activité non significatifs entre les deux types cellulaires sont notés "NS".

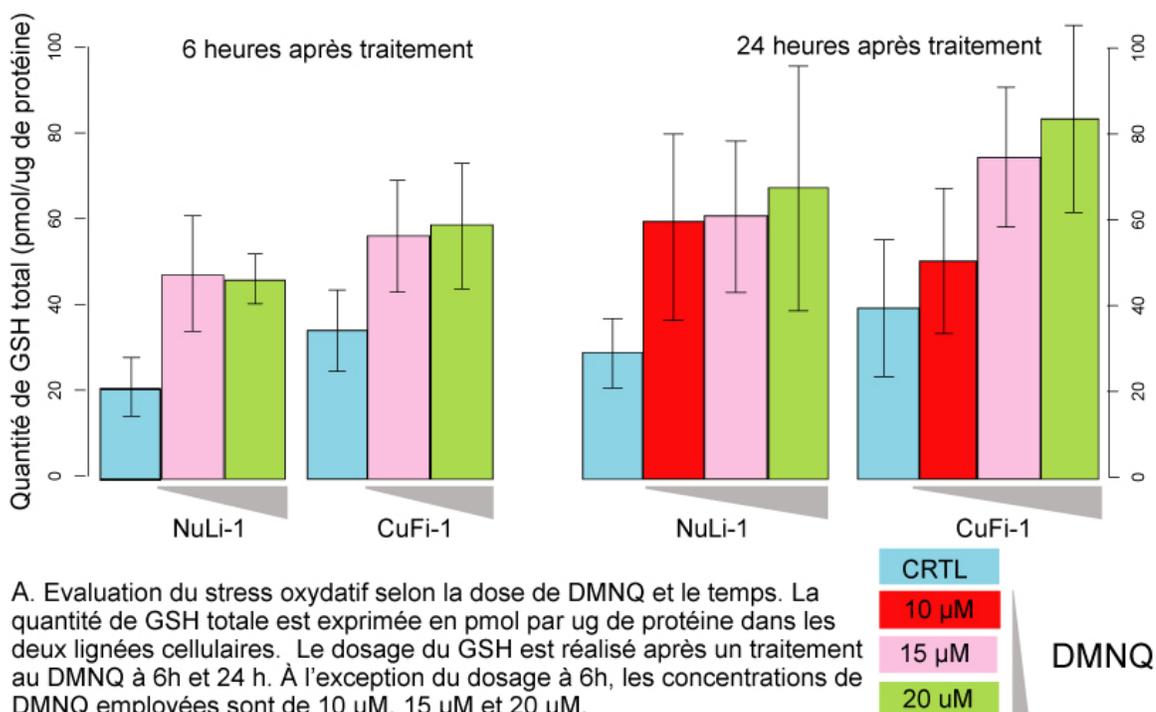
D. Traitement des 2 lignées cellulaires au DMNQ

Pour déterminer la dose de DMNQ à utiliser dans notre étude afin de générer un stress oxydatif dans nos deux modèles cellulaires, nous mesurons la production de glutathion (GSH) à différentes concentrations de DMNQ. La partie A de la figure 11 (page 88) qui suit présente les mesures réalisées dans

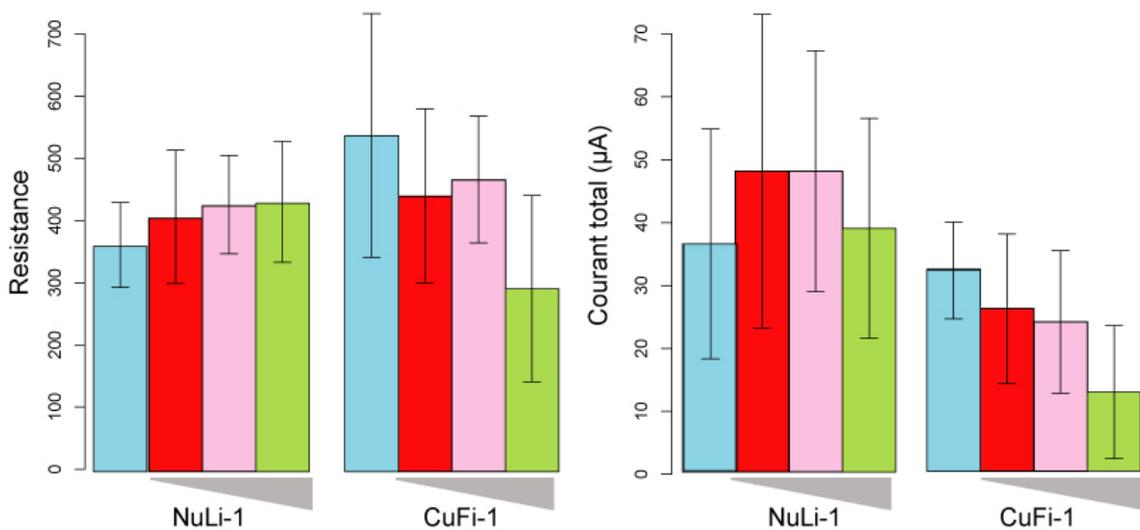
les cellules NuLi-1 et CuFi-1, 6 heures et 24 heures après le traitement à différentes concentrations de DMNQ. La partie A montre que le DMNQ aux différentes concentrations induit une augmentation de quantité de glutathion par rapport au contrôle dans les 2 lignées cellulaires que ce soit 6 heures ou 24 heures après le traitement. De plus, d'une façon générale, la quantité de glutathion mesurée est plus importante dans les cellules CuFi-1 que dans les cellules de type NuLi-1 et pour chacun des types cellulaires, les quantités de GSH mesurées à 24h sont plus importantes après 6h pour les deux lignées. Une analyse ANOVA à deux facteurs met en évidence une différence significative (Probabilité < 0.01) dans les niveaux de GSH mesurés dans les deux lignées entre les contrôles et les différentes concentrations de DMNQ à 6 heures et à 24 heures.

Parce que nous souhaitons étudier les réponses physiologiques induites par le stress oxydatif, il est important d'évaluer la toxicité de chacune des doses de DMNQ employées. Pour cela, nous mesurons la résistance cellulaire laquelle reflète la présence de jonctions serrées donc de la formation d'une monocouche cellulaire intacte. Parallèlement nous mesurons le courant total transmembranaire qui rend compte de l'activité ionique cellulaire et l'intégrité cellulaire. La partie B de la figure 11 (page 88) présente les résistances mesurées dans les deux lignées cellulaires 24 heures après le traitement au DMNQ aux différentes concentrations. Ces données indiquent que la résistance mesurée chez les cellules CuFi-1 est plus importante que chez les cellules NuLi-1 mais qu'une dose de 20 μM de DMNQ affecte la résistance chez les CuFi-1 et pourrait affecter l'intégrité de la monocouche, suggérant potentiellement une atteinte cellulaire. De plus, des concentrations de DMNQ de 10 μM et 15 μM n'affectent pas la résistance des cellules NuLi-1. De façon similaire, la partie C indique que les différentes concentrations chez les NuLi-1 n'ont pas d'effet sur le courant total, contrairement au CuFi-1, dont le courant total décroît avec des doses élevées de DMNQ. Ces résultats nous suggèrent

qu'une concentration de 15 μM est adéquate pour induire un stress oxydatif cellulaire sans affecter l'intégrité cellulaire.



A. Evaluation du stress oxydatif selon la dose de DMNQ et le temps. La quantité de GSH totale est exprimée en pmol par ug de protéine dans les deux lignées cellulaires. Le dosage du GSH est réalisé après un traitement au DMNQ à 6h et 24 h. À l'exception du dosage à 6h, les concentrations de DMNQ employées sont de 10 μ M, 15 μ M et 20 μ M.



B. Résistances mesurées dans les deux lignées cellulaires 24 heures après un traitement au DMNQ aux concentrations de 10 μ M, 15 μ M et 20 μ M.

C. Courants totaux mesurés dans les deux lignées cellulaires 24 heures après un traitement au DMNQ aux concentrations de 10 μ M, 15 μ M et 20 μ M.

Figure 11 : Détermination de la dose de DMNQ à utiliser et son impact sur l'intégrité de la monocouche cellulaire.

E. Résultats de l'analyse de l'impact du stress oxydatif sur le transcriptome d'une cellule FK

1. Un sous-groupe des transcripts testés présente un intérêt en termes de variance dans notre analyse

Basé sur la variance des données d'expression, il est possible de distinguer deux types de transcripts: ceux avec une variation informative, induite par les paramètres contrôlables de l'expérience, c'est-à-dire les conditions expérimentales, et ceux avec une variation non-informative due à des sources non-contrôlables, comme la variabilité biologique et technique. Il est important de discerner les deux groupes afin de concentrer notre analyse sur les transcripts dits informatifs, les autres étant considérés comme une source de bruit et ne présentent que très peu d'intérêt dans le cadre de notre analyse. Pour cela nous filtrons les données d'expression en fonction de cette variabilité.

Nous avons choisi l'IQR comme mesure de dispersion des données d'expression, pour effectuer le filtrage des données d'expression informatives. La figure 17 (page 90) présente la distribution de l'IQR des 54 675 sondes étudiées au cours de notre expérience. Celle-ci montre que 50 825 (93 %) d'entre elles présentent un IQR inférieur à 0.5 reflétant de faibles variations des niveaux d'expression au cours de l'expérience.

Parce que nous souhaitons que nos résultats et nos interprétations biologiques ne soient pas influencés par le choix de l'IQR, nous exploiterons un intervalle de cette mesure, comprise entre 0.5 à 1. Ceci nous permet de sélectionner six groupes selon le seuil d'IQR choisi (0.5, 0.6, 0.7, 0.8, 0.9, 1). Ces groupes inclusifs sont respectivement composés de 3 850, 2 583, 1 818, 1 370, 1 047, 819 éléments. Dans la suite, ces groupes sont qualifiés par le

terme «groupe.IQR.0.5 », «groupe.IQR.0.6», «groupe.IQR.0.7»,
«groupe.IQR.0.8», «groupe.IQR.0.9», «groupe.IQR.1».

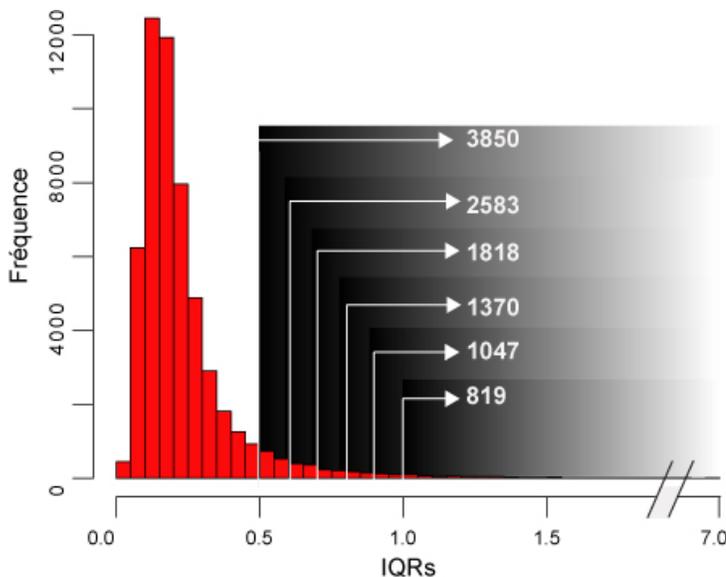


Figure 17: Histogramme des Inter-Quartile Range

Pour l'ensemble des sondes composant notre expérience (54675), nous calculons l'IQR, basé sur les données d'expression des conditions expérimentales.

Une majorité des transcrits (50825) ont un IQR inférieur à 0.5. À partir de cet histogramme, nous pouvons visualiser les groupes de transcrits à des seuils d'IQR variables: 0.5, 0.6, 0.7, 0.8, 0.9, 1 constitués respectivement de 3850, 2583, 1818, 1370, 1047, 819 transcrits.

2. Une analyse ANOVA détectent les effets significatifs de la mutation et de l'oxydant mais manque de puissance pour détecter l'interaction des deux facteurs

Basé sur les 6 groupes de transcrits sélectionnés, nous réalisons une analyse ANOVA multivariable sur l'ensemble des sondes appartenant à un groupe sélectionné, pour évaluer la capacité de ce test à détecter les effets de chacune des conditions expérimentales ainsi que leurs interactions.

Les résultats de notre analyse ANOVA sont présentés dans le tableau V (page 91) sous la forme de probabilités (obtenues par 999 permutations des données d'expression). La première constatation est l'effet significatif de la mutation et de l'oxydant sur les données d'expression et la seconde constatation est l'absence de détection significative de l'interaction entre les deux conditions expérimentales, quelque soit le groupe.IQR considéré. Le test ANOVA détecte une interaction significative quand l'effet d'un facteur change, de façon significative, selon le niveau de l'autre facteur. Plusieurs travaux ont corrélé le

niveau de stress oxydatif avec la sévérité de la maladie, par conséquent, nous nous attendons que le DMNQ influence le niveau d'expression de gènes déjà modulés par la mutation: les 2 facteurs pourraient interagir sur le niveau d'expression de gènes impliqués dans la physiopathologie de la maladie. Ainsi, nous pouvons conclure (1) que la majorité des sondes n'est pas affectée par une potentielle interaction entre la mutation et l'oxydant et/ou (2) que notre design statistique manque de puissance statistique. Ces deux raisons peuvent expliquer l'absence de détection significative de l'interaction entre les deux conditions expérimentales.

groupe.IQR	0.5	0.6	0.7	0.8	0.9	1
Nombre de transcripts	3850	2583	1818	1370	1047	819
Probabilités:						
effet mutation	0.003	0.005	0.001	0.002	0.003	0.002
effet oxydant	0.002	0.006	0.002	0.004	0.002	0.003
effet interaction	0.340	0.331	0.376	0.322	0.311	0.330

Tableau V : Résumé des probabilités issues du test ANOVA multivariable à 2 facteurs pour chaque groupe de sonde sélectionné dans l'étape de filtrage.

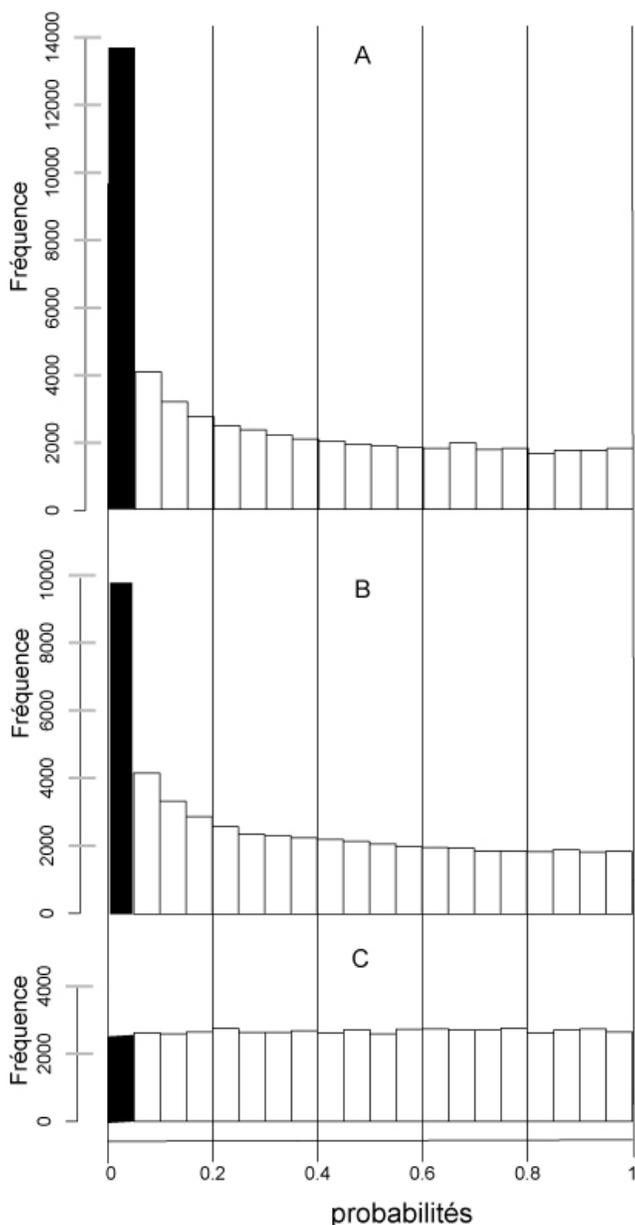


Figure 18 : Histogramme des probabilités issues d'analyses ANOVA univariées à deux facteurs réalisées sur chacun des 54675 transcripts

Les probabilités sont ajustées par le FDR. L'histogramme A représente la distribution des probabilités de l'effet mutation, L'histogramme B correspond aux probabilités de l'effet oxydant. L'histogramme C correspond aux probabilités de l'effet interaction. Dans chaque histogramme, la partie en noire représente l'ensemble des transcripts avec des probabilités inférieures à 0.05.

Dans un second temps, nous évaluons de façon indépendante, l'effet de chaque condition sur chaque sonde en prenant soin de corriger les probabilités à l'aide d'une méthode de correction pour tests multiples (FDR). Les histogrammes A et B de la figure 18 (page 92) montrent que les expressions de nombreux transcripts sont modulées sous l'effet de la mutation ou de l'oxydant, contrairement à l'effet de l'interaction qui n'est pas souvent détecté

dans l'ensemble de 54 675 sondes (histogramme C). En effet, dans l'histogramme C, 2500 sondes présentent une probabilité inférieure à 0.05, soit 5 % des 54 675 sondes testées, niveau attendu de l'erreur de type 1 pour des tests au niveau de signification de 5%. Par conséquent, nous pouvons conclure que le test n'a pas détecté d'interaction significative au-delà du niveau attendu par l'erreur de type 1. Finalement, nous concluons que le test ANOVA n'a pas mis en évidence l'interaction entre le stress oxydatif et la mutation dans notre expérience.

3. Une analyse de redondance sélectionne 687 transcripts dont la variance est expliquée par l'effet conjoint de la mutation et du stress oxydatif

L'analyse canonique de redondance nous permet une étude de la variance présente dans les données transcriptomiques d'intérêt et ainsi de pallier aux défauts du test de l'ANOVA. Cette approche permet une évaluation quantitative des effets, induits par la mutation et le stress oxydatif sur la variance des niveaux d'expression.

Le tableau VI (page 94) décrit les résultats issus de l'analyse de redondance pour chacun des groupes. IQR en termes de variance. Ces résultats montrent le pourcentage de la variance présente dans les données pouvant être expliquée par les deux facteurs ainsi que la part de variance inexpliquée. Quelle que soit la taille du groupe, la variance totale peut être expliquée en majorité par la condition mutation (entre 63.16 et 71.5 %) tandis que la condition oxydant en explique une plus petite fraction (entre 17.4 et 15 %). Parce que la variabilité observée ne peut être expliquée totalement par les facteurs étudiés, une part de celle-ci reste inexpliquée et s'élève de 19.4 à 13.5 %. De plus, étant donné l'absence de détection de l'effet de l'interaction

(tableau V, page 91), la portion de variance associée à l'interaction n'est pas quantifiée dans l'analyse de redondance.

groupe.IQR	0.5	0.6	0.7	0.8	0.9	1
Nombre de transcripts	3850	2583	1818	1370	1047	819
Variance totale	1493	1306	1146	1026	924	837
Variance expliquée						
par la mutation (%)	63.16	65.42	67.62	69.13	70.20	71.50
par l'oxydant (%)	17.43	16.90	16.26	15.90	15.70	15.00
Variance non expliquée (%)	19.41	17.68	16.12	14.97	14.10	13.50
Nombre de transcripts avec une variance combinée	687	414	258	164	107	78

Tableau VI : Résumé des résultats produits par l'analyse canonique de redondance sous forme de variance

La variance combinée est définie comme étant la variance expliquée par au moins 20 % par chacune des conditions expérimentales (au moins 20 % de la variance expliquée par la mutation et au moins 20 % de la variance expliquée par l'oxydant).

L'analyse canonique de redondance permet la décomposition de la variance pour chacun des transcripts considérés. Ainsi, leur ordination dans un espace euclidien où chacun des axes est défini par le pourcentage de variance (R^2) expliqué par une des deux conditions expérimentales met en évidence l'effet de ces conditions.

La figure 19 (page 96) montre cette ordination pour chacun des groupes.IQR considéré. Chaque graphique peut-être divisé en quatre secteurs (A, B, C, D) caractérisant la nature de la variance expliquée. Le secteur A regroupe les transcripts dont la variance s'explique essentiellement par l'effet de l'oxydant (au moins 20 % de l'effet est imputable à l'oxydant). De façon opposée, le secteur B regroupe les transcripts avec une variance s'expliquant essentiellement par l'effet de la mutation (au moins 20 % de l'effet est imputable à la mutation). Le secteur C regroupe les transcripts dont la variance est majoritairement non-expliquée par les conditions expérimentales et leurs

variances peuvent être expliquées au maximum à 40 % par les conditions étudiées. Finalement, le secteur D regroupe les transcripts avec une variance combinée puisqu'elle résulte de la combinaison de l'effet des deux conditions expérimentales. Selon la nature de la variance, un groupe.IQR peut être décomposé en quatre secteurs : A, B, C, D. Cette notation sera conservée par la suite.

Cette approche nous permet la sélection de transcripts avec une variance expliquée par un effet combiné des deux conditions expérimentales. Nous choisissons 20 % comme un seuil minimum d'effet pour chacune des conditions. Pour l'exemple, dans le groupe.IQR.0.5 nous isolons un sous-ensemble de 687 transcripts dont la variance est expliquée au minimum à 40 % par les conditions expérimentales. Cette variance se décompose par au moins 20 % de la variance expliquée par la mutation et au moins 20 % de la variance expliquée par le stress oxydatif. Une sélection similaire dans le groupe.IQR.0.6, quant à elle, isole 414 transcripts (inclut dans la sélection de 687 du groupe.IQR.0.5).

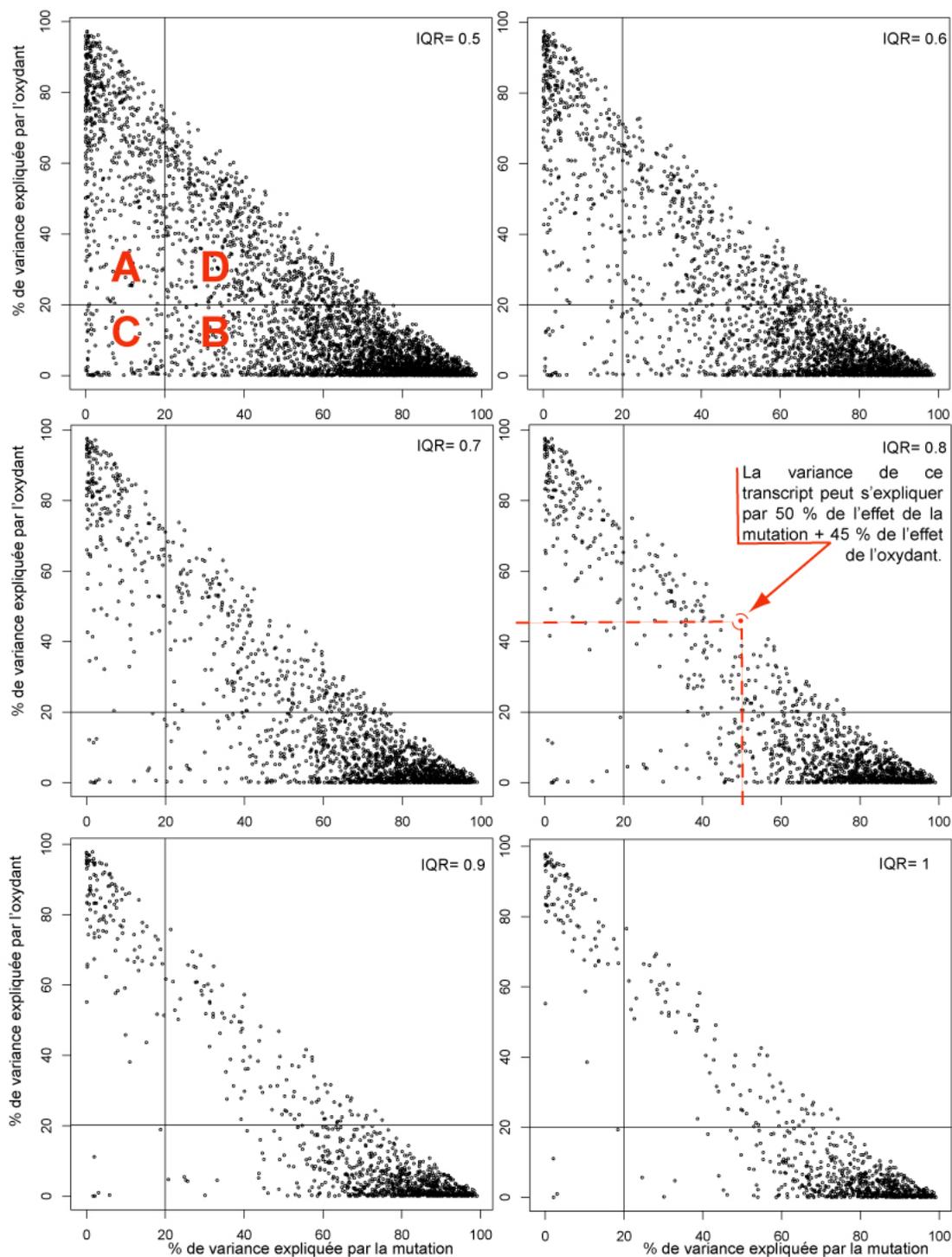


Figure 19 : Répartition des transcripts en fonction de la variance expliquée par les conditions expérimentales
 Pour chaque groupe, nous répartissons les transcripts dans un graphique où l'axe des abscisses représente le % de variance expliquée par la mutation et l'axe des ordonnées représente le % de la variance expliquée par l'oxydant. Chaque graphique peut-être divisé en 4 secteurs selon la nature de la variance des transcripts : A, variance majoritairement expliquée par l'oxydant, B, variance majoritairement expliquée par la mutation, C, variance peu expliquée par les conditions et D, variance expliquée par au moins 40 % des conditions expérimentales.

4. L'apoptose est un processus significativement surreprésenté dans les transcrits dont la variance est expliquée par la mutation et le stress oxydant.

Nous réalisons deux analyses ontologiques dont les résultats sont présentés dans la figure 20 (page 100). Dans un premier temps, nous réalisons une analyse ontologique sur l'ensemble des sondes composant chacun des groupe.IQR (réunissant les sondes des secteurs A, B, C et D) et nous nous intéressons qu'aux termes GO communs significativement surreprésentés (Probabilité<0.05) dans chaque des groupe.IQR. Cette analyse identifie 63 termes GO surreprésentés (communs aux 6 analyses) et sont présentés dans la partie A de la figure 20 sous la forme de deux quantités pour chaque terme GO : le nombre de gènes observés et le nombre de gènes espérés. Ensuite, une seconde analyse ontologique teste uniquement les sondes du secteur D dans chacun des groupe.IQR, et ne s'intéresse qu'aux termes GO communs significativement surreprésentés (Probabilité<0.05) dans chacune des analyses. Ainsi, nous identifions 11 termes GO surreprésentés que nous présentons dans la partie B de la figure 20. Par conséquent ce qui distingue les deux analyses est la nature des données puisque la première analyse s'intéresse, sans distinction, aux sondes des secteurs A, B, C et D et la seconde analyse aux sondes du secteur D. Cette comparaison vise à déterminer si la surreprésentation des fonctions observées dans les sondes du secteur D (dont la variance est associée à un effet combiné de la mutation et de l'oxydant) est dépendante de la nature de la variance.

L'analyse ontologique des sondes du secteur D montre qu'entre autre, la mort cellulaire (GO:0008219) et l'apoptose (GO:0006915) sont des processus biologiques significativement surreprésentés (figure 20, partie B, page 100).

Toutefois, ces processus apparaissent également surreprésentés dans l'analyse considérant l'ensemble des sondes (A, B, C, D). Ce résultat pourrait

nous laisser croire que la surreprésentation de ces fonctions n'est pas spécifique à la catégorie D.

Pour vérifier cela, nous comparons les proportions de deux fonctions biologiques dans chacune des analyses. Le tableau VII (page 102) rapporte les résultats d'un test binomial de proportion. Pour être plus explicite, ce test compare la proportion de gènes associés à une fonction biologique dans deux analyses distinctes. Par exemple, dans notre première analyse, 1 850 gènes sont testés et nous y avons identifié 52 gènes associés à l'apoptose soit 2,8 % des gènes testés. La seconde analyse exploite 389 identifiants et caractérise 26 associés à l'apoptose soit 6.7 % des gènes testés. Comme le montre le tableau VII (page 102), le test binomial conclut que ces deux proportions sont significativement différentes (Probabilité = 0.003) et que par conséquent, la surreprésentation de l'apoptose est spécifique à la nature de la variance des sondes du secteur D. Par contre, ce test n'a pas montré de différence significative dans les proportions des sondes associées à la fonction de mort cellulaire.

Dans le reste de notre analyse, nous considérons l'ensemble des 40 sondes associées à la mort cellulaire et à l'apoptose, car 25 d'entre eux ont été caractérisées comme participant à la mort cellulaire et à l'apoptose et cela nous laisse penser que les 15 sondes associées au processus de mort cellulaire pourraient certainement avoir un rôle à jouer dans les perturbations transcriptomiques et qu'il est intéressant de les considérer comme un tout.

À la vue de ces résultats, nous déduisons que la surreprésentation observée dans notre expérience est spécifique à la nature des sondes du secteur D et que par conséquent, l'apoptose pourrait être un processus affecté par le stress oxydatif présent dans une cellule fibro-kystique. De plus, afin de comprendre les perturbations de l'apoptose induites par nos conditions expérimentales,

nous appuyons notre discussion sur une sélection de 40 gènes présentée dans la figure 21 (page 101).

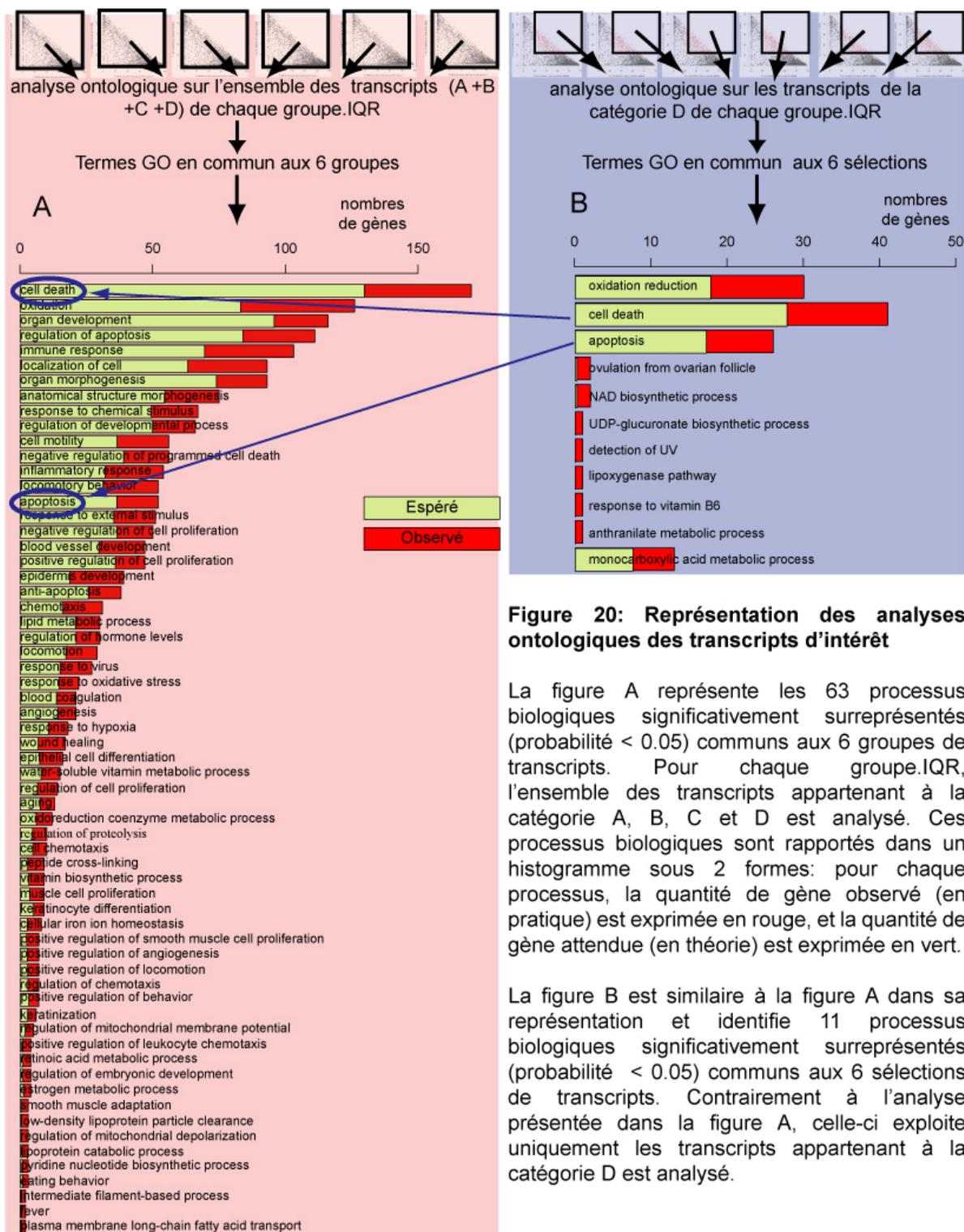


Figure 20: Représentation des analyses ontologiques des transcrits d'intérêt

La figure A représente les 63 processus biologiques significativement surreprésentés (probabilité < 0.05) communs aux 6 groupes de transcrits. Pour chaque groupe.IQR, l'ensemble des transcrits appartenant à la catégorie A, B, C et D est analysé. Ces processus biologiques sont rapportés dans un histogramme sous 2 formes: pour chaque processus, la quantité de gène observé (en pratique) est exprimée en rouge, et la quantité de gène attendue (en théorie) est exprimée en vert.

La figure B est similaire à la figure A dans sa représentation et identifie 11 processus biologiques significativement surreprésentés (probabilité < 0.05) communs aux 6 sélections de transcrits. Contrairement à l'analyse présentée dans la figure A, celle-ci exploite uniquement les transcrits appartenant à la catégorie D est analysé.

IQR	Gene	Description
1	DAPK1	Death-associated protein kinase 1
	PDCD4	Programmed cell death protein 4
	SFN	Stratifin, Epithelial cell marker protein 1
	ID3	DNA-binding protein inhibitor ID-3
	GULP1	PTB domain-containing engulfment adapter protein , Cell death protein 6 homolog
	TRIB3	Tribbles homolog 3, Neuronal cell death-inducible putative kinase, p65-interacting inhibitor of NF-kappa-B
	BCL2A1	Bcl-2-related protein A1, Bcl-2-like protein 5, Hemopoietic-specific early response protein, Protein GRS
	SOX9	Transcription factor SOX-9
	ZBTB16	Zinc finger and BTB domain-containing protein 16, Promyelocytic leukemia zinc finger protein
	NUAK2	NUAK family SNF1-like kinase 2, SNF1/AMP kinase-related kinase
	CD36	Platelet glycoprotein 4, Platelet glycoprotein IV, Glycoprotein IIIb, Leukocyte differentiation antigen CD36, Platelet collagen receptor,Fatty acid translocase, Thrombospondin receptor
0.9	CLN8	Protein CLN8
	STAT1 AIFM2	Signal transducer and activator of transcription 1-alpha/beta, Transcription factor ISGF-3 components p91/p84 Apoptosis-inducing factor 2, Apoptosis-inducing factor-like mitochondrion-associated inducer of death, Apoptosis-inducing factor homologous mitochondrion-associated inducer of death, p53-responsive gene 3 protein
0.8	RFFL	E3 ubiquitin-protein ligase rififylin, RING finger and FYVE-like domain-containing protein 1, FYVE-RING finger protein Sakura, Caspases-8 and -10-associated RING finger protein 2, CARP-2,Caspase regulator CARP2,
	MAP3K5	Mitogen-activated protein kinase kinase kinase 5, MAPK/ERK kinase kinase 5
	SERPIN9	Serpin B9, Cytoplasmic antiproteinase ,Proteinase inhibitor 9
0.7	CDKN2C	Cyclin-dependent kinase 4 inhibitor C, Cyclin-dependent kinase 6 inhibitor
	ALDH1A3	Aldehyde dehydrogenase family 1 member A3, Aldehyde dehydrogenase 6, Retinaldehyde dehydrogenase 3
	RNF144B	E3 ubiquitin-protein ligase RNF144B, RING finger protein 144B,BR domain-containing protein 2, p53-inducible RING finger protein
	MAL	Myelin and lymphocyte protein, T-lymphocyte maturation-associated protein
	PTRH2	Peptidyl-tRNA hydrolase 2, mitochondrial Precursor, Bcl-2 inhibitor of transcription 1
	RARB	Retinoic acid receptor beta, Nuclear receptor subfamily 1 group B member 2
	VEGFA	Vascular endothelial growth factor A Precursor, Vascular permeability factor
TNFSF9	Tumor necrosis factor ligand superfamily member 9	
TNFRSF10A	Tumor necrosis factor receptor superfamily member 10A Precursor,TNF-related apoptosis-inducing ligand receptor 1, TRAIL receptor 1	
0.6	ERBB3	Receptor tyrosine-protein kinase erbB-3 Precursor, Tyrosine kinase-type cell surface receptor
	NR4A1	Nuclear receptor subfamily 4 group A member 1, Orphan nuclear receptor HMR, Early response protein NAK1
	EAF2	ELL-associated factor 2, Testosterone-regulated apoptosis inducer and tumor suppressor protein
	TNFRSF10D	Tumor necrosis factor receptor superfamily member 10D Precursor, Decoy receptor 2, TNF-related apoptosis-inducing ligand receptor 4, TRAIL receptor 4, CD264 antigen
	CD14	Monocyte differentiation antigen CD14 Precursor,Myeloid cell-specific leucine-rich glycoprotein, CD14 antigen.
HIP2	Homeodomain-interacting protein kinase 2	
0.5	TICAM1	TIR domain-containing adapter molecule 1, Proline-rich, vinculin and TIR domain-containing protein B, Toll-interleukin 1 receptor domain-containing adapter protein inducing interferon-beta, TIR domain-containing adapter protein inducing IFN-beta, Putative NF-kappa-B-activating protein 502H
	EYA2	Eyes absent homolog 2
	NUP62	Nuclear pore glycoprotein p62
	BID	BH3-interacting domain death agonist (p22 BID)(BID) [Contains BH3-interacting domain death agonist p15(p15 BID),BH3-interacting domain death agonist p13(p13 BID),BH3-interacting domain death agonist p11(p11 BID)]
	TNFRSF12A	Tumor necrosis factor receptor superfamily member 12A Precursor,Fibroblast growth factor-inducible immediate-early response protein 14, FGF-inducible 14, Tweak-receptor, CD266 antigen MAPK1
	PTPRH	Mitogen-activated protein kinase 1 , Extracellular signal-regulated kinase 2, Mitogen-activated protein kinase 2,MAP kinase 2 MAPK8 Mitogen-activated protein kinase 8 , Stress-activated protein kinase JNK1, c-Jun N-terminal kinase 1,JNK-46
	CASP1	Receptor-type tyrosine-protein phosphatase H Precursor,Transmembrane-type protein-tyrosine phosphatase type H, Stomach cancer-associated protein tyrosine phosphatase 1
	CARD17	Caspase-1 Precursor, Interleukin-1 beta convertase, Interleukin-1 beta-converting enzyme,IL-1 beta-converting enzyme, p45 Caspase-1 Precursor (CASP-1)(Interleukin-1 beta convertase)(Interleukin-1 beta-converting enzyme) (IL-1 beta-converting enzyme)(ICE)(p45) [Contains Caspase-1 subunit p20-Caspase-1 subunit p10]

Figure 21 :Description des 40 gènes identifiés dans notre analyse associés au processus de la mort cellulaire

Ceux-ci sont présentés en fonction de leurs IQR. Les gènes en rouge sont associés au processus de l'apoptose.

groupe.IQR	0.5	0.6	0.7	0.8	0.9	1	
ANALYSE A+B+C+D	Nombre de transcripts	3850	2583	1818	1370	1047	819
	Nombre d'identifiant de gène utilisé par l'analyse ontologique	1850	1244	898	692	534	423
	Nombre d'identifiant de gène associé à l'apoptose (GO:0006915)	52	36	27	21	18	17
ANALYSE SECTEUR D	Nombre de transcripts avec une variance combinée	687	414	258	164	107	78
	Nombre d'identifiant du gène utilisé par l'analyse ontologique	389	250	159	107	75	52
	Nombre d'identifiant du gène associé à l'apoptose (GO:0006915)	26	18	15	11	13	11
P-Value	0.003	0.001	0.0003	0.0009	1.10⁻⁶	3.5.10⁻⁶	

Tableau VII : Résultats du test binomial de proportion

Pour vérifier la surreprésentation de la fonction apoptose dans le secteur D pour rapport à l'ensemble des transcripts sélectionnés pour un groupe.IQR considéré, nous comparons la proportion des gènes associés à l'apoptose dans les deux analyses ontologiques et testons sa significativité par un test binomial de proportion.

5. La classification des profils d'expression des transcripts du secteur D montrent quatre grandes tendances de modulation vis-à-vis des conditions expérimentales.

Basé sur une classification non-supervisée des données transcriptomiques des 687 transcripts du groupe.IQR.0.5, nous mettons en évidence quatre tendances de comportement vis-à-vis des conditions expérimentales.

Ces tendances, présentées dans la figure 22 (page 104), montrent que le DMNQ n'induit pas de changement dans le sens de la modulation des transcripts, c'est-à-dire que les transcripts présentant une modulation entre les conditions FK et non FK en absence de DMNQ (soit sur-expression ou sous-expression), ont une modulation similaire en présence de DMNQ. D'autre part, nous notons que le DMNQ influence le niveau basal d'expression des conditions FK et non FK dans les profils A et B. Cet effet est beaucoup moins

prononcé dans les profils C et D. Par conséquent, nous pouvons discerner deux grands types de comportement vis-à-vis du traitement au DMNQ: une tendance qui est plus dépendante du DMNQ avec un profil d'expression plus sensible à l'effet du stress oxydatif qu'à la mutation et une tendance qui est plus dépendante de l'effet de la mutation que de l'impact du stress oxydatif.

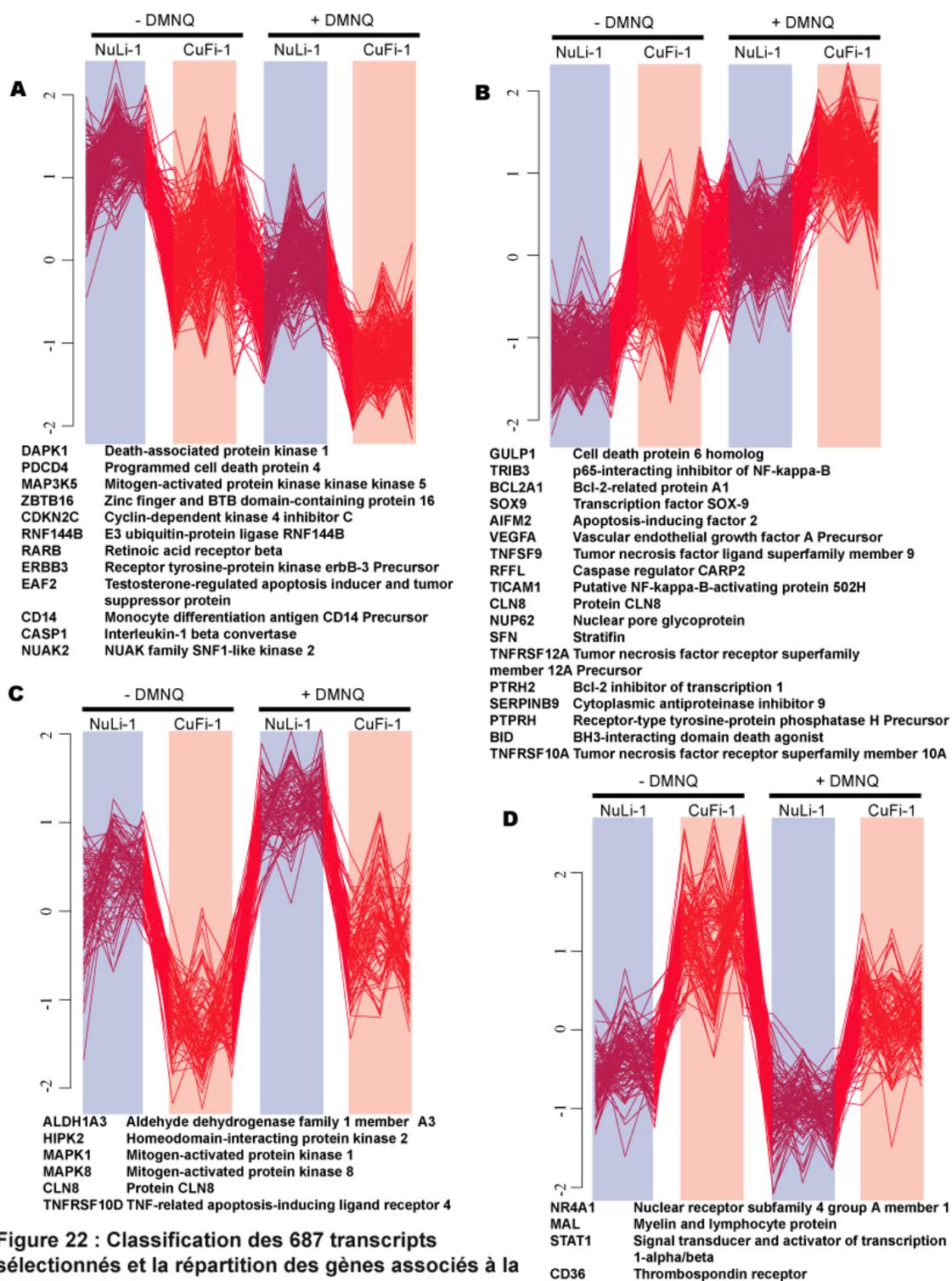


Figure 22 : Classification des 687 transcrits sélectionnés et la répartition des gènes associés à la mort cellulaire

Les données d'expression sont centrées-réduites puis, classifiées par l'algorithme FCM, en 4 classes de tendances induites par les deux conditions expérimentales.

6. De grandes classes fonctionnelles se dégagent des 40 gènes associés à la mort cellulaire et à l'apoptose

Notre analyse nous conduit à isoler 40 transcripts distincts annotés lesquels pourraient intervenir dans un dysfonctionnement de l'apoptose dans la fibrose kystique puisque leurs modulations sont dépendantes de l'effet des deux conditions expérimentales. Exploitant leurs annotations génomiques, nous pouvons arbitrairement classer ces transcripts en grandes familles fonctionnelles révélant ainsi leurs implications dans l'apoptose et sa régulation. Certaines de ces fonctions sont présentées dans le tableau VIII (page 107).

(1) les kinases

Parmi les kinases isolées et impliquées dans la mort cellulaire, trois d'entre elles appartiennent à la famille des MAPK et sont sous-modulés dans la lignée FK: MAPKKK5, MAPK1, MAPK8.

D'autres kinases d'intérêt, impliquées dans l'apoptose sont également sous-modulées: CDKN2C (p18) est une kinase qui régulent indirectement le cycle cellulaire, et NUAK2, ERBB3 et DAPK1, HIPK2 sont quatre kinases aux fonctions anti-apoptotiques. La seule kinase qui soit sur-exprimée dans notre expérience est TRIB3.

(2) les récepteurs membranaires

La seconde classe fonctionnelle présente dans les gènes modulés est composée de 11 transcripts codant pour des récepteurs membranaires.

Parmi ces gènes sur-modulés par la mutation FK, nous relevons la présence de 3 récepteurs de la famille du TNF: TNFRSF12A, TNFRSF10D, et TNFRSF10A. TICAM1 est un récepteur Toll/IL1R (TIR), CD36 est le récepteur de thrombospondine 1, PTPRH et NR4a1.

3 autres récepteurs sont sous-modulés dans le modèle FK par rapport au contrôle: CD14, RARB et ERBB3.

(3) gènes reliés au facteur de transcription NF- κ B et sa cascade

Parmi notre sélection de gènes apoptotiques, plusieurs d'entre eux sont reliés à la cascade de NF- κ B. TRIB3 est une protéine kinase induite par NF- κ B et qui agit comme un régulateur de cette cascade et intervient dans une sensibilisation de l'apoptose (Wu, Xu et al. 2003). BCL2A1 est également induite par NF- κ B (Wang, Guttridge et al. 1999). L'interaction exclusive de TICAM1 avec le récepteur TLR3 induit une activation de NF- κ B (Oshiumi, Matsumoto et al. 2003). Quant à TNFRSF10D et TNFRSF10A, ils sont des récepteurs membranaires qui induisent cette voie mais avec des conséquences différentes sur l'apoptose puisque le TNFRSF10D protège contre la cytotoxicité de TNFRSF10A (Chaudhary, Eby et al. 1997; Degli-Esposti, Dougall et al. 1997). De plus, NUP62 et STAT1 apparaissent comme étant des inhibiteurs de cette cascade contrairement à CASP1, qui agit comme un activateur de celle-ci.

(4) gènes pro- et anti-apoptotiques :

Basé sur les termes GO, GO:0043066 et GO:0043065, respectivement les processus identifiés par le GO Consortium comme étant «*negative regulation of apoptosis*» et «*positive regulation of apoptosis*» nous pouvons catégoriser une partie des gènes selon leurs actions apoptotiques. Ainsi, quelle que soit la classe de modulation considérée, il semble que globalement les transcripts pro-apoptotiques soient plus nombreux que les transcripts anti-apoptotiques. Néanmoins, dans la classe de comportement B du tableau VIII (page 107), nous pouvons noter un équilibre entre les deux catégories de gènes.

	gene symbol	anti-apoptotique GO:0043066	pro-apoptotique GO:0043065	référence	cycle cellulaire GO:0007049	référence	NF-KB GO:0007249	référence
A	PDCD4				X(nég.rég)	(Kang, Ahn et al. 2002)		
	CARD17		X	IEA				
	CASP1		X	IEA			x (activation)	(Matsuda, Suzuki et al. 2003)
	CDKN2C		X	(Schreiber, Muller et al. 1999)	X	(Schreiber, Muller et al. 1999)		
	DAPK1	X	X	IEA				
	ERBB3	X		ISS				
	MAP3K5		X	(Ichijo, Nishida et al. 1997)				
	NUAK2	X		(Legembre, Schickel et al. 2004)				
	RARB		X	IEA				
	ZBTB16		X	IEA				
B	TNFRSF10D	X		(Degli-Esposti, Dougali et al. 1997)				
	AIFM2		X	(Wu, Xu et al. 2002)				
	BCL2A1	X		(Karsan, Yee et al. 1996)				
	VEGFA	X		IEA				
	BID		X	(Luo, Budihardjo et al. 1998)				
	CLN8	X		IEA				
	SERPINB9	X		(Bruneel, Labas et al. 2005)				
	NUP62	X		IEA			x (inhibition)	(Sanz, Sanchez et al. 1999)
	SFN		X	(Samuel, Weber et al. 2001)	X	IEA		
	TICAM1		X	IEA			x (activation)	(Matsuda, Suzuki et al. 2003)
C	TNFRSF10A		X	(Walczak, Degli-Esposti et al. 1997)			x (activation)	(Chaudhary, Eby et al. 1997)
	ALDH1A3		X	IEA				
	HIPK2	X	X	(Di Stefano, Blandino et al. 2004)				
	MAPK1		X	(Gabai, Yaglom et al. 2000)	X	IEA		
	MAPK8	X	X	(Yu, Minemoto et al. 2004)				
D	MAL		X	(Kohler, Hakansson et al. 1999)				
	NR4A1		X	IEA				
	STAT1		X	(Huang, Li et al. 2000)			x (inhibition)	(Wang, Wu et al. 2000)

Tableau VIII : Gènes apoptotiques organisés selon 4 termes GO : anti-apoptotique (GO:0043066), pro-apoptotique (GO:0043065), cycle cellulaire (GO:0007049), NF-κB (GO:0007249).

IEA: Inferred from Electronic Annotation, ISS: Inferred from Sequence or Structural Similarity

7. Le DMNQ induit une diminution de l'activité des caspases 3/7 dans la lignée porteuse de la mutation $\Delta F508$ par rapport à la lignée contrôle.

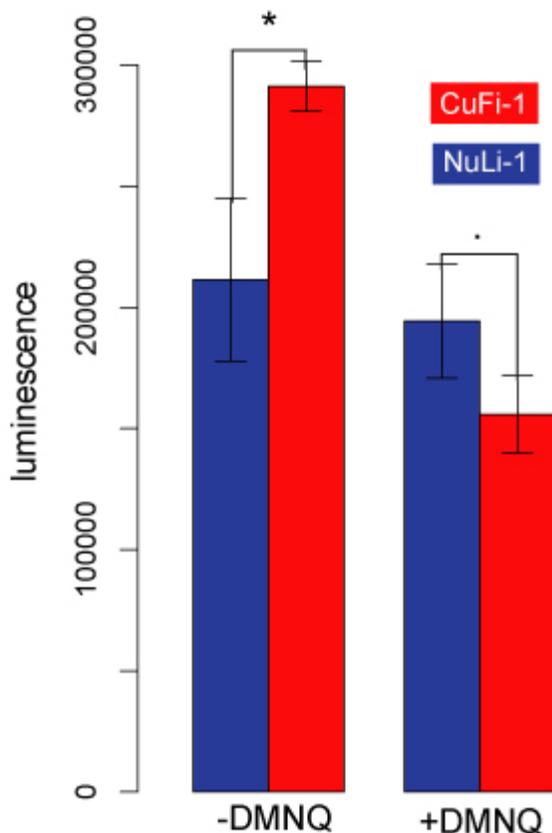


Figure 23 : Mesure de l'activité des caspases 3/7 dans les deux lignées cellulaires

Nous mesurons la luminescence 30 minutes après l'ajout de substrat dans chaque condition expérimentale testée. Les probabilités sont calculées avec le classique test t. Les probabilités inférieures à 0.05 sont notées "*", les probabilités inférieures à 0.1 sont notées ".".

La figure 23 (page 108) montre les luminescences mesurées dans les deux lignées cellulaires en absence et en présence de DMNQ. En absence de

La méthodologie d'analyse de notre expérience transcriptomique nous à orienter à mettre en évidence la mort cellulaire et l'apoptose comme étant des processus surreprésentés dans notre type cellulaire CuFi-1 par rapport au type NuLi-1. Pour aller plus loin et déterminer si nos découvertes ont un impact sur le phénotype apoptotique, nous évaluons l'activité des caspases 3/7 dans les deux lignées.

DMNQ, l'activité des caspases 3/7 est significativement supérieure dans le type CuFi-1 par rapport au NuLi-1 (Probabilité = 0.04). En présence de DMNQ, bien que la probabilité ne soit pas significative (Probabilité = 0.08), nous constatons un changement de phénotype : l'activité des caspases est plus faible dans le type CuFi-1 par rapport au type NuLi-1. De plus, le test ANOVA à 2 facteurs démontre une interaction significative entre la mutation et le traitement au DMNQ (Probabilité = 0.0007). Ces résultats suggèrent que l'apoptose pourrait être un processus perturbé dans les cellules FK et qu'en présence de stress oxydatif, les cellules développeraient une résistance à l'apoptose.

Discussion

A. Une vision globale du transcriptome

La technologie des puces à ADN est un outil de la biologie moléculaire très puissant, car elle est capable en une seule expérience d'évaluer l'activité transcriptionnelle d'environ 38000 gènes dans plusieurs conditions expérimentales. Son interprétation résulte de l'analyse produite, elle-même dépendante de nombreux critères. Il est important d'en souligner les différentes caractéristiques afin de mieux comprendre leurs potentielles implications sur les résultats et leurs interprétations.

1. Apports de l'expérience transcriptomique

Pour une meilleure compréhension des maladies génétiques, il est fondamentale d'étudier l'impact du défaut génétique sur l'intégralité du transcriptome afin de mieux comprendre la physiopathologie de la maladie. En outre, intégrer des variables qualitatives (sexe, âge, absence ou présence de traitement,...) dans le design expérimental permet de juger des interactions entre les différents facteurs et cela est particulièrement intéressant dans un contexte où la sévérité de la maladie peut être associée à plusieurs facteurs génétiques ou environnementaux. Par conséquent, la technologie qu'offre Affymetrix®, au travers de ces puces d'expression, s'avère être un outil adéquat pour ces études puisqu'elle offre à la fois une vision à la fois globale (au niveau des mécanismes) et détaillée (au niveau des gènes) du fonctionnement cellulaire dans plusieurs conditions expérimentales.

2. Limites de l'expérience transcriptomique

Toutefois, il est nécessaire de limiter le champ d'investigation de la technologie d'Affymetrix®. Dans notre cas, la puce à ADN, HG-U133 Plus 2.0 est une plateforme pangénomique, c'est-à-dire capable d'hybrider l'ensemble des transcripts polyadénylés produits par la cellule. Par conséquent, notre étude se

limite exclusivement à la modulation de ces transcripts. Basé sur ces modulations, les expériences de puce à ADN conduisent à émettre un ensemble d'hypothèses sur les processus cellulaires concernant l'action de certains facteurs de transcription (comme NF- κ B), les modifications épigénétiques, l'activation de voies de signalisation ou de voies métaboliques. Néanmoins, cette interprétation ne doit pas être surestimée, car l'expérience transcriptomique ne présume en rien de l'activité de protéome.

3. Nécessité de la validation expérimentale

Pour cette raison, il est indispensable de clarifier cette vision transcriptomique en la complétant avec des expériences plus spécifiques s'attachant à valider l'activation ou l'inactivation protéique.

Dans notre protocole d'étude du stress oxydatif, nous avons montré que le processus de l'apoptose est associé à un ensemble de gènes identifiés selon leurs variances. Toutefois, seule l'étude de l'activité des caspases 3/7 nous a amené à observer un potentiel défaut de l'apoptose dans les cellules FK par rapport aux cellules contrôles, en présence de stress oxydatif.

Notre analyse de l'étude inflammatoire nous suggère l'éventuelle implication de modifications épigénétiques dans la régulation de l'inflammation dans la cellule FK. Ceci ne restera effectivement qu'une hypothèse tant qu'elle ne sera pas validée, par exemple, avec une étude de *Chip and Chip* (Zheng, Barrera et al. 2007) de nos deux lignées, facilitant la mise en évidence de l'implication de méthyltransférases, ou d'acétylases dans l'expression de gènes inflammatoires situés sur les loci d'intérêt.

Parmi l'ensemble des études transcriptomiques présentées dans le tableau IX (page 112), une seule intègre dans son approche une étude d'activation protéique qui renforce considérablement ses conclusions biologiques par rapport aux autres (Verhaeghe, Remouchamps et al. 2007).

	Xu	Wright	Gambardella	Verhaeghe	Adib-Conquy
Interleukines	IL 1 β (\uparrow)	IL8(\uparrow)		IL8(\uparrow) IL6(\uparrow)	IL8(\downarrow) IL2(\downarrow) IL1 β (\downarrow)
Calgranulines	S100A8(\uparrow) S100A9(\uparrow)	S100A16(\downarrow)	S100A11(\uparrow) S100A8(\uparrow) S100A9(\uparrow)		
Chemokines			CXCL1(\uparrow) CXCL2(\uparrow) CXCL3(\uparrow)		CXCL9(\uparrow) CXCL10(\uparrow) CXCL11(\uparrow) CXCL12(\uparrow) CXCL5(\downarrow)
Autres	FOS(\uparrow) IRF1(\downarrow) TAP1(\uparrow)	STAT1(\uparrow)	STAT1(\uparrow)	ICAM1(\uparrow) TNFAIP6(\uparrow) PLAU(\uparrow) PLAUR(\uparrow) IL1RAP(\downarrow) IL1R(\uparrow) TLR4(\downarrow)	THBS1(\uparrow) FOS(\uparrow) IRF1(\uparrow)

Tableau IX : Modulations observées dans les autres expériences de puces à ADN des 128 gènes identifiés dans notre étude

Une description plus complète de ces expériences est présentée dans le tableau I

En résumé, les expériences de puce à ADN ne se suffisent pas à elles-mêmes, et doivent orienter le chercheur vers des expériences plus adaptées qui visent à confirmer des hypothèses, permettant d'affiner notre compréhension de la physiopathologie de la maladie. De cette façon, l'implication de plusieurs transcrits d'intérêt peut être validée dans la physiopathologie si et seulement si le mécanisme auquel ils sont associés est impliquée dans la maladie.

B. Le design expérimental

Trois éléments composent le design expérimental : le niveau de réplication, le modèle cellulaire, les conditions expérimentales. Ils participent pleinement au succès d'une expérience transcriptomique « grande échelle » et il est

nécessaire de tenir compte de leurs caractéristiques dans l'analyse des résultats.

1. Le niveau de réplication

Le niveau de réplication permet d'estimer la variabilité biologique présente dans notre expérience et ainsi de distinguer les variations liées des conditions expérimentales de celles dépendantes de la variabilité biologique (le bruit de fond). Dans notre cas, nous utilisons trois réplicats biologiques. Bien que cela puisse paraître un faible niveau de réplication, nous obtenons une bonne reproductibilité des niveaux d'expression dans les puces d'une même condition sans doute en raison de l'usage de lignées transformées.

2. Les modèles cellulaires

Les modèles utilisés dans notre expérience sont issus de cellules épithéliales pulmonaires humaines et ils ont été choisis dans le cadre de notre étude pour plusieurs raisons.

- (1) Ces modèles s'avèrent pertinents, car ils ont été développés pour l'étude de fonctions physiologiques de ce type cellulaire (Zabner, Karp et al. 2003).
- (2) L'entretien et la manipulation de lignées cellulaires sont plus aisés que des cellules primaires.
- (3) Leurs réplications biologiques présentent moins de variabilité que dans une étude utilisant des cellules primaires, étant donné que ce sont des cellules issues d'un même individu. Ceci a pour avantage de minimiser le nombre de puce à ADN à hybrider et d'obtenir une expression des transcrits consistante entre les différentes puces hybridées.
- (4) le modèle CuFi-1 est porteur de la mutation la plus fréquente dans la population atteinte par la maladie : la mutation $\Delta F508$.

Néanmoins, comme tous les systèmes biologiques, ces lignées possèdent des défauts.

(1) Le fond génétique des deux lignées cellulaires est différent, car issues de deux personnes distinctes.

(2) Leurs caryotypes sont différents. En effet, les cellules CuFi-1 possèdent les chromosomes sexuels mâles et les NuLi-1, les chromosomes sexuels femelles. Ceci a pour conséquence de biaiser l'analyse différentielle car même si nous prenons soin d'écartier de notre analyse les transcrits codés par les chromosomes sexuels, nous ne pouvons pas présumer des conséquences liées aux mécanismes de transmission et de modifications épigénétiques, comme le « *genomic imprinting* », phénomène génétique où des gènes autosomaux s'expriment différemment chez les hommes et les femmes (Reik et Walter 2001).

(3) Ces cellules transformées possèdent les caractéristiques de cellules cancéreuses et bien que cela affecte les deux lignées utilisées cela pourrait avoir un impact différent sur la modulation des gènes.

(4) Elles ne permettent que de faire des interprétations dans un contexte *in vitro* et leurs modulations transcriptionnelles peuvent être dépendantes des conditions de culture.

3. Les conditions expérimentales

Dans notre design, nous avons inclus deux conditions: la mutation et le stress oxydatif. Dans chacune des deux lignées cellulaires, ce stress est induit par une quinone, le DMNQ. Nous avons choisi cette molécule, car sa cytotoxicité est obtenue par le stress oxydatif induit via le cycle d'oxydoréduction (Gant, Rao et al. 1988), système impliquant le glutathion, qui est très proche des conditions de gestion de stress *in vivo*. Malgré cela, il est important de considérer que cette molécule exogène, à priori absente des voies aériennes

des patients FK, a un impact spécifique sur la modulation des transcripts. Pour minimiser cet impact, nous optimisons l'emploi d'une dose et d'un temps de traitement pour générer un stress oxydatif suffisamment élevée (en termes de molécule de glutathion totale intracellulaire) et cela sans affecter à l'intégrité structurelle de la cellule.

C. Le prétraitement et le filtrage des données d'expression

De l'hybridation à la production de données d'expression, un grand nombre d'étapes techniques se succèdent. À chacune d'entre elles, les données d'expression sont traitées et subissent des modifications nécessaires pour améliorer l'analyse.

Le prétraitement des données d'expression comprend trois parties :

- (1) Le contrôle qualité a pour but de vérifier que les différentes étapes techniques n'ont pas altérées ces données.
- (2) La normalisation, quant à elle, s'assure que les données d'expression issues de différentes puces à ADN soient comparables (notre analyse comprend 12 puces et il est important de les comparer sur une même base).
- (3) Finalement, le filtrage permet d'éliminer les sondes non pertinentes dans notre analyse différentielle.

1. Le contrôle qualité

Le contrôle qualité des données d'expression est indispensable, car (1) écarte les données de faible qualité, (2) maximise la valeur scientifique de l'information issue de cette technologie et (3) permet de mettre en avant les modulations biologiques. Certains auteurs présument que 20 % des données issues d'expériences transcriptomiques enregistrées dans des bases de données publiques, largement disponibles par la communauté scientifique,

sont de qualité questionable (Larsson, Wennmalm et al. 2006). Plusieurs études ont démontré l'importance de l'étape du contrôle qualité dans l'analyse de ces expériences (Larkin, Frank et al. 2005; Heber et Sick 2006).

Les paramètres de contrôle qualité mis au point par Affymetrix® se révèlent, à priori, efficaces pour la détection de données biaisées (documentation Affymetrix®, Inc. (2003)) et des librairies de R offrent des outils de contrôle qualité pour cette technologie (Wilson et Miller 2005).

Après un examen rigoureux des critères de contrôles qualités définis par Affymetrix®, nous démontrons la bonne qualité des ARNm utilisés et une bonne efficacité de chaque étape du protocole d'hybridation. Ceci nous autorise à utiliser l'intégralité des puces hybridées pour nos analyses ultérieures.

2. La normalisation.

Le contrôle qualité réalisé au préalable élimine les biais les plus importants mais ne détecte pas les différences systématiques subtiles qui existent entre les différentes puces composant l'expérience. Ces différences ne sont pas liées à des véritables variations biologiques et peuvent être atténuées par l'étape de normalisation permettant de rendre les données de chaque puce comparable entre elles. Bien qu'il existe plusieurs méthodes de normalisation (lowess, quantile,...), nous choisissons la méthode RMA (Robust MultiChips Analysis), laquelle produirait une normalisation de meilleure qualité. Une étude de comparaison de trois méthodes classiques de normalisation sur un jeu synthétique montre que RMA fournit une estimation précise de l'expression en minimisant les biais et se révélerait être le meilleur choix (Bolstad, Irizarry et al. 2003). Bien qu'elle ait été développée pour des données d'Affymetrix®, cette méthode n'utilise pas pleinement l'information disponible. En effet, l'intensité fournit par les sondes MM (MisMatch), ne sont pas exploitées dans la

normalisation des données d'expression et l'auteur argumente ce point par le fait que tenir compte de ces données génère plus de bruit que d'information exploitable (Irizarry, Bolstad et al. 2003; Irizarry, Hobbs et al. 2003).

3. Le filtrage des données

La dernière étape avant l'analyse différentielle est le filtrage des données d'expression. Celle-ci n'est pas obligatoire et peut prendre plusieurs formes en fonction de la méthodologie d'analyse.

Dans nos analyses, nous procédons à deux types de filtrage sur nos données. Le premier est lié à un biais de notre design expérimental, puisqu'il élimine les transcrits codés par les chromosomes sexuels basé sur les annotations génomiques, évitant ainsi leurs analyses. Le second type de filtrage est beaucoup plus courant dans les expériences transcriptomiques et vise à réduire l'information pertinente à analyser. De cette façon, dans notre seconde méthode d'analyse, le filtre utilise la variance des données d'expression induite par les conditions expérimentales et écarte les transcrits avec de faibles changements dans leur niveau d'expression. Pour réaliser cela, différentes méthodes ont été proposées, basées sur la moyenne d'expression, sur la variance ou sur les plus grandes variations des données. Les recherches sur le sujet n'ont pas montré l'efficacité d'une méthode par rapport aux autres mais qu'elle pouvait tout de même participer à optimiser le résultat final (Hackstadt et Hess 2009). Bien souvent, le choix de la mesure à considérer et le seuil choisi peuvent paraître arbitraire.

Le type de mesure utilisé pour le filtrage et le seuil utilisé ont un effet sur le résultat final, comme par exemple l'usage de la variance pourrait induire une sélection biaisée en raison de la violation des présomptions nécessaires à son application.

L'IQR (Inter-Quartile Range) est une mesure de dispersion plus robuste que la variance. Elle permet de caractériser au mieux les niveaux d'expression d'un transcript puisqu'elle tient compte de 50 % des données ordonnées autour de la médiane, n'est pas biaisée par des données extrêmes présentes et non sensible à l'absence de normalité ou d'hétérogénéité de la variance. Par conséquent, dans notre approche nous choisissons l'IQR comme mesure de filtrage. Pour notre analyse nous choisissons les transcripts avec un IQR supérieur ou égale à 0.5. Cela nous permis de réduire la quantité d'information d'un facteur 14 par rapport à l'ensemble initial de 52 908 sondes (excluant les sondes hybridant les transcripts codées par les chromosomes sexuels).

Le choix du seuil de l'IQR est conditionné par la littérature précisant que 0.5 est « un choix raisonnable » pour Robert Gentleman, chercheur très productif dans le domaine de la biologie computationnelle. Celui-ci précise, dans les documentations de librairie *genefilter* (*Gentleman*) que le seuil de 0.5 permet de retenir des gènes à haut niveaux d'expression. Toutefois, nous ajoutons à cela, qu'à ce seuil, sont sélectionnés des transcripts dont 50 % des données d'expression présentent une variation au minimum d'un facteur 1,4 (2 exposant 0.5, car les données d'expression sont exprimées en \log_2). Par cette approche, nous sommes intéressés par une vision globale des perturbations cellulaires donc même si notre sélection est bruitée pour le choix d'un IQR trop large, ceci ne peut pas totalement affecter la vue d'ensemble que nous offre cette sélection. Néanmoins, pour obtenir un signal plus spécifique de notre panorama biologique, nous produisons, en parallèle, des analyses similaires sur cinq autres groupes de transcripts sélectionnés selon 5 seuils d'IQR plus restreints : 0.6, 0.7, 0.8, 0.9, 1. Ceci a pour conséquence de créer des groupes de tailles plus petits composés de transcripts avec une variabilité d'expression plus élevée. Le seuil limite de 1 nous assure que la sélection ne contient que des transcripts dont la modulation est au minimum d'un facteur 2 (2 exposant 1) entre les conditions expérimentales. A ce seuil limite de 1, nous

sélectionnons 78 transcripts Sélectionner des transcripts avec des IQR supérieurs à 1 engendre un groupe trop petit de transcripts pour permettre une interprétation biologique fiable.

De cette façon, les six groupes de transcripts sélectionnés selon un intervalle d'IQR de 0.5 à 1, possèdent une information similaire en terme de processus biologiques à ceci près qu'elle est plus ou moins bruitée. En d'autres mots, le groupe le plus important composés de 687 éléments présente un bruit important et apporte une vision globale des processus perturbés. A l'autre extrême, le groupe composée de 78 transcripts à variance élevée comporte un bruit de fond très faible et apporte une vision plus restreinte des processus perturbées.

Une analyse en parallèle des ces groupes nous permet de juger de l'impact de ce bruit sur la détection de fonctions biologiques significativement surreprésentées et de cette façon de mettre en évidence les processus biologiques qui apparaissent communs dans ces différentes analyses.

D. Les développements et limites de nos analyses différentielles

Les méthodes d'analyse différentielle se sont développées les 2 dernières décennies et se sont ajustées selon un compromis entre la spécificité et la sensibilité de la détection (Pan 2002; Weeraratna et Taub 2007). Néanmoins, quelle que soit la méthode employée, un biais existe toujours et l'investigateur de l'expérience doit en avoir conscience afin d'estimer son impact et permettre la bonne interprétation des tests statistiques. Dans cette section, nous détaillons les méthodes utilisées dans l'analyse univariée, pour l'analyse de design comprenant une condition à un niveau (cas de notre première analyse), et multivariée pour les designs à deux facteurs à plusieurs niveaux (cas de notre seconde analyse).

1. Aspects statistiques de l'analyse univariée

La plupart des expériences transcriptomiques réalisées étudient une seule condition expérimentale, de ce fait, la communauté scientifique a mis beaucoup d'effort dans le développement de méthodes statistiques univariées. Historiquement, la première méthode proposée pour sélectionner des transcrits d'intérêt dans une expérience est la méthode dite des ratios. Encore largement exploitée pour son côté simple et intuitif, celle-ci compare les niveaux d'expression d'un transcript dans deux conditions expérimentales et fait la différence des moyennes des réplicats de chaque condition. Les valeurs d'expression étant généralement exprimées en log de base 2, cette différence des moyennes équivaut à leur ratio et informe sur une modulation éventuelle du transcript. Ainsi, pour l'ensemble d'une expérimentation, cette analyse met en évidence les transcrits fortement modulés et d'autres très faiblement modulés. Toutefois, il est important de se demander jusqu'à quel point ce ratio est intéressant en termes de modulation. Pour répondre à cette question, la façon la plus radicale est de s'intéresser uniquement aux transcrits avec un ratio supérieur à 1 entre les deux conditions (Draghici 2002). Cette approche, connue sous le nom de "*two fold-change*", a été utilisée dans de nombreuses études (DeRisi, Iyer et al. 1997; Sudarsanam, Iyer et al. 2000; Jiang, Tsien et al. 2001) mais, en soi, n'est pas un test statistique et présente deux inconvénients majeurs : elle dépend d'une valeur arbitraire et elle sélectionne un excès de sonde de faible intensité car leurs ratios ont une plus large variance que les gènes à forte intensité (Newton, Kendzioriski et al. 2001). Ainsi, cette technique sélectionne des sondes avec de bas niveaux d'expression et qui sont déclarées différentiellement modulées alors qu'elles ne le sont pas, qualifiées de faux-positif, et des sondes ayant des hauts niveaux d'expression déclarées non différentiellement modulées alors qu'elles le sont, qualifiées de faux-négatif (Miller, Galecki et al. 2001). Pour améliorer le pouvoir de détection de cette méthode, il est donc essentiel d'y intégrer la

variabilité présente dans les données d'expression. Cette variabilité, représentée par la variance, représente le bruit de notre expérience et en tenir compte lors de la sélection des sondes différentiellement exprimées augmente considérablement la sensibilité de la détection d'une modulation.

La variance observée dans la mesure de l'expression d'un gène peut être séparée en deux origines: la variabilité biologique et la variation technique. La variabilité biologique est inhérente à la cellule, à son patrimoine génétique, à sa capacité à réagir à son environnement. Quant à la variation technique, elle est induite par chaque étape du protocole expérimental.

Le test de Student (Student 1908) intègre la variance présente dans les données d'expression. Il compare la différence des moyennes des deux groupes par rapport au bruit de l'expérience, représenté par la variance des deux groupes. En d'autres termes, la statistique « t » de Student, peut-être vu comme une amélioration de la méthode du « *fold-change* » puisqu'elle est définie comme étant un ratio entre la différence des moyennes des deux groupes et leurs variabilités.

Cependant, son interprétation peut-être biaisée pour de nombreuses raisons :

(1) Les expériences de puce à ADN pouvant être couteuses (cela devient de moins en moins vrai), peu de répétition par condition sont employées, ainsi l'estimation de la variance repose sur un petit nombre d'échantillon (au minimum 3). Ainsi, pour l'ensemble des transcrits analysés dans une expérience, certaines variances sont alors sous estimées et d'autres surestimées, conduisant respectivement à des faux-positifs et des faux-négatifs.

(2) Une interprétation valide du test de Student présume que les variances estimées pour chaque gène composant une expérience sont les mêmes. Cela ne semble pas être le cas en pratique. Hung a montré qu'il existe une relation entre la variance et le niveau d'expression

mesuré : des gènes avec des niveaux similaires présentent des variances similaires et plus leur niveau d'expression est élevé, plus leur variance l'est également (Hung, Baldi et al. 2002).

(3) En utilisant la distribution de Student à n_1+n_2-2 degrés de liberté (n_1 et n_2 étant le nombre de réplication dans les deux conditions n), la statistique t calculée, peut-être transformée en une probabilité de rejeter l'hypothèse nulle alors qu'il faudrait l'accepter. En d'autres termes, cela nous donne la probabilité de déclarer un gène modulé alors qu'il ne l'est pas. Cette probabilité est connue sous le nom de Probabilité. Toutefois, le bon usage de cette distribution présume une indépendance des observations. Dans notre cas, où il existe un degré de dépendance entre les données (de part leurs implications dans des voies métaboliques communes, leurs régulations similaires) cette présomption n'est pas respectée.

(4) Les données d'expression, de part leur nature biologique, ne sont pas distribuées normalement. Afin d'assurer la validité d'un test paramétrique, ces données sont transformées et pour cela, plusieurs méthodes ont été développées comme Lowess, MAS5 ou RMA (Steinhoff et Vingron 2006). Toutefois, l'effet de ces transformations sur l'analyse différentielle a des conséquences qui pourraient affecter l'interprétation des résultats (Karen Kafadar 2003) mais en absence d'un jeu de donnée de référence, l'impact de l'étape de normalisation sur le résultat d'une analyse différentielle n'est pas connue, n'y documentée.

Pour toutes ces raisons, des améliorations du test t ont été proposées par certains auteurs comme par exemple, Tusher et al en 2001 (Tusher, Tibshirani et al. 2001), Efron et al en 2002 (Efron et Tibshirani 2002). L'un des développements le plus intéressant en la matière fut les approches probalistiques adaptées à partir du test de Student (Baldi et Long 2001; Huber, von Heydebreck et al. 2002; Durbin et Rocke 2004). Dans la version classique

du test de Student, la variance et la moyenne sont fixées, correspondant à une estimation basée sur un échantillon. Ces valeurs sont ensuite utilisées par établir la statistique d'intérêt. Cette approche est qualifiée de fréquentiste car elle se base sur la fréquence d'apparition d'un événement donné laquelle se stabilise lorsque l'événement est réalisé un grand nombre de fois. L'approche probabilistique, quant à elle, repose sur la connaissance des données d'expression, et informe sur la vraisemblance des paramètres calculés (moyenne, écart-type), sous la forme de probabilité. En d'autres termes, elle mesure la crédibilité d'une proposition donnée, compte tenue des informations disponibles. Ainsi, cette approche utilisant les mathématiques bayésiennes, permet l'estimation de la variance de donnée d'expression, présumant que celle-ci suit une certaine distribution dont on connaît la moyenne et l'écart-type. Cette distribution est qualifiée de « connaissance à priori ». En fonction de l'observation des données expérimentales, on détermine la vraisemblance de la variance. Ainsi on peut déterminer une distribution de la probabilité à posteriori, définie comme la combinaison entre la vraisemblance et la connaissance à priori de la variance.

Basé sur cette subtilité pour estimer la variance et pondérer le dénominateur dans le calcul de la variable t , plusieurs variantes du test de Student, communément appelés « test t modifié » ont été développées. En 2001, Pierre Baldi présente sa version (Baldi et Long 2001; Huber, von Heydebreck et al. 2002; Durbin et Rocke 2004), ensuite Lönnstedt et Speed en 2002 propose un modèle combinant une approche linéaire à une estimation de la variance par une méthode bayésienne. De plus ils définissent une nouvelle statistique, B ou « *log posterior odds ratios* » comme étant le ratio entre la probabilité qu'un gène donné soit différentiellement exprimé sur la probabilité que ce gène ne soit pas différentiellement exprimé (Speed 2002). Une autre variante de cette solution, adaptée à la problématique des puces à ADN, a ensuite été

développée par Smyth en 2004 (Smyth 2004). Cette version est implémentée dans le package Limma de Bioconductor (Smyth 2005)

Cette statistique t , corrigée par les modifications apportées, ne résout que le problème de l'estimation de la variance mais n'améliore pas le problème de l'interprétation de cette statistique par rapport à la distribution de Student à n_1+n_2-2 degrés de liberté. Pour cette raison, une autre méthode d'estimation de la probabilité est indispensable. Lorsque nous testons une différence d'expression entre deux conditions expérimentales, la probabilité, issue du test statistique quel qu'il soit, représente la probabilité que la différence de moyenne observée soit due à la chance. De cette façon, pour l'exemple, une expression différentielle avec une probabilité égale à 0.048 devra être interpréter comme suit « il y a 4,8 % de chances que la différence observée soit due au hasard » ou « il y a 4,8 % de chances de se tromper en déclarant ce gène différentiellement exprimé » (Pounds 2006). Sans faire appel à la distribution de Student, nous pouvons déterminer la probabilité par la méthode des permutations (Westfall P.H. 1993). Toutefois, l'inconvénient majeur de cette méthode, est la nécessité d'avoir un niveau de réplication suffisamment grand afin de permettre un grand nombre de permutation. Par exemple, si $n = 3$ pour chacune des deux conditions, nous pouvons permuer les données d'expression de 20 façons $(3 \times 2)! / (3! \times 3!)$, la statistique t calculée pourra être comparée uniquement à une distribution composée 20 pseudo statistiques t différentes. Cette approche par permutation présente l'avantage de transformer un test paramétrique (ANOVA ou Student), où la statistique calculée est comparée à une distribution de référence en un test non paramétrique puisque la comparaison se fait par rapport à une distribution de statistiques calculées t issues du ré-échantillonnage des données d'expression.

Un autre aspect à considérer dans l'analyse statistique d'une expérience de puce à ADN est le problème des comparaisons multiples. Pour chaque sonde de l'expérience nous testons l'hypothèse nulle, c'est-à-dire l'absence de différence d'expression entre les deux conditions. Aussi, chaque test réalisé s'accompagne d'une prise de risque, désignée par « α », et bien arbitrairement associée à la valeur 0.05 ou 0.01. Concrètement, cela signifie que chaque fois que nous testons l'hypothèse nulle, nous avons 5 % de chance de la rejeter, alors qu'il faudrait l'accepter. Puisqu'une expérience de puce est composée d'un nombre important de sondes (dans notre cas, 54 675 sondes reflétant le niveau d'expression de 38 000 transcripts environ), le test statistique est répété autant de fois. Par conséquent, il est nécessaire d'estimer la quantité de faux-positifs générée par ces tests consécutifs, afin de juger de la qualité du résultat statistique produit et de fournir une interprétation qui en tient compte. Par exemple, si le résultat d'une analyse statistique nous indique que 100 sondes sont différentiellement exprimées entre les deux conditions, nos conclusions sur une interprétation biologique n'ont pas la même pertinence si nous savons que ce groupe contient 10 % ou 50 % de faux-positifs.

Pour cela, plusieurs méthodes ont été développées (Šidák 1967 ; Holm 1979; Hochberg 1988; Westfall 1993). Celles-ci cherchent à ajuster la probabilité obtenue à partir du test statistique en fonction du nombre de faux positif potentiellement présent. Néanmoins, ces méthodes sont très conservatrices, c'est-à-dire que l'ajustement des probabilités ne produit plus de résultats significatifs d'un point de vue statistique. Parmi ces méthodes, celle proposée par Benjamini et Hochberg en 1995 (Benjamini 1995), le *False Discovery Rate* (FDR) est devenue un standard dans la communauté des puces à ADN, pour sa méthode de calcul simple mais également parce qu'il a été démontré qu'elle reste efficace avec des données présentant un certain niveau de dépendance dans les données (Benjamini Y 2001).

Dans la première partie de notre analyse, nous comparons deux conditions expérimentales: la condition NuLi-1 (absence de mutation) *versus* la condition CuFi-1 (présence de la mutation $\Delta F508$). Par conséquent l'usage d'un test univarié s'applique. Après une normalisation quantile des données, nous testons l'expression différentielle entre les 2 conditions expérimentales en utilisant la version modifiée du test t proposée par Smyth en 2004. Pour pallier au problème des comparaisons multiples, nous utilisons la méthode développée par Benjamini et Hochberg en 1995. L'ensemble des caractéristiques statistiques décrites précédemment sont incluses dans le package Limma et pour cette raison nous l'utilisons dans notre protocole d'analyse.

2. Caractéristiques de notre analyse multivariée

Les méthodes statistiques décrites dans la section précédente sont utilisées dans le cadre d'analyse univariée et ne sont pas applicables pour l'analyse de design exploitant plusieurs conditions expérimentales. De plus, les designs expérimentaux à une condition et à deux conditions visent à répondre à des questions biologiques d'intérêts différents : le design univarié pose la question : « quels sont les transcripts significativement modulés entre les deux groupes expérimentaux testés ? ». Tandis que le design multivarié cherche à isoler des groupes de transcripts avec des modulations similaires reflétant une co-régulation, et à identifier une interaction entre les conditions.

La classification est la première méthode utilisée par l'analyse de design à plusieurs conditions. Basé sur une mesure de ressemblance, celle-ci regroupe les gènes avec des comportements similaires. Ainsi, en 1998, Khan et al, montra que des tissus d'origine histologiquement différents présentent des regroupements de gènes à modulations similaires (Khan, Simon et al. 1998). De la même façon, le *clustering* hiérarchique, un algorithme de classification, a

permis la découverte de deux sous types des lymphomes des cellules B, indétectables par les méthodes traditionnelles de classification histologiques (Alizadeh, Eisen et al. 2000).

Bien qu'elles se soient révélées fructueuses, ces méthodes de classification numérique présentent quelques défauts :

(1) Le résultat issu de ces regroupements est grandement conditionné par les choix qui définissent les paramètres d'analyse: l'algorithme de classification (séquentiel, agglomératif, hiérarchique, probabiliste..), la mesure de ressemblance utilisée (similarité, distance, dépendance..), le seuil de regroupement et le nombre de classes à créer.

(2) Selon l'algorithme utilisé, celui-ci pourrait ne pas converger vers une solution unique, c'est-à-dire que la répétition de 100 procédures de classification, basée sur les mêmes paramètres peut produire 100 résultats différents.

(3) Les algorithmes de classification sont très sensibles au bruit de fond généré par les données elles-mêmes, c'est-à-dire qu'ils distinguent difficilement les profils d'expression propres aux données (bruit de fond) et les profils d'expression induits par les conditions expérimentales (structures d'intérêt).

Parce que les outils de classification permettaient de répondre à des questions d'ordre biologiques, de grands efforts ont été mis pour savoir quels étaient les algorithmes les plus adaptés à l'étude des données transcriptomiques (Eisen, Spellman et al. 1998). Mais étonnement, peu d'intérêt a été porté à la validité statistique des résultats produits (Bittner, Meltzer et al. 2000; Kerr et Churchill 2001).

Il pourrait être tentant d'utiliser le test ANOVA, l'équivalent multivarié du test t, pour juger de la différence de modulation des transcripts entre les groupes. Toutefois, celui-ci nécessite des suppositions qui ne sont pas toujours

satisfaites, ce qui par conséquent, rend son interprétation discutable. Malgré cela, cette approche statistique a été largement utilisée dans l'analyse de puce à ADN (Pavlidis 2003; Churchill 2004), et comme le test t, le test ANOVA a subi quelques améliorations mais, néanmoins sans se défaire de ses principaux défauts (Kerr et Churchill 2001). Toutefois, une analyse ANOVA à deux facteurs réalisée sur nos données d'expression permet de juger de l'impact global des conditions expérimentales.

Une autre approche a été proposée pour l'étude de ces designs complexes : l'ordination en espace réduit. Les méthodes d'ordination présentent l'avantage de réduire la taille des données d'expression, et d'exprimer les données selon les principales tendances de variabilité. Plusieurs techniques ont été développées et utilisées dans l'analyse de données multifactorielles (Fellenberg, Hauser et al. 2001; Yeung et Ruzzo 2001; Culhane, Perriere et al. 2002). La plus connue de ces méthodes est l'analyse en composante principale (ACP) et s'est révélée efficace dans l'examen d'expérience de puce à ADN pour étudier les associations existantes entre les gènes. Ainsi, en comparant plusieurs expériences étudiant des tumeurs, celles-ci ont pu être classées en fonction du comportement de groupes de gènes. De la même façon que les techniques de classification, cette approche a permis la découverte de sous-classes de tumeurs (Crescenzi et Giuliani 2001). L'ACP a également été employée dans l'étude des expériences en « *time-course* » pour isoler les gènes qui sont co-régulés dans le temps (Raychaudhuri, Stuart et al. 2000) . En pratique, l'ACP ressemble aux méthodes de classification des données présentées mais utilisent comme critère la variance des données d'expression. Ainsi, l'ACP cherche à résumer les données en plusieurs variances orthogonales (l'orthogonalité entre les variances traduit l'absence de corrélation linéaire entre elles). Les variances, sont appelées des composantes principales et définissent les axes dans une représentation graphique plane, permettant de dégager visuellement des structures dans les données.

Dans notre analyse, nous utilisons l'analyse canonique de redondance (ACR), méthode développée par Van den Wollenberg (van den Wollenberg 1977). Cette méthode, similaire à l'analyse en composante principale, tient compte de deux matrices : la matrice réponse (les données d'expression des transcrits dans chaque condition) et la matrice explicative (les conditions expérimentales). Elle cherche à étudier la relation existante entre les deux matrices et tente d'expliquer les données de la matrice réponse en fonction de la matrice explicative. Ainsi la variance des données d'expression de notre expérience peut être expliquée (et décomposée) selon les deux conditions expérimentales, la mutation et le stress oxydatif.

Cette approche nous permet de pallier aux problèmes des méthodes paramétriques, où des présomptions sont nécessaires. De plus, elle permet une réduction des données qui fournit une représentation visuelle des données d'expression explicite permettant une interprétation de l'effet des deux variables sur les données d'expression. Néanmoins, cette approche nécessite une présomption qu'il est nécessaire de préciser. La première étape de l'ACR consiste à faire une régression multiple des variables réponses sur les variables explicatives, modélisant de façon linéaire les relations qui existent entre les deux variables. Ceci suppose, en effet qu'il existe une relation linéaire entre les variables réponses et explicatives, ce qui est rarement observées dans des processus biologiques. Legendre et Makarenkov (Legendre 1999) ont développé une méthode polynomiale qui permet de modéliser des relations non linéaires, et concluent que le résultat de cette méthode est dépendante de la nature du jeu de donnée et ne démontre pas qu'elle soit supérieure au modèle linéaire proposé par l'ACR. Pour ces raisons, nous n'exploitons pas cette méthode. Toutefois, il sera intéressant de comparer les résultats produits par cette approche non-linéaire car l'un des ces avantages est qu'elle permet d'expliquer une partie de la variance, qui apparaît inexpliquée dans l'ACR.

E. Les facteurs impliqués dans l'interprétation biologique

Le prétraitement, le filtrage et l'analyse différentielle des données d'expression s'intéressent aux données numériques et cherchent ainsi à optimiser la découverte d'information pertinente. Cette information pertinente, généralement présentée, dans un premier temps, sous la forme d'une liste de sondes différentiellement modulées ou caractérisées par la nature de leurs variances, doit être transformée en information biologique. Ainsi, basé sur les annotations génomiques, chaque sonde peut être identifiée et associée à une ou plusieurs fonctions biologiques et doivent conduire à une interprétation biologique de nos données transcriptomiques.

1. De la sélection de transcripts à l'information biologique

A l'aide d'une méthode de classification, nous pouvons interpréter les modulations de notre sélection vis-à-vis des conditions expérimentales. Ceci est une étape courante dans les analyses de puces à ADN pour identifier les co-régulations géniques. La méthode de classification utilisée dans ce travail est le *fuzzy c-means* (FCM) (Bezdek 1981) et été largement employée dans l'analyse transcriptomique (Gasch et Eisen 2002). Cet algorithme de partitionnement sépare l'ensemble des données en k groupes (k étant désigné au préalable par l'utilisateur) rassemblant des profils d'expression similaires. L'avantage de FCM par rapport à d'autres algorithmes est sa flexibilité dans le processus de partitionnement. Dans des méthodes comme le *k-means* ou SOM (*Self-Organizing Map*), un élément ne peut appartenir qu'à un groupe et à un seul et cette appartenance est liée à la similarité de son profil d'expression avec d'autres éléments. Par conséquent, ils ne peuvent pas identifier les gènes dont l'expression est similaire à plusieurs éléments appartenant à des groupes différents, masquant de ce fait les multiples relations existantes. L'algorithme FCM autorise l'appartenance d'un élément à

plusieurs classes permettant ainsi de considérer de multiples relations existantes entre les transcripts. Cette notion d'appartenance multiple est importante à considérer dans un système biologique où les transcripts interagissent entre eux, sont co-régulés.

Comme tous les algorithmes de classification, il est nécessaire de choisir des paramètres qui peuvent conditionner le résultat final. Dans l'usage de FCM, deux paramètres sont à définir: k , le nombre de classe à créer et m , le paramètre d'appartenance floue à une classe.

Le paramètre m détermine le degré d'appartenance à un groupe : une valeur proche de 0 indique une faible association au groupe et une valeur proche de 1 indique une forte association au groupe. Si ce paramètre est égal à 1, alors l'algorithme FCM se comporte comme un *k-means* et chaque élément est associé à un groupe et à un seul.

Le choix du paramètre m n'est pas aisé. Après avoir appliqué, sans succès, la méthode proposée par Matthias E. Futschik dans sa thèse (<http://itb.biologie.hu-berlin.de/~futschik/PhD/soft.pdf>), pour déterminer le paramètre m , nous nous sommes intéressés à la stabilité des classes déterminées en fonction de ce paramètre. Cette analyse nous a permis de déterminer que les valeurs de m dans un intervalle de 1.05 à 2.55 retournent toujours la même composition de classe, donc assure une stabilité du résultat de la classification. Pour des valeurs de m supérieures à 2.55, la composition des classes devient très variable. Pour ces raisons notre analyse, nous avons choisi une valeur de m égale à 1.25.

De façon similaire, le choix de la valeur k , le nombre de classe à créer n'est pas simple. Après avoir testé quelques méthodes de prédiction de k sur nos données d'expression, lesquelles se sont révélées infructueuses, nous avons utilisé une approche empirique pour déterminer ce paramètre. Tout d'abord, nous avons réalisé notre analyse avec $k= 16$, en considérant que chaque conditions expérimentales peut induire deux comportements (sur-

expression/sous-expression), nous pouvons espérer, en théorique, obtenir 16 profils d'expression différents (2^4). Les résultats de cette analyse nous ont montré que les 16 profils d'expression peuvent être classés en 4 classes de comportements.

Repartis en 4 groupes, les profils des d'expressions nous conduisent à deux conclusions vis-à-vis des conditions expérimentales (figure 22):

(1) Le DMNQ n'influence pas la modulation induite par la mutation, c'est-à-dire que les transcripts présentant une modulation entre les conditions FK et non FK (soit sur-expression soit sous-expression), ont une modulation similaire en présence de stress oxydatif.

(2) Le DMNQ influence le niveau basal d'expression des conditions FK et non FK. Cet effet est nettement accentué dans les groupes deux des quatre groupes (figure 22 A et B), et moins prononcé dans les deux autres (figure 22 C et D).

En considérant le comportement du transcriptome dans sa totalité, cela suggère que le stress oxydatif en présence de la mutation influence le niveau d'expression des gènes soit par une diminution (figure 22 A et D) soit par une augmentation (figure 22 B et C) du niveau basal. Par conséquent, le stress oxydatif influence l'activité de la machine transcriptionnelle.

2. Les annotations génomiques

Pour produire une interprétation biologique à partir d'un ensemble de sondes sélectionnées, il est indispensable d'utiliser des annotations génomiques. Parce qu'il existe des discordances dans les nombreuses bases de données, nous nous avons choisi de vérifier ces informations utilisées dans notre analyse par trois bases des données : NetAFFX, Ensembl, Resnet. Étant donné les efforts pour assurer un contrôle des annotations génomiques, les discordances existantes entre NetAFFX et Ensembl ont tendance à diminuer

considérablement ces dernières années. L'usage de Resnet dans ce contrôle nous permet d'assurer une existence protéique au transcript, et par conséquent lui assurer une fonction. Toutefois, sur l'ensemble de 9 760 sondes isolées dans notre première approche, 6 478 d'entre elles ont des identifiants protéiques dans Resnet. Ceci souligne que seule une partie de l'information issue de la plateforme protéique est exploitable à un niveau protéique. Par conséquent, une interprétation future de l'expérience avec des annotations plus complète sera certainement fructueuse.

3. L'analyse ontologique

L'analyse ontologique d'un groupe de gènes sélectionnés permet de déterminer quels sont les processus biologiques surreprésentés dans le groupe et laissent présager des fonctions biologiques modulées par les conditions expérimentales, et par extension permet de produire des hypothèses pour expliquer la physiopathologie de la maladie.

L'analyse ontologique est largement utilisée par la communauté scientifique pour l'interprétation biologique de résultats de puces à ADN. Toutefois, celle-ci présente des limites qu'il est nécessaire de cerner afin de pouvoir être critique sur les résultats qui en découlent.

De nombreux outils d'analyse ontologique ont été développés les 10 dernières années. Ils sont tous plus ou moins basés sur la même approche : pour un processus donné, comparer en termes de proportion sa présence dans une sélection de transcripts d'intérêt par rapport à sa présence dans l'intégralité des transcripts testés. Néanmoins, parce que chaque méthode possède des subtilités qui tiennent compte des caractéristiques des données d'expression, celles-ci peuvent donner des résultats variables mais qui ne doivent pas être contradictoires.

Avant tout, les différents modèles statistiques utilisés (hypergéométrique, binomiale, chi carré ou test de Fisher exact) produisent des résultats convergents quand l'analyse est basée sur un nombre de transcripts suffisamment importants. Les analyses basées sur une faible quantité de transcripts sont difficilement exploitables car elles produisent une statistique faussement significative qui n'a pas beaucoup de sens biologique. Ainsi la taille des données influence considérablement l'analyse ontologique. L'usage des modèles statistiques précités nécessite également des présomptions, lesquelles ne sont pas respectées dans les données transcriptomiques et dans les relations qui existent entre les différents termes ontologiques (organisées en arborescence).

Comme pour l'analyse différentielle univariée, les corrections pour les tests multiples sont également nécessaires, et la méthode du FDR est la plus appropriée étant donné la dépendance existante entre les données. Dans de nombreuses méthodes d'analyses cette correction n'est pas implémentée (cas de DAVID ou de GOMiner). De façon identique au calcul des probabilités de l'analyse différentielle, celles-ci peuvent être établies par permutation mais ne sont exploitables seulement avec un grand nombre de permutations.

Dans notre analyse, nous utilisons l'approche implémentée dans la librairie Gostat pour plusieurs raisons :

- (1) Cette librairie s'inscrit dans un pipeline d'analyse automatisé écrit en R.
- (2) Cette méthode comprend plusieurs des paramètres statistiques décrits.
- (3) Son résultat permet de générer des graphes ontologiques qui aident pour interpréter le résultat de l'ontologie.

Les annotations génomiques exploitées dans l'analyse ontologique sont également à considérer avec beaucoup de prudence en raison de leurs caractéristiques :

(1) Le Gene Ontology consortium est chargé de l'organisation des annotations ontologiques (29 429 termes au 25/01/10) et propose un vocabulaire structuré et une arborescence représentant les liens entre ces annotations. Ainsi un gène est associé à trois catégories des GO termes : processus biologique, fonction moléculaire, localisation cellulaire. Cette association est réalisée au travers d'un ensemble des connaissances produit par la recherche. Cette association est codifiée par un sigle qui qualifie la nature de l'association. Par exemple, «EXP » est le code pour indiquer que l'association résulte d'une expérience, et peut être enrichie du code « IDA » pour « *inferred from direct Assay* » ou « IPI » pour « *protein interaction* ». Le code « IEA » signifie « *inferred from electronic annotation* » indique que l'association a été réalisée mais est en cours de validation. Ce code n'indique pas la qualité de l'annotation mais seulement son statut dans le traitement de l'information. Selon les sources de Gene Ontology Annotation database environ 40 % des termes GO possèdent le code IEA et seulement 8 % sont inférés de manière expérimentale par la suite. Le tableau VIII (page 107) décrit, pour un sous-ensemble de gènes apoptotiques sélectionnés, la nature de ces associations dont dépendent nos interprétations. Par conséquent, il est nécessaire de comprendre que la nature des annotations a un biais sur l'analyse ontologique produite (Rogers et Ben-Hur 2009).

(2) Le second aspect concernant les annotations est leur organisation puisque celle-ci tient compte de plusieurs niveaux hiérarchiques (du plus ou moins spécifiques) qui est difficile de prendre en compte dans les approches classiques d'analyse ontologique et qui est une vision

subjective de l'organisation des processus cellulaires. Par exemple, le terme GO « induction hormonale de l'apoptose » dérive du terme « induction de l'apoptose » qui lui-même dérive du terme « apoptose ». Dans notre exemple, le terme « apoptose » représente le plus grand niveau d'abstraction (très générale) et « induction hormonale de l'apoptose » le plus bas niveau (très spécifique). L'annotation ontologique des gènes cherche à les annoter du plus bas niveau d'abstraction au plus haut niveau. Ainsi, un gène connu pour induire l'apoptose en réponse à l'hormone sera annoté comme tel, mais ne sera pas annoté par les termes GO « induction de l'apoptose » et « apoptose ». Les outils d'analyse ontologique intègre cette notion d'abstraction des termes GO de façon différente et il n'a pas été démontré de modèles efficaces pour aborder ce problème. Dans notre cas, notre analyse réalisée avec la librairie Gostat ne tient pas compte de l'abstraction des termes GO, ce qui nous oblige à être prudent dans l'interprétation.

(3) La plupart des outils ontologiques sont capables d'exploiter plusieurs identifiants (Affymetrix®, Refseq, UnigeneCluster, EntrezGene..). Si nous considérons 1000 transcrits, et les analysons selon différents identifiants, cela produira des résultats variables (mais pas fondamentalement différents) car la correspondance entre les différents identifiants et les fonctions biologiques n'est pas toujours correctement établie, en raison de l'ampleur de la tâche et du manque de moyen des organismes responsable de la gestion des annotations.

De plus, de nombreuses applications d'analyse ontologiques exploitent des versions différentes des annotations, en raison de mise à jour déficiente pouvant créer les différentes observées entre les applications. Un autre aspect qui est sous-estimé dans les associations ontologiques est la présence de l'épissage alternatif des gènes car de nombreux

identifiants comme Unigene n'en tiennent pas compte (Schuler 1997). Ceci a par conséquent de biaiser l'interprétation des termes GO et prend tout son sens si on considère que de 60 % des gènes subissent un épissage alternatif (Koscielny, Le Texier et al. 2009).

4. La réduction de l'information

Une des caractéristiques importantes dans une analyse de puce à ADN, est l'application de la réduction d'information. Cette réduction peut viser plusieurs buts selon le niveau ou elle est appliquée :

(1) Au niveau oligonucléotide et au niveau sonde, la réduction d'information élimine le bruit de fond, en optimisant la sensibilité et la spécificité de la détection. La technologie d'Affymetrix® utilise un système de sondes multiples pour évaluer le niveau de transcription des gènes. Sur la plateforme utilisée, plus de 3 millions d'oligonucléotides sont utilisées, réparties en 54 675 sondes, lesquelles hybrident environ 38 000 transcrits connues. Nous rappelons que dans la technologie d'Affymetrix® plusieurs sondes peuvent hybrident le même transcript et réagissent comme une réplique technique. La transformation « oligonucléotides → sondes » se fait par l'étape « *summarization* » de la méthode RMA, laquelle n'exploite que les sondes PM. La transformation « sondes → gènes » est réalisée dans notre cas, au niveau de l'analyse ontologique. Au total, de l'hybridation des puces à ADN à l'interprétation des fonctions biologiques, une réduction d'un facteur 80 est appliquée.

(2) Au niveau de l'analyse, la réduction d'information focalise sur un groupe de sonde d'intérêt. Certains choix techniques relevant des caractéristiques de l'analyse (seuil de détection, méthode de filtrage, outils d'analyse) réduisent l'information transcriptomique. Arbitrairement,

dans les analyses transcriptomiques, nous distinguons deux types de données : informatives et non-informatives. Dans notre approche sur l'étude du stress oxydatif nous filtrons les 54 675 sondes testées dans les conditions expérimentales avec une mesure de la variation (l'IQR) ayant un seuil au minimum de 0.5 (c'est-à-dire présentant des modulations entre deux conditions expérimentales supérieures à 1.4). De cette façon, nous éliminons de notre analyse 50 825 (54 675-3 850) sondes qui ont des petites modulations. Dans cette étape, nous présumons que dans un système biologique, les modulations importantes ont plus de sens que les petites modulations. Malgré les efforts, ces petites modulations sont difficiles exploitables car elles se confondent avec le bruit de fond. Néanmoins, il faut être conscient que celles-ci pourrait être importantes dans un contexte cellulaire où tout est une question d'équilibre où les petites modulations ont certainement une importance.

(3) notre second protocole utilise une approche multivariée qui exploite la variance présente dans les données d'expression. L'analyse canonique de redondance nous permet une réduction de l'espace : de 12 dimensions (1 pour chaque puce à ADN) à deux dimensions (1 pour chaque condition expérimentale). Cette réduction, qualifiée de « transformation sans perte de généralité », rend le traitement de l'information transcriptomique plus aisé en présentant les données sous une forme qui explique le mieux les modulations observées. De plus, cette transformation s'accompagne d'une perte d'information jugée non suffisamment représentatives des données.

(4) De la même façon, puisque nous sommes intéressé à produire de l'information biologique, la plupart des études transcriptomiques exploitent exclusivement les sondes associées à des annotations génomiques. Ainsi à partir des 3 850 transcrits sélectionnés par le

groupe.IQR.0.5, l'analyse ontologique exploite uniquement 1 850 identifiants qui sont associées à une annotation ontologique soit environ 48 % des transcripts sélectionnés.

Globalement, l'ensemble des caractéristiques présentées précédemment participe à l'interprétation biologique. Dans une étude où chaque étape de l'analyse peut influencer ou orienter cette interprétation, il est important de maîtriser, voire d'optimiser les différents paramètres qui finalement permettent des inférences biologiques de bonne qualité.

F. L'interprétation biologique de nos résultats

1. Caractérisation de l'inflammation de cellule fibro-kystique

La perte de la fonctionnalité du canal CFTR entraîne d'importantes conséquences physiopathologiques chez le patient FK et l'inflammation est une caractéristique fondamentale de la maladie. Cette inflammation est caractérisée par sa précocité, sa sévérité et sa persistance. En effet, celle-ci est détectée très tôt, bien avant qu'une infection soit détectée; elle implique de nombreux facteurs inflammatoires présents à des concentrations élevées telles que les molécules emblématiques IL1 β , IL6, IL8. En outre, la sur-expression de leurs transcripts est soutenue par une hyperactivité de voies de signalisations pro-inflammatoires telles que la voie NF- κ B.

Malgré les efforts entrepris pour caractériser cette inflammation, trop peu d'éléments concernant son dysfonctionnement ont été mis à jour pour permettre le développement de thérapies anti-inflammatoires.

Basé sur une approche comparant le transcriptomique d'une cellule porteuse de la mutation Δ F508 par rapport à une cellule contrôle, nous identifions 128 transcripts dont l'expression est significativement sur-modulée dans les

cellules CuFi-1 *versus* NuLi-1, lesquels sont associés à 8 fonctions inflammatoires surreprésentées dans notre analyse. La liste complète des 128 transcripts ainsi que leurs caractéristiques génomiques sont disponibles en annexe 10.

a. *Cinq grandes familles inflammatoires*

À partir des 128 gènes, l'étude des annotations génomiques permet d'identifier 33 gènes appartenant à cinq grandes familles de gènes inflammatoires : les chemokines de type CXC, les interleukines, les calgranulines, les *Guanylate Binding Protein* (GBP) et les *2',5' Oligoadenylate Synthetase* (2'5'OAS) représentant respectivement des groupes de 10, 11, 4, 4, 4 transcripts. Les membres des cinq familles de gènes sont présentés dans le tableau III (page 82). Ces familles sont tous composées de gènes connus, annotés et sur-exprimés dans le type CuFi-1 *versus* le type NuLi-1.

Les chemokines et les interleukines (ainsi que les interférons) composent la famille des cytokines, un ensemble de molécules inflammatoires. Nous précisons que certaines cytokines peuvent appartenir à deux familles différentes, ce qui est le cas par exemple de l'interleukine 8 (IL8) qui est connue comme étant le membre 8 de la famille de chemokine de type CXC. De façon similaire, l'interleukine 6 (IL6), est membre de la famille des interférons (IFN β 2).

Les chemokines sont des molécules solubles synthétisées par les cellules du système immunitaire mais également par d'autres types cellulaires comme les cellules épithéliales pulmonaires. Elles interagissent avec des récepteurs membranaires spécifiques, et permettent une communication intra-cellulaire par une action autocrine ou une communication inter-cellulaire par une action paracrine, endocrine ou juxtacrine. Si le rôle précis de chacune des chemokines n'est pas complètement déterminé, nous savons qu'elles

interviennent dans la régulation de nombreux processus cellulaires et jouent un rôle important dans la réponse inflammatoire en activant des signaux de transduction tels que NF- κ B, AP-1 ou les voies dépendantes des IFN. Le rôle majeur des chemokines est d'attirer les cellules inflammatoires vers le site inflammatoire afin de répondre à un stimulus. Dans un contexte non pathologique, les leucocytes, monocytes et neutrophiles recrutés, analysent, réagissent pour résoudre l'inflammation. En outre, elles agissent également sur les cellules épithéliales pour stimuler l'angiogénèse (Strieter, Burdick et al. 2005) et l'apoptose (Matute-Bello, Winn et al. 2001) dans un but commun: la réparation cellulaire.

Les interleukines, dont l'expression est ubiquitaire, sont également des molécules assurant la communication cellulaire. Elles sont impliquées dans un grand nombre de processus biologique, et la réponse inflammatoire est en grande partie conditionnée par leurs actions. En effet, elles participent (avec d'autres molécules comme les membres de la famille des IFN) à la réponse immune dite Th1, qui est associées à une défense inflammatoire anti-infectieuse et à la réponse et Th2, qui favorisent la production d'anticorps immunoglobulines de type A et E, et correspond à une réponse allergique. Toutefois, une vision plus manichéenne, classe les interleukines en 2 catégories: pro-inflammatoires et anti-inflammatoires qui définissent l'homéostasie inflammatoire et immunitaire.

Dans notre analyse, 10 chemokines et 11 interleukines sont présentes. Parmi elles, nous relevons la présence des membres la famille des GRO (*growth-related oncogene*) : CXCL1 (GRO- α), CXCL2 (GRO- β), CXCL3 (GRO- ζ), démontré comme sur-exprimés dans le type FK par Verhaeghe (Verhaeghe, Remouchamps et al. 2007) comme le montre le tableau IX (page 112).

Ces 3 membres sont des chimio-attractants de cellules inflammatoires : CXCL1, de neutrophiles (Moser, Clark-Lewis et al. 1990; Schumacher, Clark-

Lewis et al. 1992), CXCL2, de leucocytes, de polynucléaires et de cellules souches hématopoïétiques (Wolpe, Sherry et al. 1989; Iida et Grotendorst 1990; Pelus et Fukuda 2006) et CXCL3, de monocytes (Smith, Galkina et al. 2005).

D'autres chemokines présentes dans notre analyse agissent ou sont impliquées dans la chimio-attraction des cellules inflammatoires : PF4, CXCL10, CXCL11 (Lasagni, Francalanci et al. 2003), CXCL5 (Persson, Monsef et al. 2003), CXCL6 (Wuyts, Van Osselaer et al. 1997).

Par conséquent, la sur-expression de ces molécules contribue fortement au phénomène inflammatoire en stimulant la présence de cellules inflammatoires dans les tissus pulmonaires.

Les interleukines IL8, IL6, IL1 β sont considérés comme les molécules « centrales » de la réponse inflammatoire, car elles ont été observées dans plusieurs contextes inflammatoires (Baggiolini et Clark-Lewis 1992; Smirnova, Kiselev et al. 2002). Présentant des niveaux d'expression et protéiques élevés chez les patients FK par rapport aux contrôles, des nombreuses études y ont porté de l'intérêt (Dean, Dai et al. 1993; Bonfield, Konstan et al. 1999; Schmitt-Grohe, Naujoks et al. 2005; Adib-Conquy, Pedron et al. 2008). Nos résultats confirment également la sur-expression de ces molécules et confirment leurs modulations observées dans les autres expériences de puce à ADN présentées dans le tableau IX (page 112).

Parmi les interleukines de notre sélection, il est intéressant de noter la prédominance des molécules appartenant à la famille des interleukines 1 puisqu'à l'exception d'IL6 et IL8, toutes sont des membres de la famille de IL1 soit comme ligand, soit comme récepteur; IL18R1 faisant également partie de cette famille (Parnet, Garka et al. 1996). Ceci souligne l'importance de cette famille dans la réponse inflammatoire.

Les membres des interleukines 1 composent une large famille de molécules comprenant des ligands et des récepteurs lesquels peuvent définir une activité pro ou anti-inflammatoire. Ainsi plusieurs ligands de cette famille présents dans notre analyse sont soit pro-inflammatoires (IL1 β , IL1 α , IL1F9), soit anti-inflammatoires (IL1F5). De la même façon, la molécule IL1RAP forme un complexe avec le récepteur IL1R2 pour induire une réponse inflammatoire tandis que IL1RN (IL1Ra) est un inhibiteur de l'activité de IL1 β et IL1 α (Dinarello 2009). Ces transcripts sont sur-exprimés dans notre analyse et étant donné l'état hyperinflammatoire observé chez les patients FK, nous pouvons penser que l'activité anti-inflammatoire des membres de la famille des IL1 fait défaut et perturbe la balance inflammatoire en faveur d'une activité pro-inflammatoire.

Un autre transcript peut-être important à considérer dans l'analyse de l'inflammation est la sur-expression du TLR4 dans les CuFi-1 par rapport au NuLi-1. Ce récepteur est impliqué dans la détection du LPS présent chez les bactéries Gram négatif tels que *P.aeruginosa*, pathogène responsable d'infections chroniques des voies aériennes chez le patient FK. Son activation conduit au recrutement de MyD88 (*Myeloid differentiation factor 88*) lequel est présent dans notre liste de transcripts inflammatoires et dont la modulation est confirmée par qPCR (tableau IV, page 83). Par l'intermédiaire de IRAK1 et IRAK4, MyD88 produit la translocation de NF- κ B (Creagh et O'Neill 2006). Néanmoins, l'activation de MyD88 n'étant pas exclusive à la voie de signalisation dépendante du TLR4, cela ne permet pas de conclure sur l'activation spécifique des récepteurs des IL1 ou du TLR4. Plusieurs études ont montré l'implication du TLR4 dans l'activation de la réponse inflammatoire chez le patient FK (Ernst, Hajjar et al. 2003; Buchanan, Ernst et al. 2009). En outre, l'homologie entre les récepteurs des membres de la famille des interleukines 1 et des *Toll-like receptor* pourrait suggérer une activation de ce dernier par les ligands des IL1 mais cela n'a jamais été démontré. Toutefois,

d'autres molécules présentes dans notre analyse, sont connues pour agir comme des ligands endogènes du récepteur TLR4 (Ehrchen, Sunderkotter et al. 2009) : la calprotectine formée par S100A8 et S100A9, deux transcripts également sur-exprimés dans notre analyse.

Les calgranulines (S100), quant à elles, forment une famille composée de 21 membres, et sont caractérisées par 2 sites de fixations du calcium dans leurs structures protéiques tertiaires. Elles sont associées à des contextes inflammatoires rencontrés dans les infections virales, l'arthrite rhumatoïde, et la fibrose kystique (Thomas, Costelloe et al. 2000; Lorenz, Muhlebach et al. 2008)

L'implication des calgranulines dans l'inflammation nécessiterait la formation de calprotectine, un hétérocomplexe formé par les membres S100A8 et S100A9, lequel pourrait contribuer à la fixation ou au relargage des acides gras polyinsaturés tels que l'acide arachidonique chez les neutrophiles. Ce complexe acide arachidonique – calprotectine, pourrait contribuer à l'initiation et à la régulation de la réponse inflammatoire (Nacken, Roth et al. 2003).

S100A12 est une molécule connue dans l'inflammation chez les patients fibro-kystiques (Foell, Seeliger et al. 2003) et le blocage de son récepteur inhibe le processus inflammatoire (Hofmann, Drury et al. 1999), ce qui pourrait en faire un élément clé de la réponse inflammatoire (Pietzsch et Hoppmann 2009). Dans notre analyse, nous avons confirmé la modulation de ce transcript par qPCR (tableau IV page 83). Comme le montre le tableau IX (page 112), la modulation de ce transcript n'a pas été observée dans les expériences transcriptomiques d'autres chercheurs. Ceci pourrait suggérer une expression spécifique à notre modèle cellulaire.

La famille des 2',5' Oligoadenylate Synthétase (2'5'OAS) comprend 4 membres, lesquels apparaissent sur-exprimés dans notre expérience de puce à ADN : OAS1, 2,3 et L. Ils sont impliqués dans la réponse antivirale de

l'interféron par induction d'IRF-1, molécules présentes dans notre analyse et les expériences transcriptomiques du tableau IX (page 112). Ces molécules agissent en synthétisant des oligomères d'adénosine possédant des liaisons 2'-5' (2-5A) lesquels sont intrinsèquement capables de fixer et d'activer une RNase cellulaire, bloquant ainsi la synthèse protéique. Seul OAS1 a été décrit comme modulé dans la fibrose kystique mais a été perçu comme une conséquence de la sur-expression de STAT1 (Zheng, De et al. 2003).

Des *Guanylate Binding Protein (GBP)* sont également modulées dans notre expérience et sont représentés par quatre transcripts : GBP 1, 3, 5, 6 et à notre connaissance, ceux-ci n'ont pas été décrites comme étant modulées dans les études sur la fibrose kystique. Ces molécules ont la capacité de fixer spécifiquement la guanosine monophosphate (GMP), la guanosine diphosphate (GDP), et la guanosine triphosphate (GTP) et présentent une activité GTPase. Celles-ci ont été démontrées comme étant des marqueurs de l'induction de l'interféron participant à la potentialisation de l'interféron γ (Haller et Kochs 2002), mais les membres de cette famille peuvent être également stimulé par d'autres cytokines comme IL1 β , et IL1 α (Guenzi, Topolt et al. 2001; Nguyen, Hu et al. 2002) et par conséquent apporter une synergie à la réponse induite par l'interféron (Guenzi, Topolt et al. 2001). Toutefois, une étude plus approfondie de ces molécules pourrait apporter un complément de connaissance sur le processus inflammatoire observé chez le patient FK.

Toutefois, autour de ces cinq familles inflammatoires, gravitent de nombreux gènes inflammatoires largement identifiés chez les patients FK comme étant sur-exprimés et/ou présentes à des concentrations élevées dans les lavages broncho-alvéolaires des malades: STAT1 présent dans les études transcriptomiques présentées dans le tableau IX (page 112), mais aussi ICAM1 (Mrugacz, Zak et al. 2007), DEFB4 (Mrugacz, Zak et al. 2007).

b. Deux systèmes de régulation géniques.

La comparaison avec des expériences de puces à ADN étudiant l'impact du NF- κ B (Hinata, Gervin et al. 2003), et l'impact de l'interféron α et γ (Browne, Wing et al. 2001; Radaeva, Jaruga et al. 2002; Sana, Janatpour et al. 2005) nous aide à identifier un ensemble de 22(/128) et 32(/128) transcripts comme étant respectivement régulé par NF- κ B, et par IFN.

Bien que cette comparaison soit basée sur des types cellulaires différents, nous pensons que ce résultat est peu (ou pas) biaisé car ces deux systèmes de régulation géniques sont des voies de signalisation fondamentales des défenses immunitaires dont le mode d'action est similaire dans l'ensemble de la variabilité cellulaire. En outre, ces expériences exploitent des versions de puces à ADN moins complètes que notre plateforme d'analyse, en termes de transcripts présents. Ceci suggère que le résultat de notre comparaison pourrait être sous-estimé.

Dans plusieurs études, le système NF- κ B a été démontré comme étant impliqué dans l'hyperinflammation observée chez les patients FK (Blackwell, Stecenko et al. 2001; Eidelman, Srivastava et al. 2001; Li, Johnson et al. 2003; Joseph, Look et al. 2005; Carrabino, Carpani et al. 2006; Nichols, Chmiel et al. 2008). Toutefois, si le mode d'action de NF- κ B est bien déterminé, son activation et sa régulation au sein d'un système FK n'est pas encore bien cernée. En effet, NF- κ B est un dimère composée d'une combinaison des sous-unités p50, p52, p65 (RelA), c-Rel, RelB, la forme canonique du facteur de transcription étant p50/p65. Ce dimère est retenu physiquement dans le cytoplasme par une famille d'inhibiteurs : I κ Bs qui phosphorylés par le complexe I κ K (I κ Bkinase), entraîne la translocation nucléaire du dimère NF- κ B, qui fixe sa séquence consensus présente sur les promoteurs de gènes cibles pour induire leurs transcriptions. Ces gènes cibles participent à l'inflammation

(IL1 β , TNF α , IL8), au cycle cellulaire (p21waf1/cip1, cyclines), à l'apoptose et au métabolisme oxydatif (Vallabhapurapu et Karin 2009).

La mesure de translocation nucléaire des sous-unités, p50 et p65, de NF- κ B présenté dans la figure 16 (page 85) ne montre pas de différence entre les lignées, CuFi-1 et NuLi-1. Par contre, la stimulation au LPS et au TNF montre une augmentation significative de la translocation de p65 (mais pas de p50) dans les CuFi-1 par rapport au NuLi-1, correspondant à une activité plus importante de NF- κ B. Contrairement à nous, les travaux de Verhaeghe (Verhaeghe, Remouchamps et al. 2007) montrent une activation constitutive de NF- κ B plus importante dans ses lignées FK par rapport aux contrôles. Les travaux de Becker en 2004 suggèrent également qu'il n'y a pas de différence dans l'activité de NF- κ B entre son modèle HTBE (*Human Tracheal Bronchial Epithelial*) et son contrôle (Becker, Sauer et al. 2004). Dans sa conclusion, Becker ajoute que l'activation de NF- κ B chez les patients FK, et par conséquent, l'induction de l'expression de gène pro-inflammatoire serait conditionnelle à certains stimuli comme par exemple la présence de *P.aeruginosa*.

L'ensemble de ces observations montre l'implication de NF- κ B dans la sur-expression des gènes inflammatoires, mais que celle-ci serait dépendante d'une stimulation via des récepteurs membranaires tels que les TLR, ou des récepteurs du TNF. De plus, l'absence d'activité de NF- κ B entre les nos deux modèles cellulaires pourrait évoquer que la réponse inflammatoire observée dans notre expérience de puce à ADN (en absence de quelconque stimulation) ne serait pas uniquement dépendante de l'action de NF- κ B, et que d'autres voies de signalisation seraient susceptibles d'intervenir.

La présence de nombreux gènes inflammatoires associés au système des IFN pourrait laisser penser que celui-ci contribuerait à la réponse inflammatoire excessive. Les interférons font partis de la famille des cytokines et sont

composés de trois classes majeures: IFN- α ; IFN- β ; et IFN- γ . Elles interviennent dans la réponse immunitaire face aux infections virales. Leurs impacts sur l'hyperinflammation FK a été très peu étudiés. Néanmoins, de nombreuses études relèvent la présence d'une activité de ce système. Adib-Conquy, dans son étude transcriptomique de neutrophiles de patient FK (Adib-Conquy, Pedron et al. 2008), relève la modulation de quelques gènes reliés à l'interféron gamma et qui apparaissent dans notre analyse, tels que CXCL10, IRF1. L'IFN- γ qui a déjà été détecté dans les crachats de patients FK (Moser, Jensen et al. 2005) et sa diminution a été observée après stimulation par *P.aeruginosa*. L'auteur suggère dans ses conclusions que l'IFN- γ pourrait contribuer au débalancement entre la réponse Th1 et Th2, participant, de cette façon à la déficience de la réponse inflammatoire. L'un des marqueurs de la réponse à l'interféron est STAT1. Comme le montre le tableau IX (page 112), cette molécule est sur-exprimée dans les cellules FK par rapport au non FK dans au moins trois expériences transcriptomiques (incluant nos observations). Nous confirmons et validons cette modulation en qPCR (tableau IV, page 83). STAT1 est un effecteur de la voie de signalisation de l'interféron : sous la forme d'homodimère, dans la voie de l'interféron de type 2 et sous la forme d'un hétérodimère STAT1/STAT2 combinée à IRF9, dans la voie de l'interféron de type 1. La phosphorylation du dimère conduit à sa translocation nucléaire et à sa fixation à une séquence nucléaire spécifique, induisant l'expression de gènes cibles.

Dans nos études plusieurs molécules participant à la réponse dépendante des interférons sont présentes. STAT1, IRF1, IRF7, IL6 (interféron β 2), ainsi que les 4 membres des 2',5' Oligoadénylate Synthétase (OAS1, 2, 3, L), pour ne citer qu'eux, sont sur-exprimés dans notre modèle FK par rapport au contrôle. Au total, ce sont 32 transcrits qui semblent sous le contrôle du système des membres de la famille des interférons.

c. *La comparaison avec des expériences similaires.*

Le tableau IX (page 112) présente les différentes expériences de puce à ADN réalisées dans l'étude de la fibrose kystique. Bien que toutes comparent le transcriptome de cellules FK et non FK, il est difficile d'en déduire un consensus clair sur les perturbations transcriptomiques (et inflammatoires) dans une cellule mutante. En effet, devant la diversité des modèles, la méthodologie d'analyse utilisée, le centre d'intérêt de l'étude, rien ne permet d'identifier une convergence vers des processus définis comme perturbés dans la maladie. Toutefois, basé sur les listes de gènes inflammatoires présentés comme modulés dans les différentes expériences, nous mettons en évidence les points communs avec notre liste de transcripts permettant de définir une tendance dans l'état inflammatoire. Ainsi, comme décrit auparavant, les sur-expressions d'IL8, IL1 β , de STAT1, de nombreuses calgranulines sont souvent relevées. Ceci nous permet de soumettre l'idée que la modulation observée de ces gènes n'est pas dépendante de conditions expérimentales mais plus liée à la condition de la mutation du CFTR et qu'ils définissent les éléments centraux de la réponse inflammatoire dans un contexte FK. De plus, de nombreuses études (soit de cellules épithéliales pulmonaires, soit de lavages broncho-alvéolaires) ont mis en avant, des niveaux protéiques d'IL8, IL1 β , de STAT1 plus importants chez les patients FK que chez les contrôles (Kelley et Elmer 2000) ; (Brown, Morgan et al. 1991; Koller, Nething et al. 1997; McGarvey, Dunbar et al. 2002; Becker, Sauer et al. 2004). De la même façon, la sur-expression des membres de la famille des calgranulines (à l'expression de S100A16) est observée dans quatre des expériences de puces à ADN produites dont la nôtre confirmant ainsi la modulation de S100A8 et S100A9 (tableau IX, page 112). Une étude protéomique de lavages broncho-alvéolaires, note également une concentration plus importante chez les patients FK par rapport aux contrôles de S100A8, S100A9, S100A12

(MacGregor, Gray et al. 2008). Par conséquent, il serait intéressant d'investiguer plus en avant sur le rôle de ces molécules inflammatoires pour permettre de mieux comprendre la réponse inflammatoire.

Les différences de détection des molécules inflammatoires entre notre expérience et les 8 autres études transcriptomiques FK peuvent s'expliquer de plusieurs façons. Premièrement, la comparaison des résultats est basée sur des listes de gènes sélectionnés par l'auteur lesquels sont restreintes à ce que l'auteur souhaite mettre nous montrer. Par exemple, les résultats présentés par Zabner sont issus d'une analyse très stringente sélectionnant uniquement 22 gènes (Zabner, Scheetz et al. 2005). Deuxièmement, la diversité des modèles utilisés pourraient suggérer des différences de modulation dans le processus inflammation où des subtilités moléculaires auraient leurs places. En effet, les cellules pancréatiques, pulmonaires, intestinales doivent probablement développer des mécanismes inflammatoires différents ayant la même finalité. Troisièmement, les études réalisées par d'Adib-Conquy (Adib-Conquy, Pedron et al. 2008) et Gambardella (Gambardella, Biancolella et al. 2006) exploitent des puces à ADN maison qui ne s'intéressent qu'à un groupe de gènes d'intérêt qui ne comprennent pas obligatoirement les nôtres. Toutefois, parmi toutes ces études, celle d'Adib-Conquy présente le plus de points communs avec nos résultats avec 11 gènes identifiées (tableau IX, page 112).

En outre, le résultat de cette comparaison devrait être pris avec prudence car elle exploite uniquement les résultats publiés des huit études transcriptomiques et est issues de méthodologie et de plateformes d'analyses différentes. Une façon optimisée de réaliser cette comparaison consisterait à appliquer notre protocole d'analyse (normalisation, analyse différentielle) sur les différents jeux de données et d'évaluer le comportement de 128 gènes d'intérêt dans les huit expériences. Hormis le fait que ce travail de comparaison est en projet en soi, la façon de produire intelligemment une

comparaison de plusieurs expériences de puce à ADN n'est pas clairement définie (Shi, Reid et al. 2006).

Toutefois notre comparaison simpliste, souligne la présence commune de certains facteurs de l'inflammation tels que les interleukines 1 β et 8, ainsi les membres de la famille des calgranulines S100A8, S100A9. Ces deux dernières protéines formant la calprotectine, pourrait apporter un nouvelle aspect à inflammation en agissant comme un activateur endogène inflammatoire (Bianchi 2007).

Les approches exploitées dans ces différentes études de puce à ADN sont orientées vers une analyse globale des perturbations induites par la mutation du CFTR et à l'exception d'une seule, aucune n'a choisi de présenter dans leurs travaux une étude plus poussée d'un processus perturbé. Cette exception est l'étude de Catherine Verhaege, qui basée sur les résultats transcriptomiques de lignées épithéliales humaines (CFT2, Hela Δ F508, 16HBE14o-), et un ensemble d'expériences complémentaires (validation par PCR et par *Western Blot*, étude de la phosphorylation, et l'usage d'inhibiteurs spécifiques) identifie 2 voies de signalisation perturbées chez les cellules FK : NF- κ B et AP-1. De plus, leurs travaux montrent qu'IL1 β et β FGF pourraient être des activateurs intrinsèques de ces voies de signalisation (Verhaeghe, Remouchamps et al. 2007).

d. Une régulation génique au niveau chromosomique.

Le tableau III (page 82) décrivant les cinq familles sur lesquelles nous avons porté notre intérêt, présente également la localisation chromosomique des transcrits. Cette localisation permet de mettre en avant des loci associés aux familles de transcrits. Ces observations nous laissent penser que la co-expression de gènes situés dans une même région chromosomique serait liée à une co-régulation faisant intervenir un contrôle d'une large portion de l'ADN

(de quelques milliers de paires de bases à quelques millions). A ce jour, seules les modifications épigénétiques peuvent expliquer une régulation aussi étendue.

Dans bien des études sur la régulation de la transcription des gènes eucaryotes, l'intérêt s'est porté vers une « mécanique génétique » où pour expliquer la régulation génique, l'idée d'un promoteur proximal et de régions cis-régulatrices dominant. Ainsi, les co-régulations observées ont été expliquées par l'activation des promoteurs partageant des éléments de régulations communs. Toutefois, à un niveau de régulation génique supérieur, il apparaît que l'activation de ces zones régulatrices dépend de modifications épigénétiques (Jiang, Bressler et al. 2004).

Les modifications épigénétiques peuvent-être définies comme toutes informations héréditaires, expliquant le niveau d'expression génique et qui ne sont pas codées par l'ADN (Egger, Liang et al. 2004). Ces modifications telles que l'acétylation/désacétylation (d'histones ou de facteurs de transcription), la méthylation de l'ADN, conditionnent la décondensation (ouverture) ou la condensation (fermeture) de la double hélice, influençant de cette façon, l'action de facteurs de transcription à une grande échelle. Il a été montré que ce phénomène intervient notamment pendant le développement et en réponse à des changements environnementaux (Haberland, Montgomery et al. 2009). De plus, des études inter-espèces ont montré que les gènes situés dans une même région chromosomique peuvent être co-régulés (Cho, Campbell et al. 1998; Caron, van Schaik et al. 2001; Boutanaev, Kalmykova et al. 2002; Roy, Stuart et al. 2002; Li, Lee et al. 2005) témoignant ainsi de l'importance de ce processus.

Bien que nous n'avons pas encore bien cerné l'importance de cette régulation dans l'homéostasie, de nombreuses études rapportent son implication dans le contrôle de l'inflammation (Turunen, Aavik et al. 2009), et dans l'activité de NF-

κ B (Ghosh et Karin 2002; Vanden Berghe, Ndlovu et al. 2006). Il a déjà été suggéré que l'inflammation pulmonaire observée dans des maladies comme le MPOC (*maladie pulmonaire obstructive chronique*) ou l'asthme pourrait résulter d'un défaut enzymatique de HDAC (Histone déacétylase), responsable de la désacétylation, dont l'action réprime la transcription des gènes (Ito, Ito et al. 2005). D'autres études récentes ont également montré que l'inhibition chimique de HDAC est impliquée dans l'inflammation des voies aériennes (Adcock, Tsaprouni et al. 2007) et que certaines modifications épigénétiques pourraient être spécifiques aux cellules épithéliales pulmonaires humaines (Ito, Barnes et al. 2000).

Des études récentes se sont intéressées aux relations entre le NF- κ B et les enzymes responsables de la désacétylation dans plusieurs types cellulaires. Celles-ci ont montré une influence de l'un envers l'autre puisque l'inhibition des désacétylases conduit à une augmentation de l'expression des gènes inflammatoires dépendant de NF- κ B (Ito, Barnes et al. 2000; Ashburner, Westerheide et al. 2001; Chen, Fischle et al. 2001; Zhong, May et al. 2002) Néanmoins, le mécanisme d'interaction reste encore hypothétique car il n'a pas été démontré que les HDAC avait une interaction physique avec NF- κ B au niveau de l'ADN ou si les HDAC interagissent avec l'inhibiteur de NF- κ B (Viatour, Legrand-Poels et al. 2003).

De la même façon, les modifications épigénétiques interviennent dans les voies de signalisation dépendant de l'interféron puisque l'activation de STAT1 nécessiterait la présence d'acétylase (Klampfer, Huang et al. 2004)

Notre analyse inflammatoire met en évidence la présence d'une désacétylase : HDAC9, dont l'expression est plus importante chez les CuFi-1 *versus* les NuLI-1. Ce gène n'a jamais été décrit comme étant modulé dans la fibrose kystique. Une étude exploitant un modèle murin HDAC9^{-/-} a mis en évidence la sur-expression de HSP70 (*Heat Shock Protein 70*), mettant ainsi en avant la

régulation de HDAC9 sur HSP70 (de Zoeten, Wang et al. 2010). HSP70 est associé à un rôle de protection contre l'inflammation (Jacquier-Sarlin, Fuller et al. 1994) et de plus, il a été montré que HSP70 est impliqué dans l'adressage du CFTR à la membrane (Choo-Kang et Zeitlin 2001) Ce qui en fait une molécule centrale dans deux problématiques qui touchent la fibrose kystique. Le mécanisme de régulation de HDAC9 sur HSP70 n'a pas été mis à jour et il sera nécessaire d'investiguer plus en avant.

L'ensemble de nos résultats sont basés sur une exploitation très ciblée de nos données d'expression de puce à ADN de transcrits inflammatoires sur-exprimés dans les cellules CuFi-1 par rapport au NuLi-1. Notre étude tente de mettre en avant les caractéristiques inflammatoires de notre modèle FK. À partir d'une liste de 128 gènes inflammatoires, nous avons tenté d'en extraire les grandes tendances : la présence de familles inflammatoires, deux systèmes de régulation géniques et potentiellement des éléments qui nous laissent penser que les modifications épigénétiques pourraient intervenir dans la réponse inflammatoire.

Si de nombreuses études se sont intéressées au rôle des modifications épigénétiques dans l'asthme (Bhavsar, Ahmad et al. 2008), dans la MPOC (Rajendrasozhan, Yao et al. 2009) ou l'arthrite rhumatoïde (Sanchez-Pernaute, Ospelt et al. 2008), aucune d'étude n'a encore porté son intérêt sur leurs implications dans le dysfonctionnement de la réponse inflammation observée dans un contexte FK.

Pour investiguer les potentielles modifications épigénétiques suggérées dans notre étude, nous pourrions nous intéresser aux loci mis en évidence dans notre analyse et investiguer si ces régions sont soumises à certaines modifications et quelles en seraient l'impact sur l'action de NF- κ B, et des IFN.

2. Impact du stress oxydatif sur une cellule fibro-kystique

L'analyse canonique de notre expérience de puce à ADN montre l'implication de mécanisme moléculaire associé à la mort cellulaire et à l'apoptose dans notre modèle FK soumis à un stress oxydatif. Nous identifions plusieurs gènes reliés à ces fonctions par leurs annotations génomiques et dont les modulations s'expliquent par l'effet conjugué des deux conditions expérimentales: la mutation et le stress oxydatif. En outre, pour juger de l'état d'apoptose dans les deux lignées cellulaires en présence du DMNQ, nous évaluons l'activité protéique des caspases 3/7. Cette expérience complémentaire à notre approche transcriptomique, nous permet de conclure que les cellules FK en présence de DMNQ seraient moins sensibles à l'apoptose que les cellules contrôles.

a. Dysfonctionnement de la prolifération et de la mort cellulaire dans notre modèle FK

La mort cellulaire programmée, ou l'apoptose, est un processus physiologique normal impliqué dans le renouvellement des tissus épithéliaux. Ce renouvellement s'établit par un équilibre entre la prolifération et l'élimination des cellules. Dans un contexte normal, l'épithélium pulmonaire participe à l'élimination des cellules apoptotiques lors du développement des poumons car les cellules épithéliales ont la capacité de phagocyter les cellules sénescents, ou considérées anormales (Kresch, Christian et al. 1998).

De part leur intérêt à comprendre le défaut observé dans la réparation tissulaire chez les patients fibro-kystiques, plusieurs études se sont concentrées sur la prolifération cellulaire et l'apoptose de cellules épithéliales pulmonaires. Certaines d'entre elles ont mis en évidence une prolifération cellulaire plus importante chez les cellules porteuses de la mutation $\Delta F508$ que

chez les cellules contrôles. (Koller, Kim et al. 1991; Leigh, Kylander et al. 1995; Hassid, Degaute et al. 1997).

L'étude de l'apoptose chez les cellules épithéliales FK a également suscité beaucoup d'intérêt sans toutefois aboutir à un consensus sur son dysfonctionnement. En 1996, Gottlieb explore le rôle du pH intracellulaire sur l'inhibition du processus de mort cellulaire programmée dans des cellules épithéliales mammaires de souris $\Delta F508$ (Gottlieb et Dosanjh 1996). Ces résultats sont contredits par Maiuri en 1997, qui rapportent, une augmentation de la fragmentation de l'ADN, et de l'expression de Fas et FasL, médiateurs du signal apoptotique dans des cellules de l'épithélium bronchiques et du duodénum issues de patients FK (Maiuri, Raia et al. 1997). Durieu confirme le haut niveau d'expression de FasL dans des cellules de l'épithélium bronchique humain obtenues par biopsie de patient FK, ainsi que dans une lignée de cellules trachéales humaines FK (HTEC) comparé aux contrôles (Durieu, Amsellem et al. 1999).

Jungas est le premier à s'intéresser au rôle du stress oxydatif sur le processus de l'apoptose dans le contexte de la fibrose kystique. Celui-ci met en évidence une résistance à l'apoptose induite par l'hyperoxie. Cette résistance apparaît comme plus importante dans les cellules épithéliales IB3-1, porteuses de la mutation $\Delta F508$ que dans les cellules contrôles Hela, et cellules primaires pulmonaires non FK (Jungas, Motta et al. 2002). Plus récemment, le Dr Boncoeur décrit la diminution de l'activité des caspases 3/7, dans des conditions hyperoxiques, chez un modèle épithélial cellulaire FK, humain (IB3-1) et murin ($CFTR^{-/-}$) et montre la dépendance de l'activité des caspases 3/7 au stress oxydatif (Boncoeur, Tabary et al. 2006; Boncoeur, Roque et al. 2008).

Les études menées par l'équipe de Vandivier permettent de constater une augmentation de cellules apoptotiques dans les crachats et les tissus de

patients fibro-kystiques (Vandivier, Fadok et al. 2002). Ils démontrent que l'activité de l'élastase produite par les neutrophiles, inhibe l'efferocytose, le processus d'élimination des cellules apoptotiques, en clivant un signal d'apoptose présent à la surface des cellules apoptotiques et des phagocytes : la calreticuline. Récemment, ils ont démontré le rôle inhibiteur partiel de CFTR dans des cellules épithéliales pulmonaires humaines FK et dans la lignée IB3-1 sur l'efferocytose. Pour aller plus loin, ceux-ci précisent que cette inhibition implique la kinase RhoA (*Ras homolog A*), et le canal ENaC dont la régulation par le CFTR est connue (Vandivier, Richens et al. 2009).

Bien qu'il y ait des contradictions dans les activités apoptotiques observées dans l'ensemble de ces travaux, ceux-ci montrent son dysfonctionnement dans les tissus de patients FK ou des lignées cellulaires. Il semble que le stress oxydatif, facteur aggravant de la maladie, apporterait aux cellules FK une résistance à l'apoptose. D'après les travaux récents de Vandivier, cette résistance serait reliée à l'efferocytose.

b. Vers un modèle d'étude de l'apoptose

La mesure de l'activité des caspases 3/7 dans nos conditions expérimentales (figure 23, page 108), nous montrent la diminution de l'apoptose dans les cellules CuFi-1 par rapport à NuLi-1 en présence de DMNQ. Cette observation est en accord avec les travaux de Jungas et de Boncoeur. En absence de DMNQ, dans notre expérience, l'activité des caspases 3/7 est plus importante dans les cellules CuFi-1 par rapport à NuLi-1. Cette observation ne confirme pas les résultats du Dr Boncoeur (Boncoeur, Tabary et al. 2006) qui, par deux méthodes de dosage différentes (dosage d'incorporation de l'Annexine V et le dosage d'activité des caspases 3/7), ne montre pas de différence dans l'activité apoptotique entre son contrôle et son modèle FK (IB3-1). Contrairement à notre modèle CuFi-1 ($\Delta F508/\Delta F508$) où la protéine CFTR

est correctement produite mais présente un défaut de maturation, la lignée IB3-1 utilisée par Boncoeur, est porteuse hétérozygote de la mutation $\Delta F508$ ($\Delta F508/W1282X$) et produit une protéine altérée. Cette différence dans la production de CFTR pourrait expliquer la divergence des résultats observés.

Pour expliquer le phénotype apoptotique observé dans nos deux types cellulaires en présence et absence de stress oxydatif, nous proposons un modèle d'état apoptotique. Ce type de modèle est largement utilisé afin d'améliorer la connaissance de nombreux processus biologiques, comme l'inflammation ou l'oxydo-réduction. Celui-ci modélise simplement un équilibre entre deux populations de molécules avec des actions opposées. Cet équilibre dynamique, l'homéostasie, est sous la contrainte de facteurs qui permet la viabilité des processus biologiques et en cas de défaillance de cette stabilité les pathologies apparaissent. Ainsi dans le contexte de l'apoptose, son insuffisance entraîne le cancer, des maladies auto-immunes et son excessivité, des pathologies dégénératives.

Notre modèle d'état apoptotique est basé sur un équilibre entre des molécules, pro-apoptotiques et anti-apoptotiques et permet d'expliquer le changement de phénotype observé chez les CuFi-1 dans l'activité des caspases 3/7 après l'ajout du DMNQ. En effet, l'induction du stress oxydatif chez les cellules CuFi-1 modifie le rapport entre des molécules pro-apoptotiques et anti-apoptotiques. En considérant l'activité des caspases 3/7 chez le type NuLi-1 comme le niveau basal, ce rapport devient inversé chez les CuFi-1 en présence de stress oxydatif.

D'après ce modèle, cela suggère un débalancement de l'action des gènes pro-apoptotiques et des gènes anti-apoptotiques. Ainsi, en absence de DMNQ, l'action des gènes pro-apoptotiques domine et, en présence de DMNQ, l'action des gènes anti-apoptotiques apparaît plus importante que l'effet des gènes pro-apoptotiques.

Pour expliquer la résistance de l'apoptose observée chez les CuFi-1 en présence de stress oxydatif, trois possibilités existent : soit une inhibition de l'activité pro-apoptotique, soit une activation de l'activité anti-apoptotique, soit une combinaison des deux. Pour explorer ces hypothèses, les données transcriptomiques, reflétant une balance apoptotique, peuvent nous apporter des éléments de compréhension.

c. *La balance apoptotique au niveau transcriptomique*

Notre analyse transcriptomique nous a permis d'isoler un ensemble de transcrits liés à la mort cellulaire et à l'apoptose et dont la modulation est dépendante des effets conjoints des conditions expérimentales (la mutation et le stress oxydatif). Ces transcrits pourraient en partie expliquer les phénotypes apoptotiques observés dans l'expérience d'activité des caspases 3/7.

Parmi notre sélection de transcrits d'intérêt, nous identifions cinq gènes sous-modulés dans le type CuFi-1 (en absence et présence de DMNQ), qui sont associés à la cascade des MAPK : MAPK1, MAPK8, MAP3K5, PDCD4, HIPK2.

La voie de signalisation MAPK est un « carrefour » de régulation biologique, car elle répond à des nombreux stimuli (facteur de croissance, hormones, cytokines, stress). Elle peut-être décomposée en 6 voies de signalisations distinctes donc les trois plus connues sont : ERK (*Extracellular signal-Regulated Kinase*), p38, et JNK (*Jun N-terminal Kinase*). Celles-ci sont régulées, de façon très fine et flexible, par une cascade de phosphorylation catalysée par une série de kinases (MAPK, MEK, MEKK) et interviennent dans la transduction du signal de la prolifération, de la différenciation et de la mort cellulaire.

Plusieurs études ont montré l'activation de la voie ERK dans l'induction de l'apoptose dans des modèles de cellules épithéliales alvéolaires en culture (Buckley, Barsky et al. 2005; Papaiahgari, Zhang et al. 2006), en réponse à un stress induit par l'hyperoxie, ainsi que l'implication de la voie JNK dans l'induction de l'apoptose chez les cellules épithéliales pulmonaires (Romashko, Horowitz et al. 2003). Zhang et al ont montré que la voie ERK régule l'expression des caspases (Zhang, Shan et al. 2003) et que l'inhibition de cette voie conduit à l'inactivation de l'apoptose (Zhang, Leonard et al. 2003) chez des lignées épithéliales pulmonaires (MLE12, A549). De la même façon, Romashko met en évidence que l'inhibition de l'apoptose par une inactivation de la voie JNK (Romashko, Horowitz et al. 2003)

Les 3 transcripts des protéines kinases sous-modulées dans notre expérience, MAPK1 (alias: ERK2), MAPK8 (alias: JNK1) et MAP3K5 (alias: MEKK5, ASK1) interviennent dans la cascade de phosphorylation des MAPK. Comme le montre leurs profils d'expression (figure 22, page 104), la MAP3K5 semble la plus affectée par la présence de DMNQ que les 2 autres kinases, lesquelles semblent moins sensibles au stress oxydatif.

L'activité protéique de MAP3K5, agit en amont de la voie des MAPK, et pourrait être un facteur limitant de cette cascade. Des études rapportent que cette protéine agit comme un élément important dans la régulation de l'activation des voies JNK et p38 afin d'induire l'apoptose suite à la production intracellulaire d'oxydant comme les ROS (*reactive oxygen species*) (Sumbayev 2008). De plus, MAP3K5 est connue pour s'associer à la thioredoxine, protéine impliquée dans la balance d'oxydoréduction en exerçant une activité complémentaire à celle du glutathion dans la détoxification. Il a été montré que la thioredoxine est un inhibiteur de MAP3K5 (Saitoh, Nishitoh et al. 1998).

MAP3K5 joue également un rôle critique dans la réponse immune innée induite par la voie de signalisation du TLR4. La famille des TLR, est un ensemble de récepteur membranaire, qui par reconnaissance d'une variété

d'antigène, permet une immunité innée. Le TLR4 est reconnue, entre autre par le LPS, un élément partagé par les bactéries Gram négatif. La diminution de son activité a été montré chez des cellules FK par rapport au contrôle (John, Yildirim et al. 2009).

Les deux transcripts sous-exprimées PDCD4 (*programmed cell death 4*) et HIPK2 (*Homeodomain-Interacting Protein Kinase-2*) interviennent dans la régulation des MAPK. Bien que PDCD4 n'ai pas été décrit clairement dans la littérature pour son action dans l'apoptose, plusieurs éléments montre qu'il pourrait participer à un processus pro-apoptotique. (1) Sa perte d'expression corrèle avec la progression de la tumeur dans le cancer du poumon (Chen, Knosel et al. 2003). (2) Il est surexprimé pendant l'apoptose (Shibahara, Asano et al. 1995; Chauhan, Li et al. 2003). (3) Chez les cellules épithéliales pulmonaires, il est impliqué dans l'activité pro-apoptotique du complexe AP-1, fin du signal de transduction des MAPK. (Romashko, Horowitz et al. 2003) HIPK2 dont le rôle pro-apoptotique est bien documentée, interagit avec Tp53 (*Tumor proteine 53*), un inducteur bien connu de l'apoptose et active la voie JNK potentialisée par TGF- β (Hofmann, Stollberg et al. 2003). Ces deux actions pro-oncogènes conduisent à la mort cellulaire.

L'ensemble de ces résultats, nous permet d'émettre l'hypothèse que la sous-activité de la voie des MAPK dans la cellule FK en présence de stress oxydatif participe à l'état anti-apoptotique observé. Cette sous-activité se traduit au niveau transcriptionnel par une diminution de l'activité des kinases impliquées dans le type CuFi-1 comparé au NuLi-1 (MAPK1, MAPK8, MAP3K5) et également par la sous-expression de transcripts pro-apoptotiques (PDCD4, HIPK2) impliqués dans le signal de transduction de la voie des MAPK. Le profil d'expression de ces 5 gènes (figure 22, page 104) montre que leurs sous-expressions sont indépendantes de la présence du stress oxydatif, néanmoins,

le niveau d'expression des transcrits des gènes PDCD4 et MAP3K5 est affecté par le DMNQ.

Dans notre modèle apoptotique, PDCD4 et MAP3K5 pourraient participer à l'activité pro-apoptotique en fonction de leurs niveaux de transcription. Ainsi des concentrations élevées de ces gènes permettraient une activité des MAPK conduisant à l'apoptose et des concentrations plus faibles influenceraient de façon plus radicale l'activité apoptotique de cette voie.

d. Le contrebalancement anti-apoptotique

Précédemment, nous avons présenté la diminution de l'activité des MAPK comme étant une composante de notre modèle. Toutefois, nos résultats montrent également l'augmentation du niveau d'expression de transcrits décrits comme anti-apoptotiques (figure 22 page 104). Nous ne souhaitons pas décrire, de façon exhaustive l'ensemble des transcrits sur-exprimés dans les conditions CuFi-1 *versus* NuLi-1. Toutefois deux de ces transcrits anti-apoptotiques retiennent notre attention, parce que leurs activités apoptotiques sont bien documentées, et parce que leurs niveaux d'expression et leurs variances élevées font d'eux, de bons candidats pour expliquer notre modèle apoptotique : BCL2A1, et VEGFA.

BCL2A1 est un membre de la famille de BCL, famille de protéines dont certains membres inhibent l'apoptose (Bcl-2, Bcl-xL) et d'autres l'activent (Bik, Bax, Bak) (Choi, Park et al. 1995; D'Sa-Eipper, Subramanian et al. 1996). Son activité anti-apoptotique se traduit par le blocage de l'activation des caspases et de la libération mitochondriale de cytochrome c, médiateur de l'apoptose. Il est également impliqué dans la cytoprotection induite par IL11 et VEGF contre l'hyperoxie dans le poumon (He, Waxman et al. 2005). BCL2A1 est le seul membre de la famille des BCL associé à un contexte inflammatoire, facteur aggravant de la physiopathologie de la fibrose kystique. De plus, sa

transcription est activée par NFAT1 (*nuclear factor of activated T cells*) et est non dépendante de NF- κ B (Ulleras, Karlberg et al. 2008). Ceci pourrait expliquer la sur-expression de BCL2A1 dans le type CuFi-1 par rapport au NuLi-1. En effet, il a été proposé que le canal CFTR soit un modulateur de canaux Ca^{2+} et qu'en absence de CFTR fonctionnel, l'augmentation intracellulaire de calcium induirait une augmentation de la concentration nucléaire de NFAT. (Allard, Poynter et al. 2006).

Dans notre modèle, BCL2A1 participerait à l'action anti-apoptotique observée dans notre expérience en présence de DMNQ. Toutefois, son action serait conditionnelle (1) à sa concentration dans le milieu intracellulaire, (2) à un contexte moléculaire anti-apoptotique.

VEGFA fait partie la famille de ligand antigénique VEGF (*Vascular Endothelial Growth Factor*) et il est considéré comme un élément de régulation central dans l'angiogénèse des cellules endothéliales. Sa sur-expression a été rapportée dans de nombreuses études sur le cancer du poumon (Carrillo de Santa Pau, Arias et al. 2009; Pau, Arias et al. 2009) ainsi que dans des lignées trachéales fœtales humaines (CFT-2) (Verhaeghe, Tabruyn et al. 2007). Sa régulation s'établit par un effet autocrine rétroactif d'une de ses formes épissées. Ainsi 2 variants de VEGFA participent à un équilibre qui orchestre l'angiogénèse (Harper et Bates 2008). Il a été démontré que sa transcription est dépendante (1) de l'activation de la voie ERK, (2) des facteurs de transcription comme Sp1 (activé par ERK) et Sp3 (comme un potentiel régulateur de Sp1) (Suske 1999). Dans des conditions de stress, il semble que seul Sp1 voit son activité augmentée (Schafer, Cramer et al. 2003). Parce que l'activité de Sp1 est régulée par la phosphorylation de sa protéine, nous ne notons pas de modulation de son transcript dans notre expérience. Son récepteur est la neuropiline-1, et son inhibition conduit à une activité apoptotique (Ochiumi, Kitadai et al. 2006) par la voie des MAPK (Wey, Gray et

al. 2005). Bien que nous n'observions pas de modulation de ce transcript dans notre expérience, il a été montré qu'en condition hypoxique, son expression est réduite. Dans notre expérience, sa sur-expression chez les CuFi-1 par rapport à nos contrôles est en accord avec les observations faites chez le Dr Verhaeghe (Verhaeghe, Tabruyn et al. 2007), qui propose que ce facteur antigénique participe à l'angiogénèse excessive chez les cellules FK, nous proposons que VEGF-A pourrait également avoir un rôle anti-apoptotique par une activation autocrine du récepteur neuropiline-1.

L'ensemble de nos interprétations reposent sur deux expérimentations dans des conditions similaires : (1) l'activité des caspases 3/7 (reflétant l'état apoptotique) et (2) l'activité transcriptomique. Toutes les deux soulignent, de façon différente, un défaut dans le processus apoptotique et sa relation avec la mutation et le stress oxydatif. Pour interpréter ces observations, nous nous appuyons sur un modèle d'équilibre entre les molécules pro-apoptotiques et anti-apoptotiques. Celui-ci propose une action pro-apoptotique plus importante que l'action anti-apoptotique en absence de stress oxydatif dans les cellules CuFi-1 par rapport au NuLi-1 et ces actions s'inversent en présence d'oxydant. Par conséquent, en considérant les profils d'expression de gènes apoptotiques, ces phénotypes pourraient s'expliquer par les niveaux d'expression de ces transcripts qui influenceraient le devenir de l'apoptose.

Notre méthodologie d'analyse transcriptomique identifie un groupe de gènes apoptotiques qui pourraient participer à ce déséquilibre et, exploitant la connaissance de certains gènes, nous proposons plusieurs hypothèses d'études :

(1) La condition de stress oxydatif chez des cellules FK, inhibe l'activité de la voie des MAPK. Il a été montré qu'en condition *in vitro* l'activité (protéique) des MAPK est plus élevée dans les cellules FK que les cellules non FK (Verhaeghe, Remouchamps et al. 2007). Ceci est en accord avec l'état

apoptotique observé dans notre expérience d'activité des caspases en absence de DMNQ, bien que nous notions une sous-expression de certains transcripts associés à cette voie. Parce que la présence de DMNQ inverse le phénotype et induit une résistance des cellules CuFi-1 à l'apoptose, ceci suggèrerait une diminution de l'activité pro-apoptotique dépendante de la voie des MAPK, qui semble sous-activée, en raison d'inhibition de ses kinases, comme l'inhibition de MAP3K5 par la thioredoxine, ou d'une diminution de la stimulation de cette voie, par l'inhibition de ces récepteurs comme la neuropiline-1.

(2) Le niveau de transcription de ces gènes participant à la balance apoptotique est sous l'influence de facteurs de transcription dépendant du stress oxydatif comme NFAT.

Les molécules inhibitrices de la voie des MAPK sont sous l'action de facteurs de transcriptions sensibles au stress comme Sp-1 et NFAT. En effet, il a été montré que NFAT participe à la voie de signalisation calcineurine-NFAT, qui régule MAP3K5 (Liu, Wilkins et al. 2006). Ces facteurs de transcription pourraient avoir une activité protéique dysfonctionnelle en raison de la mutation FK.

Conclusion

Près de 50 ans après la découverte de la molécule d'ADN par Watson et Crick en 1953 et 10 ans après le premier séquençage du génome humain, nous sommes à un tournant de la recherche en biologie moléculaire. Les nouvelles technologies comme les puces à ADN et de nouveaux acteurs moléculaires, tels que les ARN non codants, incitent les chercheurs à devoir innover dans leurs expériences et à améliorer leurs expertises dans de nouveaux domaines.

La technologie des puces à ADN est le premier outil de biotechnologie à autoriser une étude approfondie des phénomènes cellulaires en permettant une vision complète et globale de la vie du transcriptome. De plus, elle permet d'intégrer dans son design expérimental, des caractéristiques génétiques et/ou environnementales, autorisant l'étude des interactions, qui semble désormais inévitables pour comprendre la physiopathologie des maladies. Bien que largement utilisée dans le typage des tumeurs, elle n'a pas encore trouvé sa pleine maturité dans l'étude des maladies génétiques. Ceci s'explique, en parti par le fait que leurs exploitations exigent une complète maîtrise de nombreux aspects de l'analyse et de son interprétation.

Toutefois au regard de nos connaissances et de notre capacité à distinguer les modulations induites dans deux approches, notre analyse transcriptomique nous a permis de mieux caractériser les processus dysfonctionnels impliqués dans la physiopathologie de la fibrose kystique. Effectivement, nos découvertes soulignent différents types de régulation inflammatoire encore mal compris dans la maladie, comme la présence potentielle de modifications épigénétiques. De plus, en accord avec des résultats d'autres équipes, nous avons mis en évidence un potentiel défaut dans l'apoptose présent chez le type FK en présence de stress oxydatif. Ces résultats sont associés à des listes de gènes d'intérêt dans l'inflammation et l'apoptose et ceux-ci devront être à la base d'expérience plus spécifique dans le but de valider des

hypothèses suggérées par l'interprétation de notre étude transcriptomique et de permettre de détailler les mécanismes impliqués.

Parallèlement à nos interprétations biologiques, nous avons mis à l'étude une nouvelle approche d'analyse différentielle exploitant la variance des données d'expression. Cette méthodologie, adaptée pour l'analyse de design multifactoriel, s'avère efficace pour identifier des processus biologiques spécifiquement perturbés par les conditions expérimentales de notre expérience. À notre connaissance, une telle approche analytique n'a pas été exploitée dans des expériences multifactorielles.

Afin de bien comprendre l'importance du tournant qu'entreprend la biologie moléculaire au XXI^{ème} siècle, nous pouvons faire un rapprochement avec l'histoire de la physique. Celui-ci reflète l'état de nos connaissances en biologie moléculaire puisqu'il est proche de celui d'Isaac Newton lorsque celui-ci regarda la pomme tombée de l'arbre. Cela suggère que les développements les plus importants dans le domaine de la biologie sont devant nous et que nous devons nous attendre à de grandes avancées dans les moyens d'investigations de la biologie moléculaire et cellulaire.

La vision traditionaliste de la biologie moléculaire du XX^{ème} siècle a cherché à détailler les mécanismes moléculaires dans le but de comprendre des processus cellulaires plus élaborés. Étant donné le développement de nouvelles techniques à venir pour l'étude du protéome, de l'interactome, du kinome, du phosphatome et bien d'autres « -ome » qui échappent encore à notre imagination, il sera nécessaire de considérer des nouvelles approches aux visions plus généralistes, où l'ensemble des données devront être considéré comme un tout. Ces approches devront intégrer des modèles mathématiques et statistiques pour saisir la dynamique des événements cellulaires et des algorithmes issus de l'intelligence artificielle pour permettre des extrapolations sur le fonctionnement cellulaire dans sa globalité. La

technologie des puces à ADN, ainsi que notre méthodologie d'analyse développée pour les designs multifactoriels s'inscrit dans ces objectifs et une bonne expertise dans leurs interprétations facilitera les approches futures.

Bibliographie

A

Adcock, I. M., L. Tsaprouni, et al. (2007). "Epigenetic regulation of airway inflammation." *Curr Opin Immunol* 19(6): 694-700.

Adib-Conquy, M., T. Pedron, et al. (2008). "Neutrophils in cystic fibrosis display a distinct gene expression pattern." *Mol Med* 14(1-2): 36-44.

Alexis, N. E., M. S. Muhlebach, et al. (2006). "Attenuation of host defense function of lung phagocytes in young cystic fibrosis patients." *J Cyst Fibros* 5(1): 17-25.

Alizadeh, A. A., M. B. Eisen, et al. (2000). "Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling." *Nature* 403(6769): 503-511.

Allard, J. B., M. E. Poynter, et al. (2006). "Aspergillus fumigatus generates an enhanced Th2-biased immune response in mice with defective cystic fibrosis transmembrane conductance regulator." *J Immunol* 177(8): 5186-5194.

Anderson, M. P., D. P. Rich, et al. (1991). "Generation of cAMP-activated chloride currents by expression of CFTR." *Science* 251(4994): 679-682.

Aridor, M., S. I. Bannykh, et al. (1995). "Sequential coupling between COPII et COPI vesicle coats in endoplasmic reticulum to Golgi transport." *J Cell Biol* 131(4): 875-893.

Ashburner, B. P., S. D. Westerheide, et al. (2001). "The p65 (RelA) subunit of NF-kappaB interacts with the histone deacetylase (HDAC) corepressors HDAC1 and HDAC2 to negatively regulate gene expression." *Mol Cell Biol* 21(20): 7065-7077.

B

Baechler, E. C., F. M. Batliwalla, et al. (2003). "Interferon-inducible gene expression signature in peripheral blood cells of patients with severe lupus." *Proc Natl Acad Sci U S A* 100(5): 2610-2615.

Baggiolini, M. and I. Clark-Lewis (1992). "Interleukin-8, a chemotactic and inflammatory cytokine." *FEBS Lett* 307(1): 97-101.

Baldi, P. and A. D. Long (2001). "A Bayesian framework for the analysis of microarray expression data: regularized t -test and statistical inferences of gene changes." *Bioinformatics* 17(6): 509-519.

Becker, M. N., M. S. Sauer, et al. (2004). "Cytokine secretion by cystic fibrosis airway epithelial cells." *Am J Respir Crit Care Med* 169(5): 645-653.

Benjamini Y, Y. D. (2001). "The control of the false discovery rate in multiple testing under dependency." *Ann Stat* 29: 1165-1188.

Benjamini, Y. a. H., Y (1995). "Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing." *Journal of the Royal Statistical Society B* 57: 289-300.

Bennett, L., A. K. Palucka, et al. (2003). "Interferon and granulopoiesis signatures in systemic lupus erythematosus blood." *J Exp Med* 197(6): 711-723.

Bertucci, F., R. Houlgatte, et al. (2000). "Gene expression profiling of primary breast carcinomas using arrays of candidate genes." *Hum Mol Genet* 9(20): 2981-2991.

Bezdek, J. C. (1981). "Pattern Recognition with Fuzzy Objective Function Algorithms." Plenum Press New York.

Bhavsar, P., T. Ahmad, et al. (2008). "The role of histone deacetylases in asthma and allergic diseases." *J Allergy Clin Immunol* 121(3): 580-584.

Bianchi, M. E. (2007). "DAMPs, PAMPs and alarmins: all we need to know about danger." *J Leukoc Biol* 81(1): 1-5.

Bittner, M., P. Meltzer, et al. (2000). "Molecular classification of cutaneous malignant melanoma by gene expression profiling." *Nature* 406(6795): 536-540.

Biwersi, J., N. Emans, et al. (1996). "Cystic fibrosis transmembrane conductance regulator activation stimulates endosome fusion in vivo." *Proc Natl Acad Sci U S A* 93(22): 12484-12489.

Blackwell, T. S., A. A. Stecenko, et al. (2001). "Dysregulated NF-kappaB activation in cystic fibrosis: evidence for a primary inflammatory disorder." *Am J Physiol Lung Cell Mol Physiol* 281(1): L69-70.

Bolstad, B. M. (1998). "Low-level Analysis of High-density Oligonucleotide Array Data: Background,

Bolstad, B. M., R. A. Irizarry, et al. (2003). "A comparison of normalization methods for high density oligonucleotide array data based on variance and bias." *Bioinformatics* 19(2): 185-193.

Boncoeur, E., O. Tabary, et al. (2006). "Oxidative stress response results in increased p21WAF1/CIP1 degradation in cystic fibrosis lung epithelial cells." *Free Radic Biol Med* 40(1): 75-86.

Boncoeur, E., T. Roque, et al. (2008). "Cystic fibrosis transmembrane conductance regulator controls lung proteasomal degradation and nuclear

factor-kappaB activity in conditions of oxidative stress." *Am J Pathol* 172(5): 1184-1194.

Boncoeur, E., V. S. Criq, et al. (2008). "Oxidative stress induces extracellular signal-regulated kinase 1/2 mitogen-activated protein kinase in cystic fibrosis lung epithelial cells: Potential mechanism for excessive IL-8 expression." *Int J Biochem Cell Biol* 40(3): 432-446.

Bonfield, T. L., M. W. Konstan, et al. (1999). "Altered respiratory epithelial cell cytokine production in cystic fibrosis." *J Allergy Clin Immunol* 104(1): 72-78.

Boucher, R. C., T. Chinet, et al. (1991). "Ion transport in normal and CF airway epithelia." *Adv Exp Med Biol* 290: 105-115; discussion 115-108.

Boutanaev, A. M., A. I. Kalmykova, et al. (2002). "Large clusters of co-expressed genes in the *Drosophila* genome." *Nature* 420(6916): 666-669.

Bradbury, N. A., T. Jilling, et al. (1992). "Regulation of plasma membrane recycling by CFTR." *Science* 256(5056): 530-532.

Broackes-Carter, F. C., S. H. Williams, et al. (2003). "Alternative splicing of the ovine CFTR gene." *Mamm Genome* 14(11): 778-787.

Brown, M. A., W. J. Morgan, et al. (1991). "Circulating levels of tumor necrosis factor and interleukin-1 in cystic fibrosis." *Pediatr Pulmonol* 10(2): 86-91.

Browne, E. P., B. Wing, et al. (2001). "Altered cellular mRNA levels in human cytomegalovirus-infected fibroblasts: viral block to the accumulation of antiviral mRNAs." *J Virol* 75(24): 12319-12330.

Buchanan, P. J., R. K. Ernst, et al. (2009). "Role of CFTR, *Pseudomonas aeruginosa* and Toll-like receptors in cystic fibrosis lung inflammation." *Biochem Soc Trans* 37(Pt 4): 863-867.

Buckley, S., L. Barsky, et al. (2005). "In vivo inosine protects alveolar epithelial type 2 cells against hyperoxia-induced DNA damage through MAP kinase signaling." *Am J Physiol Lung Cell Mol Physiol* 288(3): L569-575.

C

Cannon, C. L., M. P. Kowalski, et al. (2003). "Pseudomonas aeruginosa-induced apoptosis is defective in respiratory epithelial cells expressing mutant cystic fibrosis transmembrane conductance regulator." *Am J Respir Cell Mol Biol* 29(2): 188-197.

Caron, H., B. van Schaik, et al. (2001). "The human transcriptome map: clustering of highly expressed genes in chromosomal domains." *Science* 291(5507): 1289-1292.

Carrabino, S., D. Carpani, et al. (2006). "Dysregulated interleukin-8 secretion and NF-kappaB activity in human cystic fibrosis nasal epithelial cells." *J Cyst Fibros* 5(2): 113-119.

Carrillo de Santa Pau, E., F. C. Arias, et al. (2009). "Prognostic significance of the expression of vascular endothelial growth factors A, B, C, and D and their receptors R1, R2, and R3 in patients with nonsmall cell lung cancer." *Cancer* 115(8): 1701-1712.

Chappe, V., T. Irvine, et al. (2005). "Phosphorylation of CFTR by PKA promotes binding of the regulatory domain." *EMBO J* 24(15): 2730-2740.

Chaudhary, P. M., M. Eby, et al. (1997). "Death receptor 5, a new member of the TNFR family, and DR4 induce FADD-dependent apoptosis and activate the NF-kappaB pathway." *Immunity* 7(6): 821-830.

Chauhan, D., G. Li, et al. (2003). "Identification of genes regulated by 2-methoxyestradiol (2ME2) in multiple myeloma cells using oligonucleotide arrays." *Blood* 101(9): 3606-3614.

Chen, L., W. Fischle, et al. (2001). "Duration of nuclear NF-kappaB action regulated by reversible acetylation." *Science* 293(5535): 1653-1657.

Chen, Y., T. Knosel, et al. (2003). "Loss of PDCD4 expression in human lung cancer correlates with tumour progression and prognosis." *J Pathol* 200(5): 640-646.

Cheng, S. H., D. P. Rich, et al. (1991). "Phosphorylation of the R domain by cAMP-dependent protein kinase regulates the CFTR chloride channel." *Cell* 66(5): 1027-1036.

Chmiel, J. F. and P. B. Davis (2003). "State of the art: why do the lungs of patients with cystic fibrosis become infected and why can't they clear the infection?" *Respir Res* 4: 8.

Chmiel, J. F., M. W. Konstan, et al. (1999). "IL-10 attenuates excessive inflammation in chronic *Pseudomonas* infection in mice." *Am J Respir Crit Care Med* 160(6): 2040-2047.

Cho, R. J., M. J. Campbell, et al. (1998). "A genome-wide transcriptional analysis of the mitotic cell cycle." *Mol Cell* 2(1): 65-73.

Choi, S. S., I. C. Park, et al. (1995). "A novel Bcl-2 related gene, Bfl-1, is overexpressed in stomach cancer and preferentially expressed in bone marrow." *Oncogene* 11(9): 1693-1698.

Choo-Kang, L. R. and P. L. Zeitlin (2001). "Induction of HSP70 promotes DeltaF508 CFTR trafficking." *Am J Physiol Lung Cell Mol Physiol* 281(1): L58-68.

Churchill, G. A. (2004). "Using ANOVA to analyze microarray data." *Biotechniques* 37(2): 173-175, 177.

Creagh, E. M. and L. A. O'Neill (2006). "TLRs, NLRs and RLRs: a trinity of pathogen sensors that co-operate in innate immunity." *Trends Immunol* 27(8): 352-357.

Crescenzi, M. and A. Giuliani (2001). "The main biological determinants of tumor line taxonomy elucidated by a principal component analysis of microarray data." *FEBS Lett* 507(1): 114-118.

Culhane, A. C., G. Perriere, et al. (2002). "Between-group analysis of microarray data." *Bioinformatics* 18(12): 1600-1608.

Cutting, G. R. (2005). "Modifier genetics: cystic fibrosis." *Annu Rev Genomics Hum Genet* 6: 237-260.

D

De Braekeleer, M., C. Allard, et al. (1998). "Correlation of sweat chloride concentration with genotypes in cystic fibrosis patients in Saguenay Lac-Saint-Jean, Quebec, Canada." *Clin Biochem* 31(1): 33-36.

de Zoeten, E. F., L. Wang, et al. (2010). "Inhibition of HDAC9 increases T regulatory cell function and prevents colitis in mice." *Gastroenterology* 138(2): 583-594.

Dean, T. P., Y. Dai, et al. (1993). "Interleukin-8 concentrations are elevated in bronchoalveolar lavage, sputum, and sera of children with cystic fibrosis." *Pediatr Res* 34(2): 159-161.

Degli-Esposti, M. A., W. C. Dougall, et al. (1997). "The novel receptor TRAIL-R4 induces NF-kappaB and protects against TRAIL-mediated apoptosis, yet retains an incomplete death domain." *Immunity* 7(6): 813-820.

Denning, G. M., M. P. Anderson, et al. (1992). "Processing of mutant cystic fibrosis transmembrane conductance regulator is temperature-sensitive." *Nature* 358(6389): 761-764.

Der, S. D., A. Zhou, et al. (1998). "Identification of genes differentially regulated by interferon alpha, beta, or gamma using oligonucleotide arrays." *Proc Natl Acad Sci U S A* 95(26): 15623-15628.

DeRisi, J. L., V. R. Iyer, et al. (1997). "Exploring the metabolic and genetic control of gene expression on a genomic scale." *Science* 278(5338): 680-686.

Dieckgraefe, B. K., W. F. Stenson, et al. (2000). "Analysis of mucosal gene expression in inflammatory bowel disease by parallel oligonucleotide arrays." *Physiol Genomics* 4(1): 1-11.

DiMango, E., A. J. Ratner, et al. (1998). "Activation of NF-kappaB by adherent *Pseudomonas aeruginosa* in normal and cystic fibrosis respiratory epithelial cells." *J Clin Invest* 101(11): 2598-2605.

Dinarello, C. A. (2009). "Immunological and inflammatory functions of the interleukin-1 family." *Annu Rev Immunol* 27: 519-550.

Divangahi, M., H. Balghi, et al. (2009). "Lack of CFTR in skeletal muscle predisposes to muscle wasting and diaphragm muscle pump failure in cystic fibrosis mice." *PLoS Genet* 5(7): e1000586.

Draghici, S. (2002). "Statistical intelligence: effective analysis of high-density microarray data." *Drug Discov Today* 7(11 Suppl): S55-63.

D'Sa-Eipper, C., T. Subramanian, et al. (1996). "bfl-1, a bcl-2 homologue, suppresses p53-induced apoptosis and exhibits potent cooperative transforming activity." *Cancer Res* 56(17): 3879-3882.

Du, K., M. Sharma, et al. (2005). "The DeltaF508 cystic fibrosis mutation impairs domain-domain interactions and arrests post-translational folding of CFTR." *Nat Struct Mol Biol* 12(1): 17-25.

Dunn, J. C. (1973). "A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters." *Journal of Cybernetics* 3: 32-57.

Durbin, B. P. and D. M. Rocke (2004). "Variance-stabilizing transformations for two-color microarrays." *Bioinformatics* 20(5): 660-667.

Durieu, I., C. Amsellem, et al. (1999). "Fas and Fas ligand expression in cystic fibrosis airway epithelium." *Thorax* 54(12): 1093-1098.

E

Efron, B. and R. Tibshirani (2002). "Empirical bayes methods and false discovery rates for microarrays." *Genet Epidemiol* 23(1): 70-86.

Egger, G., G. Liang, et al. (2004). "Epigenetics in human disease and prospects for epigenetic therapy." *Nature* 429(6990): 457-463.

Ehrchen, J. M., C. Sunderkotter, et al. (2009). "The endogenous Toll-like receptor 4 agonist S100A8/S100A9 (calprotectin) as innate amplifier of infection, autoimmunity, and cancer." *J Leukoc Biol* 86(3): 557-566.

Eidelman, O., M. Srivastava, et al. (2001). "Control of the proinflammatory state in cystic fibrosis lung epithelial cells by genes from the TNF- α /NF κ B pathway." *Mol Med* 7(8): 523-534.

Eisen, M. B., P. T. Spellman, et al. (1998). "Cluster analysis and display of genome-wide expression patterns." *Proc Natl Acad Sci U S A* 95(25): 14863-14868.

Ellgaard, L. and A. Helenius (2001). "ER quality control: towards an understanding at the molecular level." *Curr Opin Cell Biol* 13(4): 431-437.

Ernst, R. K., A. M. Hajjar, et al. (2003). "Pseudomonas aeruginosa lipid A diversity and its recognition by Toll-like receptor 4." *J Endotoxin Res* 9(6): 395-400.

Estivill, X., C. Bancells, et al. (1997). "Geographic distribution and regional origin of 272 cystic fibrosis mutations in European populations. The Biomed CF Mutation Analysis Consortium." *Hum Mutat* 10(2): 135-154.

F

Fadok, V. A., A. de Cathelineau, et al. (2001). "Loss of phospholipid asymmetry and surface exposure of phosphatidylserine is required for phagocytosis of apoptotic cells by macrophages and fibroblasts." *J Biol Chem* 276(2): 1071-1077.

Fadok, V. A., D. R. Voelker, et al. (1992). "Exposure of phosphatidylserine on the surface of apoptotic lymphocytes triggers specific recognition and removal by macrophages." *J Immunol* 148(7): 2207-2216.

Fellenberg, K., N. C. Hauser, et al. (2001). "Correspondence analysis applied to microarray data." *Proc Natl Acad Sci U S A* 98(19): 10781-10786.

Fisher, R. A. (1954). "Statistical Methods for Research Workers." Oliver and Boyd.

Foell, D., S. Seeliger, et al. (2003). "Expression of S100A12 (EN-RAGE) in cystic fibrosis." *Thorax* 58(7): 613-617.

Futschik, M. (2009). "Mfuzz: Soft clustering of time series gene expression data.".

G

Gabriel, S. E., L. L. Clarke, et al. (1993). "CFTR and outward rectifying chloride channels are distinct proteins with a regulatory relationship." *Nature* 363(6426): 263-268.

Gambardella, S., M. Biancolella, et al. (2006). "Gene expression profile study in CFTR mutated bronchial cell lines." *Clin Exp Med* 6(4): 157-165.

Gan, K. H., W. P. Geus, et al. (1995). "Genetic and clinical features of patients with cystic fibrosis diagnosed after the age of 16 years." *Thorax* 50(12): 1301-1304.

Gant, T. W., D. N. Rao, et al. (1988). "Redox cycling and sulphydryl arylation; their relative importance in the mechanism of quinone cytotoxicity to isolated hepatocytes." *Chem Biol Interact* 65(2): 157-173.

Gao, L., K. J. Kim, et al. (1999). "Abnormal glutathione transport in cystic fibrosis airway epithelia." *Am J Physiol* 277(1 Pt 1): L113-118.

Gasch, A. P. and M. B. Eisen (2002). "Exploring the conditional coregulation of yeast gene expression through fuzzy k-means clustering." *Genome Biol* 3(11): RESEARCH0059.

Gelman, M. S., E. S. Kannegaard, et al. (2002). "A principal role for the proteasome in endoplasmic reticulum-associated degradation of misfolded intracellular cystic fibrosis transmembrane conductance regulator." *J Biol Chem* 277(14): 11709-11714.

Gentleman, R. C., V. J. Carey, et al. (2004). "Bioconductor: open software development for computational biology and bioinformatics." *Genome Biol* 5(10): R80.

Gentleman, R., Carey, V., Huber, W. and Hahne, F. " genefilter: methods for filtering genes from microarray experiments."

Gentleman., S. F. a. R. (2007). " Using GOstats to test gene lists for GO term association." *Bioinformatics*, 23: 257-258.

Gething, M. J. and J. Sambrook (1992). "Protein folding in the cell." *Nature* 355(6355): 33-45.

Ghosh, S. and M. Karin (2002). "Missing pieces in the NF-kappaB puzzle." *Cell* 109 Suppl: S81-96.

Golub, T. R., D. K. Slonim, et al. (1999). "Molecular classification of cancer: class discovery and class prediction by gene expression monitoring." *Science* 286(5439): 531-537.

Gottlieb, R. A. and A. Dosanjh (1996). "Mutant cystic fibrosis transmembrane conductance regulator inhibits acidification and apoptosis in C127 cells: possible relevance to cystic fibrosis." *Proc Natl Acad Sci U S A* 93(8): 3587-3591.

Greene, C. M. and N. G. McElvaney (2005). "Toll-like receptor expression and function in airway epithelial cells." *Arch Immunol Ther Exp (Warsz)* 53(5): 418-427.

Guenzi, E., K. Topolt, et al. (2001). "The helical domain of GBP-1 mediates the inhibition of endothelial cell proliferation by inflammatory cytokines." *EMBO J* 20(20): 5568-5577.

Guilbault, C., J. P. Novak, et al. (2006). "Distinct pattern of lung gene expression in the Cftr-KO mice developing spontaneous lung disease compared with their littermate controls." *Physiol Genomics* 25(2): 179-193.

Guillot, L., R. Le Goffic, et al. (2005). "Involvement of toll-like receptor 3 in the immune response of lung epithelial cells to double-stranded RNA and influenza A virus." *J Biol Chem* 280(7): 5571-5580.

H

Haardt, M., M. Benharouga, et al. (1999). "C-terminal truncations destabilize the cystic fibrosis transmembrane conductance regulator without impairing its biogenesis. A novel class of mutation." *J Biol Chem* 274(31): 21873-21877.

Haberland, M., R. L. Montgomery, et al. (2009). "The many roles of histone deacetylases in development and physiology: implications for disease and therapy." *Nat Rev Genet* 10(1): 32-42.

- Hackstadt, A. J. and A. M. Hess (2009). "Filtering for increased power for microarray data analysis." *BMC Bioinformatics* 10: 11.
- Haller, O. and G. Kochs (2002). "Interferon-induced mx proteins: dynamin-like GTPases with antiviral activity." *Traffic* 3(10): 710-717.
- Harper, S. J. and D. O. Bates (2008). "VEGF-A splicing: the key to anti-angiogenic therapeutics?" *Nat Rev Cancer* 8(11): 880-887.
- Hassid, S., M. P. Degaute, et al. (1997). "Determination of proliferative activity in nasal polyps." *J Clin Pathol* 50(11): 923-928.
- He, C. H., A. B. Waxman, et al. (2005). "Bcl-2-related protein A1 is an endogenous and cytokine-stimulated mediator of cytoprotection in hyperoxic acute lung injury." *J Clin Invest* 115(4): 1039-1048.
- Heber, S. and B. Sick (2006). "Quality assessment of Affymetrix GeneChip data." *OMICS* 10(3): 358-368.
- Hedenfalk, I. A., M. Ringner, et al. (2002). "Gene expression in inherited breast cancer." *Adv Cancer Res* 84: 1-34.
- Hill, A. E., J. S. Hong, et al. (2006). "Micro-RNA-like effects of complete intronic sequences." *Front Biosci* 11: 1998-2006.
- Hinata, K., A. M. Gervin, et al. (2003). "Divergent gene regulation and growth effects by NF-kappa B in epithelial and mesenchymal cells of human skin." *Oncogene* 22(13): 1955-1964.
- Hochberg, Y. (1988). "A sharper bonferroni procedure for multiple tests of significance." *Biometrika* 75 800-802.807.

Hofmann, M. A., S. Drury, et al. (1999). "RAGE mediates a novel proinflammatory axis: a central cell surface receptor for S100/calgranulin polypeptides." *Cell* 97(7): 889-901.

Hofmann, T. G., N. Stollberg, et al. (2003). "HIPK2 regulates transforming growth factor-beta-induced c-Jun NH(2)-terminal kinase activation and apoptosis in human hepatoma cells." *Cancer Res* 63(23): 8271-8277.

Holland, K. A. and I. B. Holland (2005). "Adventures with ABC-proteins: highly conserved ATP-dependent transporters." *Acta Microbiol Immunol Hung* 52(3-4): 309-322.

Holm, S. (1979). "A simple sequentially rejective multiple test procedure." *Scand. J. Statist.* 6: 65-70.66.

Huber, W., A. von Heydebreck, et al. (2002). "Variance stabilization applied to microarray data calibration and to the quantification of differential expression." *Bioinformatics* 18 Suppl 1: S96-104.

Hubert, D., T. Bienvenu, et al. (1996). "Genotype-phenotype relationships in a cohort of adult cystic fibrosis patients." *Eur Respir J* 9(11): 2207-2214.

Hung, S. P., P. Baldi, et al. (2002). "Global gene expression profiling in *Escherichia coli* K12. The effects of leucine-responsive regulatory protein." *J Biol Chem* 277(43): 40309-40323.

I

Iida, N. and G. R. Grotendorst (1990). "Cloning and sequencing of a new gro transcript from activated human monocytes: expression in leukocytes and wound tissue." *Mol Cell Biol* 10(10): 5596-5599.

Irizarry, R. A., B. Hobbs, et al. (2003). "Exploration, normalization, and summaries of high density oligonucleotide array probe level data." *Biostatistics* 4(2): 249-264.

Irizarry, R. A., B. M. Bolstad, et al. (2003). "Summaries of Affymetrix GeneChip probe level data." *Nucleic Acids Res* 31(4): e15.

Ito, K., M. Ito, et al. (2005). "Decreased histone deacetylase activity in chronic obstructive pulmonary disease." *N Engl J Med* 352(19): 1967-1976.

Ito, K., P. J. Barnes, et al. (2000). "Glucocorticoid receptor recruitment of histone deacetylase 2 inhibits interleukin-1beta-induced histone H4 acetylation on lysines 8 and 12." *Mol Cell Biol* 20(18): 6891-6903.

J

Jacquier-Sarlin, M. R., K. Fuller, et al. (1994). "Protective effects of hsp70 in inflammation." *Experientia* 50(11-12): 1031-1038.

Jiang, C. H., J. Z. Tsien, et al. (2001). "The effects of aging on gene expression in the hypothalamus and cortex of mice." *Proc Natl Acad Sci U S A* 98(4): 1930-1934.

Jiang, Y. H., J. Bressler, et al. (2004). "Epigenetics and human disease." *Annu Rev Genomics Hum Genet* 5: 479-510.

John, G., A. O. Yildirim, et al. (2009). "Toll-like Receptor (TLR)-4 Mediated Innate Immunity is Reduced in Cystic Fibrosis Airway Cells." *Am J Respir Cell Mol Biol*.

Joseph, T., D. Look, et al. (2005). "NF-kappaB activation and sustained IL-8 gene expression in primary cultures of cystic fibrosis airway epithelial cells

stimulated with *Pseudomonas aeruginosa*." *Am J Physiol Lung Cell Mol Physiol* 288(3): L471-479.

Jungas, T., I. Motta, et al. (2002). "Glutathione levels and BAX activation during apoptosis due to oxidative stress in cells expressing wild-type and mutant cystic fibrosis transmembrane conductance regulator." *J Biol Chem* 277(31): 27912-27918.

K

Karen Kafadar , T. P. (2003). "Transformations, background estimation, and process effects in the statistical analysis of microarrays." *Computational Statistics & Data Analysis* 44 313 - 338.

Kaur, S., O. Norkina, et al. (2004). "Acidic duodenal pH alters gene expression in the cystic fibrosis mouse pancreas." *Am J Physiol Gastrointest Liver Physiol* 287(2): G480-490.

Kelley, T. J. and H. L. Elmer (2000). "In vivo alterations of IFN regulatory factor-1 and PIAS1 protein levels in cystic fibrosis epithelium." *J Clin Invest* 106(3): 403-410.

Kelley, T. J. and M. L. Drumm (1998). "Inducible nitric oxide synthase expression is reduced in cystic fibrosis murine and human airway epithelial cells." *J Clin Invest* 102(6): 1200-1207.

Kerr, M. K. and G. A. Churchill (2001). "Bootstrapping cluster analysis: assessing the reliability of conclusions from microarray experiments." *Proc Natl Acad Sci U S A* 98(16): 8961-8965.

Khan, J., R. Simon, et al. (1998). "Gene expression profiling of alveolar rhabdomyosarcoma with cDNA microarrays." *Cancer Res* 58(22): 5009-5013.

Klampfer, L., J. Huang, et al. (2004). "Requirement of histone deacetylase activity for signaling by STAT1." *J Biol Chem* 279(29): 30358-30368.

Klein, U., Y. Tu, et al. (2001). "Gene expression profiling of B cell chronic lymphocytic leukemia reveals a homogeneous phenotype related to memory B cells." *J Exp Med* 194(11): 1625-1638.

Kogan, I., M. Ramjeesingh, et al. (2003). "CFTR directly mediates nucleotide-regulated glutathione flux." *EMBO J* 22(9): 1981-1989.

Koller, B. H., H. S. Kim, et al. (1991). "Toward an animal model of cystic fibrosis: targeted interruption of exon 10 of the cystic fibrosis transmembrane regulator gene in embryonic stem cells." *Proc Natl Acad Sci U S A* 88(23): 10730-10734.

Koller, D. Y., I. Nething, et al. (1997). "Cytokine concentrations in sputum from patients with cystic fibrosis and their relation to eosinophil activity." *Am J Respir Crit Care Med* 155(3): 1050-1054.

Koller, D. Y., R. Urbanek, et al. (1995). "Increased degranulation of eosinophil and neutrophil granulocytes in cystic fibrosis." *Am J Respir Crit Care Med* 152(2): 629-633.

Konstas, A. A., J. P. Koch, et al. (2003). "cAMP-dependent activation of CFTR inhibits the epithelial sodium channel (ENaC) without affecting its surface expression." *Pflugers Arch* 445(4): 513-521.

Koscielny, G., V. Le Texier, et al. (2009). "ASTD: The Alternative Splicing and Transcript Diversity database." *Genomics* 93(3): 213-220.

Kowalski, M. P. and G. B. Pier (2004). "Localization of cystic fibrosis transmembrane conductance regulator to lipid rafts of epithelial cells is

required for *Pseudomonas aeruginosa*-induced cellular activation." *J Immunol* 172(1): 418-425.

Kresch, M. J., C. Christian, et al. (1998). "Ontogeny of apoptosis during lung development." *Pediatr Res* 43(3): 426-431.

Kunzelmann, K., R. Schreiber, et al. (2000). "Control of epithelial Na⁺ conductance by the cystic fibrosis transmembrane conductance regulator." *Pflugers Arch* 440(2): 193-201.

Kunzelmann, K., R. Schreiber, et al. (2001). "Mechanisms of the inhibition of epithelial Na⁽⁺⁾ channels by CFTR and purinergic stimulation." *Kidney Int* 60(2): 455-461.

L

Larkin, J. E., B. C. Frank, et al. (2005). "Independence and reproducibility across microarray platforms." *Nat Methods* 2(5): 337-344.

Larsson, O., K. Wennmalm, et al. (2006). "Comparative microarray analysis." *OMICS* 10(3): 381-397.

Lasagni, L., M. Francalanci, et al. (2003). "An alternatively spliced variant of CXCR3 mediates the inhibition of endothelial cell growth induced by IP-10, Mig, and I-TAC, and acts as functional receptor for platelet factor 4." *J Exp Med* 197(11): 1537-1549.

LeBlanc, A. C. (2003). "Natural cellular inhibitors of caspases." *Prog Neuropsychopharmacol Biol Psychiatry* 27(2): 215-229.

Lee, M. G., W. C. Wigley, et al. (1999). "Regulation of Cl⁻/ HCO₃⁻ exchange by cystic fibrosis transmembrane conductance regulator expressed in NIH 3T3 and HEK 293 cells." *J Biol Chem* 274(6): 3414-3421.

Legendre, P. and L. Legendre (1998). *Numerical ecology*. New York, Elsevier.

Legendre, P. M. J. A. (1999). "Distance-based redundancy analysis: testing multi-species responses in multi-factorial ecological experiments." *Ecological Monographs* 69 (1): 1-24.

Legendre, V. M. e. P. (1999). "UNE MÉTHODE D'ANALYSE CANONIQUE NON-LINÉAIRE ET SON APPLICATION À DES DONNÉES BIOLOGIQUES." *Math. Inf. Sci. Hum* 147: 135-147.

Leigh, M. W., J. E. Kylander, et al. (1995). "Cell proliferation in bronchial epithelium and submucosal glands of cystic fibrosis patients." *Am J Respir Cell Mol Biol* 12(6): 605-612.

Lewis, H. A., X. Zhao, et al. (2005). "Impact of the deltaF508 mutation in first nucleotide-binding domain of human cystic fibrosis transmembrane conductance regulator on domain folding and structure." *J Biol Chem* 280(2): 1346-1353.

Li, J., X. D. Johnson, et al. (2003). "Signaling intermediates required for NF-kappa B activation and IL-8 expression in CF bronchial epithelial cells." *Am J Physiol Lung Cell Mol Physiol* 284(2): L307-315.

Li, Q., B. T. Lee, et al. (2005). "Genome-scale analysis of positional clustering of mouse testis-specific genes." *BMC Genomics* 6(1): 7.

Liechti-Gallati, S., I. Bonsall, et al. (1992). "Genotype/phenotype association in cystic fibrosis: analyses of the delta F508, R553X, and 3905insT mutations." *Pediatr Res* 32(2): 175-178.

Linsdell, P. and J. W. Hanrahan (1998). "Glutathione permeability of CFTR." *Am J Physiol* 275(1 Pt 1): C323-326.

Linsdell, P., J. A. Tabcharani, et al. (1997). "Permeability of wild-type and mutant cystic fibrosis transmembrane conductance regulator chloride channels to polyatomic anions." *J Gen Physiol* 110(4): 355-364.

Liu, D., C. Li, et al. (2004). "Nuclear import of proinflammatory transcription factors is required for massive liver apoptosis induced by bacterial lipopolysaccharide." *J Biol Chem* 279(46): 48434-48442.

Liu, G., A. E. Loraine, et al. (2003). "NetAffx: Affymetrix probesets and annotations." *Nucleic Acids Res* 31(1): 82-86.

Liu, Q., B. J. Wilkins, et al. (2006). "Direct interaction and reciprocal regulation between ASK1 and calcineurin-NFAT control cardiomyocyte death and growth." *Mol Cell Biol* 26(10): 3785-3797.

Livak, K. J. and T. D. Schmittgen (2001). "Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method." *Methods* 25(4): 402-408.

Lorenz, E., M. S. Muhlebach, et al. (2008). "Different expression ratio of S100A8/A9 and S100A12 in acute and chronic lung diseases." *Respir Med* 102(4): 567-573.

Loussouarn, G., S. Demolombe, et al. (1996). "Expression of CFTR controls cAMP-dependent activation of epithelial K⁺ currents." *Am J Physiol* 271(5 Pt 1): C1565-1573.

Lukacs, G. L., A. Mohamed, et al. (1994). "Conformational maturation of CFTR but not its mutant counterpart (Δ F508) occurs in the endoplasmic reticulum and requires ATP." *EMBO J* 13(24): 6076-6086.

Lyczak, J. B., C. L. Cannon, et al. (2002). "Lung infections associated with cystic fibrosis." *Clin Microbiol Rev* 15(2): 194-222.

M

MacGregor, G., R. D. Gray, et al. (2008). "Biomarkers for cystic fibrosis lung disease: application of SELDI-TOF mass spectrometry to BAL fluid." *J Cyst Fibros* 7(5): 352-358.

Maiuri, L., V. Raia, et al. (1997). "DNA fragmentation is a feature of cystic fibrosis epithelial cells: a disease with inappropriate apoptosis?" *FEBS Lett* 408(2): 225-231.

Mall, M., M. Bleich, et al. (1998). "The amiloride-inhibitable Na⁺ conductance is reduced by the cystic fibrosis transmembrane conductance regulator in normal but not in cystic fibrosis airways." *J Clin Invest* 102(1): 15-21.

Matute-Bello, G., R. K. Winn, et al. (2001). "Fas (CD95) induces alveolar epithelial cell apoptosis in vivo: implications for acute pulmonary inflammation." *Am J Pathol* 158(1): 153-161.

McGarvey, L. P., K. Dunbar, et al. (2002). "Cytokine concentrations and neutrophil elastase activity in bronchoalveolar lavage and induced sputum from

patients with cystic fibrosis, mild asthma and healthy volunteers." *J Cyst Fibros* 1(4): 269-275.

Meyer, K. C. and J. J. Zimmerman (2002). "Inflammation and surfactant." *Paediatr Respir Rev* 3(4): 308-314.

Miller, R. A., A. Galecki, et al. (2001). "Interpretation, design, and analysis of gene array expression experiments." *J Gerontol A Biol Sci Med Sci* 56(2): B52-57.

Moser, B., I. Clark-Lewis, et al. (1990). "Neutrophil-activating properties of the melanoma growth-stimulatory activity." *J Exp Med* 171(5): 1797-1802.

Moser, C., P. O. Jensen, et al. (2005). "Serum concentrations of GM-CSF and G-CSF correlate with the Th1/Th2 cytokine response in cystic fibrosis patients with chronic *Pseudomonas aeruginosa* lung infection." *APMIS* 113(6): 400-409.

Mourad, T., K. L. Min, et al. (2000). "Measurement of oxidized glutathione by enzymatic recycling coupled to bioluminescent detection." *Anal Biochem* 283(2): 146-152.

Mrugacz, M., J. Zak, et al. (2007). "ICAM-1 expression on conjunctival epithelial cells in patients with cystic fibrosis." *Cytometry B Clin Cytom* 72(3): 204-208.

Muhlebach, M. S., P. W. Stewart, et al. (1999). "Quantitation of inflammatory responses to bacteria in young cystic fibrosis and control patients." *Am J Respir Crit Care Med* 160(1): 186-191.

Muir, A., G. Soong, et al. (2004). "Toll-like receptors in normal and cystic fibrosis airway epithelial cells." *Am J Respir Cell Mol Biol* 30(6): 777-783.

Murphy, D. (2002). "Gene expression studies using microarrays: principles, problems, and prospects." *Adv Physiol Educ* 26(1-4): 256-270.

N

Nacken, W., J. Roth, et al. (2003). "S100A9/S100A8: Myeloid representatives of the S100 protein family as prominent players in innate immunity." *Microsc Res Tech* 60(6): 569-580.

Newton, M. A., C. M. Kendzioriski, et al. (2001). "On differential variability of expression ratios: improving statistical inference about gene expression changes from microarray data." *J Comput Biol* 8(1): 37-52.

Nguyen, T. T., Y. Hu, et al. (2002). "Murine GBP-5, a new member of the murine guanylate-binding protein family, is coordinately regulated with other GBPs in vivo and in vitro." *J Interferon Cytokine Res* 22(8): 899-909.

Nichols, D., J. Chmiel, et al. (2008). "Chronic inflammation in the cystic fibrosis lung: alterations in inter- and intracellular signaling." *Clin Rev Allergy Immunol* 34(2): 146-162.

Nie, H., P. B. Neerincx, et al. (2009). "Microarray data mining using Bioconductor packages." *BMC Proc* 3 Suppl 4: S9.

Norkina, O., S. Kaur, et al. (2004). "Inflammation of the cystic fibrosis mouse small intestine." *Am J Physiol Gastrointest Liver Physiol* 286(6): G1032-1041.

O

Ochiumi, T., Y. Kitadai, et al. (2006). "Neuropilin-1 is involved in regulation of apoptosis and migration of human colon cancer." *Int J Oncol* 29(1): 105-116.

Oshiumi, H., M. Matsumoto, et al. (2003). "TICAM-1, an adaptor molecule that participates in Toll-like receptor 3-mediated interferon-beta induction." *Nat Immunol* 4(2): 161-167.

Ott, C. J., M. Suszko, et al. (2009). "A complex intronic enhancer regulates expression of the CFTR gene by direct interaction with the promoter." *J Cell Mol Med* 13(4): 680-692.

P

Pan, W. (2002). "A comparative review of statistical methods for discovering differentially expressed genes in replicated microarray experiments." *Bioinformatics* 18(4): 546-554.

Papaiahgari, S., Q. Zhang, et al. (2006). "Hyperoxia stimulates an Nrf2-ARE transcriptional response via ROS-EGFR-PI3K-Akt/ERK MAP kinase signaling in pulmonary epithelial cells." *Antioxid Redox Signal* 8(1-2): 43-52.

Parnet, P., K. E. Garka, et al. (1996). "IL-1Rrp is a novel receptor-like molecule similar to the type I interleukin-1 receptor and its homologues T1/ST2 and IL-1R AcP." *J Biol Chem* 271(8): 3967-3970.

Pau, E. C., F. C. Arias, et al. (2009). "Vascular Endothelial Growth Factor (VEGF) Serum Levels Are Associated With Survival in Early Stages of Lung Cancer Patients." *Cancer Invest.*

Pavlidis, P. (2003). "Using ANOVA for gene selection from microarray studies of the nervous system." *Methods* 31(4): 282-289.

Pelus, L. M. and S. Fukuda (2006). "Peripheral blood stem cell mobilization: the CXCR2 ligand GRObeta rapidly mobilizes hematopoietic stem cells with enhanced engraftment properties." *Exp Hematol* 34(8): 1010-1020.

Perez, A., A. C. Issler, et al. (2007). "CFTR inhibition mimics the cystic fibrosis inflammatory profile." *Am J Physiol Lung Cell Mol Physiol* 292(2): L383-395.

Perez-Iratxeta, C. and M. A. Andrade (2005). "Inconsistencies over time in 5% of NetAffx probe-to-gene annotations." *BMC Bioinformatics* 6: 183.

Persson, T., N. Monsef, et al. (2003). "Expression of the neutrophil-activating CXC chemokine ENA-78/CXCL5 by human eosinophils." *Clin Exp Allergy* 33(4): 531-537.

Pietzsch, J. and S. Hoppmann (2009). "Human S100A12: a novel key player in inflammation?" *Amino Acids* 36(3): 381-389.

Pinet, C., M. Cassart, et al. (2003). "Function and bulk of respiratory and limb muscles in patients with cystic fibrosis." *Am J Respir Crit Care Med* 168(8): 989-994.

Pounds, S. B. (2006). "Estimation and control of multiple testing error rates for microarray studies." *Brief Bioinform* 7(1): 25-36.

Prince, L. S., R. B. Workman, Jr., et al. (1994). "Rapid endocytosis of the cystic fibrosis transmembrane conductance regulator chloride channel." *Proc Natl Acad Sci U S A* 91(11): 5192-5196.

R

R. Gentleman, V. C., S. Dudoit, R. Irizarry, W. Huber (2005). "Limma: linear models for microarray data."

Radaeva, S., B. Jaruga, et al. (2002). "Interferon-alpha activates multiple STAT signals and down-regulates c-Met in primary human hepatocytes." *Gastroenterology* 122(4): 1020-1034.

Raia, V., L. Maiuri, et al. (2005). "Inhibition of p38 mitogen activated protein kinase controls airway inflammation in cystic fibrosis." *Thorax* 60(9): 773-780.

Rajendrasozhan, S., H. Yao, et al. (2009). "Current perspectives on role of chromatin modifications and deacetylases in lung inflammation in COPD." *COPD* 6(4): 291-297.

Rao, C. R. (1964). "The use and interpretation of principal component analysis in applied research." *Sankhyaá, Ser. A* 26: 329-358.

Raychaudhuri, S., J. M. Stuart, et al. (2000). "Principal components analysis to summarize microarray experiments: application to sporulation time series." *Pac Symp Biocomput*: 455-466.

Reik, W. and J. Walter (2001). "Genomic imprinting: parental influence on the genome." *Nat Rev Genet* 2(1): 21-32.

Ren, Y. G., K. W. Wagner, et al. (2004). "Differential regulation of the TRAIL death receptors DR4 and DR5 by the signal recognition particle." *Mol Biol Cell* 15(11): 5064-5074.

Riordan, J. R., J. M. Rommens, et al. (1989). "Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA." *Science* 245(4922): 1066-1073.

Rogers, M. F. and A. Ben-Hur (2009). "The use of gene ontology evidence codes in preventing classifier assessment bias." *Bioinformatics* 25(9): 1173-1177.

Romashko, J., 3rd, S. Horowitz, et al. (2003). "MAPK pathways mediate hyperoxia-induced oncotic cell death in lung epithelial cells." *Free Radic Biol Med* 35(8): 978-993.

Rommens, J. M., M. C. Iannuzzi, et al. (1989). "Identification of the cystic fibrosis gene: chromosome walking and jumping." *Science* 245(4922): 1059-1065.

Rosenberg, M. F., R. Callaghan, et al. (2005). "Three-dimensional structure of P-glycoprotein: the transmembrane regions adopt an asymmetric configuration in the nucleotide-bound state." *J Biol Chem* 280(4): 2857-2862.

Rosenwald, A., A. A. Alizadeh, et al. (2001). "Relation of gene expression phenotype to immunoglobulin mutation genotype in B cell chronic lymphocytic leukemia." *J Exp Med* 194(11): 1639-1647.

Roum, J. H., R. Buhl, et al. (1993). "Systemic deficiency of glutathione in cystic fibrosis." *J Appl Physiol* 75(6): 2419-2424.

Roy, P. J., J. M. Stuart, et al. (2002). "Chromosomal clustering of muscle-expressed genes in *Caenorhabditis elegans*." *Nature* 418(6901): 975-979.

Rozmahel, R., M. Wilschanski, et al. (1996). "Modulation of disease severity in cystic fibrosis transmembrane conductance regulator deficient mice by a secondary genetic factor." *Nat Genet* 12(3): 280-287.

S

Sadikot, R. T., J. W. Christman, et al. (2004). "Molecular targets for modulating lung inflammation and injury." *Curr Drug Targets* 5(6): 581-588.

Sagel, S. D., J. F. Chmiel, et al. (2007). "Sputum biomarkers of inflammation in cystic fibrosis lung disease." *Proc Am Thorac Soc* 4(4): 406-417.

Saitoh, M., H. Nishitoh, et al. (1998). "Mammalian thioredoxin is a direct inhibitor of apoptosis signal-regulating kinase (ASK) 1." *EMBO J* 17(9): 2596-2606.

Sana, T. R., M. J. Janatpour, et al. (2005). "Microarray analysis of primary endothelial cells challenged with different inflammatory and immune cytokines." *Cytokine* 29(6): 256-269.

Sanchez-Pernaute, O., C. Ospelt, et al. (2008). "Epigenetic clues to rheumatoid arthritis." *J Autoimmun* 30(1-2): 12-20.

Savill, J., N. Hogg, et al. (1992). "Thrombospondin cooperates with CD36 and the vitronectin receptor in macrophage recognition of neutrophils undergoing apoptosis." *J Clin Invest* 90(4): 1513-1522.

Schafer, G., T. Cramer, et al. (2003). "Oxidative stress regulates vascular endothelial growth factor-A gene transcription through Sp1-and Sp3-dependent activation of two proximal GC-rich promoter elements." *Journal of Biological Chemistry* 278(10): 8190-8198.

Schagat, T. L., J. A. Wofford, et al. (2001). "Surfactant protein A enhances alveolar macrophage phagocytosis of apoptotic neutrophils." *J Immunol* 166(4): 2727-2733.

Schekman, R. and L. Orci (1996). "Coat proteins and vesicle budding." *Science* 271(5255): 1526-1533.

Schena, M., D. Shalon, et al. (1995). "Quantitative monitoring of gene expression patterns with a complementary DNA microarray." *Science* 270(5235): 467-470.

Schmitt-Grohe, S., C. Naujoks, et al. (2005). "Interleukin-8 in whole blood and clinical status in cystic fibrosis." *Cytokine* 29(1): 18-23.

Schreiber, R., R. Greger, et al. (1997). "Cystic fibrosis transmembrane conductance regulator activates water conductance in *Xenopus* oocytes." *Pflugers Arch* 434(6): 841-847.

Schroeder, T. H., M. M. Lee, et al. (2002). "CFTR is a pattern recognition molecule that extracts *Pseudomonas aeruginosa* LPS from the outer membrane into epithelial cells and activates NF-kappa B translocation." *Proc Natl Acad Sci U S A* 99(10): 6907-6912.

Schuler, G. D. (1997). "Pieces of the puzzle: expressed sequence tags and the catalog of human genes." *J Mol Med* 75(10): 694-698.

Schumacher, C., I. Clark-Lewis, et al. (1992). "High- and low-affinity binding of GRO alpha and neutrophil-activating peptide 2 to interleukin 8 receptors on human neutrophils." *Proc Natl Acad Sci U S A* 89(21): 10542-10546.

Schwiebert, E. M., M. M. Morales, et al. (1998). "Chloride channel and chloride conductance regulator domains of CFTR, the cystic fibrosis transmembrane conductance regulator." *Proc Natl Acad Sci U S A* 95(5): 2674-2679.

Schwiebert, E. M., T. Flotte, et al. (1994). "Both CFTR and outwardly rectifying chloride channels contribute to cAMP-stimulated whole cell chloride currents." *Am J Physiol* 266(5 Pt 1): C1464-1477.

Shi, L., L. H. Reid, et al. (2006). "The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements." *Nat Biotechnol* 24(9): 1151-1161.

Shibahara, K., M. Asano, et al. (1995). "Isolation of a novel mouse gene MA-3 that is induced upon programmed cell death." *Gene* 166(2): 297-301.

Short, D. B., K. W. Trotter, et al. (1998). "An apical PDZ protein anchors the cystic fibrosis transmembrane conductance regulator to the cytoskeleton." *J Biol Chem* 273(31): 19797-19801.

Šidák, Z. (1967). "Rectangular confidence regions for the means of multivariate normal distributions." *Journal of the American Statistical Association* 62 626-633.

Smirnova, M. G., S. L. Kiselev, et al. (2002). "Role of the pro-inflammatory cytokines tumor necrosis factor-alpha, interleukin-1 beta, interleukin-6 and interleukin-8 in the pathogenesis of the otitis media with effusion." *Eur Cytokine Netw* 13(2): 161-172.

Smith, A. N., M. L. Barth, et al. (1996). "A regulatory element in intron 1 of the cystic fibrosis transmembrane conductance regulator gene." *J Biol Chem* 271(17): 9947-9954.

Smith, D. F., E. Galkina, et al. (2005). "GRO family chemokines are specialized for monocyte arrest from flow." *Am J Physiol Heart Circ Physiol* 289(5): H1976-1984.

Smyth, G. K. (2004). "Linear models and empirical bayes methods for assessing differential expression in microarray experiments." *Stat Appl Genet Mol Biol* 3: Article3.

Smyth, G. K. (2005). "Limma: linear models for microarray data. ." *Bioinformatics and Computational Biology Solutions using R and Bioconductor* ,Springer, New-york: 397-420.

Speed, I. L. o. a. T. (2002). "Replicated microarray data." *Statistica Sinica* 12: 31-46.

Stalteri, M. A. and A. P. Harrison (2007). "Interpretation of multiple probe sets mapping to the same gene in Affymetrix GeneChips." *BMC Bioinformatics* 8: 13.

Steinhoff, C. and M. Vingron (2006). "Normalization and quantification of differential expression in gene expression microarrays." *Brief Bioinform* 7(2): 166-177.

Strieter, R. M., M. D. Burdick, et al. (2005). "CXC chemokines in angiogenesis." *Cytokine Growth Factor Rev* 16(6): 593-609.

Student (1908). "The probable error of mean." *Biometrika* 6.

Sudarsanam, P., V. R. Iyer, et al. (2000). "Whole-genome expression analysis of *snf/swi* mutants of *Saccharomyces cerevisiae*." *Proc Natl Acad Sci U S A* 97(7): 3364-3369.

Sumbayev, V. V. (2008). "PI3 kinase and direct S-nitrosation are involved in down-regulation of apoptosis signal-regulating kinase 1 during LPS-induced Toll-like receptor 4 signalling." *Immunol Lett* 115(2): 126-130.

Suske, G. (1999). "The Sp-family of transcription factors." *Gene* 238(2): 291-300.

Suzuki, M., T. Morita, et al. (2006). "Diversity of Cl(-) channels." *Cell Mol Life Sci* 63(1): 12-24.

T

Tabary, O., H. Corvol, et al. (2006). "Adherence of airway neutrophils and inflammatory response are increased in CF airway epithelial cell-neutrophil interactions." *Am J Physiol Lung Cell Mol Physiol* 290(3): L588-596.

Tabary, O., S. Escotte, et al. (1999). "Genistein inhibits constitutive and inducible NFkappaB activation and decreases IL-8 production by human cystic fibrosis bronchial gland cells." *Am J Pathol* 155(2): 473-481.

Tabcharani, J. A., J. M. Rommens, et al. (1993). "Multi-ion pore behaviour in the CFTR chloride channel." *Nature* 366(6450): 79-82.

Thomas, G. R., E. A. Costelloe, et al. (2000). "G551D cystic fibrosis mice exhibit abnormal regulation of inflammation in lungs and macrophages." *J Immunol* 164(7): 3870-3877.

Thomas, P. J., Y. H. Ko, et al. (1992). "Altered protein folding may be the molecular basis of most cases of cystic fibrosis." *FEBS Lett* 312(1): 7-9.

Thornberry, N. A. and Y. Lazebnik (1998). "Caspases: enemies within." *Science* 281(5381): 1312-1316.

Tirouvanziam, R., S. de Bentzmann, et al. (2000). "Inflammation and infection in naive human cystic fibrosis airway grafts." *Am J Respir Cell Mol Biol* 23(2): 121-127.

Townsend, D. M., K. D. Tew, et al. (2003). "The importance of glutathione in human disease." *Biomed Pharmacother* 57(3-4): 145-155.

Turunen, M. P., E. Aavik, et al. (2009). "Epigenetics and atherosclerosis." *Biochim Biophys Acta* 1790(9): 886-891.

Tusher, V. G., R. Tibshirani, et al. (2001). "Significance analysis of microarrays applied to the ionizing radiation response." *Proc Natl Acad Sci U S A* 98(9): 5116-5121.

U

Ulleras, E., M. Karlberg, et al. (2008). "NFAT but not NF-kappaB is critical for transcriptional induction of the prosurvival gene A1 after IgE receptor activation in mast cells." *Blood* 111(6): 3081-3089.

Ussing, H. H. and K. Zerahn (1951). "Active transport of sodium as the source of electric current in the short-circuited isolated frog skin." *Acta Physiol Scand* 23(2-3): 110-127.

V

Vallabhapurapu, S. and M. Karin (2009). "Regulation and function of NF-kappaB transcription factors in the immune system." *Annu Rev Immunol* 27: 693-733.

van den Wollenberg, A. L. (1977). "Redundancy analysis, an alternative for canonical correlation analysis." *Psychometrika* 42: 207-219.

Vanden Berghe, W., M. N. Ndlovu, et al. (2006). "Keeping up NF-kappaB appearances: epigenetic control of immunity or inflammation-triggered epigenetics." *Biochem Pharmacol* 72(9): 1114-1131.

Vandivier, R. W., T. R. Richens, et al. (2009). "Dysfunctional cystic fibrosis transmembrane conductance regulator inhibits phagocytosis of apoptotic cells with proinflammatory consequences." *Am J Physiol Lung Cell Mol Physiol* 297(4): L677-686.

Vandivier, R. W., V. A. Fadok, et al. (2002). "Elastase-mediated phosphatidylserine receptor cleavage impairs apoptotic cell clearance in cystic fibrosis and bronchiectasis." *J Clin Invest* 109(5): 661-670.

Verhaeghe, C., C. Remouchamps, et al. (2007). "Role of IKK and ERK pathways in intrinsic inflammation of cystic fibrosis airways." *Biochem Pharmacol* 73(12): 1982-1994.

Verhaeghe, C., S. P. Tabruyn, et al. (2007). "Intrinsic pro-angiogenic status of cystic fibrosis airway epithelial cells." *Biochem Biophys Res Commun* 356(3): 745-749.

Viatour, P., S. Legrand-Poels, et al. (2003). "Cytoplasmic I κ B α increases NF- κ B-independent transcription through binding to histone deacetylase (HDAC) 1 and HDAC3." *J Biol Chem* 278(47): 46541-46548.

W

Walsh, D. E., B. J. Harvey, et al. (2000). "CFTR regulation of intracellular calcium in normal and cystic fibrosis human airway epithelia." *J Membr Biol* 177(3): 209-219.

Wang, C. Y., D. C. Guttridge, et al. (1999). "NF- κ B induces expression of the Bcl-2 homologue A1/Bfl-1 to preferentially suppress chemotherapy-induced apoptosis." *Mol Cell Biol* 19(9): 5923-5929.

Wang, X., J. Matteson, et al. (2004). "COPII-dependent export of cystic fibrosis transmembrane conductance regulator from the ER uses a di-acidic exit code." *J Cell Biol* 167(1): 65-74.

Ward, C. L., S. Omura, et al. (1995). "Degradation of CFTR by the ubiquitin-proteasome pathway." *Cell* 83(1): 121-127.

- Watt, A. P., J. Courtney, et al. (2005). "Neutrophil cell death, activation and bacterial infection in cystic fibrosis." *Thorax* 60(8): 659-664.
- Weeraratna, A. T. and D. D. Taub (2007). "Microarray data analysis: an overview of design, methodology, and analysis." *Methods Mol Biol* 377: 1-16.
- Welsh, M. J. and A. E. Smith (1993). "Molecular mechanisms of CFTR chloride channel dysfunction in cystic fibrosis." *Cell* 73(7): 1251-1254.
- West, S. D. and N. J. Brunskill (2002). "Complications associated with influenza infection." *Postgrad Med J* 78(916): 100, 107-108.
- Westfall P.H., Y. S. S. (1993). "Resampling-based Multiple Testing." Wiley New York.
- Westfall, P. H., Young, S. S (1993). "Resampling-based multiple testing: Examples and methods for P-Value adjustment." John Wiley & Sons.
- Wey, J. S., M. J. Gray, et al. (2005). "Overexpression of neuropilin-1 promotes constitutive MAPK signalling and chemoresistance in pancreatic cancer cells." *Br J Cancer* 93(2): 233-241.
- Wilson, C. L. and C. J. Miller (2005). "Simpleaffy: a BioConductor package for Affymetrix Quality Control and data analysis." *Bioinformatics* 21(18): 3683-3685.
- Wolpe, S. D., B. Sherry, et al. (1989). "Identification and characterization of macrophage inflammatory protein 2." *Proc Natl Acad Sci U S A* 86(2): 612-616.
- Wright, J. M., C. A. Merlo, et al. (2006). "Respiratory epithelial gene expression in patients with mild and severe cystic fibrosis lung disease." *Am J Respir Cell Mol Biol* 35(3): 327-336.

Wu, M., L. G. Xu, et al. (2003). "SINK is a p65-interacting negative regulator of NF-kappaB-dependent transcription." *J Biol Chem* 278(29): 27072-27079.

Wuyts, A., N. Van Osselaer, et al. (1997). "Characterization of synthetic human granulocyte chemotactic protein 2: usage of chemokine receptors CXCR1 and CXCR2 and in vivo inflammatory properties." *Biochemistry* 36(9): 2716-2723.

X

Xu, Y., C. Liu, et al. (2006). "Functional genomic responses to cystic fibrosis transmembrane conductance regulator (CFTR) and CFTR(delta508) in the lung." *J Biol Chem* 281(16): 11279-11291.

Xu, Y., D. Mayhugh, et al. (2003). "Design and synthesis of a potent and selective triazolone-based peroxisome proliferator-activated receptor alpha agonist." *J Med Chem* 46(24): 5121-5124.

Y

Yeung, K. Y. and W. L. Ruzzo (2001). "Principal component analysis for clustering gene expression data." *Bioinformatics* 17(9): 763-774.

Z

Zabner, J., P. Karp, et al. (2003). "Development of cystic fibrosis and noncystic fibrosis airway cell lines." *Am J Physiol Lung Cell Mol Physiol* 284(5): L844-854.

Zabner, J., T. E. Scheetz, et al. (2005). "CFTR DeltaF508 mutation has minimal effect on the gene expression profile of differentiated human airway epithelia." *Am J Physiol Lung Cell Mol Physiol* 289(4): L545-553.

Zhang, X., P. Shan, et al. (2003). "Reactive oxygen species and extracellular signal-regulated kinase 1/2 mitogen-activated protein kinase mediate hyperoxia-induced cell death in lung epithelium." *Am J Respir Cell Mol Biol* 28(3): 305-315.

Zhang, Z., S. S. Leonard, et al. (2003). "Role of reactive oxygen species and MAPKs in vanadate-induced G(2)/M phase arrest." *Free Radic Biol Med* 34(10): 1333-1342.

Zheng, M., L. O. Barrera, et al. (2007). "ChIP-chip: data, model, and analysis." *Biometrics* 63(3): 787-796.

Zheng, S., B. P. De, et al. (2003). "Impaired innate host defense causes susceptibility to respiratory virus infections in cystic fibrosis." *Immunity* 18(5): 619-630.

Zhong, H., M. J. May, et al. (2002). "The phosphorylation status of nuclear NF-kappa B determines its association with CBP/p300 or HDAC-1." *Mol Cell* 9(3): 625-636.

Annexe

ANNEXE1: composition des suppléments et antibiotiques du milieu de culture

ANNEXE 2: librairies de R utilisées dans nos analyses

ANNEXE 3 : code 1 : *Preprocessing* des données d'expression

ANNEXE 4 : code 2 : contrôle qualité des fichiers

ANNEXE 5 : code 3 : l'analyse différentielle CuFi-1 par rapport au NuLi-1

ANNEXE 6 : code 4 : Filtration des données d'expression en fonction des fonctions de l'interquartile range des sondes

ANNEXE 7 : code 5 : Analyse de variance des données d'expression filtrées

ANNEXE8 : code 6 : Analyse de canonique de redondance des données d'expression filtrées

ANNEXE 9 : code 7 : Classification des données d'expression selon leurs profils d'expression

ANNEXE 10: Caractéristiques des 128 gènes inflammatoires sur-exprimés dans les cellules CuFi-1 versus NuLi-1

ANNEXE 11: Description des sondes sens et antisens utilisées dans la quantification des transcrits inflammatoires d'intérêts.

ANNEXE 1 : composition des suppléments et antibiotiques du milieu de culture

Pour le milieu de culture DMEM+F-12, sont ajoutés les antibiotiques suivant pour une bouteille de 500 ml :

- 2,5 ml de pen-strep Gibco #15140-122 (final 50 µg/ml)
- 500 µl de gentamycine 50mg/ml Gibco # 10131-035 (final 50 µg/ml)
- 500 µl de fluconazole 2 mg/ml de la pharmacie(final 2 µg/ml)
- 2,5 ml d'amphotéricine B 250 µg/ml (final 1.25 ug/ml)

Pour milieu de culture BEGM (Bronchial Epithelial Cell Growth Media), sont ajoutés les antibiotiques suivants pour une bouteille de 500 ml:

- 2,5 ml de pen-strep Gibco #15140-122 (final 50 µg/ml)
- 500 µl de gentamicin 50 mg/ml Gibco # 10131-035 (final 50 µg/ml)
- 500 µl de fluconazole 2 mg/ml (final : 2 µg/ml)

Les suppléments suivants sont ajoutés pour une bouteille de 500 ml:

- 0,5 ml d'hydrocortisone 0,5 mg/ml
- 2,0 ml de bovine pituitary extract (BPE)
- 0,5 ml d'epidermal growth factor (EGF) 0,50 UG/ml
- 0,5 ml de transferrine 10 mg/ml
- 0,5 ml de bovine insulin 5 mg/ml
- 0,5 ml de triiodothyronine 6,51 UG/ml
- 0,5 ml d'épinéphrine 0,5 mg/ml
- 0,5 ml d'acide rétinolique 0,1 UG/ml

ANNEXE 2 : librairies de R utilisées dans nos analyses

fonction	librairie	Version
Lecture les fichiers CEL	affy	affy_1.22.1
Analyse ontologique	GO.db, GOstats	GO.db_2.2.11 GOstats_2.10.0
Contrôle qualité	simpleaffy	simpleaffy_2.20.0
Gestion des annotations génomique de la puce utilisée	HGU133.2.plus.db	hgu133plus2.db_2.2.11
Analyse différentielle des données d'expression	limma	limma_2.18.3
Gestion des annotations	biomaRt	biomaRt_2.0.0
Clustering des données d'expression	Mfuzz	Mfuzz_2.2.0
Analyse par redondance	vegan	vegan_1.15-4

Remarque: le fonctionnement de ces librairies est dépendent d'autres librairies :

gcrma_2.16.0	GO.db_2.2.11	preprocessCore_1.6.0
Biostrings_2.12.10	RSQLite_0.7-3	RBGL_1.20.0
IRanges_1.2.3	DBI_0.2-4	RCurl_0.98-1
genefilter_1.24.3	AnnotationDbi_1.6.1	splines_2.9.2
e1071_1.5-20	affy_1.22.1	survival_2.35-4
class_7.2-48	Biobase_2.4.1	tools_2.9.2
GOstats_2.10.0	affyio_1.12.0	XML_2.5-1
graph_1.22.5	annotate_1.22.0	xtable_1.5-6
Category_2.10.1	GSEABase_1.6.1	

ANNEXE 3 : code 1 : *Preprocessing* des données d'expressionDescription:

Permet le « *preprocessing* des données d'expression ». A partir des fichiers CEL et du fichier descripteur de l'expérience, les données sont normalisées par la méthode RMA. Tous les objets générés par ce code sont sauvegardés dans l'objet "preprocessing.RData". Cet objet est utilisé dans les codes suivants.

Code:

```
#librairies
library(affy)
library(limma)
targets_CUFI_NULI_DMNQ = readTargets("target_CUFI_NULI_DMNQ.txt", sep="")
abatch_CUFI_NULI_DMNQ = ReadAffy(filenamees =
targets_CUFI_NULI_DMNQ$FileName)
#normalisation
eset = rma(abatch_CUFI_NULI_DMNQ)
#sauvegarde des objets
save(targets_CUFI_NULI_DMNQ, eset, file = "preprocessing.RData")
```

ANNEXE 4 : code 2 : contrôle qualité des fichiers

Description:

Permet de générer le contrôle qualité des fichiers CEL. Génère les fichiers image à partir du fichier CEL, présente la distribution des intensités par chaque des puces et la distribution des intensités par chaque des puces, génère des métriques d'Affymetrix® et les graphes de la dégradation des ARNm, inspecte la qualité des contrôles étalons.

Code:

```
#librairie
library(affy)
library(simpleaffy)
load("preprocessing.RData")
#métriques d'affymetrix
abatch.qc = qc(abatch_CUFI_NULI_DMNQ)
avbg(abatch.qc)
sfs(abatch.qc)
percent.present(abatch.qc)
metric_AFF=data.frame(avbg(abatch.qc),
sfs(abatch.qc),percent.present(abatch.qc))
par(mfcol=c(3,2), mar=c(1,2,2,2))
palette.grey <- c(rep(gray(0:10/10), times = seq(1, 41, by = 4)))
or (loop_count in 1:length(abatch_CUFI_NULI_DMNQ)){
  image(abatch_CUFI_NULI_DMNQ[loop_count,], transfo = function (x)
log2(x), col=palette.grey)
}
#####
#genere un graphe pour les contrôles internes
abatch.qc@bioBCalls
spikeInProbes(abatch.qc)
concentration <- log(c(1.5, 5, 25, 100))
x_values <- array(concentration, c(4, length(abatch_CUFI_NULI_DMNQ)))
x_values <- t(x_values)
y_values <- spikeInProbes(abatch.qc)
# plot the values in a scatterplot of y as function of x
```

```

plot(x_values, y_values, col=1:12, main="Spike-in performance", xlab="log
( concentration in pM)", ylab="log2(expression)")
legend(legend=sampleNames(abatch_CUFI_NULI_DMNQ), x=3.5, y=10, lty=1,
      col=1:12, cex=0.75)
for (loop_count in 1:length(abatch_CUFI_NULI_DMNQ))
  {y_values <- spikeInProbes(abatch.qc) [loop_count,]
  lm_spike <- lm(y_values~concentration)
  slope <- coef(lm_spike) [2]
  intercept <- coef(lm_spike) [1]
  abline(intercept, slope, col=loop_count)}
#####
#genere un graphe pour la degradation des RNAs.
RNAdeg = AffyRNAdeg(abatch_CUFI_NULI_DMNQ)
summaryAffyRNAdeg(RNAdeg)
par(mfcol=c(3,2), mar=c(1,2,2,2))
palette.grey <- c(rep(gray(0:10/10), times = seq (1, 41, by = 4)))
image(abatch_CUFI_NULI_DMNQ, transfo = function (x) log2(x),
col=palette.grey)

```

ANNEXE 5 : code 3 : l'analyse différentielle CuFi-1 par rapport au NuLi-1

Description:

Permet de produire l'analyse différentielle entre les conditions expérimentales CuFi-1 et NuLi-1. Isole l'information significative

Code:

```
#librairies
library(affy)
library(limma)
library(hgu133plus2.db)
#analyse differentielle de L'EXPERIENCE :
design = cbind (est_nuli = c(1,1,1,1,1,1),est_cufi= c(0,0,0,1,1,1))
fit = lmFit(eset, design)
fit2 = eBayes(fit)
table_coef2 = topTable( fit2 , coef=2 , adjust="fdr", sort.by = "B",
number = 54675)
hist(table_coef2$P.Value, nclass =100, col= "red", ylim = c(0,5500))
table_coef.sig0.05 = table_coef2[table_coef2$P.Value < 0.05,]
table_coef.sig0.05.UP = table_coef.sig0.05[table_coef.sig0.05$logFC>0,]
table_coef.sig0.05.DOWN = table_coef.sig0.05[table_coef.sig0.05$logFC<0,]
write.table(table_coef.sig0.05$ID, "table_coef.sig0.05.txt", quote= FALSE,
row.names= FALSE, col.names= FALSE)
table_coef.sig0.01 = table_coef2[table_coef2$P.Value < 0.01,]
table_coef.sig0.01.UP = table_coef.sig0.01[table_coef.sig0.01$logFC>0,]
table_coef.sig0.01.DOWN = table_coef.sig0.01[table_coef.sig0.01$logFC<0,]
#####
##histogramme des pvalue pv
pv =table_coef.cufi.nuli$P.Value
hist(pv, nclass =100, col= "red")
#histogramme des pvalue RESAMPLER
pv_RESAMPLER =table_coef.cufi.nuli_RESAMPLER$P.Value
hist(pv_RESAMPLER, nclass =100, col= "green")
#####
#
#Volcano-plot
A = table_coef.cufi.nuli$AveExpr
M = table_coef.cufi.nuli$logFC
```

```

plot( A, M)
pv = table_coef.cufi.nuli$P.Value
FC = table_coef.cufi.nuli$logFC
plot(FC, -log(pv), cex= 3, pch= ".")
abline(h= -log10(0.05))
#####
##PERMUTATION DE LA MATRICE
#a partir de l'objet ExpressionSet (54675 probesets x 12 conditions)
#pour chaque gene (de 1 a 54675) , on fait une permutation
#(on echange un element d'un groupe traite pour le mettre dans le groupe
controle)
# matrice -> expressionRandom ->limma -> B => histogramme des B random
#resampling de l'objet eset=
#in = objet eset
#out= un objet resampler expressionSet
resamplingEset = fonction (eset) {
                esetMatrice = as.matrix (eset)
                for (i in 1:dim(eset.temp)[1]){
                    esetMatrice[i,]<-
ample(esetMatrice[i,])
                }
                esetResampler= new ("ExpressionSet", exprs =
esetMatrice,annotation = "hgu133plus2")
                }
#####
probeset.chrom.sex=as.matrix(read.table("probesets-dependent-sexual-
chromosome.txt"))
table_coef2.sex=
able_coef2[is.element(table_coef2$ID,probeset.chrom.sex),]
hist(table_coef2.sex$P.Value, nclass =100, col= "blue")
sum(table_coef2.sex$P.Value <0.05)
compteur =c()
for (i in 1:10000){
    print(i)
    sample.id = sample(table_coef2$ID, length(probeset.chrom.sex),
replace= FALSE)
    table_coef2.sex.SAMPLE =
table_coef2[is.element(table_coef2$ID,sample.id),]
    compteur = append(compteur,
sum(table_coef2.sex.SAMPLE$P.Value<0.05))
}
sum(compteur>sum(table_coef2.sex$P.Value <0.05))/10000

```


ANNEXE 6 : code 4 : Filtration des données d'expression en fonction des fonctions de l'interquartile range des sondes

Description:

Détermine les IQR de chaque sonde. Teste la normalité des données d'expression.

Teste l'homogénéité de la variance.

Crée un objet « eset.filtered.IQR.RData » qui comprend les données d'expression filtrées selon plusieurs seuils d'IQR.

Code:

```
load("preprocessing.RData")
library(affy)
eset.exprs = exprs(as.matrix(eset))
IQRs <- apply(eset,1,IQR)
eset.IQRs = as.data.frame(cbind(eset, IQRs))
hist(IQRs, nclass = 200, col = "red")
hist(IQRs, nclass = 200, col = "red" ,xlim = c(0,1))
abline(v= 0.5)
abline(v= 0.6)
abline(v= 0.7)
abline(v= 0.8)
abline(v= 0.9)
abline(v= 1)
variances <- apply(eset.exprs,1,var)
eset.variances = as.data.frame(cbind(eset.exprs, variances))
hist(variances, nclass = 200, col = "blue" ,xlim = c(0,2))
correlation.variances.iqr = cor(as.vector(IQRs), as.vector(variances))
var.iqr = data.frame(IQRs,variances)
var.iqr.order = var.iqr[order(var.iqr$IQRs),]
# la recherche de outliers
valeur.outlier = IQRs*1.5
eset.outlier = as.data.frame(cbind(eset.exprs, valeur.outlier))
rs.boolean = (eset.outlier$valeur.outlier> eset.outlier[,1:12])
rs.boolean <- apply(rs.boolean,1,sum)
eset.outlier.interet = eset.exprs[rs.boolean != 0,]
iqr =apply(eset.outlier.interet,1,IQR)
vars =apply(eset.outlier.interet,1,var)
```

```

eset.outlier.interet = data.frame (eset.outlier.interet,iqrs, vars,
iqrs*1.5)
##teste la normalité de la distribution.
test.normal = fonction(vecteur){
  return(shapiro.test(vecteur)$p.value) }
rs.shapiro.test = apply(eset.exprs,1,test.normal)
hist(rs.shapiro.test, nclass = 100)
eset.shapiro.test = as.data.frame(cbind(eset.exprs, rs.shapiro.test))
sum(rs.shapiro.test<0.05)
conditions<-factor(rep(c("ctr","mut","ctr.DMNQ","mut.DMNQ"),c(3,3,3,3)))
test.variance.fligner = fonction(vecteur, condition){
  return(fligner.test(vecteur~conditions)$p.value) }
rs.fligner.test = apply(eset.exprs,1,test.variance.fligner)
hist(rs.fligner.test, nclass = 100)
sum(rs.fligner.test<0.05)
test.variance.bartlett = fonction(vecteur, condition){
  return(bartlett.test(vecteur~conditions)$p.value) }
rs.bartlett.test = apply(eset.exprs,1,test.variance.bartlett)
hist(rs.bartlett.test, nclass = 50)
sum(rs.bartlett.test<0.05)
commun =intersect(
rownames(eset.exprs[rs.shapiro.test<0.05,]) ,
rownames(eset.exprs[rs.bartlett.test<0.05,]) )
eset.filtered.IQR.0.5 = eset.exprs[eset.IQRs$IQRs > 0.5,]
dim(eset.filtered.IQR.0.5) [1]
eset.filtered.IQR.0.6 = eset.exprs[eset.IQRs$IQRs > 0.6,]
dim(eset.filtered.IQR.0.6) [1]
eset.filtered.IQR.0.7 = eset.exprs[eset.IQRs$IQRs > 0.7,]
dim(eset.filtered.IQR.0.7) [1]
eset.filtered.IQR.0.8 = eset.exprs[eset.IQRs$IQRs > 0.8,]
dim(eset.filtered.IQR.0.8) [1]
eset.filtered.IQR.0.9 = eset.exprs[eset.IQRs$IQRs > 0.9,]
dim(eset.filtered.IQR.0.9) [1]
eset.filtered.IQR.1 = eset.exprs[eset.IQRs$IQRs > 1,]
dim(eset.filtered.IQR.1) [1]
#####
save( eset.filtered.IQR.0.5, eset.filtered.IQR.0.6,
      eset.filtered.IQR.0.7, eset.filtered.IQR.0.8,eset.filtered.IQR.0.9,
eset.filtered.IQR.1, file = "eset.filtered.IQR.RData")

```

ANNEXE 7 : code 5 : Analyse de variance des données d'expression filtréesDescription:

A partir de l'objet « eset.filtered.IQR.RData », on teste l'effet de la mutation, de l'oxydant, de l'interaction.

L'ensemble des résultats est sauvegardé dans l'objet « anova.global.RData »

Code:

```

library(vegan)
load("eset.filtered.IQR.X.X.RData")
#####
facteur_mutation = gl(2,3, length = 12)
facteur_oxidant = gl(2,6)
helmert = model.matrix (~facteur_mutation*facteur_oxidant, contrast =
list(facteur_mutation = "contr.helmert", facteur_oxidant= "contr.helmert"))
#####
# on teste le facteur mutation : le facteur_mutation est-il significatif.
mutation.rda = rda(t(eset.filtered.IQR.0.4), helmert[,2],helmert[,3:4])
anova(mutation.rda, step =1000, perm.max = 1000)
#####
# on teste le facteur oxidant : le facteur_oxidant est-il significatif.
oxidant.rda = rda(t(eset.filtered.IQR.0.4), helmert[,3],helmert[,c(2,4)])
anova(oxidant.rda, step =1000, perm.max = 1000)
#####
# on teste l'interaction entre les 2 facteurs :
interaction.rda = rda(t(eset.filtered.IQR.0.4), helmert[,4],helmert[,2:3])
anova(interaction.rda, step =1000, perm.max = 1000)

save(list = ls(all=TRUE), file = "anova.global.RData")

```

ANNEXE 8 : code 6 : Analyse de canonique de redondance des données d'expression filtrées.

Description:

A partir de l'objet « eset.filtered.IQR.RData », on utilise la fonction `rdaTest` pour produire une analyse canonique de redondance.

A partir de chaque objet de type `rda`, on extrait l'objet « *cumulative fit of species* », correspondant pour chaque sonde à la décomposition de la variance en fonction de la matrice helmert.

De cet objet, on sélectionne les sondes présentant une variance > 20 % pour chacune des conditions expérimentales.

L'ensemble des résultats est sauvegardé dans l'objet «variance.partage.RData»

Code:

```
library(vegan)
load("eset.filtered.IQR.X.X.RData")
source("rdaTest/rdaTest.R")
source("rdaTest/graph.rdaTest.R")
#####
facteur_mutation = gl(2,3, length = 12)
facteur_oxidant = gl(2,6)
facteur_replication = as.factor(rep(1:3, 4))
helmert = model.matrix
(~facteur_mutation*facteur_oxidant*facteur_replication, contrast =
list(facteur_mutation = "contr.helmert", facteur_oxidant= "contr.helmert",
facteur_replication = "contr.helmert" ))
v.e = helmert[,2:3]
#####
result.0.5.rda= rdaTest(t(eset.filtered.IQR.0.5), v.e, testF=TRUE,
nperm=999)
graph.rdaTest(result.0.5.rda, graph.type="F", label.spc = FALSE, xlim =
c(-5,5), ylim= c(-5,5))
pcVarParProbeset = data.frame(VAR1= result.0.5.rda$FitSpe[,1], VAR2=
result.0.5.rda$FitSpe[,2]- result.0.5.rda$FitSpe[,1])
pcVarParProbeset = pcVarParProbeset[order(pcVarParProbeset[,1],
decreasing= TRUE),]
```

```
plot(pcVarParProbeset[,1],pcVarParProbeset[,2], xlab=" % de la variance
expliqué par la mutation",ylab=" % de la variance expliqué par l'oxidant")
abline(v=20)
abline(h=20)
variance.selection.0.5.20.20 = pcVarParProbeset[pcVarParProbeset$VAR1>20 &
pcVarParProbeset$VAR2>20 ,]
write.table(rownames(variance.selection.0.5.20.20),
"variance.selection.0.5.20.20.txt", quote = FALSE, col.names = FALSE,
row.names = FALSE)
#####
save(list = ls(all=TRUE), file = "variance.partage.RData")
```

ANNEXE 9 : code 7 : Classification des données d'expression selon leurs profils d'expression

Description:

A partir de l'objet «variance.partage.RData» on utilise l'algorithme Fuzzy c-means pour produire un *clustering* des données d'expression basé sur 2 et 4 clusters.

L'ensemble des résultats est sauvegardé dans l'objet «clustering.RData»

Code:

```
library(Mfuzz)
load("variance.partage.RData")
#####
e.s.0.5.20.20 = standardise(new("ExpressionSet", exprs =
eset.filtered.IQR.0.5[rownames(variance.selection.0.5.20.20),]))
e.s.0.5.20.20.avec2clusters <- mfuzz(e.s.0.5.20.20, c = 2, m = 1.25)

mfuzz.plot(e.s.0.5.20.20, cl = e.s.0.5.20.20.avec2clusters, mfrow=c(1,2) )

clusters.e.s.0.5.20.20.avec2clusters=
as.matrix(e.s.0.5.20.20.avec2clusters$cluster)
cluster.A.0.5.20.20=
clusters.e.s.0.5.20.20.avec2clusters[clusters.e.s.0.5.20.20.avec2clusters[
,1] == 1,]
cluster.B.0.5.20.20=
clusters.e.s.0.5.20.20.avec2clusters[clusters.e.s.0.5.20.20.avec2clusters[
,1] == 2,]
write.table(names(cluster.A.0.5.20.20), "cluster.A.0.5.20.20.txt", quote =
FALSE, col.names = FALSE, row.names = FALSE)
write.table(names(cluster.B.0.5.20.20), "cluster.B.0.5.20.20.txt", quote =
FALSE, col.names = FALSE, row.names = FALSE)
save(list = ls(all=TRUE), file = "clustering.RData")
cl.e.s.0.5.20.20 <- mfuzz(e.s.0.5.20.20, c = 4, m = 1.25)
mfuzz.plot(e.s.0.5.20.20, cl = cl.e.s.0.5.20.20, mfrow=c(2,2) )
clusters.e.s.0.5.20.20 = as.matrix(cl.e.s.0.5.20.20$cluster)
cluster.1.0.5.20.20=clusters.e.s.0.5.20.20[clusters.e.s.0.5.20.20[,1]==
1,]
```

```
cluster.2.0.5.20.20= clusters.e.s.0.5.20.20[clusters.e.s.0.5.20.20[,1]==
2,]
cluster.3.0.5.20.20=clusters.e.s.0.5.20.20[clusters.e.s.0.5.20.20[,1]==
3,]
cluster.4.0.5.20.20=clusters.e.s.0.5.20.20[clusters.e.s.0.5.20.20[,1]==
4,]
write.table(names(cluster.1.0.5.20.20), "cluster.1.0.5.20.20.txt", quote =
FALSE, col.names = FALSE, row.names = FALSE)
write.table(names(cluster.2.0.5.20.20), "cluster.2.0.5.20.20.txt", quote =
FALSE, col.names = FALSE, row.names = FALSE)
write.table(names(cluster.3.0.5.20.20), "cluster.3.0.5.20.20.txt", quote =
FALSE, col.names = FALSE, row.names = FALSE)
write.table(names(cluster.4.0.5.20.20), "cluster.4.0.5.20.20.txt", quote =
FALSE, col.names = FALSE, row.names = FALSE)
```

ANNEXE 10 : code 8 : Analyse ontologique des données d'expression sélectionnée

Dans la méthodologie 1

Description:

A partir de la sélection de sonde issue de la méthodologie 1, nous produisons une analyse ontologique. Nous utilisons une fonction `determine.GO ()`, créé par nos soins

Les annotations sont obtenues par la librairie `biomaRt` en interrogeant la base de données d'Ensembl (www.ensembl.org)

Rem : le même code adapté est utilisé dans la méthodologie 2

Code:

```
library(GOstats)
library(annotate)
library(hgu133plus2.db)
library(biomaRt)
###creation de l'univers pour faire le test hypergeometrique.
univ <- ls(hgu133plus2GO)
hasUnivGO <- sapply(mget(univ, hgu133plus2GO),
function(ids) if(!is.na(ids) && length(ids) > 1) TRUE else FALSE)
univ <- univ[hasUnivGO]
univ <- unique(getEG(univ, "hgu133plus2"))
rs.go.0.05 = determine.GO(table_coef.sig0.05$ID,0.05)
rs.go.0.05.UP = determine.GO(table_coef.sig0.05.UP$ID,0.05)
write.table(summary(rs.go.0.05.UP[[1]]), "rs.go.0.05.UP.txt", quote=
FALSE, row.names= FALSE, sep= ";")
rs.go.0.05.DOWN = determine.GO(table_coef.sig0.05.DOWN$ID,0.05)
write.table(summary(rs.go.0.05.DOWN[[1]]), "rs.go.0.05.DOWN.txt", quote=
FALSE, row.names= FALSE, sep= ";")
rs.go.0.01 = determine.GO(table_coef.sig0.01$ID,0.05)
rs.go.0.01.UP = determine.GO(table_coef.sig0.01.UP$ID,0.05)
listMarts()
ensembl = useMart("ensembl")
ensembl = useDataset("hsapiens_gene_ensembl", mart = ensembl)
filters = listFilters(ensembl)
attributes = listAttributes(ensembl)
```

```
gene.0.05 = getBM(attributes = c("affy_hg_u133_plus_2", "hgnc_symbol",
"description"), filters = "affy_hg_u133_plus_2", values =
table_coef.sig0.05$ID, mart = ensembl)
length(unique(gene.0.05$hgnc_symbol))
gene_0009615 = getBM(attributes = c("affy_hg_u133_plus_2", "hgnc_symbol",
"description"), filters = "entrezgene", values =
as.matrix(geneIdsByCategory(rs.go.0.05.UP[[1]])["GO:0009615"][[1]]), mart
= ensembl)
gene_0045087 = getBM(attributes = c("affy_hg_u133_plus_2", "hgnc_symbol",
"description"), filters="entrezgene", values =
as.matrix(geneIdsByCategory(rs.go.0.05.UP[[1]])["GO:0045087"][[1]]), mart
= ensembl)
gene_0042742 = getBM(attributes = c("affy_hg_u133_plus_2", "hgnc_symbol",
"description"), filters = "entrezgene", values =
as.matrix(geneIdsByCategory(rs.go.0.05.UP[[1]])["GO:0042742"][[1]]), mart
= ensembl)
gene_0002250 = getBM(attributes = c("affy_hg_u133_plus_2", "hgnc_symbol",
"description"), filters = "entrezgene", values =
as.matrix(geneIdsByCategory(rs.go.0.05.UP[[1]])["GO:0002250"][[1]]), mart
= ensembl)
gene_0060326 = getBM(attributes = c("affy_hg_u133_plus_2", "hgnc_symbol",
"description"), filters = "entrezgene", values =
as.matrix(geneIdsByCategory(rs.go.0.05.UP[[1]])["GO:0060326"][[1]]), mart
= ensembl)

gene_0007249 = getBM(attributes = c("affy_hg_u133_plus_2", "hgnc_symbol",
"description"), filters = "entrezgene", values =
as.matrix(geneIdsByCategory(rs.go.0.05.UP[[1]])["GO:0007249"][[1]]), mart
= ensembl)

gene_0006935 = getBM(attributes = c("affy_hg_u133_plus_2", "hgnc_symbol",
"description"), filters = "entrezgene", values =
as.matrix(geneIdsByCategory(rs.go.0.05.UP[[1]])["GO:0006935"][[1]]), mart
= ensembl)

gene_0002286 = getBM(attributes = c("affy_hg_u133_plus_2", "hgnc_symbol",
"description"), filters = "entrezgene", values =
as.matrix(geneIdsByCategory(rs.go.0.05.UP[[1]])["GO:0002286"][[1]]), mart
= ensembl)
```

```

gene_0051092 = getBM(attributes = c("affy_hg_u133_plus_2", "hgnc_symbol",
"description"), filters = "entrezgene", values =
as.matrix(geneIdsByCategory(rs.go.0.05.UP[[1]])["GO:0051092"][[1]]), mart
= ensembl)

gene_0009615.symbol = unique(gene_0009615$hgnc_symbol)
gene_0045087.symbol = unique(gene_0045087$hgnc_symbol)
gene_0042742.symbol = unique(gene_0042742$hgnc_symbol)
gene_0002250.symbol = unique(gene_0002250$hgnc_symbol)
gene_0060326.symbol = unique(gene_0060326$hgnc_symbol)
gene_0007249.symbol = unique(gene_0007249$hgnc_symbol)
gene_0006935.symbol = unique(gene_0006935$hgnc_symbol)
gene_0002286.symbol = unique(gene_0002286$hgnc_symbol)
gene_0051092.symbol = unique(gene_0051092$hgnc_symbol)
gene.symbol =
unlist(unique(c(gene_0009615.symbol, gene_0045087.symbol, gene_0042742.symbo
l, gene_0002250.symbol, gene_0060326, gene_0007249.symbol,
gene_0006935.symbol, gene_0002286.symbol, gene_0051092.symbol)))

gene_0009615.affy_hg_u133_plus_2 =
unique(gene_0009615$affy_hg_u133_plus_2)
gene_0045087.affy_hg_u133_plus_2 =
unique(gene_0045087$affy_hg_u133_plus_2)
gene_0042742.affy_hg_u133_plus_2 =
unique(gene_0042742$affy_hg_u133_plus_2)
gene_0002250.affy_hg_u133_plus_2 =
unique(gene_0002250$affy_hg_u133_plus_2)
gene_0060326.affy_hg_u133_plus_2 =
unique(gene_0060326$haffy_hg_u133_plus_2)
gene_0007249.affy_hg_u133_plus_2 =
unique(gene_0007249$affy_hg_u133_plus_2)
gene_0006935.affy_hg_u133_plus_2 =
unique(gene_0006935$affy_hg_u133_plus_2)
gene_0002286.affy_hg_u133_plus_2 =
unique(gene_0002286$affy_hg_u133_plus_2)
gene_0051092.affy_hg_u133_plus_2 =
unique(gene_0051092$affy_hg_u133_plus_2)
gene.affy_hg_u133_plus_2 = unlist(
unique(c(gene_0009615.affy_hg_u133_plus_2, gene_0045087.affy_hg_u133_plus_2
, gene_0042742.affy_hg_u133_plus_2, gene_0002250.affy_hg_u133_plus_2, gene_00
60326, gene_0007249.affy_hg_u133_plus_2,

```

```

gene_0006935.affy_hg_u133_plus_2,gene_0002286.affy_hg_u133_plus_2,
gene_0051092.affy_hg_u133_plus_2)))

#selection sur les ids significatif

id.probeset.GO =
gene.affy_hg_u133_plus_2[is.element(gene.affy_hg_u133_plus_2,table_coef.sig0.05$ID)]

is.element( id.probeset.GO, as.matrix(data.resnet$Values))

data.resnet= read.csv("id.protein.tsv", sep = "\t")
data.affy = read.csv2("HG-U133_Plus_2.na26.annot")
name.resnet = as.matrix(data.resnet$Name)

#####
#
#on regarde dans les id.probesets des gènes sélectionnées dans les
différentes fonctions biologiques si il y a des Fonctions GO surreprésenté

id.probeset.infl = as.matrix(read.table("id.probeset.inflam.inter2.txt"))

id.probeset.infl = id.probeset.infl[1:288,]

GO.inflammation = determine.GO(id,0.05)

write.table(summary(GO.inflammation[[1]]), "GO.inflammation.txt", quote=
FALSE, row.names= FALSE, sep= ";")

gene_0043123 = getBM(attributes = c("affy_hg_u133_plus_2", "hgnc_symbol",
"description"), filters = "entrezgene", values =
as.matrix(geneIdsByCategory(GO.inflammation[[1]])["GO:0043123"][[1]]),
mart = ensembl)

#####
#
gene.inflammatoire = read.csv2("gene.inflammatoire.txt", sep =";")
table.inflammatoire = read.csv2("table.inflammatoire.txt", sep ="\t")

plot(table_coef.sig0.05$logFC,table_coef.sig0.05$AveExpr,pch = 19,cex=
0.1, xlab ="", ylab="")

```

```
points(as.matrix(table.inflammatoire$logFC),as.matrix(table.inflammatoire$
AveExpr), col= "red", pch = 19,cex= 0.5)
```

```
plot(table_coef.sig0.05$logFC,table_coef.sig0.05$P.Value,pch = 19,cex=
0.1, xlab = "", ylab="", log= "y")
```

```
points(as.matrix(table.inflammatoire$logFC),as.matrix(table.inflammatoire$
P.Value), col= "red", pch = 19,cex= 0.5, , log= "y")
```

```
#####Fonction
```

```
determine.GO
```

```
determine.GO <- function(prbs, PV) {
```

```
#creation du set a tester
```

```
hasGO <- sapply(mget(prbs, hgu133plus2GO), function(ids) if(!is.na(ids) &&
length(ids) > 1) TRUE else FALSE)
```

```
prbs <- prbs[hasGO]
```

```
prbs <- getEG(prbs, "hgu133plus2")
```

```
prbs <- prbs[!duplicated(prbs)]# remove duplicates, but keep named vector
```

```
#creation d'une instance GOHyperGParams----- BP
```

```
p <- new("GOHyperGParams", geneIds=prbs, universeGeneIds=univ, ontology=
"BP", annotation="hgu133plus2", conditional=TRUE)
```

```
pvalueCutoff(p) <- PV
```

```
rs_bp <- hyperGTest(p)
```

```
#creation d'une instance GOHyperGParams----- MF
```

```
p <- new("GOHyperGParams", geneIds=prbs, universeGeneIds=univ, ontology=
"MF", annotation="hgu133plus2", conditional=TRUE)
```

```
pvalueCutoff(p) <- PV
```

```
rs_mf <- hyperGTest(p)
```

```
#creation d'une instance GOHyperGParams----- CC
```

```
p <- new("GOHyperGParams", geneIds=prbs, universeGeneIds=univ, ontology=
"CC", annotation="hgu133plus2", conditional=TRUE)
```

```
pvalueCutoff(p) <- PV
```

```
rs_cc <- hyperGTest(p)
```

```
return(list(rs_bp, rs_mf, rs_cc))  
}
```

ANNEXE 10 : Caractéristiques des 128 gènes inflammatoires sur-exprimés dans les cellules CuFi-1 versus NuLi-1

	immune response	inflammatory response	chemotaxis	cytokine production	response to virus	regulation of immune response	positive regulation of immune system process	innate immune response	interferon systeme	NF-kb systeme	LOCUS	GENE SYMBOL	DESCRIPTION	RATIO	PROB.
			x								9q31.1	ABCA1	ATP-binding cassette, sub-family A , member 1	2,12	2,9E-05
	x				x	x					20q12-q13.11	ADA	adenosine deaminase	1,41	6,2E-03
x											1q22	AIM2	absent in melanoma 2	2,18	8,8E-03
	x										9q12-q21.2	ANXA1	Annexin A1	1,58	8,7E-05
	x										19q13.2	APOE	apolipoprotein E	1,56	9,0E-03
			x	x							19q13.1-q13.2	BCL3	B-cell CLL/lymphoma 3	1,63	2,7E-04
x	x										10q23.2-q23.33	BLNK	B-cell linker	2,11	1,8E-05
x											19p13.2	BST2	bone marrow stromal cell antigen 2	2,32	2,3E-03
						x	x	x	x		12p13	C1R	complement component 1, r subcomponent	2,57	2,5E-03
						x	x	x			12p13	C1S	complement component 1, s subcomponent	4,73	1,1E-03
x	x	x								x	2q33-q37	CCL20	chemokine (C-C motif) ligand 20	1,99	2,9E-03

immune response	inflammatory response	chemotaxis	cytokine production	response to virus	regulation of immune response	positive regulation of immune system process	innate immune response	interferon systeme	NF-kb systeme	LOCUS	GENE SYMBOL	DESCRIPTION	RATIO	PROB.
x	x	x		x						16q13	CCL22	chemokine (C-C motif) ligand 22	1,76	1,4E-04
x										9p24	CD274	CD274 molecule	1,68	1,2E-04
			x		x	x				3q21	CD86	CD86 molecule	2,94	8,1E-05
	x									3q24	CHST2	carbohydrate sulfotransferase 2	1,65	5,6E-03
					x	x	x			8p21-p12	CLU	clusterin	7,84	3,7E-05
x										11q14.1-q14.3	CTSC	cathepsin C	3,09	6,7E-06
x										11q13.1	CTSW	cathepsin W	1,31	5,8E-03
x	x	x							x	4q21	CXCL1	chemokine (C-X-C motif) ligand 1	5,74	5,4E-05
x	x	x							x x	4q21	CXCL10	chemokine (C-X-C motif) ligand 10	8,91	3,9E-04
x	x	x							x x	4q21.2	CXCL11	chemokine (C-X-C motif) ligand 11	10,70	1,8E-03
x	x	x							x	4q21	CXCL2	Chemokine (C-X-C motif) ligand 2	4,46	2,9E-04
x	x	x							x	4q21	CXCL3	chemokine (C-X-C motif) ligand 3	10,45	3,7E-05
x		x							x	4q12-q13	CXCL5	chemokine (C-X-C motif) ligand 5	9,85	5,7E-04

	immune response	inflammatory response	chemotaxis	cytokine production	response to virus	regulation of immune response	positive regulation of immune system process	innate immune response	interferon systeme	NF-kb systeme	LOCUS	GENE SYMBOL	DESCRIPTION	RATIO	PROB.
x	x	x								x	4q21	CXCL6	chemokine (C-X-C motif) ligand 6	5,94	2,8E-04
x	x	x		x							2q21	CXCR4	chemokine (C-X-C motif) receptor 4	1,65	7,6E-03
	x							x			16q24	CYBA	cytochrome b-245, alpha polypeptide	1,70	1,8E-04
	x										19p13.1	CYP4F11	cytochrome P450, family 4, subfamily F	2,32	3,1E-04
		x									1p31-p22	CYR61	cysteine-rich, angiogenic inducer, 61	2,69	1,2E-03
				x				x			9p12	DDX58	DEAD (Asp-Glu-Ala-Asp) box polypeptide 58	2,52	3,2E-03
		x						x			8p23.2-p23.1	DEFB1	defensin, beta 1	2,79	5,2E-06
x		x									8p23.1-p22	DEFB4	defensin, beta 4	2,06	1,2E-04
				x							15q15.3	DUOX2	dual oxidase 2	1,58	2,3E-03
				x							2p22-p21	EIF2AK2	eukaryotic translation initiation factor 2-alpha kinase 2	1,46	7,9E-03
x											19q13.3	FCGRT	Fc fragment of IgG, receptor, transporter, alpha	1,68	1,5E-04
					x						20p13	FKBP1A	FK506 binding protein 1A, 12kDa	2,13	5,1E-05
	x								x		14q24.3	FOS	FBJ murine osteosarcoma viral oncogene homolog	1,62	1,4E-04

immune response	inflammatory response	chemotaxis	cytokine production	response to virus	regulation of immune response	positive regulation of immune system process	innate immune response	interferon systeme	NF-kb systeme	LOCUS	GENE SYMBOL	DESCRIPTION	RATIO	PROB.
		x		x				x		11q13	FOSL1	FOS-like antigen 1	2,19	1,9E-04
x										5p13.1	FYB	FYN binding protein (FYB-120/130)	1,46	5,0E-03
					x	x				6q21	FYN	FYN oncogene related to SRC, FGR, YES	1,52	8,6E-03
x								x	x	1p22.2	GBP1	guanylate binding protein 1, interferon-inducible, 67kDa	2,93	3,7E-04
x										1p22.2	GBP3	guanylate binding protein 3	1,64	5,5E-04
x										1p22.2	GBP5	guanylate binding protein 5	1,58	1,6E-03
x										1p22.2	GBP6	guanylate binding protein family, member 6	2,84	2,5E-04
x										8q13-q21	GEM	GTP binding protein overexpressed in skeletal muscle	1,44	2,4E-03
			x							1p22.1	GLMN	Glomulin, FKBP associated protein	1,36	2,4E-03
	x									7p21.1	HDAC9	histone deacetylase 9	1,96	2,4E-03
x										6p21.3	HFE	hemochromatosis	1,33	2,6E-03
	x		x					x		14q21-q24	HIF1A	hypoxia inducible factor 1, alpha subunit	1,41	1,1E-03
					x	x		x	x	19p13.3-p13.2	ICAM1	intercellular adhesion molecule 1	1,97	9,7E-03

immune response	inflammatory response	chemotaxis	cytokine production	response to virus	regulation of immune response	positive regulation of immune system process	innate immune response	interferon systeme	NF-kb systeme	LOCUS	GENE SYMBOL	DESCRIPTION	RATIO	PROB.
				x				x		14q32	IFI27	interferon, alpha-inducible protein 27	3,60	1,6E-04
x								x		1p31.1	IFI44L	interferon-induced protein 44-like	2,83	2,8E-03
				x			x	x		2p24.3-q24.3	IFIH1	interferon induced with helicase C domain 1	2,56	8,9E-04
x								x		11p15.5	IFITM3	interferon induced transmembrane protein 3	1,89	1,0E-03
				x				x	x	21q22.1	IFNAR2	interferon (alpha, beta and omega) receptor 2	1,29	4,0E-03
				x				x		21q22.11	IFNGR2	interferon gamma receptor 2	1,40	5,8E-04
x										15q26.3	IGF1R	insulin-like growth factor 1 receptor	1,55	2,3E-03
	x									17q12-q21.1	IGFBP4	insulin-like growth factor binding protein 4	2,17	1,1E-03
							x			2q12	IL18R1	interleukin 18 receptor 1	1,88	8,3E-03
x			x					x		2q14	IL1A	interleukin 1, alpha	2,87	6,8E-05
		x	x		x	x		x	x	2q14	IL1B	interleukin 1, beta	2,46	9,9E-07

	immune response	inflammatory response	chemotaxis	cytokine production	response to virus	regulation of immune response	positive regulation of immune system process	innate immune response	interferon systeme	NF-kb systeme	LOCUS	GENE SYMBOL	DESCRIPTION	RATIO	PROB.
x											2q14	IL1F5	interleukin 1 family, member 5	1,54	1,0E-03
x											2q12-q21	IL1F9	interleukin 1 family, member 9	3,76	3,1E-08
x											2q12-q22	IL1R2	interleukin 1 receptor, type II	11,46	6,3E-05
	x							x			3q28	IL1RAP	interleukin 1 receptor accessory protein	1,50	3,6E-02
								x			2q12	IL1RL1	Interleukin 1 receptor-like 1	4,84	3,4E-03
x	x								x		2q14.2	IL1RN	interleukin 1 receptor antagonist	2,48	6,5E-05
		x	x	x	x	x		x	x		7p21	IL6	interleukin 6 (interferon, beta 2)	16,72	6,9E-06
x	x	x						x	x		4q13-q21	IL8	interleukin 8	4,95	5,5E-05
			x					x	x		5q31.1	IRF1	interferon regulatory factor 1	1,61	7,1E-04
				x				x			11p15.5	IRF7	interferon regulatory factor 7	2,37	5,0E-04
				x				x	x		15q26	ISG20	interferon stimulated exonuclease gene 20kDa	3,21	5,0E-04
x											19q13.4	KIR2DL1	Killer cell immunoglobulin-like receptor, two domains	1,27	8,3E-03
								x			2q22.2	KYNU	Kynureninase	3,05	2,2E-06

	immune response	inflammatory response	chemotaxis	cytokine production	response to virus	regulation of immune response	positive regulation of immune system process	innate immune response	interferon systeme	NF-kb systeme	LOCUS	GENE SYMBOL	DESCRIPTION	RATIO	PROB.
x											22q12.2	LIF	leukemia inhibitory factor	2,12	9,2E-06
x											19q13.4	LILRA3	leukocyte immunoglobulin-like receptor, subfamily A), member 3	1,32	1,4E-03
x											3q21-q23	LTF	lactotransferrin	3,70	4,7E-03
x	x										6p25.1	LY86	lymphocyte antigen 86	1,36	7,0E-03
				x	x	x					6p21.3	MICB	MHC class I polypeptide-related sequence B	1,78	8,6E-03
	x										22q11.23	MIF	macrophage migration inhibitory factor	1,38	4,6E-04
				x					x		21q22.3	MX1	myxovirus resistance 1, interferon-inducible protein p78	4,65	3,8E-05
				x					x		21q22.3	MX2	myxovirus resistance 2 (mouse)	5,15	1,8E-04
	x		x	x	x	x	x	x			3p22	MYD88	myeloid differentiation primary response gene (88)	1,52	5,0E-04
x										x	10q24	NFKB2	nuclear factor of kappa light polypeptide gene enhancer in B-cells	1,42	8,6E-03
x									x		12q24.1	OAS1	2',5'-oligoadenylate synthetase 1, 40/46kDa	3,63	3,2E-06

immune response	inflammatory response	chemotaxis	cytokine production	response to virus	regulation of immune response	positive regulation of immune system process	innate immune response	interferon systeme	NF-kb systeme	LOCUS	GENE SYMBOL	DESCRIPTION	RATIO	PROB.
x								x		12q24.2	OAS2	2'-5'-oligoadenylate synthetase 2, 69/71kDa	4,60	2,6E-04
x								x		12q24.2	OAS3	2'-5'-oligoadenylate synthetase 3, 100kDa	2,74	1,3E-04
x								x		12q24.2	OASL	2'-5'-oligoadenylate synthetase-like	4,64	8,4E-04
x	x									12p13.2-p12.3	OLR1	oxidized low density lipoprotein (lectin-like) receptor 1	2,66	1,3E-06
x										8q21.13	PAG1	phosphoprotein associated with glycosphingolipid microdomains 1	1,44	6,5E-03
			x							11q21	PANX1	pannexin 1	1,90	1,4E-03
			x		x	x				12q21	PAWR	PRKC, apoptosis, WT1, regulator	1,34	5,3E-03
x						x				9p24.2	PDCD1LG2	programmed cell death 1 ligand 2	1,29	4,9E-03
x		x	x							4q12-q21	PF4	platelet factor 4 (chemokine (C-X-C motif) ligand 4)	1,88	9,8E-05
		x							x	10q24	PLAU	plasminogen activator, urokinase	2,28	3,2E-04
		x						x		19q13	PLAUR	plasminogen activator, urokinase receptor	1,59	1,6E-03

immune response	inflammatory response	chemotaxis	cytokine production	response to virus	regulation of immune response	positive regulation of immune system process	innate immune response	interferon systeme	NF-kb systeme	LOCUS	GENE SYMBOL	DESCRIPTION	RATIO	PROB.
				x						3q23	PLSCR1	phospholipid scramblase 1	1,49	4,6E-03
							x			3p25	PPARG	peroxisome proliferator-activated receptor gamma	1,59	1,4E-03
x		x								4q12-q13	PPBP	pro-platelet basic protein (chemokine (C-X-C motif) ligand 7)	3,16	1,1E-05
x	x	x	x							1p35-p34.3	PTAFR	platelet-activating factor receptor	1,66	1,8E-03
x										5p13.1	PTGER4	prostaglandin E receptor 4 (subtype EP4)	2,74	1,6E-03
x	x									3q25	PTX3	pentraxin-related gene, rapidly induced by IL-1 beta	7,99	2,0E-03
					x	x	x			19q13.2	PVR	poliovirus receptor	1,28	7,2E-03
		x								7p15-p13	RALA	v-ral simian leukemia viral oncogene homolog A (ras related)	1,26	5,9E-03
					x					10q23-q24	RBP4	Retinol binding protein 4, plasma	2,90	4,4E-03
							x			14q11.2	RNASE7	ribonuclease, RNase A family, 7	1,68	4,0E-03
x				x						2p25.2	RSAD2	radical S-adenosyl methionine domain containing 2	7,20	3,4E-05

immune response	inflammatory response	chemotaxis	cytokine production	response to virus	regulation of immune response	positive regulation of immune system process	innate immune response	interferon systeme	NF-kb systeme	LOCUS	GENE SYMBOL	DESCRIPTION	RATIO	PROB.
	x									1q21	S100A12	S100 calcium binding protein A12	7,40	8,8E-07
							x	x		1q21	S100A7	S100 calcium binding protein A7	1,73	7,7E-04
	x									1q21	S100A8	S100 calcium binding protein A8	1,98	2,1E-06
	x	x								1q21	S100A9	S100 calcium binding protein A9	1,80	2,2E-05
x										20pter-q12	SAMHD1	SAM domain and HD domain 1	1,70	2,2E-02
					x	x	x			11q12-q13.1	SERPING1	serpin peptidase inhibitor, clade G (C1 inhibitor), member 1	5,90	1,2E-06
					x					5q32	SPINK5	serine peptidase inhibitor, Kazal type 5	5,22	6,5E-05
				x				x		2q32.2	STAT1	signal transducer and activator of transcription 1, 91kDa	2,50	1,4E-03

immune response	inflammatory response	chemotaxis	cytokine production	response to virus	regulation of immune response	positive regulation of immune system process	innate immune response	interferon systeme	NF-kb systeme	LOCUS	GENE SYMBOL	DESCRIPTION	RATIO	PROB.
				x						12q13.2	STAT2	signal transducer and activator of transcription 2, 113kDa	1,46	9,4E-04
x									x	6p21.3	TAP1	transporter 1, ATP-binding cassette, sub-family B (MDR/TAP)	1,63	5,7E-03
x									x	6p21.3	TAP2	transporter 2, ATP-binding cassette, sub-family B (MDR/TAP)	1,56	8,5E-05
						x				3p22	TGFBR2	transforming growth factor, beta receptor II (70/80kDa)	2,03	3,9E-04
	x									20q12	TGM2	transglutaminase 2	1,82	4,2E-03
x	x	x	x			x		x		15q15	THBS1	Thrombospondin 1	8,99	2,2E-05
	x		x				x			9q32-q33	TLR4	toll-like receptor 4	1,79	2,4E-02
	x								x	2q23.3	TNFAIP6	tumor necrosis factor, alpha-induced protein 6	7,49	5,6E-04
					x	x				13q32-34	TNFSF13B	tumor necrosis factor (ligand) superfamily, member 13b	4,33	5,9E-03

	immune response	inflammatory response	chemotaxis	cytokine production	response to virus	regulation of immune response	positive regulation of immune system process	innate immune response	interferon systeme	NF-kb systeme	LOCUS	GENE SYMBOL	DESCRIPTION	RATIO	PROB.
x	x										11p15.5	TOLLIP	Toll interacting protein	1,39	1,6E-03
x											6q25	ULBP2	UL16 binding protein 2	2,65	1,4E-06
				x							7q34	ZC3HAV1	zinc finger CCCH-type, antiviral 1	2,39	1,2E-03

ANEXE 11: Description des sondes sens et antisens utilisées dans la quantification des transcrits inflammatoires d'intérêts.

Gene Symbol	Identifiant	Sonde sens	Sonde antisens	#Universal ProbeLibrary
IL6	NM_000600.3	gatgagtacaaaagtctgatcca	ctgcagccactggtctgt	40
IL8	NM_000584.2	agacagcagagcacacaagc	atggtcctccggtggt	72
IL1 β	NM_000576.2	tacctgtcctgctgttgaa	tcttgggaattttgggatct	78
CXCL10	NM_001565.2	gaaagcagtagcaaggaaaggt	gacatatactccatgtagggaagtga	34
CXCL11	NM_005409.3	agtgtgaagggcatggcta	tctttgaacatggggaagc	76
STAT1	NM_007315.3	gagcttcactcccttagtttga	cacaacgggcagagaggt	64
MYD88	NM_002468.4	ctgctcgagctgctacca	ctttggcaatcctcctcaa	33
TNFAIP6	NM_007115.2	ggccatctcgcaacttaca	gcagcacagacatgaaatcc	34
S100A12	NM_005621.1	cacattcctgtgcattgagg	ggtgtcaaatgcccttc	31
C1S	NM_201442.2	gaccttgaggagagaagattt	ccgatctcctgcaacaaaa	38

Le design des sondes sens et antisens est réalisé avec le logiciel *ProbeFinder* (Version: 2.45), disponible sur le site www.roche-applied-science.com.