



**HAL**  
open science

# Caractérisation de l'expression génique des tumeurs des voies aérodigestives supérieures: perspectives diagnostiques et thérapeutiques

Frédéric Jean Laurent Lemaire

## ► To cite this version:

Frédéric Jean Laurent Lemaire. Caractérisation de l'expression génique des tumeurs des voies aérodigestives supérieures: perspectives diagnostiques et thérapeutiques. Life Sciences [q-bio]. INAPG (AgroParisTech), 2004. English. NNT : 2004INAP0004 . pastel-00000629

**HAL Id: pastel-00000629**

**<https://pastel.hal.science/pastel-00000629>**

Submitted on 21 Jul 2004

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



INSTITUT NATIONAL AGRONOMIQUE PARIS-GRIGNON

Ecole Doctorale ABIES

THÈSE

pour obtenir le grade de Docteur de l'Institut National Agronomique Paris-Grignon

Discipline : Biologie Moléculaire

Présentée et soutenue publiquement par :

Frédéric LEMAIRE

le 10/03/2004

Caractérisation de l'expression génique des tumeurs des voies  
aérodigestives supérieures: perspectives diagnostiques et  
thérapeutiques

Characterization of gene expression in upper aero digestive tract  
tumours: diagnostic and therapeutical perspectives

Directeur de thèse: Dr Bohdan WASYLYK

Jury

Dr Jean Benard

Dr Laurent Debussche

Dr Joseph Abecassis

Dr Laurent Bracco

Dr Bohdan Wasylyk

Pr Claude Gaillardin

Rapporteur

Rapporteur

# Remerciements

## **Equipe Wasylyk**

Au Dr Bohdan WASYLYK, pour m'avoir accueilli dans son laboratoire et avoir poursuivi ma formation scientifique et à Christine WASYLYK, pour son aide technique.

Au futur Dr Gilles BUCHWALTER, pour son calme olympien, sa patience et sa gentillesse, sans lesquelles il n'aurait pas pu me supporter sur la paillasse d'à côté, lors des périodes stressantes de ma dernière année de thèse.

Au futur Dr Anne CROMER, pour avoir supporté mes taquineries (en rougissant avec patience), mon activité parfois excessive et pour son attitude philosophe que je dois lui emprunter à l'avenir.

Au futur Dr Christian GROSS, pour avoir supporté les taquineries, d'un français de l'intérieur, sur son accent alsacien (léger, rassures toi) et avoir partagé mes jours étranges.

Thanks to Dr Gitali GANGULI, for her ever smiling face, kind mood, appeasing presence as well as scientific reliability. Namasté.

Thanks to Dr Hong ZHENG, surgeon of human soul, for sharing his views on life and Big Science.

A Patricia MARCHAL pour son aide technique et sa bonne humeur.

A Annaick CARLES pour son aide avec l'informatique.

## **Aux anciens du laboratoire:**

Au Dr Manuela ARGENTINI, pour sa rigueur scientifique, son sourire, son accent chantant, la danse africaine et la canalisation de mon énergie qu'elle a assuré pendant les dernières années de sa thèse sur la paillasse d'à côté.

Aux Docteurs Paola CRIQUI-FILIFE, Catherine DUCRET et Abdelkader AYADI, partis depuis quelques temps mais qui ont partagé une partie de mon odysée.

Thanks to Dr Julia YOUNG, for her senior help and advises on ways to approach science as well as her positive spin attitude in the DD project.

**Services communs de l'IGBMC :**

A Serge VICAIRE, pour les innombrables séquençages que nous lui avons fait subir.

A Bernard JOST, pour son aide dans la constitution de macro arrays.

A Jochen Barts et Claudine EBEL, pour leur aide dans l'utilisation du FACS.

**Centre Paul Strauss**

A Régine MILLON, à Danièle MULLER et à Christine MACABRE, pour leur assistance dans l'analyse des listes de gènes et leur implication dans les étapes de validation des clones.

Au Dr ABECASSIS pour ses explications des aspects cliniques et histopathologiques des tumeurs.

**ExonHit Therapeutics S.A.**

Au Dr Laurent BRACCO, pour m'avoir permis d'effectuer ma thèse dans le cadre d'une convention CIFRE, sur un sujet d'intérêt industriel, en accord avec mon projet professionnel.

**Remerciements également :**

A l'ADEREGEM et à la Ligue Régionale Contre le Cancer du Bas-Rhin pour leur support financier à diverses étapes de ma thèse.

Au Dr Jean BENARD pour ses nombreuses corrections apportées à ce manuscrit.

Au Dr Joseph ABECASSIS pour son aide dans la description des cancers VADS.

# Table des matières

---

<b>Remerciements</b>	<b>2</b>
<b>Table des matières</b>	<b>4</b>
<b>Table des figures</b>	<b>10</b>
<b>Abréviations</b>	<b>13</b>
<b>Introduction</b>	<b>17</b>

## **Partie I: Connaissances générales sur la carcinogenèse et les tumeurs des voies aérodigestives supérieures (VADS).**

<b>A. Connaissances générales sur le cancer</b>	<b>20</b>
1. Le Cancer : notions historiques	20
2. Causes du cancer : mutation et aneuploïdie	22
3. Causes de l'instabilité génétique et des mutations	23
4. Les gènes impliqués : Oncogènes et gènes suppresseurs de tumeurs	24
5. Un modèle de l'initiation et de la progression tumorale	29
6. Autres facteurs influençant l'expression génique dans les cancers	29
<b>B. Les cancers des voies aéro-digestives supérieures (VADS)</b>	<b>32</b>
1. Données épidémiologiques sur les cancers VADS	32
2. Facteurs étiologiques	33
3. Données anatomopathologiques sur les cancers VADS	34
4. Altérations chromosomiques et gènes cibles	37
5. Traitement des cancers VADS	40
6. Les différentes localisations des tumeurs VADS	41
<b>C. Méthodes d'analyse du transcriptome</b>	<b>43</b>
1. Les systèmes « fermés »	44
2. Analyses en masse en systèmes « ouverts »	46

## **Partie II: Objectifs et stratégie expérimentale**

<b>A. Problématique</b>	<b>52</b>
-------------------------	-----------

**B. Rappel du contexte de l'étude et stratégie expérimentale appliquée à la problématique** **53**

1. Rappel du contexte de l'étude 53
2. Stratégie expérimentale de l'étude en DD 55
3. Design expérimental global : complémentarité avec l'étude par puces à ADN Affymetrix 57
4. Adaptation aux contraintes expérimentales du post-Differential Display 58
5. Traitement des échantillons : Microdissection ou analyse de la tumeur dans sa globalité 58

**Partie III: Etude des tumeurs VADS par la technique du Differential Display ; première publication et résultats généraux** **60**

**A. La technique du Differential Display** **61**

1. Principe 61
2. Facteurs assurant le succès d'une étude en DD 63
3. Une comparaison systématique de quelques techniques ouvertes 65
4. Modifications du DD et intégration de plateformes de validation 68

**B. Résultats globaux du DD, face cachée** **73**

1. Premier criblage en Reverse Northern blot sur colonies bactériennes 73
2. Seconde série de criblages sur fragments de PCR 75
3. Etablissement d'arrays finaux pour la première publication 79
4. Validation de gènes individuels 80
5. Analyse de l'expression de la collection de gènes DD dans 15 lignées cellulaires VADS 84
6. Analyse de la valeur pronostique : macroarray haute densité 85
7. Test fonctionnel en masse : 87

**C. Base de données : collection DD dans toute son étendue** **88**

1. Identification des séquences 88
2. Confirmation du taux de couverture du génome 89
3. Localisation chromosomique 90
4. Proportion des différents profils d'expression après validation en RN 91
5. Prédications de la fonction à partir des catégories de geneOntology 92

**D. Première publication** **96**

<b>E. Première publication : commentaires additionnels</b>	<b>97</b>
1. Gènes impliqués dans les phénomènes immunologiques	97
2. Enzymes	100
3. Gènes localisés en 12p13 et 1q21-22	101
4. Comparaison avec une série d'autres études sur les cancers VADS	102
5. Fréquence élevée des pertes d'expression dans les tumeurs	106

#### **Partie IV: Comparaison des études par puces à ADN Affymetrix et par Differential**

<b><u>Display :</u></b>	<b>108</b>
-------------------------	------------

<b>A. 2ème publication</b>	<b>109</b>
<b>B. Commentaires: comparaison DD/Affymetrix</b>	<b>110</b>
1. Contribution personnelle et cadre de l'étude	110
2. Résultats résumés	110
3. Comparaison avec le DD	110

#### **Partie V: Caractérisation du candidat F1.4: SMAP31**

<b>A. Matériel et méthodes non présentés dans les articles</b>	<b>114</b>
1. Immunohistochimie	114
2. Immunocytochimie	115
3. FACS	115
4. Cultures cellulaires	116
5. Transfection	116
6. Tests de clonogénicité et clones stables	117
7. Western blotting	117
<b>B. Communication courte sur F1.4/HOP</b>	<b>119</b>
<b>C. Caractérisation du candidat F1.4 : SMAP31</b>	<b>120</b>
1. Analyse informatique de SMAP31	120
2. Caractérisation de l'expression dans les lignées cellulaires VADS	120
3. Clonage de la séquence codante	121
4. Etude de la localisation subcellulaire de la protéine transfectée par immunocytochimie avec l'anticorps Anti-Flag	121
5. Obtention d'anticorps spécifiques	123

6. Caractérisation de l'expression par Immunohistochimie	125
7. Etude de la localisation subcellulaire endogène par immunocytochimie avec l'anticorps spécifiques	127
<b>D. Etude fonctionnelle préliminaire:</b>	<b>130</b>
1. Etude par FACS	130
2. Clones stables et test de clonogénicité	131
3. Etude de l'induction par le sérum	133
4. Etude de l'implication dans la différenciation des kératinocytes	134
<b>E. Matériel biologique</b>	<b>136</b>

**Partie VI: Caractérisation du candidat B14.4 : hypothetical protein FLJ10261** **137**

<b>A. Caractérisation du candidat B14.4 : Hypothetical protein FLJ10261</b>	<b>138</b>
1. Profil DD	138
2. Identification	138
3. Validation en Virtual Northern blot	139
4. Validation en Northern Blot	141
5. Expression dans les lignées cellulaires VADS	141
6. Caractérisation de l'expression par hybridation <i>In Situ</i>	142
7. Etude de la localisation subcellulaire de la protéine transfectée par immunocytochimie avec l'anticorps Anti-Flag	145
8. Obtention d'anticorps spécifiques	146
9. Etude de la localisation subcellulaire endogène par immunocytochimie avec l'anticorps spécifique	148
10. Caractérisation de l'expression par Immunohistochimie	153
<b>B. Etude fonctionnelle préliminaire :</b>	<b>157</b>
1. Test de clonogénicité	157
2. Analyse par FACS	158
3. Etude de l'induction par le sérum	158
<b>C. Avancée des connaissances durant l'étude</b>	<b>159</b>
<b>D. Matériel biologique</b>	<b>162</b>



**Partie VII: Caractérisation du candidat 99b78** **163**

---

<b>A. Caractérisation du candidat 99b78 :</b>	<b>164</b>
1. Profil en DD	164
2. Profil en Reverse Northern blot	164
3. Identification du gène	164
4. Validation en Virtual Northern blot	167
5. Validation en Northern blot	168
6. Validation en Northern Slot blots	169
7. Etude de l'expression dans les lignées cellulaires VADS	170
8. Hybridation <i>In Situ</i>	170
9. Immunocytochimie pour le clone 99b78-Flag	171
10. Obtention et caractérisation d'anticorps polyclonaux	172
11. Immunocytochimie pour le clone 99b78	174
12. Immunohistochimie	177
<b>B. Analyse fonctionnelle préliminaire :</b>	<b>181</b>
1. Analyse par FACS	181
2. Test de clonogénicité et clones stables	181
3. Etude de la régulation par le sérum	182
4. Etude de l'induction de kinase en aval de la voie de signalisation membranaire	182
<b>C. Matériel biologique</b>	<b>183</b>

**Partie VIII: Perspectives** **184**

---

<b>A. Perspectives sur les analyses de masse</b>	<b>185</b>
1. Perspectives sur l'étude globale en Differential Display	185
2. Perspectives sur l'étude globale par puces à ADN Affymetrix	188
3 Perspectives technologiques et applications diagnostiques de nos études	189
<b>B. Perspectives sur les applications thérapeutiques potentielles : gènes cibles favoris</b>	<b>192</b>
1. Perspectives sur le gène HOP, gène suppresseur de tumeurs potentiel	192
2. Perspectives sur les oncogènes potentiels	195
3 Perspectives sur le gène FLJ10261	197
4 Perspectives sur le gène 99b78	198

5. Perspectives générales sur les traitements futurs des cancers 199

**Références bibliographiques** 202

**Abstract** 217

**Résumé** 218

# Table des figures

---

Figure I.1: Mécanismes d'altération des oncogènes et des gènes suppresseurs de tumeurs	26
Figure I.2: Les marques du cancer	28
Figure I.3: Modèle de l'initiation et de la progression tumorale des carcinomes du colon	29
Figure I.4: Localisations des tumeurs VADS	32
Figure I.5 : Aspect anatomo-pathologique de l'épithélium au cours de la progression tumorale	35
Figure I.6: Cancérisation en champ	37
Figure I.7: Modèle préliminaire de l'initiation et de la progression tumorale des carcinomes VADS	40
Figure II.1: Stratégie expérimentale globale de l'étude des cancers VADS	57
Figure III.1: Principe du Differential Display	63
Figure III.2: Procédure de vérification du profil des bandes DD et isolement des clones positifs correspondants	69
Figure III.3: Principe de la technologie SMART	71
Figure III.4: Reverse Northern blot sur pousses bactériennes	74
Figure III.5: Etapes de validation en Reverse Northern blot	80
Figure III.6: Exemples de Virtual Northern/Northern Blots différentiels	83
Figure III.7: Analyse de la collection DD par Transat	89
Figure III.8: Relation entre séquences analysées en Affymetrix et en DD	89
Figure III.9: Localisation sur le génome et existence d'ARNm dans les bases de données publiques	90
Figure III.10: Prédiction du "Process" dans geneOntology	92
Figure III.11: Prédictions "Function" dans geneOntology	93
Figure III.12: Prédictions du "Component" par geneOntology	94
Figure III.13: Exemple d'interactions fonctionnelles (CD91 / PIGR)	100
Figure III.14: Altérations communément détectées dans les diverses études des tumeurs VADS	105
Figure V.1: Structure de la protéine SMAP31	120

Figure V.2: Immunocytochimie pour HOP-Flag; lignée HaCaT	122
Figure V.3: Immunocytochimie pour HOP-Flag; lignée Fadu	122
Figure V.4: Immunocytochimie pour HOP-Flag; lignée Det562	123
Figure V.5: Caractérisation et purification des sérums Anti-HOP	124
Figure V.6: Variants putatifs de HOP	125
Figure V.7: Immunohistochimie pour HOP; tumeur 24656T	126
Figure V.8: Immunohistochimie pour HOP; tumeur 26312T	126
Figure V.9: Immunocytochimie pour HOP; lignée HaCaT	128
Figure V.10: Immunocytochimie pour HOP; lignée Fadu	128
Figure V.11: Immunocytochimie pour HOP; lignée Det562	129
Figure V.12: Tests de clonogénicité pour HOP	132
Figure V.13: Induction de HOP par le sérum dans les cellules HaCaT	133
Figure V.14: Expression de HOP au cours de la différenciation induite par le CaCl <sub>2</sub> dans les cellules HaCaT	135
Figure V.15: Matériel biologique obtenu pour l'étude de HOP	136
Figure VI.1: Analyse bioinformatique initiale de FLJ10261	139
Figure VI.2: Virtual Northern blot pour B14.4	140
Figure VI.3: Northern Blot pour FLJ10261	141
Figure VI.4: Analyse de l'expression de FLJ10261 dans des lignées cellulaires par PCR semi-quantitative	142
Figure VI.5: Hybridation <i>In Situ</i> pour FLJ10261; tumeur 840T et 347T	143
Figure VI.6: Hybridation <i>In Situ</i> pour FLJ10261; tumeur 413T	144
Figure VI.7: Hybridation <i>In Situ</i> pour FLJ10261; tumeur 771T	144
Figure VI.8: Immunocytochimie pour FLJ10261-Flag; lignée HaCaT	145
Figure VI.9: Immunocytochimie pour FLJ10261-Flag; lignée Fadu	146
Figure VI.10: Caractérisation et purification de sérums Anti FLJ10261	147
Figure VI.11: Immunocytochimie pour FLJ10261; lignée HaCaT	148
Figure VI.12: Immunocytochimie pour FLJ10261; lignée Det562	151
Figure VI.13: Immunocytochimie pour FLJ10261; lignée Fadu	152
Figure VI.14: Immunocytochimie pour FLJ10261; lignée Fadu	152
Figure VI.15: Immunohistochimie pour FLJ10261; tumeur 24189T	153
Figure VI.16: Immunohistochimie pour FLJ10261; tumeur 31073	154
Figure VI.17: Immunohistochimie pour FLJ10261; tumeur 35574T	154
Figure VI.18: Immunohistochimie pour FLJ10261; tumeur 29335T	155

Figure VI.19: Position des épitopes et orientation dans la cellule	156
Figure VI.20: Test de clonogénicité pour FLJ10261	157
Figure VI.21: Structure du locus CCND1-EMS1 en 11q13	159
Figure VI.22: Gene FLJ10261 humain et transcrits	160
Figure VI.23: Structure putative des protéines de la famille FLJ10261	161
Figure VI.24: Matériel biologique obtenu pour l'étude de FLJ10261	162
Figure VIII.1: Isoformes possibles et structure de 99b78	167
Figure VII.2: Virtual Northern blot pour 99b78	168
Figure VII.3: Northern blot pour 99b78	168
Figure VII.4: Northern slot blot pour 99b78	169
Figure VII.5: Analyse de l'expression de FLJ10261 dans des lignées cellulaires par PCR semi-quantitative	170
Figure VII.6:Hybridation <i>In Situ</i> pour 99b78; tumeur 245T	171
Figure VII.7: Immunocytochimie pour 99b78-Flag; lignée HaCaT	172
Figure VII.8: Caractérisation et puification des sérums Anti-99b78	173
Figure VII.9: Position des épitopes et structure de 99b78	174
Figure VII.10: Immunocytochimie pour 99b78; lignée HaCaT	175
Figure VII.11: Immunocytochimie pour 99b78; lignée Fadu	175
Figure VII.12: Immunocytochimie pour 99b78; lignée Det562	176
Figure VII.13: Immunocytochimie pour 99b78; lignée Det562	176
Figure VII.14: Immunohistochimie pour 99b78; tumeur 31680T	178
Figure VII.15: Immunohistochimie pour 99b78; tumeur 27847T	178
Figure VII.16: Immunohistochimie pour 99b78; tumeur 25004T	179
Figure VII.17: Immunohistochimie pour 99b78; tumeur 25005T	179
Figure VII.18: Tests de clonogénicité pour 99b78	181
Figure VII.19: Induction par le sérum dans les cellules Det562	182
Figure VII.20: Matériel biologique obtenu pour l'étude de 99b78	183

## Abréviations

---

uPAR	Urokinase plasminogen activator receptor
5-FU	5-Fluorouracile
A2M	Alpha 2 macroglobuline
Abl	Abelson murine leukemia virus gene
ADN	Acide déoxyribonucléique
ADNc	ADN complémentaire
AFLP	Amplified fragment length polymorphism
APC	adenomatous polyposis coli gene
APC	Antigen presenting cell
APP	amyloid precursor protein
AP-PCR	Arbitrarily primed PCR
ARN	Acide ribonucléique
ARNm	ARN messenger
ATRS	adenosine/thymidine -rich sequence
BBS	BES-buffered solution
BRCA1	Breast cancer associated gene 1
BRCA2	Breast cancer associated gene 2
BSA	Bovine serum albumine
CAPN7	Calpain like protease 7
CCND1	Cycline D1
CE	Cornified envelope
CGH	Comparative genomic hybridization
CIN	Chromosomal instability
CLECSF2	C-type lectin superfamily member 2
CMV	Cytomegalovirus
CSRP2	Glycine cysteine rich protein 2
CUL2	Culline 2
DATAS	Differential analysis of transcripts with alternative splicing
DD	Differential Display
DMBA	7,12-dimethylbenz-[alpha]-anthracene
DMEM	Dulbecco's minimum essential medium
DMR	Differentially methylated region
Dnmt1	DNA (cytosine-5-)-methyltransferase 1
DRG1	Differentiation related gene 1
E	Early
EDC	Epidermal differentiation complex
EGF	Epidermal growth factor
EGFR	EGF receptor
EGR1	early growth response 1
EI	sonde tumorale Early, de type linéaire
ELISA	Enzyme-Linked Immunosorbent Assay
EMP1	Epithelial membrane protein 1
EMS-1	EMS1 sequence
ERAD	Endoplasmic reticulum associated degradation
ErbB	epidermal growth factor receptor (synonym)
ES	Electronic subtraction, soustraction électronique
EST	Expressed sequence tag
FACS	Fluorescence activated cell sorting
FADD	Fas-associating protein with death domain
FAP	Familial adenomatous polyposis

FB	Sérum final bleed (saignée finale)
FCS	Fetal calf serum
FGF	Fibroblast growth factor
FISH	Fluorescence in situ hybridization
FITC	Fluorescein (reactive isothiocyanate form)
G	Ganglion
GAPDH	glyceraldehyde-3-phosphate dehydrogenase
GARB	Goat anti rabbit biotinylated
GFP	Green fluorescent protein
GJB2	Gap junction protein beta 2/ connexin 26
GSE	Genetic suppressor elements
HAP	HindIII arbitrary primer
HAT	Histone acetyl transferase
HDAC	Histone deacetylase
HEK-a	Human epithelial keratinocyte
HOP	Homeodomain only protein
HPRT	Hypoxanthine phosphoribosyltransferase
HPV	Human papilloma virus
HRP	Horse radish peroxydase
HSCP150	ubiquitin conjugating enzyme E2
HT11	HindIII poly-d(11)T primer
ICAM	Intercellular adhesion molecule
IFN	Interferon
Ig	Immunoglobuline
IGBMC	Institut de Génétique et de Biologie Moléculaire et Cellulaire
IGF	Insulin-like growth factor
IHC	Immuno histo chemistry, immunohistochimie
IL	Interleukine
ISH	In Situ Hybridization, hybridation in situ
ITGA6	Intégrine alpha 6
JNK	Jun N-terminal kinase
K5	Keratin 5
KGFR	Keratinocyte growth factor receptor
KLK-L4	Kallikrein-like L4
LAGY	Lung cancer associated protein Y
LCM	Laser capture microdissection
LOH	Loss of heterozygosity, perte d'hétérozygotité
LPRP	Lacrimal proline rich
LTB4	Leukotriene B4 receptor
LTBP1	Latent TGF beta binding protein 1
M	Metastasis prone
MAP kinase	Mitogen activated protein kinase
MCP1	Macrophage chemotactic protein 1
MDM2	Mouse double minute 2
MEK	MAP/ERK kinase
MIN	Microsatellite instability
MMLV	Moloney murine leukemia virus
MMP	Matrix metalloproteinase
MTDN2	NADH dehydrogenase
Myc	myelocytomatosis viral oncogene homolog
N	Tissu normal
NB	Northern blot
NE	Normal early

NECC1	Not expressed in CC1 (choriocarcinoma cell line)
Nel	sonde normale Early, de type linéaire
Net	New ets transcription factor
NIN	Nucleotide instability
NM	Non metastasis prone
NS	Normal stable
NU	Normal Unstable
ORAOV1	Overexpressed in oral cancer 1
ORF	Open reading frame, phase ouverte de lecture
ORL	Otorhinolaryngé
PAFah	Plasminogen activator factor acetylhydrolase
PAGE	Polyacrylamide gel electrophoresis
PAI1	Plasminogen activator inhibitor 1
PBS	Phosphate buffered saline
PBST	PBS tween
PCR	Polymerase chain reaction
PDGF	Platelet derived growth factor
PI	Sérum pré immun
PIGR	Polymeric immunoglobulin receptor
PLA2G7	Phospholipase A2 group VII
PON2	Paraoxanase 2
PPFAI1	PTRF interacting coiled-coil protein
PS1	gamma secretase
PSA	Prostate specific antigen
PSMB8	Proteasome subunit beta type 8
PTCH2	Novel human patch-like gene
PTMS	PBS tween milk solution
Pur	Sérum purifié
RACE	Rapid Amplification of cDNA Ends
Ras	rat sarcoma viral oncogene
Rb	Retinoblatoma gene
RDA	Representational differences analysis
READS	Restriction enzyme analysis of differentially expressed sequences
RFLP-DD	Restriction fragment length polymorphism DD
RN	Reverse Northern blot
RNAi	RNA interference, ARN interférence
RP	Ribosomal protein
RTK	Receptor tyrosine kinase
RT-PCR	Reverse transcription PCR, PCR avec transcription inverse
S	Stable
SABRE	Selective amplification via-biotin and restriction-mediated enrichment
SAGE	Serial analysis of gene expression
SAM	Statistical analysis of microarray
SC	Secretory component
SELDI-TOF	Surface enhanced laser desorption/ionization time-of-flight mass spectrometry
SH	Subtractive hybridization, hybridation soustractive
SIgA	Secreted IgA
SKY	Spectral karyotyping
SMART-DD	Sondes complexes ( marquage focalisé sur les clones isolés en DD par utilisation des amorces HAP



SOD2	Manganese superoxyde dismutase
SP1	Serine protease Site 1
SP2	Metalloprotease Site protease 2
SPRR	Small proline rich protein
SPT	Second primary tumour, récurrence locale
src	rous sarcoma virus
SREBP	Sterol regulatory element binding protein
SRF	Serum response factor
SSH	Suppression subtractive hybridization, hybridation soustractive avec suppression
STAT1	signal transducer and activator of transcription 1
T	Tissu tumoral
TACE	Tumour necrosis factor alpha converting enzyme
TALEST	Tandem arrayed ligation of expressed sequence tags
TAM	Tumour associated macrophage
TAOS1	tumor amplified and overexpressed sequence 1
TBP	Tata Box Protein
TCF	Ternary complex factor
TGF	Transforming growth factor
TNF	Tumour necrosis factor
TOGA	Total gene expression analysis
TPA	12-O-tetradecanoylphorbol-13-acetate
TRA1	Tumour rejection antigen 1
TRANSAT	Transcriptome analysis tool
U	Unstable
UV	Ultraviolets
VADS	Voies aérodigestives supérieures
VEGF	Vascular endothelial growth factor
VHL	Von Hippel Lindau
VN	Virtual Northern blot
WT1	Wilms tumour gene 1

# **Introduction**

En première partie de cette thèse, je vais procéder à un rappel de grandes notions sur la carcinogenèse : des notions générales, des données sur la carcinogenèse des cancers des voies aréodigestives supérieures (VADS) et plus particulièrement des tumeurs hypopharyngées. Diverses méthodes d'analyse du transcriptome vont être succinctement décrites, avec leurs avantages et leurs inconvénients. L'objectif de ma thèse (2<sup>ème</sup> partie) est l'analyse simultanée de l'expression d'un grand nombre de gènes, afin de déterminer une signature des tumeurs VADS. Dans ce but, notre stratégie expérimentale a intégré une étude par la technique du Differential Display (DD) à large échelle, part essentielle de mon projet de thèse, ainsi que d'autres techniques de biologie moléculaire (puces à ADN, DATAS). En troisième partie de ma thèse, je présenterai la technique du Differential Display ainsi que les résultats généraux obtenus, tant au niveau de notre 1<sup>ère</sup> publication que de la collection globale des clones isolés en DD. Des commentaires additionnels sur la liste de gènes isolés dans notre publication seront présentés. En quatrième partie, j'intégrerai l'article traitant de l'analyse des mêmes tumeurs par puces à ADN Affymetrix, menée par Anne Cromer (2<sup>ème</sup> publication en tant que 6<sup>ème</sup> auteur), ainsi qu'une brève comparaison des résultats obtenus avec ces deux techniques. Enfin, les études préliminaires de 3 clones favoris feront l'objet de 3 parties spécifiques. Les perspectives de poursuite de l'analyse de ces clones seront évoquées (huitième partie), en particulier dans l'optique d'applications diagnostiques et thérapeutiques.

**Partie I:**  
**Introduction sur les tumeurs des voies  
aérodigestives supérieures (VADS) et les  
méthodes d'analyse du transcriptome.**

## A. Connaissances générales sur le cancer

---

### 1. Le Cancer : notions historiques

Dans cette introduction je vais récapituler certaines données de base en Oncologie, l'étude scientifique des tumeurs. En grec Onkos signifie masse et, c'est ainsi qu'ont été initialement définies les excroissances tumorales. Le terme néoplasie, désignant une croissance anormale détectable, signifie nouvelle croissance. Le cancer est une maladie connue depuis l'antiquité. On a trouvé des déformations attribuables à des carcinomes nasopharyngés sur des crânes égyptiens datant de plus de 3000 ans avant Jésus-Christ. Cependant les cancers oraux n'ont été cliniquement décrits qu'au 17<sup>ème</sup> siècle. Le cancer a longtemps été considéré comme une maladie rare, sans doute en lien avec une espérance de vie plus faible lors des siècles passés. Le cancer est, en effet, une maladie associée la plupart du temps au vieillissement, sauf dans le cas des tumeurs pédiatriques d'origine héréditaire.

Un organisme adulte, constitué d'environ  $10^{15}$  cellules, maintient son intégrité, en une masse cellulaire constante, grâce à la prolifération et la différenciation très étroitement régulées d'un contingent de cellules souches pluripotentes. Ce phénomène de renouvellement cellulaire, appelé homéostasie, dépend de l'équilibre entre la différenciation, la mort cellulaire programmée (apoptose) et la désquamation des cellules, d'une part, et, la prolifération cellulaire d'autre part (environ  $10^{12}$  divisions cellulaires/jour chez l'homme). Les cellules cancéreuses échappent aux contraintes de l'homéostasie, poursuivant leur prolifération jusqu'à former des tumeurs.

L'étude des tumeurs, après que les progrès techniques en chirurgie ont permis l'exérèse des tumeurs, s'est initialement focalisée sur l'analyse de l'activité enzymatique des cellules tumorales. Ces études ont conclu à des altérations du métabolisme et des enzymes de réparation de l'ADN, convergentes vers un état cancéreux.

Un pathologiste allemand, David Von Hanseman, a observé dès 1890 la présence de mitoses asymétriques dans les cellules cancéreuses, suggérant le rôle d'une ségrégation anormale des chromosomes dans le développement des cancers. L'origine potentielle des différences de profil enzymatique des cellules cancéreuses avait été anticipée dès 1914 par Boveri. Selon lui le cancer découle d'une mauvaise combinaison des chromosomes apparaissant dans une cellule somatique. Cette altération est transmise de la cellule initiale à toutes ses cellules

filles. Dès 1960 une première aberration chromosomique est associée au cancer: un petit chromosome surnuméraire, appelé chromosome de Philadelphie, est détecté chez les patients atteints de leucémie myéloïde chronique (Nowell & Hungerford, 1960). Le développement de techniques de coloration de chromosomes en bandes (Caspersson et al., 1970; Drets & Shaw, 1971) a permis de mettre en évidence de nombreuses anomalies chromosomiques récurrentes dans les tumeurs hématologiques. Rowley démontre en 1973 que le chromosome de Philadelphie résulte d'une translocation équilibrée entre les chromosomes 9 et 22 dans un peu moins de 90% des leucémies myéloïdes chroniques, et, en 1975, que la majorité des leucémies aiguës présentent un chromosome 8 additionnel (Rowley, 1973; Rowley, 1975).

La caractérisation d'altérations chromosomiques communes dans la majorité des populations cellulaires tumorales a conforté le modèle d'évolution clonale des populations cellulaires tumorales (Nowell, 1976). Dans ce modèle, l'acquisition de caractéristiques nouvelles par les cellules cancéreuses leur confère un avantage sélectif dans le cadre de la sélection naturelle proposée par Darwin dans sa théorie de l'évolution des espèces. Certaines mutations rares (principe de variation), confèrent à certaines cellules des capacités de survie ou de reproduction accrues (principe de sélection), qui seront transmises à leurs cellules filles (principe d'hérédité) constituant une sous population cellulaire. De mutation en mutation une sélection des cellules aptes à survivre s'opère, menant ultimement à la formation d'une tumeur maligne, voir ayant acquis des capacités métastatiques. Cette sélection correspond à une accumulation, en plusieurs étapes successives, d'altérations génétiques rares au sein d'une population {multi step carcinogenesis, (Ames et al., 1993a; Ames et al., 1993b; Land et al., 1983)}. Cette population doit acquérir, à chaque étape, une masse critique rendant l'apparition d'une nouvelle mutation statistiquement probable (fréquence de mutation de  $10^{-6}$  pour un allèle particulier,  $10^{-10}$  pour la seule sous population des cellules souches en prolifération, nécessitant une masse critique de  $10^6$  à  $10^8$  cellules). Ce processus nécessite plusieurs mois ou plusieurs années (Nowell, 1976). D'après l'analyse du taux d'incidence de chaque type de cancer au cours de la vie, une modélisation mathématique a indiqué le nombre d'évènements indépendants, correspondant aux altérations génétiques nécessaires à l'apparition de formes cliniques de cancers (Renan, 1993). Pour les formes familiales de cancers, apparaissant souvent chez l'enfant ou le jeune adulte, 3 ou 4 évènements indépendants sont nécessaires. Pour la majorité des cancers, d'apparition sporadique à un âge généralement plus avancé et spécifique du type de cancer considéré, 6 à 12 évènements paraissent requis. Dans le cas du cancer du colon, Vogelstein (Cho & Vogelstein, 1992a; Cho & Vogelstein, 1992b) a démontré la nécessité d'accumuler un minimum de 5 mutations.

Nowell (Nowell, 1976) propose que l'acquisition de l'instabilité génétique soit l'altération initiale dans le cancer, catalysant les mutations subséquentes. L'instabilité génétique constituerait un phénotype mutateur acquis par une mutation précoce déstabilisant l'ensemble du génome (Cairns, 1975; Loeb, 2001). Vogelstein a élucidé des mécanismes impliqués dans l'apparition de l'instabilité génétique dans les cancers colorectaux. Dans 15% des tumeurs colorectales une instabilité moléculaire est détectée, soit dite de type MIN (Microsatellite Instability), associée à un défaut de réparation des mésappariements de bases nucléotidiques, soit de type NIN (Nucleotide Instability), associée à un défaut de réparation par excision des nucléotides (Lengauer et al., 1997b). Le reste des tumeurs présente une forte aneuploïdie, associée à une instabilité chromosomique, CIN {Chromosomal Instability, (Lengauer et al., 1997b)}.

Un modèle a été établi distinguant deux types de gènes critiques dans la carcinogenèse (Kinzler & Vogelstein, 1997) : les « caretakers », maintenant l'intégrité du génome dans les cellules normales, dont l'inactivation précoce dans la carcinogenèse induit l'instabilité génétique, et, l'accumulation catalytique des mutations du second type de gènes; les « gatekeepers » contrôlant la croissance tumorale par régulation de la prolifération cellulaire et de l'apoptose.

## **2. Causes du cancer : mutation et aneuploïdie**

Toutes les cellules somatiques normales sont diploïdes et contiennent deux allèles d'un même gène. La généralisation des techniques de caryotypage a montré que la quasi-totalité des tumeurs solides présentent des caryotypes anormaux, avec isolement de gains ou de pertes au niveau des chromosomes. On observe également une représentation anormale des bandes cytogénétiques, repères de couleur variable au sein des chromosomes en lien avec des variations de densité et de structure de la chromatine. Cette représentation anormale de chromosomes dans le caryotype, nommée aneuploïdie, peut expliquer une partie des variations de l'expression génique observées dans les cancers. Généralement considéré comme résultant d'une mutation somatique, le cancer a été expliqué par deux théories concurrentes :

- l'hypothèse d'un rôle prépondérant de la mutation au niveau des gènes.
- l'hypothèse de l'aneuploïdie comme cause du cancer ; l'évènement initial de la carcinogenèse serait la génération de cellules aneuploïdes, qui provoquerait une instabilité génétique accrue de façon autocatalytique. L'aneuploïdie semble bien associée à la

carcinogénèse puisqu'elle est détectée dans 29 à 90% des cancers VADS (Kaplan et al., 1986; Sakr et al., 1989; Wennerberg et al., 1998). Certaines études ont associé un haut degré d'aneuploïdie à un mauvais pronostic (Sudbo et al., 2001). Cependant l'aneuploïdie ne constitue pas l'unique cause de la carcinogénèse, attendu que des modèles de carcinogénèse sans aneuploïdie ont été décrits (Hahn et al., 1999; Zimonjic et al., 2001).

Quelque soit la cause de l'initiation de la carcinogénèse au niveau de l'ADN, les variations d'expression des gènes qui y sont associées peuvent être, entre autres, détectées par profilage du transcriptome des tumeurs. Bien que l'analyse puisse être menée au niveau des protéines, nous n'aborderons pas l'analyse protéomique des cancers, en essor depuis l'amélioration des techniques de spectrométrie de masse pour l'identification des protéines à large échelle.

### **3. Causes de l'instabilité génétique et des mutations**

#### **a. Les mutations sont induites par des carcinogènes**

D'après **Cancer Medicine 5th ed.** Bast, Robert C.; Kufe, Donald W.; Pollock, Raphael E.; Weichselbaum, Ralph R.; Holland, James F.; Frei, Emil, editors. Hamilton (Canada): [BC Decker Inc](#); c2000.

Approximativement 70% des cancers dans les populations du monde occidental peuvent être attribuées au mode de vie et au régime alimentaire, l'exposition au tabac constituant à elle seule le facteur principal (30%) (Doll & Peto, 1981). Un autre facteur de risque est une consommation insuffisante de légumes et de fruits, qui ont un effet préventif. L'existence d'empreintes moléculaires (mutations caractéristiques de p53) associées à l'exposition à des substances carcinogènes dans l'environnement a été bien caractérisée (Greenblatt et al., 1994). Les dommages spontanés à l'ADN existent mais ne suffisent pas pour expliquer l'incidence des cancers, du fait de l'existence de mécanismes de réparation de l'ADN. Un phénotype particulier de formation accrue de cancers, appelé phénotype « mutateur » est lié à l'inactivation d'un des gènes assurant la fidélité de la réplication de l'ADN (Jackson & Loeb, 1998; Loeb, 1991). Le rôle carcinogène de certaines substances chimiques a été décelé dès 1775, et, depuis, la liste des carcinogènes s'est rapidement étendue (Doll & Peto, 1981). Le métabolisme cellulaire active ces substances carcinogènes en composés hautement réactifs via des enzymes impliquées dans la détoxification des substances xénobiotiques (Miller & Miller, 1975). Les dommages à l'ADN peuvent être causés par des carcinogènes physiques comprenant les radiations ionisantes (cassures) et les



ultraviolets (liaisons covalentes, « cross-linking »). Quatre vingt dix pourcent des cancers de la peau apparaissent dans des zones exposées aux UV, et, la maladie héréditaire xeroderma pigmentosum, caractérisée par une sensibilité aux UV et le développement de cancers de la peau, résulte de la mutation d'une enzyme d'élimination et de réparation des dommages à l'ADN (van Steeg & Kraemer, 1999).

## **b. Les causes de l'instabilité génétique**

Un défaut de ségrégation des chromosomes en mitose par altération d'un gène responsable du contrôle du fuseau mitotique ou de la séparation des chromatides filles est capable d'induire l'instabilité génétique (Cahill et al., 1998; Lengauer et al., 1998; Michel et al., 2001). Toutefois, aucune altération récurrente d'un gène de contrôle de la ségrégation des chromosomes n'a été démontrée dans les tumeurs primitives humaines (Jallepalli & Lengauer, 2001). Un dysfonctionnement des télomères, séquences répétitives localisées à l'extrémité des chromosomes, maintenant leur intégrité dans les cellules à cycle réplicatif illimité (Bryan & Cech, 1999), pourrait entraîner une instabilité chromosomique par fusion des extrémités des chromosomes. Une réactivation secondaire de l'expression de la télomérase, est observée dans les tumeurs et pourrait contribuer à leur échappement à la mort cellulaire et à un potentiel prolifératif illimité (Kim et al., 1994; Maser & DePinho, 2002). En combinaison avec les causes précédemment évoquées (contraintes sur l'ADN, extrémités à réparer), des mutations d'enzymes de réparation de l'ADN pourraient favoriser des fusions de chromosomes. L'inactivation de BRCA1 et BRCA2 (Breast Cancer Associated gene 1&2), gènes de prédisposition au cancer du sein impliqués dans la réparation de l'ADN, entraîne une instabilité génétique chromosomique (Bouffler et al., 1995; Patel et al., 1998; Xu et al., 1999).

## **4. Les gènes impliqués: Oncogènes et gènes suppresseurs de tumeurs**

Dans les années 1970, l'étude du caryotype des leucémies et des lymphomes a confirmé que des bandes chromosomiques étaient altérées de manière systématique (Manolov & Manolova, 1972; Rowley, 1973).

### **a. La découverte des oncogènes**

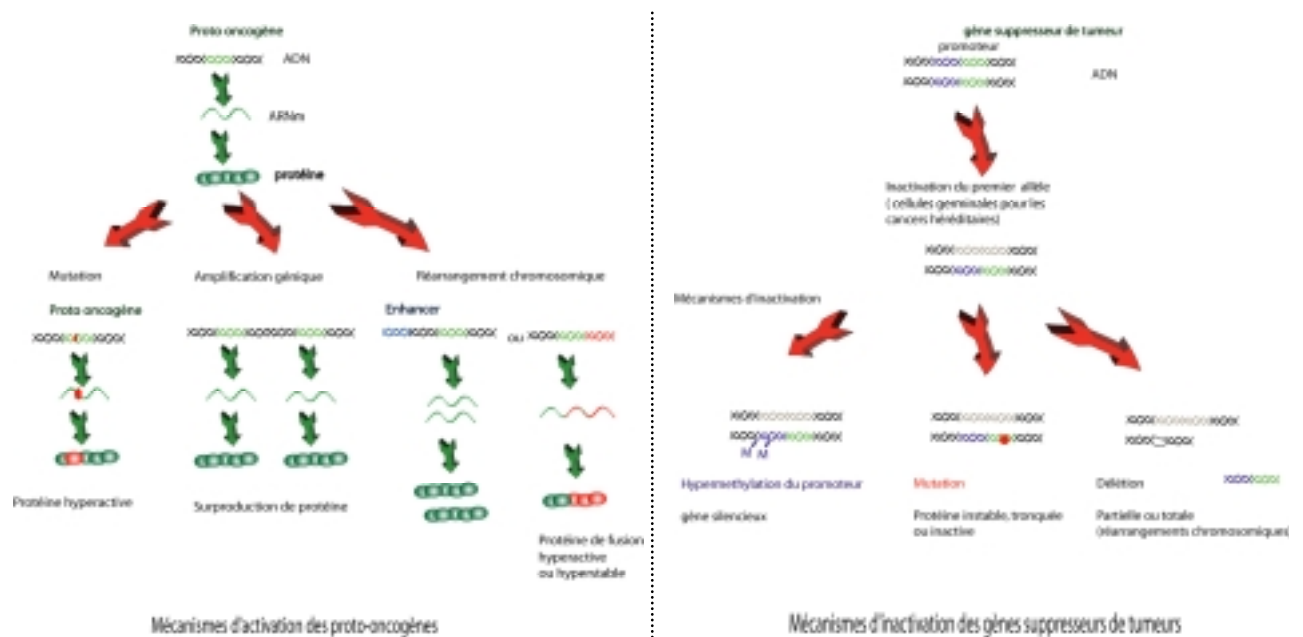
Chez l'animal, l'étude de rétrovirus transformants a permis d'isoler, au sein des génomes viraux de taille restreinte, des oncogènes, formes mutées hyperactives de gènes

présents dans les cellules hôtes. Chez le poulet, on a montré que le virus du sarcome de Rous contenait l'oncogène v-src, homologue du gène src (ou c-src) du poulet (Stehelin et al., 1976a; Stehelin et al., 1976b). L'homologue cellulaire normal de l'oncogène viral est appelé proto-oncogène. La découverte de proto-oncogènes et des oncogènes RAS, MYC ou ABL, dans le génome humain au début des années 1980 a été cruciale puisqu'elle a mis en évidence l'existence de gènes présents dans le génome des cellules normales capables de déclencher la transformation cellulaire. Une activation par des carcinogènes, comme les mutagènes chimiques ou les rayons X, est nécessaire puisque les virus du sarcome de Rous ne peuvent infecter les cellules humaines. Les proto-oncogènes sont souvent responsables de fonctions de base de la vie cellulaire, et, sont donc généralement hautement conservés d'une espèce à une autre. Le proto-oncogène est converti en oncogène par une mutation qui entraîne la production d'une oncoprotéine associée à un gain de fonction, et, qui est, de façon générale, dominante. La mutation d'un seul allèle est généralement suffisante. Il existe plusieurs mécanismes d'activation des oncogènes (FigI.1) : une mutation de séquence peut entraîner la production d'une protéine hyperactive ou ayant de nouvelles fonctions ; une amplification génique liée à des réarrangements aberrants des chromosomes peut causer une surproduction de la protéine normale ; un réarrangement chromosomique peut résulter en une production d'une protéine de fusion anormale, hyper stable hyperactive, ou, en une surproduction en changeant la transcription du gène (introduction d'un enhancer ou utilisation du promoteur plus actif de l'autre gène impliqué dans la fusion).

### **b. La découverte des gènes suppresseurs de tumeurs**

L'étude des formes familiales de cancers a permis d'identifier des gènes suppresseurs de tumeurs. Ainsi la délétion de gènes suppresseurs de tumeurs a été détectée dans le rétinoblastome {Rb, (Cavenee et al., 1983)}, dans le syndrome de Li Fraumeni {p53, (Malkin et al., 1990)}, dans les tumeurs de Wilms (WT1), dans les formes familiales des polypes du colon FAP Familial Adenomatous Polyposis (APC). Contrairement à un oncogène qui augmente les capacités de prolifération cellulaire des cellules cancéreuses, confère une résistance à l'apoptose ou stimule l'angiogenèse, un gène suppresseur de tumeur possède des propriétés inhibitrices de la prolifération cellulaire, d'induction de l'apoptose ou d'inhibition de l'angiogenèse. L'inactivation d'un gène suppresseur de tumeur correspondant à une perte de fonction, et, nécessite l'inactivation des deux allèles normaux présents dans le génome par

mutation ou par remaniement de l'ADN (perte chromosomique, remaniement du promoteur, condensation de la chromatine, voir Fig I.1).



### Mécanismes d'altération des oncogènes et des gènes suppresseurs de tumeurs

Figure I.1

A partir de l'étude du rétinoblastome, Knudson (Knudson, 1978; Knudson et al., 1975) a émis la théorie dite du « two hits »: dans les formes familiales de cancers les parents transmettent un allèle muté d'un gène suppresseur de tumeurs dans leurs cellules germinales. Un seul des deux allèles doit donc être inactivé pour progresser vers le cancer. La probabilité d'une mutation ( $10^{-6}$ , une cellule sur un million est mutée pour un allèle particulier), bien plus élevée que celle de deux mutations successives des deux allèles d'un même gène suppresseur de tumeurs dans une même cellule, peut expliquer la durée d'apparition beaucoup plus courte des cancers héréditaires. Depuis l'établissement de ce modèle général, on a également détecté l'existence de mutations dominantes négatives de gain de fonction comme dans le cas de p53.

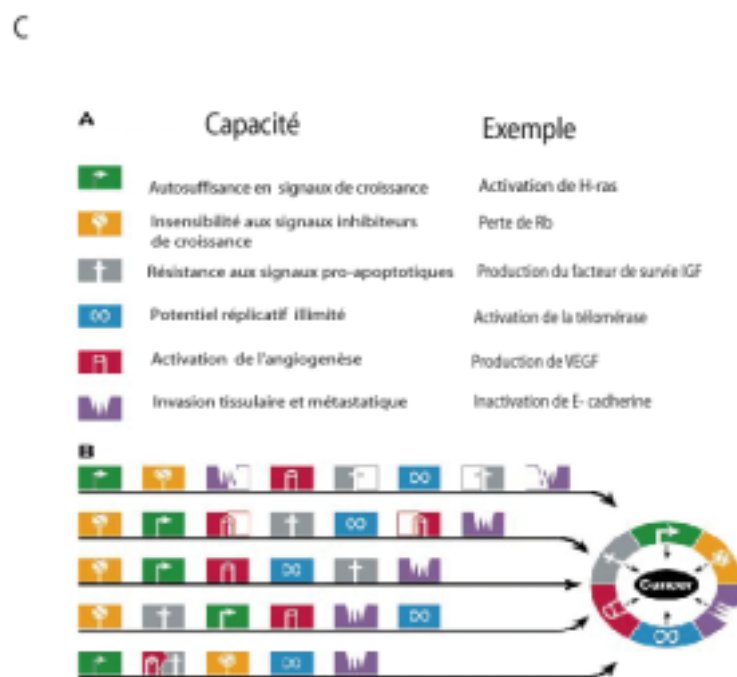
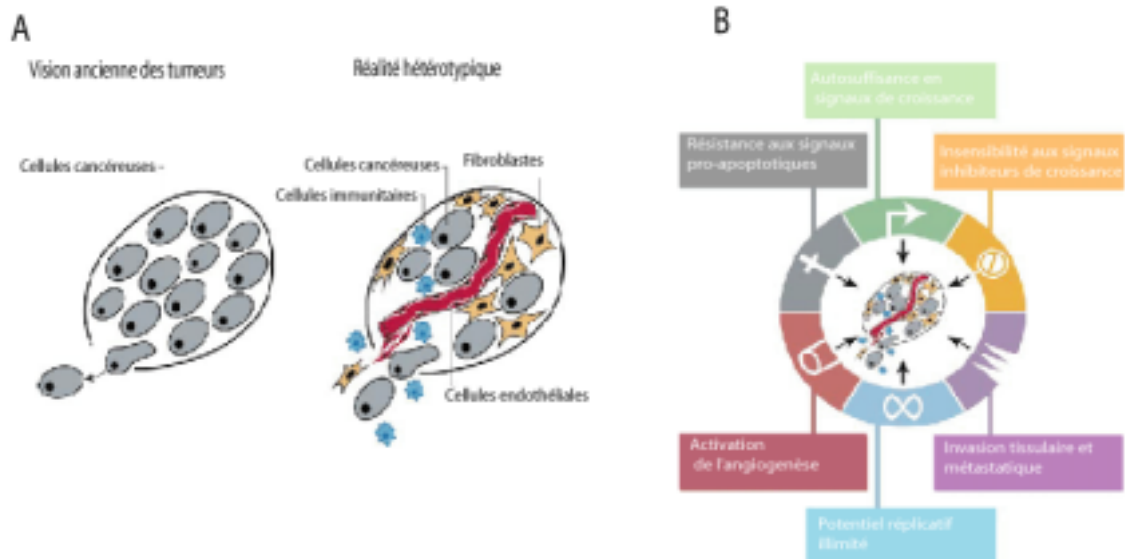
### c. Les voies métaboliques impliquées : les « marques » du cancer

Weinberg a résumé les types de fonctions altérées lors de l'initiation et de la progression tumorale en 6 propriétés importantes, les « marques » du cancer {Fig I.2,

(Hanahan & Weinberg, 2000)}. Les tumeurs malignes acquièrent presque obligatoirement au cours de leur accumulation d'altérations génétiques : la résistance vis à vis des signaux pro-apoptotiques ; l'autosuffisance en facteurs de croissance ; l'insensibilité aux signaux inhibiteurs de la croissance ; la capacité d'activation de l'angiogenèse ; un potentiel réplcatif illimité ; la capacité d'invasion tissulaire et de métastase.

On peut également ajouter à ces propriétés la capacité des cellules cancéreuses à échapper aux mécanismes de détection et d'élimination par les cellules immunitaires, et, l'évasion aux mécanismes de contrôle du cycle cellulaire. Les gènes effecteurs précis mutés, le nombre de mutations et leur séquence d'acquisition sont variables d'un type tumoral à un autre.

L'ancienne conception des biologistes de la tumeur comme masse composée uniquement de cellules cancéreuses s'est enrichie d'une vision hétéro typique des tumeurs proposée par les pathologistes {Revue (Elenbaas & Weinberg, 2001)}. La tumeur contient en effet des cellules stromales de type fibroblastique, des cellules endothéliales, des cellules de muscle lisse des vaisseaux sanguins irriguant la tumeur, ainsi que des cellules immunitaires, liées au processus inflammatoire, infiltrées dans la masse tumorale. Une des propriétés des cellules aux stades précoces de la carcinogenèse est leur capacité à recruter et à détourner de leur fonction normale les cellules du stroma. On parle de réponse stromale, dont le type le plus fréquent dans les carcinomes est la desmoplasie, caractérisée par l'acquisition d'une consistance dense et dure. Elle résulte d'une accumulation accrue de collagènes, de fibronectine, de protéoglycanes et de glycosylaminoglycanes dans la matrice extracellulaire, et, du recrutement de nouveaux vaisseaux sanguins. Au sein de la tumeur, les cellules du stroma, non cancéreuses, sont souvent stables génétiquement et sont donc une cible de choix pour la conception de nouvelles drogues. Un axe majeur de recherche est l'inhibition de la néoangiogenèse tumorale en ciblant les cellules endothéliales du stroma.



D'après Hanahan et Weinberg (2000)

Figure A: Ancienne vision des tumeurs et vision actuelle de la nature hétérogène des tumeurs

Figure B: Les six grandes capacités acquises lors de la progression tumorale

Figure C: A) Quelques exemples d'altérations génétiques et de capacités afférentes

B) Des enchainements séquentiels différents au sein de chaque tumeur mènent au statut malin de la tumeur

## Les marques du cancer

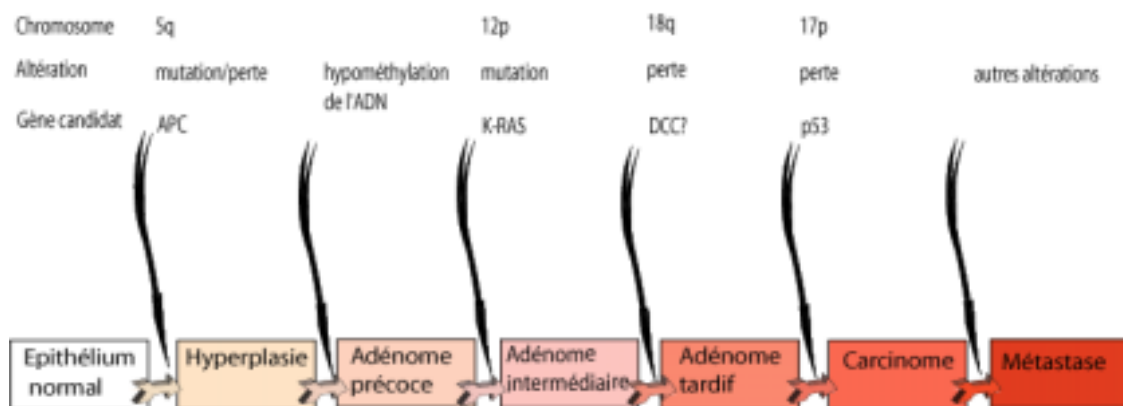
Figure I.2

Aux 6 types d'altérations acquises par les tumeurs malignes correspondent principalement 5 grands types de protéines potentiellement oncogènes : les facteurs de croissance ; les récepteurs aux facteurs de croissance et aux hormones, incluant les récepteurs

membranaires et les récepteurs intracellulaires ; les protéines assurant la transduction intracellulaire des signaux comme les tyrosine-kinases et les sérine/thréonine-kinases ; les facteurs de transcription nucléaires.

En supplément aux leucémies et aux lymphomes, les tumeurs solides doivent acquérir la capacité de proliférer indépendamment des inhibitions de contact et de survivre en suspension dans le flux sanguin lors de la métastase. On a longtemps pensé que les dernières aptitudes acquises par les cellules cancéreuses sont les capacités d'invasion et de métastase, cependant des études de profils d'expression ont démontré l'existence d'une signature « métastatique » dans des tumeurs de stade précoce (Ramaswamy et al., 2003).

## 5. Un modèle de l'initiation et de la progression tumorale



Modèle de l'initiation et de la progression tumorale des carcinomes du colon

Figure I.3

L'étude du cancer du colon a permis, grâce au nombre important de lésions présentes chez les patients souffrant de polypes, d'établir des statistiques et de mieux disséquer la séquence d'évènements menant à ce type de cancer {FigI.3, (Vogelstein et al., 1989)}.

## 6. Autres facteurs influençant l'expression génique dans les cancers

### a. Modulation épigénétique de l'expression génique

Outre les mutations et les remaniements génétiques, il existe d'autres causes influençant l'expression des gènes, sans nécessairement impliquer une altération de leur

séquence. En effet, le degré de condensation de l'ADN est variable suivant les régions du génome. On distingue l'hétérochromatine, correspondant à des régions de l'ADN condensées, dans lesquelles les gènes ne sont pas exprimés, et, l'euchromatine, correspondant aux régions de l'ADN de moindre condensation, où l'expression des gènes est possible.

La condensation de l'ADN est dépendante de son niveau de méthylation. Une méthylation différentielle de régions particulières (Differentially methylated region ; DMR), présentes dans les régions proximales de 40 à 50% des gènes humains (Bird, 1986), peut constituer une des causes de l'altération de l'expression génique dans les tumeurs. On parle de régulation épigénétique, car on constate une expression différentielle des gènes sans altération de séquence. En particulier, la méthylation de l'ADN a été associée à l'inactivation de gènes suppresseurs de tumeurs dans les cancers tandis qu'une hypométhylation généralisée sur l'ensemble du génome est associée à une instabilité chromosomique dans les cancers du colon (Lengauer et al., 1997a). Le terme « Méthylome » a été proposé pour englober l'ensemble des modifications de l'état de méthylation du génome (Feinberg, 2001). Les enzymes ADN méthyltransférases qui reconnaissent le brin hémi-méthylé de l'ADN après la réplication, maintiennent le niveau de méthylation de l'ADN. Le niveau de l'enzyme Dnmt1 (DNA cytosine-5-methyltransferase 1) est 3 fois supérieur à la normale dans les fibroblastes transformés par l'oncogène fos, et, l'inhibition par antisens de Dnmt1 reverse le phénotype des cellules transformées.

L'acétylation des histones influe également sur la structure de l'ADN. Les enzymes histones déacétylases (HDAC) et histones acétyltransférases (HAT) régulent le niveau d'acétylation du génome. L'initiation de la transcription nécessite un dépliement partiel des brins d'ADN. Ce procédé dépend de l'état d'acétylation et de phosphorylation des histones. L'acétylation de l'histone H4 permet un relâchement de l'ADN et donc la transcription tandis que sa dé-acétylation provoque un repliement local de l'ADN et empêche la transcription.

La structure de l'ADN est également modulée par des procédés de phosphorylation : en réponse aux signaux mitogéniques la kinase Rsk-2 phosphoryle transitoirement l'histone H3, facilitant la transcription.

### **b. L'épissage alternatif comme cause de l'expression différentielle**

Chez les eucaryotes, la plupart des régions codantes du génome ne sont pas continues. On distingue les séquences transcrites en ARNm, les exons, des séquences non transcrites dans l'ARNm mûr, les introns. Au cours de la maturation des ARN pré messagers, les

introns sont éliminés par excision par des facteurs d'épissage. Des séquences spécifiques délimitant introns et exons, sites d'épissage, sont reconnues et un ARNm ne contenant que les séquences codantes sera traduit en protéine. On a constaté qu'à partir d'un seul gène plusieurs ARNms pouvaient être produits par un phénomène d'épissage alternatif, plusieurs combinaisons de sites d'épissage pouvant être reconnues. Les différentes protéines codées par ces variants d'épissage alternatif peuvent présenter des fonctions, des propriétés de stabilité ou de régulation différentes, responsables de réponses cellulaires différentes. Ces phénomènes ont pris une importance particulière depuis l'estimation à la baisse du nombre total de gènes dans le génome humain, puisqu'ils contribuent à générer une importante diversité de protéines à partir d'un nombre plus restreint de gènes codants.

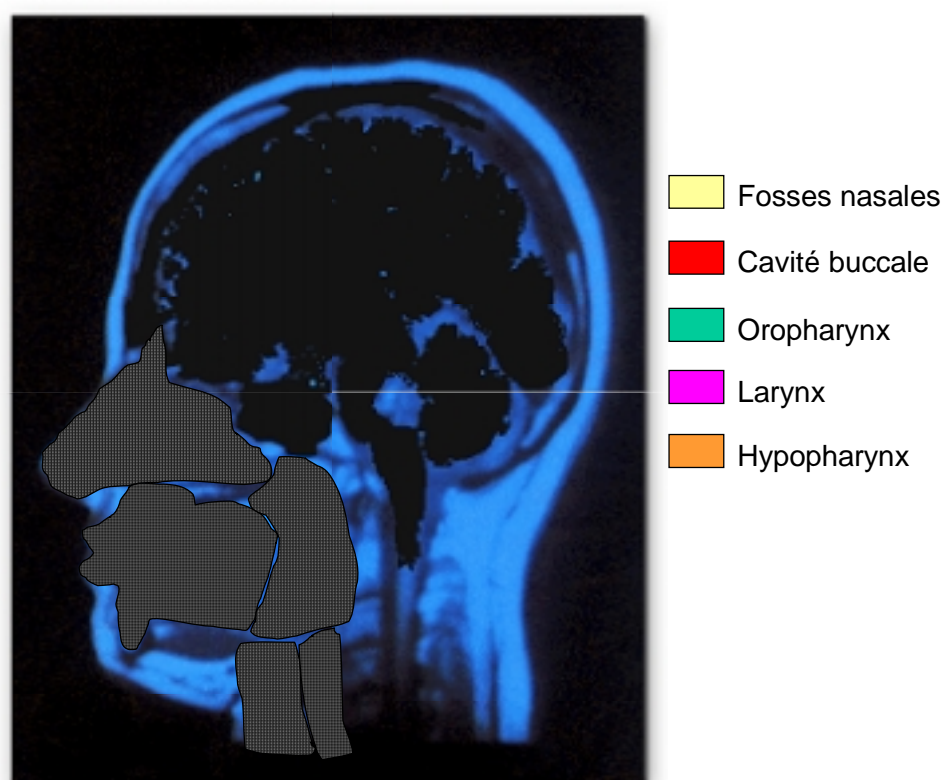


## B. Les cancers des voies aéro-digestives supérieures (VADS)

### 1. Données épidémiologiques sur les cancers VADS

D'après **Cancer, Principles, Practice of Oncology, 6th edition**. LVW Company 2001.  
Dorland's illustrated medical Dictionary, 27<sup>th</sup> edition. W.B. Saunders Company 1998.

Les cancers des voies aéro-digestives supérieures (VADS) regroupent les tumeurs originaires de la sphère ORL (oto-rhino-laryngée) qui inclut la cavité buccale (lèvres, langue, glandes salivaires, gencives et cavité buccale proprement dite), le pharynx (oropharynx, nasopharynx et hypopharynx) et le larynx (FigI.4).



Localisations des tumeurs VADS

Figure I.4

Ces cancers, essentiellement masculins, sont liés à une intoxication ethylo-tabagique. Dans 95% des cas, les cancers VADS sont des carcinomes épidermoïdes. Près des 2 tiers des

carcinomes siègent dans la filière digestive (cavité buccale, oropharynx et hypopharynx), les cancers du larynx représentant 30% des cas. L'incidence annuelle des cancers VADS est estimée à 500'000 cas nouveaux par an dans le monde, et, correspond à la cinquième position de l'ensemble des cancers (Parkin et al., 1999). En Europe occidentale, l'incidence des cancers VADS s'élevait en 1990 à 40,7 et à 6,1 cas pour 100'000 personnes pour les hommes et les femmes respectivement. La France présente la plus grande incidence de cas nouveaux, avec environ 19'500 cas nouveaux par an. Les cinq régions les plus touchées sont les départements français du Bas-Rhin, de la Somme, du Calvados, du Haut-Rhin et du Doubs (Franceschi et al., 2000; Menegoz et al., 2002; Parkin et al., 1999<sup>Ê</sup>). En France, les cancers VADS sont responsables de 12% des causes de mortalité par cancer, plaçant la France au 1<sup>er</sup> rang mondial pour les décès par cancer du larynx et au second rang mondial pour les décès par cancer du pharynx après l'Inde. Aux USA, les cancers VADS représentent 45'000 cas nouveaux par an.

Le taux de survie à 5 ans des patients atteints de tumeurs de la cavité buccale et du larynx était de 52,1% sur la période 1986-1991 et n'a guère progressé depuis 1973 (Sankaranarayanan et al., 1998). Le taux de survie à 5 ans des patients atteints de tumeurs de l'hypopharynx ne dépasse pas 30%. Les récurrences locorégionales sont très fréquentes puisqu'elles concernent 40 à 50% des patients. Les métastases à distance (pulmonaires, hépatiques, osseuses) surviennent dans 30% des cas (Vokes et al., 1993). Le terrain sur lequel se développent habituellement les tumeurs VADS, muqueuses chroniquement exposées au tabac et à l'alcool, explique en partie la fréquence des évolutions défavorables liées aux seconds cancers, développés dans 15 à 20% des cas, au niveau de la sphère ORL, de l'œsophage et des bronches.

## **2. Facteurs étiologiques**

### **a. Alcool et tabac**

Les cancers VADS, en Europe et en Amérique du nord, sont très fortement associés à une consommation importante d'alcool combinée à un tabagisme chronique. L'association ethylo-tabagique est particulièrement nocive puisqu'elle provoque un phénomène d'irritation chronique, dû aux brûlures thermiques et chimiques, ainsi qu'à l'exposition à des substances cancérigènes contenues dans la fumée de tabac, comme les hydrocarbures polycycliques. L'intoxication tabagique et l'imprégnation ethylique sont souvent associées, et, leurs effets sur le risque de cancer VADS sont multiplicatifs. Un mauvais état dentaire peut encore

amplifier les effets de l'alcool et du tabac. Seuls 3% des patients atteints de cancers VADS n'ont jamais fumé (Lingen et al., 2001).

### **b. Virus et autres facteurs alimentaires**

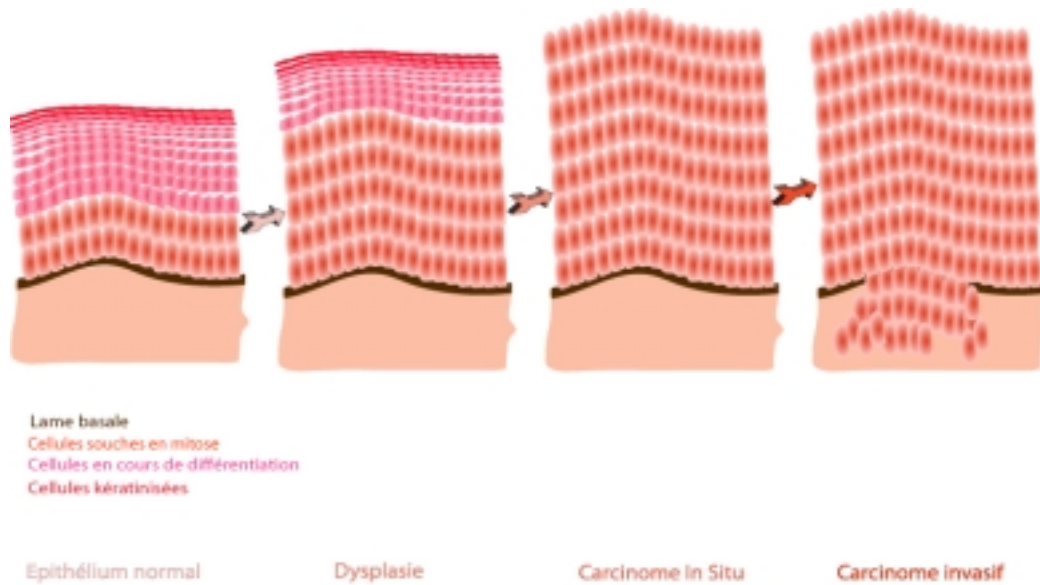
- Les papillomavirus humains semblent jouer un rôle dans certains cancers VADS (Gillison et al., 2000). La présence de HPV16, un variant à haut potentiel oncogénique, est notamment détectée dans 50% des tumeurs de l'oropharynx (Mork et al., 2001). Les virus pourraient plus particulièrement être impliqués chez les patients ne consommant pas d'alcool, ni de tabac.
- Le virus d'Epstein-Barr semble incriminé dans l'incidence très élevée des cancers du nasopharynx observée en Chine du Sud. Certains facteurs alimentaires, comme la consommation de poissons fumés riches en nitrosamines, pourraient contribuer à la survenue de ce type de cancers. En Inde, la consommation de noix de bétel accroît le risque de cancer de la cavité buccale.
- Dans le cas d'un cancer hypopharyngé de la région retro-crico-aryténoïdienne, fréquent dans les pays anglo-saxons, un déficit d'absorption de fer, dans le cadre de l'anémie sidéropénique de Plummer-Vinson, est incriminé en l'absence de toute intoxication ethylo-tabagique.

## **3. Données anatomopathologiques sur les cancers VADS**

### **a. Aspect anatomopathologique des cancers VADS**

La grande majorité des cancers développés au dépens des muqueuses VADS sont des carcinomes malpighiens: ils naissent soit directement d'une muqueuse malpighienne saine ou hyperplasique, soit d'une muqueuse de la région aérienne supérieure, après métaplasie.

Le développement d'un carcinome épidermoïde est habituellement considéré comme un processus progressif qui emprunte des stades intermédiaires avec apparition de lésions pré-malignes et de lésions pré-invasives. Au niveau histologique, on observe une séquence classique d'altérations progressives (Figure I.5).



Aspect anatomo-pathologique de l'épithélium au cours de la progression tumorale

Figure I.5

Le caractère hyperplasique conserve l'aspect d'une muqueuse saine au sein d'un épithélium épaissi, en raison d'une prolifération accrue des cellules épithéliales. L'aspect dysplasique concerne les modifications morphologiques comme la présence de mitoses et de pléïomorphismes dans la couche de cellules proliférantes. Cette dysplasie peut aboutir à un état malin confiné, dans un premier temps, au stade de carcinome *in situ*, avant l'étape invasive des structures sous-jacentes de l'épithélium. Ce schéma simplifié de la transformation maligne n'est pas obligatoire dans l'apparition des cancers VADS, et, chaque état intermédiaire peut être court-circuité jusqu'à l'éclosion apparemment directe d'un carcinome invasif. La diffusion de cellules cancéreuses vers les vaisseaux lymphatiques ou sanguins contribue à l'extension ganglionnaire et à la dissémination métastatique de la tumeur primitive initiale.

### **b. Lésions pré-cancéreuses**

Au niveau de la cavité buccale, différentes lésions bénignes ont été décrites comme potentiellement précurseurs des carcinomes épidermoïdes, principalement les leucoplasies et les erythroplasies.

- les leucoplasies correspondent à des tâches d'aspect blanchâtre, constituées de plaques épaisses kératinisées, que l'on peut retrouver sur la langue, le plancher buccal, le voile du palais ou la joue. Environ 10% des cancers évoluent sur des leucoplasies présentes.

- les erythroplasies sont des dysplasies épithéliales caractérisées par des plaques érythémateuses d'aspect lisse et brillant. Les erythroplasies sont souvent très diffuses et sont signalées dans 1 à 5% des cancers de la cavité buccale.
- la papillomatose orale floride prédominant sur les joues et les gencives, est constituée de touffes de villosités, de couleur blanche ou rosée, et, est mentionnée dans 3% des cas avant apparition d'un cancer invasif de la cavité buccale.

La détection de tels états pré-cancéreux permet soit une surveillance régulière pour prévenir une éventuelle aggravation, soit une ablation préventive dès lors que ces lésions ne sont pas diffuses, en nappes ou en foyers multiples.

### **c. Classification TNM**

La classification TNM décrit les caractéristiques physiques de la tumeur. La taille ou l'étendue de la tumeur primitive est indiquée en ordre croissant par son score T. Le degré éventuel d'extension ganglionnaire est noté par son score N. Enfin, la présence éventuelle de métastases à distance est indiquée par l'annotation M. La classification TNM est recommandée par l'American Joint Comite on Cancer Staging (AJCC) pour déterminer des stades cliniques notés de 1 à 4. Les tumeurs de stade 1 sont des tumeurs de petite taille, détectées le plus précocement pour lesquelles les chances de rémission sont les plus élevées. Les tumeurs de stade 4, de plus grande taille et présentant généralement des métastases à distance lors du diagnostic, sont pratiquement incurables.

### **d. Effet de champ**

Les cancers VADS sont généralement associés à des lésions souvent extensives et multifocales dénotant un "effet de champ" {field cancerization, FigI.6, (Ha & Califano, 2003)} par lequel un grand nombre de cellules de l'épithélium malpighien VADS sont simultanément exposées aux agents carcinogènes. L'apparition de tumeurs VADS semble correspondre à un processus de cancérisation qui serait la conséquence d'une exposition répétée aux agents carcinogènes alcool et tabac, de l'ensemble de l'épithélium VADS. Cette exposition globale et permanente augmenterait fortement les risques d'apparition de lésions multiples dysplasiques ou malignes, par induction d'altérations génétiques indépendantes. Ce

concept est compatible avec la fréquence élevée de seconds cancers, aussi bien au niveau de la sphère VADS qu'au niveau des régions voisines aériennes (bronches) et digestives (œsophage). La transformation maligne de l'épithélium du tractus aéro-digestif est le résultat d'une accumulation d'événements génétiques impliquant des mutations, délétions et gains chromosomiques, d'oncogènes et de gènes suppresseurs de tumeurs.

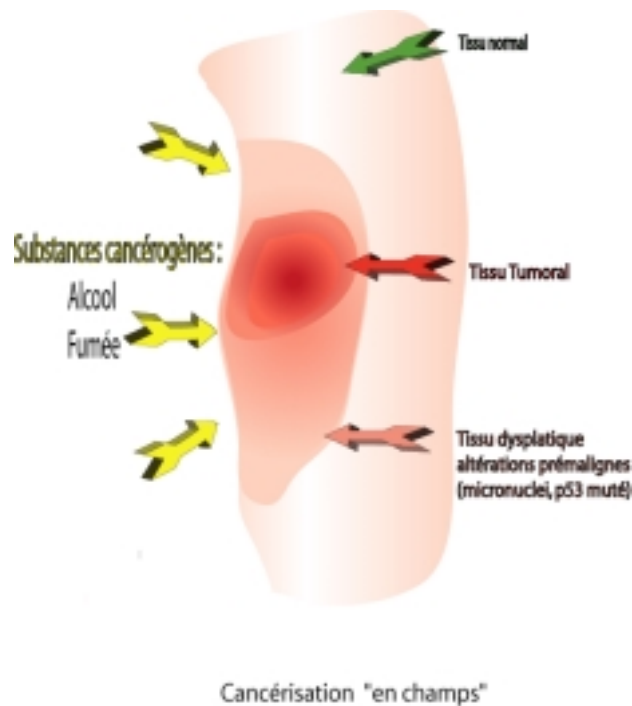


Figure I.6

#### 4. Altérations chromosomiques et gènes cibles

##### a. Altérations les plus communes dans les cancers VADS

Des altérations chromosomiques spécifiques et fréquentes ont été identifiées, notamment par CGH {Comparative Genomic Hybridization, revue par (Bockmuhl & Petersen, 2002)} dans les tumeurs VADS. Les pertes les plus fréquentes concernent les chromosomes 3p (70%), 9p (70%), 18q (43%) 8p (40%) et 17p (40%). Les gains de matériel génétique sont souvent retrouvés en 3q (70%), 8q (50%) et dans la région 11q13 (40-50%). Pour certaines aberrations comme les délétions de 9p, 17p et l'amplification en 11q13, les

gènes cibles ont été identifiés et correspondent respectivement à p16INK4a-p14ARF, p53 et Cyclin D1.

- Le gène INK4a code à la fois pour p16INK4a et p14ARF, produits résultant d'un cadre de lecture alternatif. La protéine p16INK4a régule négativement la voie pRb et p14ARF intervient dans la stabilité et l'activité de transactivation de la protéine p53. Les modalités d'inactivation de ces 2 gènes sont la délétion homozygote et l'hyperméthylation d'ilôts CpG dans la région du promoteur. Les pertes alléliques de p16 atteignent 57% des tumeurs VADS. La perte de l'expression de la protéine p16 est observée dans 55 à 80% des tumeurs VADS (Reed et al., 1996).
- Les gains de la région 11q13 sont détectés dans 30 à 50% des carcinomes épidermoïdes VADS (Muller et al., 1997). Cette dernière région contient le gène CCND1, codant pour la cycline D1, oncogène associé à la progression des carcinomes épidermoïdes (Jares et al., 1994; Motokura et al., 1991). La cycline D1 facilite le passage du point de restriction R d'entrée dans le cycle cellulaire (Malumbres & Barbacid, 2001). Plus de 10 gènes sont contenus dans les amplicons de la région 11q13, mais seuls EMS-1 (Patel et al., 1996) et TAOS1 (Huang et al., 2002) sont amplifiés et surexprimés dans les cancers VADS. Ils sont donc des oncogènes potentiels. Dans la sixième partie de cette thèse, je décrirai les résultats préliminaires de la caractérisation d'un nouveau gène (B14.4 hypothetical protein FLJ10261), situé dans l'amplicon 11q13 et surexprimé dans les tumeurs hypopharyngées. Le locus du gène de la cycline D1, en 11q13, est amplifié dans 35 à 58% des cancers VADS, avec une fréquence particulièrement élevée dans les tumeurs de l'hypopharynx (Muller et al., 1997). Cette amplification, associée à une surexpression du gène, semble être associée aux événements d'initiation tumorale puisqu'elle est détectée dès les tumeurs précoces.
- Les mutations du gène p53 représentent l'événement génétique le plus fréquemment retrouvé dans les cancers VADS et concerne 80% des cas (Kashiwazaki et al., 1997). Les mutations sont réparties dans l'ensemble du domaine de liaison à l'ADN, et, abolissent les capacités de transactivation de p53. Les mutations ponctuelles correspondent dans plus de la moitié des cas à des transversions, avec un taux plus élevé de type GC=>TA, reflétant le rôle des carcinogènes du tabac qui agissent plus particulièrement sur la substitution des guanines. Ces mutations stabilisent la protéine, qui devient alors détectable en immunohistochimie. La conservation de p53 non muté serait un marqueur de bon pronostic, dans le cadre des traitements radio-chimiothérapeutiques. Les mutations

inactivant p53, comme les mutations dans le domaine de liaison à l'ADN, ont été associées à une résistance aux traitements chimiothérapeutiques néo-adjuvants. Les différentes études pronostiques demeurent néanmoins divergentes (Temam et al., 2000).

- Le récepteur EGF (Epidermal Growth Factor) fait partie de la famille des récepteurs à activité tyrosine kinase. La liaison de facteurs de croissance, parmi lesquels EGF et TGF alpha, active de nombreuses protéines de la cascade de transmission du signal, en particulier la voie de signalisation de la phosphatidylinositol kinase, et, celle des MAP kinases. L'activation du récepteur EGF contribue à la stimulation de la prolifération cellulaire, l'inhibition de l'apoptose, l'induction de la néo-angiogenèse et l'expression de métalloprotéases matricielles impliquées dans les processus invasifs des cancers VADS (O-charoenrat et al., 2002). Le récepteur EGF est surexprimé dans 42 à 90% selon les localisations des cancers VADS, et, son expression est corrélée à une moindre survie. L'expression du récepteur EGF serait associée à une agressivité accrue des tumeurs (Ang et al., 2002). Un anticorps monoclonal humanisé dirigé contre EGFR (Cetuximab) ainsi qu'un inhibiteur de l'activité tyrosine kinase de EGFR sont en cours d'évaluation thérapeutique.

## **b. Un modèle génétique de progression tumorale des cancers VADS**

De manière similaire au modèle d'initiation et de progression des tumeurs du colon (Vogelstein et al., 1989), une analyse des pertes chromosomiques observées dans diverses études cytogénétiques des tumeurs VADS a permis au groupe de Sidransky (Califano et al., 1996) d'établir un modèle préliminaire d'initiation et de progression tumorale de ces tumeurs. Cette analyse était basée sur l'analyse de 10 régions sélectionnées du génôme pour leur implication dans l'initiation et la progression tumorale. Un approfondissement de ce modèle, basé sur l'utilisation de la CGH, avec intégration de gains chromosomiques, a été proposé (Bockmuhl & Petersen, 2002). La perte séquentielle des régions 3p, 9p, 11q, 13q, 8 et 4q est comparable au modèle établi par Califano. La perte en 14q est isolée mais associée de manière contradictoire à des phases tardives de la progression tumorale (FigI.7).



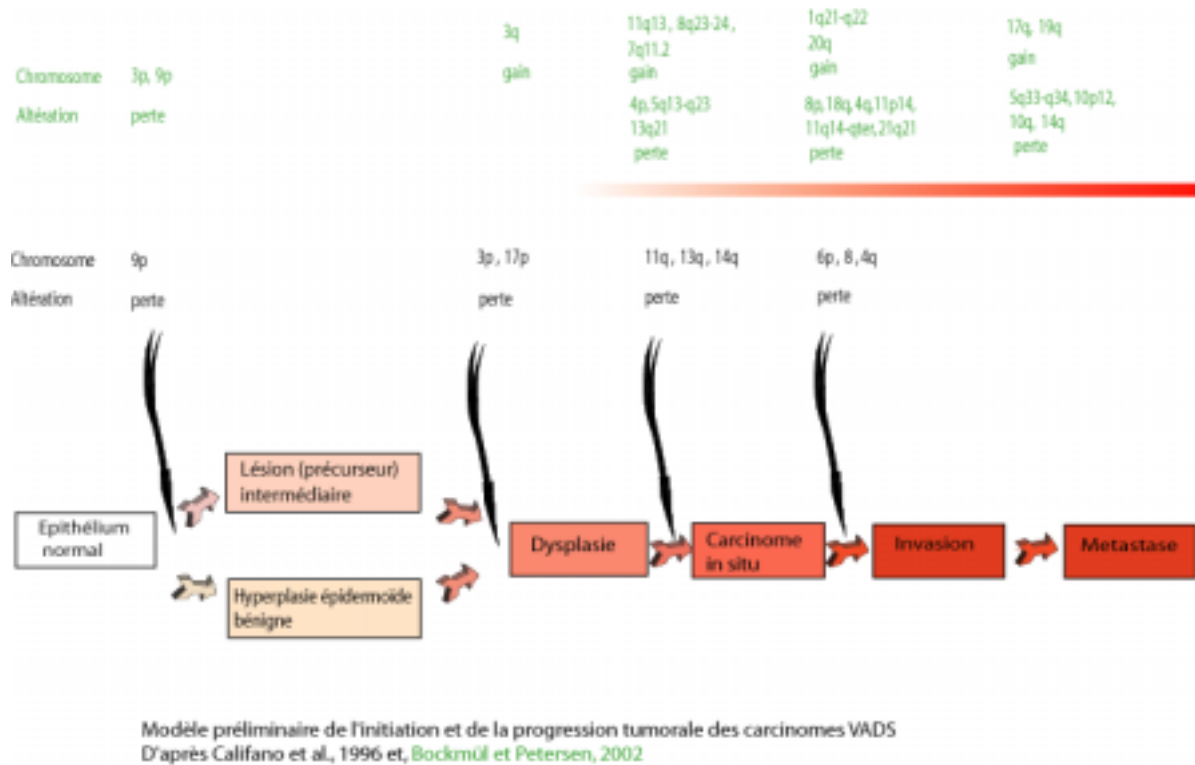


Figure I.7

## 5. Traitement des cancers VADS

D'après **Cancer Medicine 5th ed.** Bast, Robert C.; Kufe, Donald W.; Pollock, Raphael E.; Weichselbaum, Ralph R.; Holland, James F.; Frei, Emil, editors. Hamilton (Canada): **BC Decker Inc**; c2000.

Le traitement des cancers VADS reste avant tout loco-régional. Les deux piliers du traitement des cancers VADS associent la chirurgie cervico-faciale à la radiothérapie. Le diamètre et l'extension de la tumeur, la présence d'adénopathies, leur nombre et leur siège et leur fixation éventuelle sont des facteurs de choix du traitement initial. Ces critères permettent de choisir entre la chirurgie ou la radiothérapie seule pour les lésions limitées ou superficielles, et, une combinaison planifiée de ces traitements pour les tumeurs étendues et infiltrantes.

- Les progrès concomitants de la chirurgie réparatrice et de l'anesthésie-réanimation permettent une chirurgie susceptible de retirer des tumeurs plus étendues avec une marge de sécurité d'exérèse plus fiable. La meilleure connaissance des voies d'extension tumorale permet une chirurgie plus conservatrice des structures pharyngo-laryngées. L'élément

critique de la chirurgie est l'excision de la totalité des cellules malignes, la radicalité de l'exérèse déterminant fortement la survie du patient. Un examen anatomopathologique des marges de résection de la biopsie est donc nécessaire.

- La radiothérapie post-opératoire est fondée sur l'extension initiale et la prédiction du risque de récurrence loco-régionale, évaluée essentiellement à partir de l'examen des marges d'exérèse et de l'intensité de l'envahissement ganglionnaire. Une radiothérapie sur le lit d'exérèse et les aires ganglionnaires cervicales réduit le risque de récurrences locales et ganglionnaires. La radiothérapie bénéficie de progrès récents (appareillages, dosimétrie, radiobiologie) particulièrement dans des protocoles de modalités de fractionnement des doses administrées, et, de l'association chimio-radiothérapeutique.
- La chimiothérapie des cancers VADS a été longtemps réservée aux traitements des récurrences et des métastases. De nombreux agents anticancéreux comme le méthotrexate, le cisplatine, le fluorouracile, les taxanes sont des agents actifs sur les tumeurs VADS, l'association cisplatine/5-FU (5-Fluorouracile) étant la plus utilisée. Les associations incluant le cisplatine et le 5-FU ont permis d'observer des réponses complètes de l'ordre de 20%, mais sans augmentation significative de la survie. La chimiothérapie est dite néoadjuvante lorsqu'elle précède la chirurgie, de façon à réduire la masse tumorale avant l'exérèse, rendant l'opération moins mutilante. Elle permet la préservation de l'organe, en particulier le larynx, mais ne diminue que faiblement l'incidence des métastases et ne modifie pas la survie des malades. Les taxanes, nouvelle génération de drogues agissant sur le fuseau mitotique, semblent donner des résultats prometteurs sur les cancers VADS.

## **6. Les différentes localisations des tumeurs VADS**

### **a. Les cancers de la cavité buccale**

La cavité buccale est le siège de 30% des cancers VADS, représentant 5'000 cas nouveaux par an en France. Le cancer de la langue représente à lui seul 50% des cancers de la cavité buccale. Selon le stade d'évolution le traitement fait appel à la chirurgie, à la radiothérapie (tumeur très localisée), ou à la chimiothérapie (tumeur étendue). Les cancers de la base de la langue ont généralement un moins bon pronostic que les cancers de la bordure de la langue. Les cancers des joues, du plancher buccal, de la bouche et des gencives se

signalent par des symptômes variés et sont traités de manière similaire aux cancers de la langue.

### **b. Les cancers du pharynx**

En France, le pharynx est le siège de 35% des cancers VADS, représentant 7'000 cas nouveaux par an. Le pharynx est subdivisé en trois parties : le rhinopharynx en relation avec les fosses nasales; l'oropharynx en relation avec la cavité buccale, comprenant les amygdales et le voile du palais ; l'hypopharynx précédant l'oesophage et situé en arrière du larynx. La région du sinus pyriforme est fréquemment le siège de cancers, en rapport avec une exposition quotidienne à l'alcool (eaux de vie et alcools purs). Les cancers de l'hypopharynx sont les plus fréquents avec 17% des cancers VADS, après les cancers du larynx (20%), les autres cancers les plus fréquents étant les cancers de la langue et du plancher buccal (16%), et, des amygdales (11%). Les cancers de l'hypopharynx ont un taux de survie à 5 ans de 30% seulement contre 55% pour les cancers du larynx. Les tumeurs pharyngées sont généralement de taille élevée, avec des extensions ganglionnaires au moment du diagnostique. Le traitement des cancers du pharynx comprend la chirurgie (ablation d'une partie du pharynx et parfois du larynx) et/ou la radiothérapie. Cette chirurgie mutilante nécessite une reconstruction du pharynx par chirurgie plastique. Pour les tumeurs de stade avancé, une chimiothérapie néoadjuvante ou des associations chimioradiothérapiques sont proposées aux patients.

### **c. Les cancers du larynx**

Les cancers du larynx représentent plus de 30% des cancers VADS, affectant essentiellement les hommes (90% des cas). Le larynx est un organe cartilagineux assurant la protection des voies respiratoires par fermeture de la glotte lors de la déglutition. Le tabac joue un rôle prépondérant amplifié par son association avec l'alcool. La moitié des ces cancers affectent les cordes vocales tandis l'autre moitié se développe au dessus de celles-ci (épiglotte). Contrairement aux cancers du pharynx, un diagnostique rapide est possible avec une survie à 5 ans dans 55% des cas. Le traitement met en jeu la radiothérapie (tumeur localisée) ou la chirurgie. En cas de lésions très étendues la chimiothérapie peut être utilisée. La chirurgie peut être partielle ou une laryngectomie totale, mutilante, suivie d'une trachéotomie avec rééducation de la voix (voix oesophagienne).

## C. Méthodes d'analyse du transcriptome

---

Certaines approches, mettant en jeu un nombre fini de gènes (gène individuel ou plusieurs milliers de gènes disposés sur un support), limitent dès le départ le champs d'étude ; c'est le cas des puces à ADN disponibles commercialement jusqu'à une époque récente. Ces supports permettent surtout l'analyse d'un grand nombre de gènes connus. On parle de systèmes fermés.

Certaines approches sont dites ouvertes, dans lesquelles les expérimentateurs isolent, sans *a priori*, des séquences différentiellement exprimées à partir d'échantillons biologiques et les identifient *a posteriori*. Ces techniques permettent d'isoler des séquences demeurées inconnues, car correspondant à des gènes exprimés à faible nombre de copies par cellule, ou, à des variants inconnus, liés à des mutations ou des phénomènes d'épissage alternatif. De plus, ces techniques sont adaptées à l'étude d'organismes au génome mal caractérisé.

L'utilisation des systèmes fermés et ouverts est complémentaire car les séquences isolées dans des systèmes ouverts peuvent ensuite être transférées à de nouvelles générations de puces en système fermé.

Pour l'ensemble des techniques, plusieurs facteurs sont critiques :

-La quantité d'ARN nécessaire à la réalisation du protocole. Cette limitation est particulièrement importante dans le cadre de l'utilisation de biopsies, comme dans l'étude de l'expression génique de tumeurs.

-La sensibilité de la technique, définie par sa capacité à isoler de manière probable des gènes différentiellement exprimés à une faible prévalence. Cette caractéristique est particulièrement importante puisqu'on estime que 90-95% des ARNms sont exprimés à 5 copies ou moins par cellule (Alberts, 1998; Bishop et al., 1974; Bonaldo et al., 1996; Zhang et al., 1997). Si l'on recherche des ARNs encore mal caractérisés, généralement exprimés à faible nombre de copies, on doit faire appel à des techniques capables de détecter des ARNms présents à une copie pour 100'000 ARNms (Wan et al., 1996, Green et al., 2001).

-Le taux de couverture du génome, responsable de l'exhaustivité de l'étude, est également critique. Les techniques fermées disponibles pendant la durée de notre étude ne pouvaient prétendre à l'exhaustivité (12'000 gènes typiquement). Les techniques ouvertes majeures sont

censées pouvoir offrir une couverture de 90 à 99% du transcriptome, en fonction du nombre d'itérations, de réactions ou de réactions de séquençage réalisées.

## **1. Les systèmes « fermés »**

### **a. Analyse de gène individuel**

L'expression des gènes peut être analysée à titre individuel avec les techniques de Northern blot et de protection des nucléases (RNase protection assay), sans qu'aucune amplification par PCR (polymerase chain reaction) n'intervienne. Le Northern blot permet d'avoir accès à la taille du transcrit et, éventuellement, à ses différents variants, tandis que la technique de protection des nucléases permet de dresser un plan des mésappariements dans le transcrit (gene mapping).

Ces techniques permettent l'analyse des transcrits d'abondance forte à moyenne, sauf si les ARN poly-A sont sélectionnés au préalable. Elles nécessitent des quantités importantes d'ARN. Ces techniques sont moins sensibles que les techniques plus récentes de PCR semiquantitatives ou quantitatives {Realtime PCR, (Heid et al., 1996)}.

### **b. Analyse en masse selon le « Reverse Northern blot » : macro et micro arrays**

La technique du Reverse Northern blot (RN), inspirée du Northern blot, consiste à hybrider une sonde complexe à une série d'ADNcs disposés sur un support, en tant que plasmides, fragments de PCR ou encore d'oligonucléotides.

La technique du RN permet d'analyser l'expression de plusieurs centaines de gènes en parallèle, et, éventuellement d'observer l'altération d'un grand nombre de processus physiologiques. Elle permet également de réaliser une vaste étude à l'aide d'une seule synthèse de sonde, ce qui permet un travail à partir de petites quantités d'ARNm, comme celles extraites de biopsies. L'analyse des résultats nécessite un traitement informatique et peut mener à la caractérisation de signatures corrélées au pronostique (clustering) ou à diverses caractéristiques biologiques (Golub et al., 1999; Mendez et al., 2002; Ramaswamy et al., 2003; van 't Veer et al., 2002).

Les macro arrays sont des puces à ADN de moindre densité que les micro-arrays, généralement réalisés sur des membranes en nylon et fabriqués, suivant leur densité, par des

expérimentateurs (plaques 96 puits) ou par un robot de dépôt. Ce type d'arrays peut constituer la plateforme secondaire de validation en Reverse Northern blot du criblage initial en système ouvert, du fait de leur personnalisation (« customisation ») aisée. Les macro arrays commerciaux, comme les membranes Atlas (Clontech) comportant environ 600 gènes, sont par contre des systèmes fermés. Les micro-arrays commerciaux se distinguent en fonction du type d'ADN déposé sur le support : ADNc en tant que clones ou produits de PCR, ou, oligonucléotides synthétisés par photolithographie sur le support (Marshall & Hodgson, 1998). Cette dernière technique, utilisée par Affymetrix pour produire ses puces à ADN comportant environ 12'000 gènes, permet d'obtenir une grande reproductibilité et une grande homogénéité d'une puce à l'autre. La détection de transcrits effectivement différenciellement exprimés est facilitée, tout comme l'analyse comparative d'une grande série d'échantillons. Les quantités d'ARN nécessaires pour assurer la qualité de l'étude demeurent assez importantes.

Un certain nombre de critiques à l'égard des micro-arrays et de l'interprétation de leurs résultats ont été résumées par Liang, l'inventeur du Differential Display, dans sa rétrospective de 10 ans d'utilisation du DD (Liang, 2002). Liang note que beaucoup d'expérimentateurs ont un peu rapidement considéré les résultats de micro-arrays, qui sont des Reverse Northern blots, comme des preuves ultimes du profil d'expression. Or la seule preuve absolue (gold standard) d'une expression différentielle est produite en Northern blot, sans faire intervenir de PCR. On note généralement un faible taux de preuves en validation secondaire présentées dans les études par micro-arrays. La comparaison que j'ai effectué avec des études concurrentes menées par micro-arrays (Al Moustafa et al., 2002; Alevizos et al., 2001; Belbin TJ, 2002; El-Naggar et al., 2002; Leethanakul et al., 2000; Mendez et al., 2002; Xie et al., 2000) m'a permis de constater personnellement que, dans l'étude des cancers VADS, les clones utilisés pour la validation des résultats sont souvent peu nombreux (moins d'une dizaine de gènes validés pour des listes de plusieurs centaines de gènes différenciellement exprimés). Ces gènes sont généralement classiquement associés au cancer, comme les cadhérines ou les caténines. La validation est souvent assurée par des techniques hétérogènes (PCR semi quantitative, Western Blot, etc...). Apparemment les exigences de vérification des faux positifs n'ont pas été aussi sévères que lors de l'apparition du DD, bien que de nombreux problèmes techniques existent, comme pour toute technique émergente. Une forme indirecte de validation dans certaines études est l'obtention de groupes de gènes prédictifs de sous-groupes d'échantillons (clustering). Si les micro-arrays ont les limitations des systèmes fermés, une autre de leurs limites majeures est l'utilisation de sondes complexes.

Il est difficile de s'assurer de la reproductibilité des signaux. Des membres de familles de gènes ou des gènes possédant une forte homologie peuvent produire des hybridations croisées. La sensibilité des sondes et la linéarité de la détection du signal sont également des points critiques. La constitution même des puces a été, initialement au moins, sujette à des problèmes de reproductibilité d'une puce à l'autre. Affymetrix tente de s'affranchir de ces limitations en associant à chaque oligonucléotide un oligonucléotide de contrôle muté, moins spécifique, et, en synthétisant les oligonucléotides par photolithographie. Les premières études se sont voulues, à tort, exhaustives sur tout le génome alors que seulement une fraction des gènes (1/3 typiquement) était disposée sur le support. Les difficultés techniques associées à la production de puces à ADN sont illustrées par le fait qu'une des compagnies majeures dans leur développement a abandonné toute production (Knight, 2001a) et qu'Affymetrix ai dû rappeler ses puces « souris » du fait d'un taux de synthèse de 65% de séquences erronées (Marshall, 2001) et que les puces comportaient parfois 30% d'erreurs {séquences mal répertoriées sur la puce, erreurs dans lors de la synthèse de l'oligonucléotide, etc... ; (Knight, 2001b)}.

## **2. Analyses en masse en systèmes « ouverts »**

Les techniques suivantes sont toutes des techniques « ouvertes », souvent adaptées à l'isolement d'événements encore mal caractérisés. De leurs principes résultent des besoins en matériel initial, une exhaustivité et une sensibilité variables, certaines techniques n'étant que peu adaptées à la détection des ARNms peu abondants.

### **a. Criblage différentiel (Differential screening)**

Le principe de cette méthode ancienne (Sargent, 1987) est le clonage de banques d'ADNcs correspondant à chacune des situations comparées dans des bactéries, sur boîte de Petri, et, de l'hybridation de membranes répliques de ces boîtes de culture avec des sondes synthétisées à partir des ARNms correspondant à chaque situation. Les clones différemment allumés par les sondes sont différentiellement exprimés entre les deux situations. Cette technique est adaptée à l'isolement rapide de quelques clones d'intérêt. Une même séquence peut être isolée de façon très redondante. Aucune connaissance préalable du génome n'est requise.

## **b. Techniques basées sur le séquençage en masse: Soustraction Electronique (ES) et SAGE**

Le moyen le plus évident de mesurer l'expression d'un gène est de cloner et de séquencer tous les ADNcs d'un échantillon, puis de compter le nombre relatif de clones contenant une séquence précise. Le nombre de clones, ou sa fréquence de clonage, est proportionnel à l'abondance du transcrit initial. Cependant cette méthode est difficilement applicable de manière pratique, tant le nombre de clones à séquencer serait important. Une technique alternative est le séquençage d'une fraction d'une banque d'ADNc puis l'interprétation informatique de la fréquence de détection des divers transcrits. Cette technique, appelée Soustraction Electronique (Electronic Substraction, ES) est basée sur l'isolement d'EST {Expressed Sequence Tag ; (Adams et al., 1992)}. Toutes les techniques basées sur un séquençage massif de banque d'ADNcs ou d'étiquettes ont en commun un coût très important de séquençage. Aussi leur utilisation est elle limitée lorsque même les transcrits invariants sont répertoriés et constituent un poids mort dans l'étude.

Une technique alternative, appelée SAGE {Serial Analysis of Gene Expression, (Velculescu et al., 1995)}, a été développée de façon à réduire la quantité de séquençage nécessaire à l'analyse, par digestion des ADNcs par des enzymes de restriction. Des étiquettes typiquement de 10 à 14bp (ditags), permettant l'identification du gène correspondant, sont concaténées et clonées dans des plasmides en vue de séquençage. En pratique, un clone contient en moyenne 30 à 35 étiquettes-SAGE différentes, permettant une diminution importante de la tâche de séquençage à accomplir et une accélération de l'analyse par rapport à l'ES.

Cette technique demeure cependant exigeante en qualité et en quantité de séquençage, nécessite des étapes multiples, et, est lourde à mettre en oeuvre. La sensibilité et l'exhaustivité de cette technique dépend du nombre de plasmides (et donc d'étiquettes) séquencés. Pour avoir plus de 90% de chances de repérer un gène différentiellement exprimé présent à 3 copies par cellules, il est nécessaire de séquencer 300'000 étiquettes correspondant à 10'000 à 15'000 clones (Bertelsen, 1998). L'étude de cinétiques ou la comparaison d'échantillons nombreux est quasiment impossible par cette technique puisqu'un nombre très important de séquences doit être obtenu pour chaque point de comparaison.

La technique MICROSAGE est une technique dérivée du SAGE plus adaptée à l'utilisation de biopsies, grâce à la réalisation de toutes les réactions dans un tube unique et à l'utilisation



de cycles de PCR supplémentaires (Datson et al., 1999). La technique TALEST {Tandem Arrayed Ligation of Expressed Sequence Tags, (Spinella et al., 1999)} est une technique dérivée du SAGE dans laquelle la génération de ditags n'implique pas d'amplification par PCR mais uniquement une série de digestions enzymatiques.

En plus de son coût de séquençage, la technique SAGE, bien qu'ayant contribué à une accélération de la construction de bases de données sur les ESTs, présente l'inconvénient de nécessiter de nombreuses étapes de manipulation par rapport au DD et aux micro-arrays, ce qui augmente les sources potentielles d'artéfacts (Liang, 2002).

### **c. Techniques basées sur l'hybridation et la soustraction : RDA, SH, SSH**

L'hybridation soustractive (SH), méthode très utilisée, est décrite plus loin (Partie III) dans une comparaison stricte avec la technique du DD effectuée par Wan et al (Wan et al., 1996). Cette méthode nécessite une forte quantité d'ARN polyA (20µg), implique de multiples étapes d'hybridation et de soustraction. Elle est de mise en oeuvre laborieuse. Cette technique n'est pas toujours adaptée à la détection des transcrits peu abondants. Les différences d'expression isolées peuvent être quantitatives mais cette méthode permet plus efficacement la détection de différences qualitatives, de séquences spécifiques à une situation. L'analyse complète d'une comparaison nécessite de réaliser l'hybridation soustractive pour les deux situations, en deux manipulations, et, rend difficile l'analyse de larges séries d'échantillons.

La méthode RDA {Representational Differences Analysis, (Lisitsyn & Wigler, 1993)} est une technique dérivée de l'Hybridation Soustractive, mettant en jeu des amplifications par PCR pour enrichir les différences entre les deux génomes. On effectue une comparaison entre le transcriptome analysé, le « tester », et, le contrôle, le « driver ». Bien que les différences ne soient pas isolées de manière spécifiquement quantitative, les validations par Northern Blots ont montré que cette technique permet d'isoler des ADNcs différenciellement exprimés de 2 à 80 fois (Hubank & Schatz, 1994). Seulement 1% des ARNms de la cellule sont isolés. Cette méthode est difficile à mettre en oeuvre techniquement et le taux de faux positifs (et de faux négatifs) est important. Il faut répéter l'opération pour obtenir les séquences spécifiques du driver. Un des grands avantages de cette technique est son application possible aux organismes procaryotes (Bowler et al., 1999).

La technique d'Hybridation Soustractive avec Suppression {SSH, (Diatchenko et al., 1996)} basée sur la technique RDA et sur l'effet de soustraction par PCR, a été développée afin de d'isoler des séquences moins abondantes que le RDA, à partir de moins de matériel. Cette méthode combine plusieurs étapes de soustraction à une amplification par PCR, assurant une normalisation des banques d'ADNc du tester et du driver. Des amorces/adaptateurs différentes sont utilisées pour le tester et le driver. Après hybridation soustractive, une amplification par PCR est effectuée avec des amorces de PCR courtes correspondant aux séquences spécifiques des différents adaptateurs, ne permettant l'amplification que des séquences présentes dans le tester mais absente ou sous-exprimées dans le driver, portant les deux types d'extrémités amplifiables par PCR. Les fragments portant les adaptateurs issus d'un même tester sont éliminés (effet suppressif de la PCR) car la présence de séquences répétées de taille supérieure aux amorces de PCR entraîne d'hybrides en boucles plus stables que les interactions amorces/matrice (suppression). Cette technique permet d'utiliser de moindres quantités d'ARN (l'ADNc peut être utilisé également alternativement), d'enrichir les banques spécifiques de chaque situation de l'ordre de 1000 à 5000 fois, et, de détecter les transcrits d'abondance moyenne à faible. Une mise au point au niveau de la quantité de tester permet de renforcer le caractère quantitatif potentiel de cette méthode. Un seul cycle de manipulation est nécessaire.

#### **d. PCR Differential Display (PCR-DD)**

Cette technique implique la sélection de sous populations des ARNm par PCR avec des amorces poly-T combinées avec des amorces arbitraires, séparables en bandes discrètes après migration sur gel. Cette technique qui possède l'avantage de nécessiter peu d'ARNm, peu de matériel (machines à PCR et gels de séquençage), est la technique avec laquelle nous avons mené notre étude. Elle sera décrite plus précisément plus loin. Ces avantages sont sa sensibilité, son faible besoin en ARN, sa simplicité conceptuelle et le faible nombre de manipulations à effectuer avant de visualiser les différences d'expression sur gel.

Plusieurs méthodes très proches conceptuellement du DD existent, comme l'AP-PCR {Arbitrarily Primed PCR, (Navarro & Jorcano, 1999)}, l'AFLP {Amplified Fragment Length Polymorphism ; (Bachem et al., 1996)}, et, la technique SABRE {Selective Amplification via -Biotin and Restriction-mediated Enrichment ; (Lavery et al., 1997)}. Ces techniques n'ont été en comparaison du DD que peu utilisées dans l'étude de l'expression génique.

D'autres techniques conceptuellement proches du DD utilisent des enzymes de restriction pour effectuer une subdivision des ARNms, comme la technique RFLP-DD (Restriction Fragment Length Polymorphism- DD). Cette technique et les autres techniques dérivées Genecalling (Green et al., 2001), TOGA {Total Gene Expression Analysis ; (Sutcliffe et al., 2000)}, READS {Restriction Enzyme Analysis of Differentially Expressed Sequences, (Prashar & Weissman, 1996)} présentent le désavantage par rapport au DD de nécessiter plus d'étapes de manipulation, augmentant les risques d'artéfacts. La plupart de ces méthodes n'existaient pas lors du démarrage de notre étude, et, à ce jour, on ne possède que peu de recul sur leur utilité.

#### **e. DATAS**

La technique D.A.T.A.S. (Differential Analysis of Transcripts with Alternative Splicing), brevetée par *Exonhit Therapeutics S.A.*, met en jeu des étapes de soustraction et d'amplification par PCR, permettant d'isoler de façon exhaustive les évènements d'épissage alternatif. A des variants d'épissage alternatif correspondent des protéines possédant éventuellement des propriétés fonctionnelles spécifiques. L'estimation du nombre de gènes dans le génome humain après son séquençage complet a diminué de 50'000-100'000 gènes estimés à 30'000-40'000 gènes estimés (Lander et al., 2001; Venter, 2003; Venter et al., 2001). Dans cette nouvelle optique, les phénomènes d'épissage alternatif ont pris une importance croissante puisque ce mécanisme permet d'expliquer en partie le codage d'un nombre important de protéines à partir d'un nombre restreint de gènes.

## **Partie II:**

# **Objectifs et stratégie expérimentale**

## A. Problématique

---

Après que les connaissances sur le cancer ont progressé et que des altérations d'expression génétique ont été identifiées comme responsables de la carcinogénèse, un axe majeur de recherches a été l'établissement d'une signature, d'une carte d'identité des tumeurs. Ainsi peut-on envisager d'améliorer le pronostic des patients atteints de ces tumeurs et de mieux caractériser les événements clés de la carcinogénèse. Sur ces nouvelles bases, de nouvelles thérapies ciblées et rationnelles, devraient être développées, avec potentiellement une plus grande efficacité et une moindre toxicité que les traitements actuels. Étant donné les techniques à notre disposition lors du démarrage de notre étude, la nature hétérotypique des tumeurs et les caractéristiques particulières des carcinomes VADS (effet de champs, mauvais pronostic), une stratégie expérimentale a été conçue de manière à effectuer une analyse aussi exhaustive que possible de l'expression génique de ces tumeurs. Les carcinomes représentant plus de 90% des tumeurs, nos résultats pourraient être partiellement informatifs pour d'autres localisations tumorales. Ainsi une analyse complémentaire sur des tumeurs prostatiques a été entreprise à partir d'une sélection de gènes isolés dans les carcinomes VADS. Plusieurs approches ont été combinées de manière à effectuer un profilage aussi bien quantitatif que qualitatif de l'expression génique dans les tumeurs VADS. Pour ma part, j'ai principalement été impliqué dans la mise en œuvre d'un criblage par Differential Display à large échelle.

## B. Rappel du contexte de l'étude et stratégie expérimentale appliquée à la problématique

---

### 1. Rappel du contexte de l'étude

#### a. Contexte historique au niveau du laboratoire

Depuis de nombreuses années le laboratoire du Dr Bohdan Wasylyk étudie l'expression d'oncogènes (MDM2, Ras) et de gènes suppresseurs de tumeurs dans les tumeurs VADS. En particulier, une collaboration a été établie avec le Dr Ranju Ralhan, qui supervise une collection de tumeurs VADS en Inde. Ces tumeurs sont particulièrement intéressantes car la mastication de noix de bétel provoque la formation de nombreuses lésions cancéreuses et précancéreuses de la cavité buccale. Ce type de tissus permet d'avoir accès aux étapes de l'initiation tumorale.

Le laboratoire a collaboré avec Rhône Poulenc RORER (désormais Aventis) à la recherche d'inhibiteurs de l'interaction entre p53 et MDM2, ainsi qu'à la caractérisation des liens entre la signalisation par la voie Ras et le facteur de transcription Net.

L'étude de l'expression de MDM2, et de ses variants d'épissage alternatif, dans les cancers VADS a été entreprise conjointement par les équipes du Dr Joseph Abecassis et du Dr Bohdan Wasylyk.

Lorsque les Drs Bruno Tocqué, Laurent Bracco et Fabien Shweighoffer ont quitté Rhône Poulenc RORER pour fonder la compagnie *ExonHit Therapeutics SA*, les collaborations entre les laboratoires du Dr Wasylyk, du Dr Abecassis et *Exonhit Therapeutics*, ont été poursuivies par l'étude à large échelle du transcriptome des tumeurs VADS.

Dans le cadre de cette étude, j'ai été recruté en tant qu'ingénieur thésard, titulaire d'une bourse CIFRE. Le Dr Julia Young a rejoint le « projet DD » quelques mois plus tard. Christine Wasylyk a initialement effectué un cycle d'analyses en DD (« 1 round ») avant d'être remplacée dans le projet par Anne Cromer, qui a effectué son DEA sur le projet DD, et, qui a progressivement pris en charge l'étude par puces à ADN Affymetrix. Chunhua Zhao a utilisé une sélection de clones DD, validés avec des sondes complexes focalisées « DD » (voir partie III D; publication DD), et, étudié leur expression dans les tumeurs de la prostate. Le Dr Alberto Zambrano poursuit actuellement l'étude d'un gène surexprimé dans les tumeurs de la

prostate. Le Dr Senghua Hao a effectué une analyse préliminaire par DD de la chimiorésistance des tumeurs VADS. Après avoir soutenu son doctorat sur l'étude de MDM2 au laboratoire, le Dr Gitali Ganguli a rejoint le projet DD en reprenant l'étude fonctionnelle des clones favoris du Dr Young. Dans le cadre du traitement informatique des données (DD et Affymetrix), Annaick Carles a également été recrutée.

## **b. Contexte général**

Dans le cadre de ce projet, nous nous sommes proposé d'étudier les événements moléculaires permettant de mieux caractériser les cancers VADS et leur évolution. A l'issue d'un traitement par chirurgie, on observe des reprises métastatiques ou des seconds cancers d'une part, ou, une rémission d'autre part, sans disposer d'une évaluation pronostique satisfaisante à l'heure actuelle. Le but de ce projet est d'obtenir des marqueurs moléculaires à valeur pronostique, permettant une adaptation éventuelle du protocole thérapeutique. La confirmation d'un rôle physiologique de certains de ces marqueurs pourrait déboucher sur le développement de nouvelles stratégies thérapeutiques (drogues interagissant avec les voies de signalisation impliquées, thérapie génique, etc...). La carcinogenèse est associée à de nombreuses modifications de l'expression génique, aussi bien quantitatives (surexpression d'oncogènes ou répression de gènes suppresseurs de tumeur) que qualitatives (phénomènes d'épissage alternatif de p53 dans les cancers VADS). *Exonhit Therapeutics S.A.* (Paris) a donc mis en œuvre sa technique brevetée D.A.T.A.S. (Differential Analysis of Transcripts with Alternative Splicing) afin de détecter de manière systématique les phénomènes d'épissage alternatif dans les cancers VADS. Les groupes de recherche du Dr Wasyluk à l'I.G.B.M.C. (Illkirch) et du Dr Abecassis au Centre Paul Strauss (Strasbourg) ont appliqué la méthode du Differential Display sur les mêmes échantillons, afin de détecter des variations de niveau d'expression des gènes. Les profils quantitatifs et qualitatifs d'expression obtenus devraient permettre une meilleure caractérisation des tumeurs des cancers VADS et une meilleure compréhension de l'évolution tumorale. Parmi les différents types de tumeurs VADS, notre choix s'est porté sur les tumeurs de l'hypopharynx dont la fréquence importante nous a permis de mener une étude sur des groupes homogènes de tumeurs. En outre, ces tumeurs de mauvais pronostic présentent un volume relativement important, permettant d'obtenir des quantités d'ARNm compatibles avec les besoins de notre étude. Le traitement par chirurgie de ces

tumeurs nous donne accès à des biopsies, représentant l'histoire naturelle de la tumeur. Des traitements adjuvants par radiothérapie ne sont appliqués qu'après l'exérèse.

Notre étude vise à l'exhaustivité maximale pour constituer une collection importante de marqueurs. En cela, elle se distingue catégoriquement d'un projet académique de DD, consistant à isoler rapidement quelques gènes impliqués dans le cancer et leur étude fonctionnelle dans la foulée (kit de DD de 8 amorces puis sélection de quelques cibles, voir étude en DD sur des lignées cellulaires, de manière à faciliter les expériences de validation). Cette approche allonge considérablement les étapes de criblage et n'assure pas forcément un rendement maximal en terme de publications.

## **2. Stratégie expérimentale de l'étude en DD**

Trois groupes de patients présentant des tumeurs bien caractérisées et homogènes au niveau histologique (classification TNM), dont le suivi de l'évolution de la maladie était supérieur à 20 mois, ont été sélectionnés par le Dr Aecassis au Centre Paul Strauss. Une quantité suffisante de matériel tumoral, mais aussi de tissu sain de l'utérus, devait être disponible pour effectuer toutes les validations nécessaires (sélection des clones, Northern Blot, Virtual Northern blot, Macro-arrays, PCR quantitative, etc...), soit 60 µg d'ARN totaux.

- groupe E « Early »: patients présentant des tumeurs de petite taille peu différenciées (T1, T2), sans envahissement ganglionnaire, prélevées par chirurgie à un stade précoce de développement du cancer. Ces échantillons permettent la mise en évidence de modifications impliquées dans les étapes du développement des cancers VADS.

- groupe NM « non métastatiques »: patients présentant des tumeurs de taille moyenne (T2, T3), de différenciation moyenne, n'ayant donné lieu après traitement chirurgical ni à des extensions métastatiques, ni à des récurrences après une période de suivi de 32 mois minimum. Ce groupe avait été initialement désigné comme S « Stable » du fait de la survie des patients, mais cette dénomination a été changée dans notre article du fait d'une confusion possible avec les phénomènes d'instabilité génétique. Toutefois lorsque nous parlons des échantillons normaux correspondant à ce groupe, il me paraît plus aisé de parler de NS plutôt que de NNM.



- groupe M « Métastatiques »: patients présentant des tumeurs histologiquement indiscernables des tumeurs des patients du groupe NM, mais ayant développé des métastases comme premier signe d'évolution au cours de la période de suivi de 32 mois. Ce groupe a été désigné initialement comme U « Unstable ». L'appellation NM est ambiguë, aussi je lui préférerai NU.

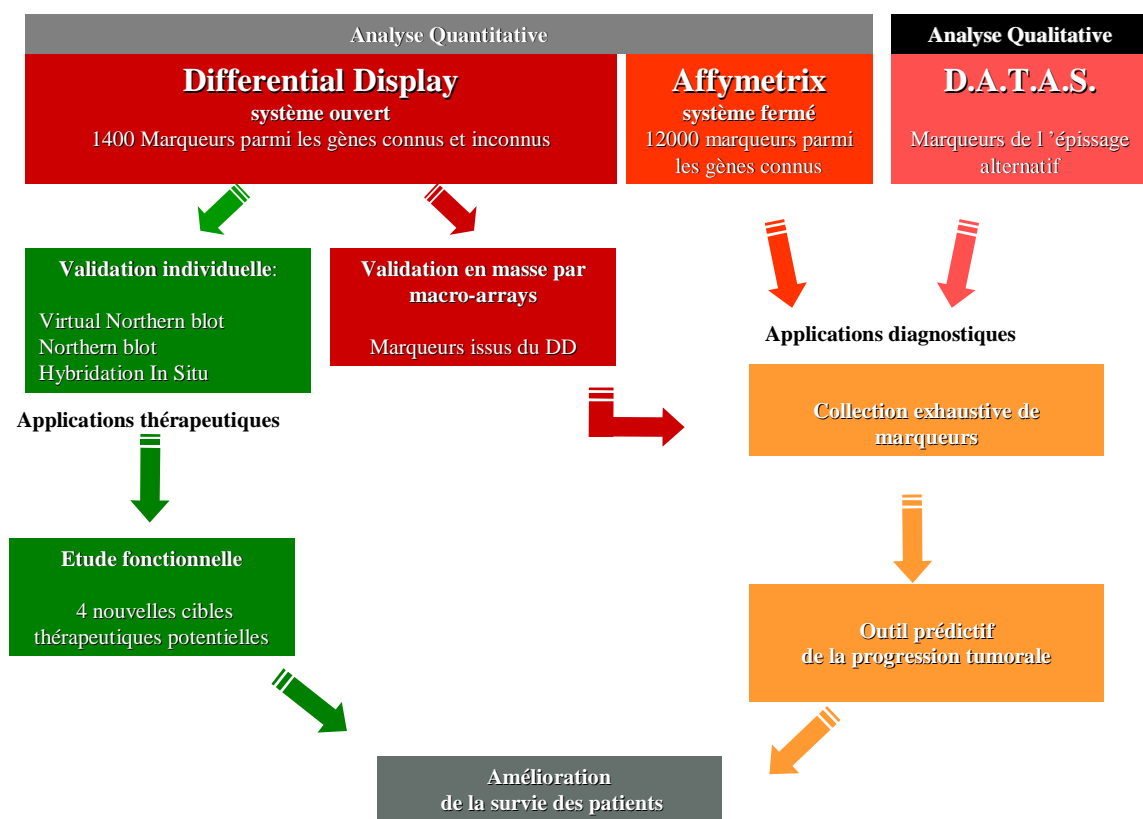
La comparaison des groupes M et NM doit permettre de corrélérer les différences de profil d'expression au pronostic tandis que la comparaison des tumeurs, en particulier précoces, avec les tissus normaux correspondants, doit permettre d'identifier de nouveaux acteurs de la carcinogenèse. Les ARNms ont été traités à la DNaseI de manière à éviter une contamination par l'ADN génomique et toutes les réactions de DD ont été réalisées en duplicat de manière à diminuer le taux de faux positifs de l'étude. Avant de commencer l'étude par DD, nous avons constitué des "pools" en regroupant les ARNms des patients présentant les mêmes caractéristiques. L'usage de pools nous permet de réduire le nombre de manipulations à effectuer, tout en évitant d'isoler des variations peu fréquentes, présentes chez un seul individu, qui introduiraient un biais dans les résultats. Pour nous affranchir des variations liées au fond génétique des individus, sans lien évident avec la carcinogenèse, nous avons comparé les divers pools (E, NM, M) avec des pools des ARNms extraits des tissus normaux (luette) des mêmes patients. Cette étude est menée sur des tumeurs non micro-disséquées, en envisageant la tumeur dans sa globalité, comme un mélange complexe de différents types cellulaires. La compréhension des profils d'expression doit donc être complétée, pour des gènes d'intérêt, par la technique d'hybridation *In Situ*.

Nous avons décidé d'effectuer l'analyse en DD (description détaillée en partie III) avec 58 amorces arbitraires HAP (42 IGBMC / 16 Centre Paul Strauss) en combinaison avec les 3 amorces HT11 (3 "rounds" de DD). Une telle combinaison d'amorces, assure théoriquement une couverture de plus de 90% des ARNms exprimés par la cellule. La sensibilité de la technique permet de repérer des événements correspondant à des gènes exprimés à divers niveaux. En particulier, elle permet, par rapport à des techniques soustractives, de repérer des gènes surexprimés à un très faible niveau tels des gènes clés dans le contrôle de la biologie cellulaire comme les facteurs de transcription ou les molécules des voies de signalisation. Il est également possible d'isoler, de manière marginale, des variations qualitatives d'expression des gènes : la présence ou l'absence d'une séquence dans un ADNc épissé de façon alternative peut déterminer l'amplification d'une bande sur gel de DD. La

qualité des échantillons est strictement examinée de manière à éviter toute dégradation des échantillons source éventuelle des faux positifs.

Nous avons développé une plateforme de validation secondaire en masse par la technique du Reverse Northern blot en combinaison avec divers types de sondes complexes (technologie SMART, marquage focalisé avec les amorces de DD). Une série de techniques de validation individuelles -Northern Blot, Virtual Northern blot et PCR quantitative- ont été employées.

### 3. Design expérimental global : complémentarité avec l'étude par puces à ADN Affymetrix



Stratégie expérimentale globale pour l'étude des cancers VADS

Figure II.1

Notre étude met en œuvre plusieurs techniques complémentaires d'analyse du transcriptome des cancers VADS (FigII.1). L'analyse quantitative est réalisée, tant en système ouvert en DD, avec la potentialité d'isoler des marqueurs parmi les gènes connus comme inconnus, qu'en système fermé avec utilisation de puces à ADN Affymetrix, assurant une

couverture complémentaire sur des gènes connus. *ExonHit Therapeutics* assure, avec sa technique DATAS, une analyse des variations qualitatives de l'expression génique de ces tumeurs. La combinaison des marqueurs isolés par ces diverses méthodes devrait permettre d'établir une collection exhaustive des altérations de l'expression génique associées à l'initiation et à la progression de ces tumeurs. Le but à terme est la constitution d'une membrane ou d'une puce regroupant toutes les signatures spécifiques des cancers VADS et son utilisation diagnostique et pronostique. De manière parallèle, un nombre restreint de gènes, cibles thérapeutiques potentielles, sera caractérisé fonctionnellement.

#### **4. Adaptation aux contraintes expérimentales du post-Differential Display**

L'étape la plus souvent limitante d'une étude en DD, en terme de quantité d'ARN et de quantité de travail, est le Post-Differential Display, l'ensemble des processus de validation, tant en masse qu'au niveau des gènes individuels, situés en aval de l'isolement de bandes différentielles sur gel. Nous avons adopté plusieurs techniques allant de l'analyse en Reverse Northern blot de clones sur membrane de nylon à la synthèse de sondes complexes, mettant en jeu divers procédés d'amplification.

#### **5. Traitement des échantillons : Choix entre microdissection ou analyse de la tumeur dans sa globalité**

Dans certains cancers, comme dans le cancer de la prostate, le développement parallèle de plusieurs lésions tumorales, repérables au niveau histologique au sein d'une même glande, peut rendre nécessaire d'isoler chacune des lésions d'origine clonale, de grade et d'agressivité différentes. Dans certaines biopsies de cancer du sein, le compartiment des cellules pré-malignes et malignes ne représente que 5% de la masse totale (Simone et al., 1998). La culture à court terme *in vitro* des cellules cancéreuses est une méthode d'enrichissement de l'échantillon mais elle peut donner lieu à une adaptation aux conditions de culture et une accumulation supplémentaire d'altérations génétiques.

Du fait de la nature hétérotypique des tumeurs, des techniques ont été développées de façon à isoler, microdisséquer les diverses composantes des tumeurs. Le but de cette microdissection est l'enrichissement en cellules cancéreuses préalable aux études de profilage, ou, l'isolement de gènes liés au recrutement des cellules normales, comme les cellules

endothéliales intra tumorales, en vue du développement de drogues les prenant spécifiquement pour cible.

Introduites en 1996, les techniques de microdissection au laser {Laser Capture Microdissection LCM ; (Emmert-Buck et al., 1996)} permettent d'isoler des groupes homogènes de cellules à partir d'un tissu hétérogène. L'étude de la contribution dans la progression tumorale des seules cellules cancéreuses mais aussi des divers types cellulaires du stroma devient possible. Un des problèmes majeurs de ces techniques est la faible quantité d'ARN extraite à partir de lames histologiques. L'utilisation de plusieurs tumeurs issues de patients différents peut être nécessaire, surtout pour les lésions précoces de petite taille. Des techniques d'amplification de l'ARN peuvent être utilisées avec des risques de distorsion du profil d'expression. En effet, pour toute technique mettant en jeu une réaction de transcription inverse, l'utilisation d'une quantité très faible d'ARN peut affecter la détection des transcrits rares. Nous avons pris le parti dans notre étude de généralement réaliser les réactions de transcription inverse avec d'ARN quantités supérieures (1µg) au minimum requis par la technique, et, d'ensuite seulement diluer les ADNcs obtenus de manière à éviter la perte des transcrits rares lors de la transcription inverse initiale. Une autre limitation historique est une éventuelle dégradation de l'ARN (Mikulowska-Mennis et al., 2002) ou surtout des protéines (Craven & Banks, 2001) pour les études protéomiques lors de la procédure de microdissection, du fait de l'échauffement provoqué par le découpage au laser.

Comme présenté dans notre article, les tumeurs hypopharyngées analysées contiennent plus de 70% de cellules cancéreuses, et la microdissection ne paraît donc pas nécessaire. Cette technique n'étant pas déjà implantée en Alsace lors de l'initiation de notre étude, nous avons donc choisi d'étudier les tumeurs dans leur globalité. De plus, les cellules stromales, non cancéreuses, apportent une contribution bien documentée au développement tumoral. Le DD, détectant les ARNms rares, nous permet de détecter les transcrits spécifiques de chaque compartiment tumoral même s'ils sont dilués dans la tumeur. Enfin, les analyses cliniques ne mettent à ce jour pas en oeuvre une microdissection systématique des tumeurs, et, il est donc important que les marqueurs tumoraux identifiés lors de notre étude présentent un profil d'expression suffisamment fort pour être détectable sur les tumeurs entières. En absence de microdissection, il est nécessaire de recourir à l'hybridation *in situ* ou à l'immunohistochimie pour affiner la caractérisation du profil d'expression.

**Partie III**  
**Etude des tumeurs VADS par la technique**  
**du Differential Display : première**  
**publication et résultats généraux**

## A. La technique du Differential Display

---

### 1. Principe

Comme présenté sommairement plus haut, la technique du Differential Display (DD) est basée sur la subdivision de la population totale des ARNms en sous populations, par amplification par PCR avec des amorces arbitraires, jusqu'à obtention de bandes discrètes sur gel de séquençage. Le profil d'expression est directement lisible par comparaison des profils de bandes obtenus. Le concept du DD intègre 3 des techniques les plus puissantes de la biologie moléculaire que sont la RT-PCR, l'électrophorèse sur gel de séquençage et le clonage de l'ADN. La simplicité du protocole, jusqu'à obtention de différences observables sur gel, contribue au succès et à la robustesse de cette technique.

Le matériel de laboratoire nécessaire à sa mise en oeuvre, sa capacité à isoler de nouveaux gènes de fonction inconnue et ses faibles contraintes en quantité d'ARN (utilisation de la PCR) ont contribué à son large usage. Cette technique peut être utilisée avec un nombre restreint d'amorces de PCR, pour isoler rapidement quelques séquences nouvelles régulées dans un processus précis, ou, elle peut être utilisée avec un nombre important d'amorces, afin de mener une étude exhaustive sur l'ensemble du transcriptome.

La technique du Differential Display (DD, FigIII.1) est une des plus anciennes techniques ouvertes d'analyse du transcriptome (Liang & Pardee, 1992). Dix ans après sa publication, cette technique a contribué à plus de 3'800 études, ce qui représente plus de publications que pour les techniques de micro-arrays, d'hybridation soustractive, de SAGE et RDA réunies (Liang, 2002). Cette forte utilisation du DD atteste de sa capacité à isoler des gènes différenciellement exprimés.

Après traitement à la DNaseI, éliminant tout ADN génomique contaminant les préparations d'ARN, une réaction de transcription inverse par l'enzyme MMLV (Moloney Murine Leukemia Virus) est réalisée avec une amorce Poly-d(11)T, complémentaire des queues poly-A présentes en 3' des ARNm (polyadénylation), se terminant par une des bases A, G ou C. On parle d'amorces ancrées. Ces amorces portent, en sus de cette séquence, un site

de digestion pour l'enzyme HindIII en 5'. Initialement ces sites étaient utilisés pour cloner les ADNc. Depuis les fragments sont clonés directement dans des vecteurs du type pGEM-T, mettant à profit l'ajout fréquent d'une base A par la Taq Polymérase lors des réactions subséquentes d'amplification par PCR. On parle d'amorce HT11A, HT11G et HT11C, H codant pour HindIII. Ces sites introduits par incorporation dans les amorces de PCR peuvent toutefois être utilisés pour vérifier la taille et la présence d'inserts dans le vecteur de clonage. A l'utilisation de chaque amorce HT11 (ancrée par un A, un G ou un C) correspond un « round » de DD, et la sélection d'un tiers des ARNm.

Après transcription inverse, une nouvelle sélection d'une sous population des ADNc va être effectuée en réalisant une amplification par PCR avec une amorce arbitraire HAP (HindIII arbitrary primer), capable typiquement d'amplifier 50 à 100 transcrits. Lors de la PCR, le nucléotide d'ATP radioactif ( $P^{33}$ ) va être incorporé dans l'ADNc double brin et permettre la visualisation de bandes discrètes, après migration sur gel de poly-acrylamide dénaturant similaire aux gels de séquençage. Le  $P^{33}$  est utilisé plutôt que le  $P^{32}$  de façon à générer des bandes très finement résolues sur gel, clairement définies après séchage et exposition sur film. L'affinité de l'amorce HAP pour un transcrit particulier dépend de sa séquence, et, est relativement indépendante de l'abondance absolue, du niveau d'expression du transcrit. Le biais introduit par l'utilisation d'amorces HAP permet de détecter des séquences exprimées à faible niveau, demeurées inconnues à ce jour. Ces transcrits rares peuvent coder pour des protéines régulatrices clés des processus cellulaires. Les bandes produites sur gel sont comparées d'un échantillon à l'autre et les bandes différenciellement représentées, qui doivent typiquement être de l'ordre de 5 à 10% du total des bandes sur gel, sont découpées, éluées du gel, ré amplifiées puis sous clonées afin de subir une validation secondaire. Le taux de bandes différentielles permet de mener une analyse critique précoce du design expérimental. Une étape de sous clonage est nécessaire puisque plusieurs fragments peuvent migrer à la même taille. Il faut donc isoler les clones possédant une séquence expliquant le profil observé avec une seconde étape de criblage, qui confirme le profil DD. Les clones sont disposés sur membrane de nylon permettant une hybridation avec des sondes complexes. Les diverses modalités de constitution de membranes (pousse bactérienne, dépôt de fragments amplifiés en PCR) et de synthèse de sondes que nous avons dû mettre en oeuvre seront décrits plus bas.

Des variations sur le thème du DD comprennent des techniques assurant la subdivision des ARNm par digestion enzymatique (RFLP-DD, Genecalling, TOGA citées plus haut) ou

encore des DD « ciblés ». L'utilisation d'amorces spécifiques d'une famille de gènes permet de cibler leur expression. Cependant les amorces doivent être allongées car les amorces HAP sont sélectionnées pour leur faible sélectivité. Des stratégies ont été conçues de façon à isoler les parties N terminales des protéines proches des séquences Kozak en amont des promoteurs. L'utilisation d'amorces homologues des séquences d'adressage (séquence correspondant au peptide signal) permet de focaliser l'étude sur les protéines sécrétées {Revue par (Liang, 2002)}.

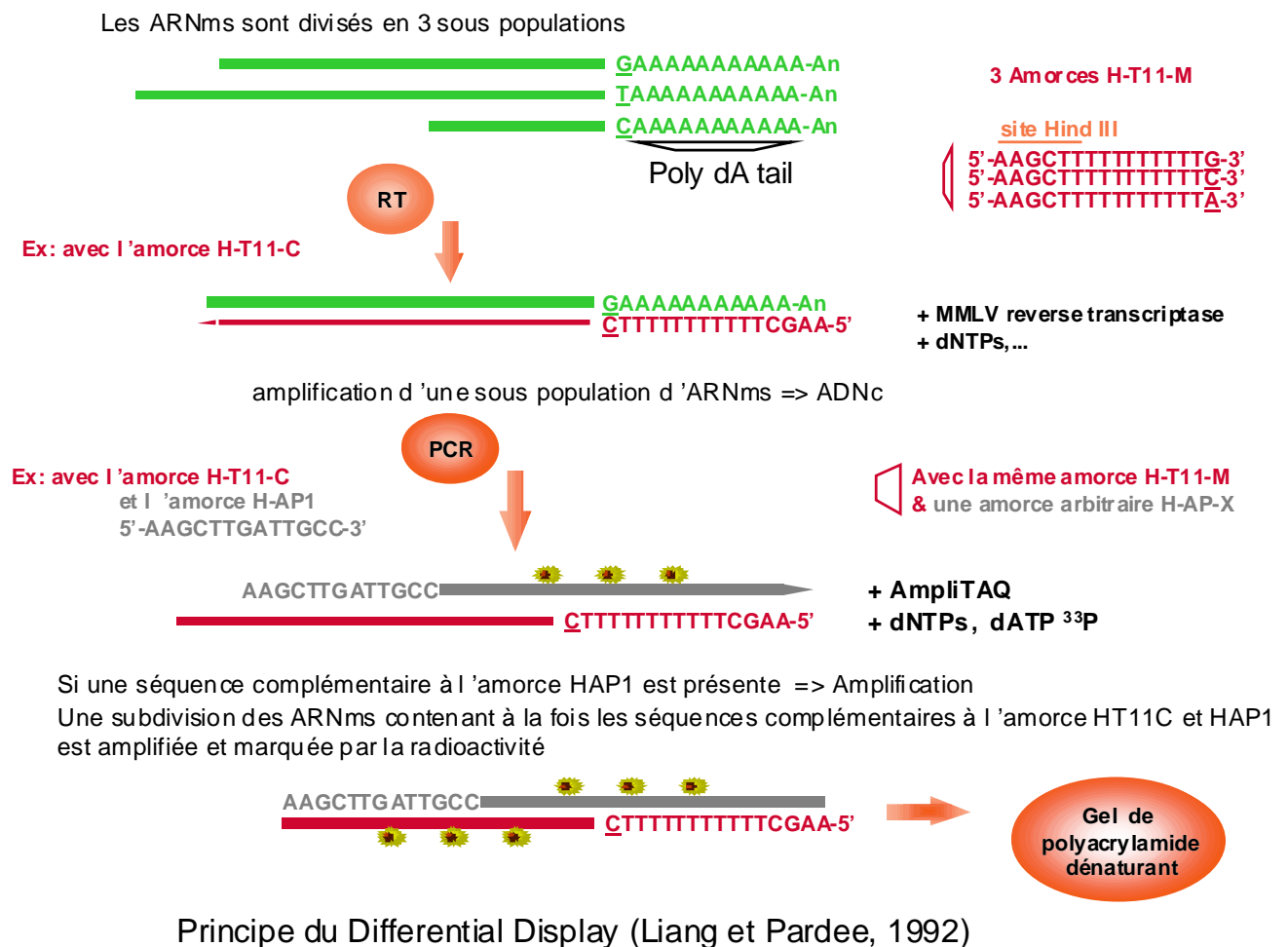


Figure III.1

## 2. Facteurs assurant le succès d'une étude en DD

Un certain nombre d'améliorations, de mises au point et de conseils permettent d'utiliser la technique du DD avec succès (Zhao et al., 1995; Liang, 1998).



### **a. Amorces PolyT**

Historiquement les amorces poly-(11)dT étaient ancrées par deux bases, les amorces ancrées par une seule base ont depuis démontré une meilleure reproductibilité (Liang et al., 1994). De plus les amorces HT11, plus longues et ancrées par une seule base, réduisent la redondance de l'amorçage, la production d'un bruit de fond (« smear ») et le nombre de réactions de transcription inverse, de 12 à 3 cycles (« rounds ») de DD, ce qui diminue donc la quantité d'ARN nécessaire au protocole. Des erreurs d'amorçage avec les amorces poly-dT ancrées par deux bases peuvent réduire l'efficacité du DD (Frost & Guggenheim, 1999).

### **b. Amorces HAP**

La taille optimale des amorces HAP est déterminée d'après des considérations statistiques. De manière à reconnaître 50 à 100 ARNms, les amorces doivent hybrider en tant que 6 à 8-mères. En pratique, les amorces de taille inférieure à 9 paires de bases ne produisent pas d'amplification par PCR du fait de la nécessité d'une surface de contact minimum entre l'ADN double brin et la Taq polymérase. Bien que des amorces de 10 paires de bases aient initialement été utilisées, elles ont généré des signaux faibles, un nombre faible de bandes et une faible reproductibilité. Ces limitations sont critiques dans des conditions sub-optimales liées à la qualité des ARNs, des amorces, du lot de transcriptase inverse, de dCTP\*, etc... Les amorces ne s'hybrident essentiellement qu'en tant que 6 à 9 mères en fonction du contenu en GC, sur les bases situées en extrémité 3'. Des considérations théoriques indiquent que la sélection de 50 à 100 ARNms est assurée par 7 nucléotides, que l'amorce doit mesurer au minimum 9 bases, et, que 8 à 9 nucléotides apportent une stabilité optimum. Des amorces trop longues nuisent à la redondance et à la reproductibilité des amplifications (dégénérescence). Des amorces de 20 bases peuvent être utilisées, mais elles favorisent l'amplification très spécifique d'un nombre restreint de gènes. Or le but du DD est de couvrir un nombre important de gènes par amorce HAP. En conséquence, les amorces ont donc été rallongées d'une taille minimum de 9 bases à une taille optimum 13 bases. Les 7 dernières bases, assurant la reconnaissance des ARNms, sont dessinées de manière à être les plus différentes possibles au sein des amorces HAP. Les 5 premières bases sont fixes et ne contribuent que peu à la sélectivité (1,5 paire de base participe à la reconnaissance en moyenne). 8 à 9 bases participent effectivement à l'hybridation. Une séquence de 7 bases représente 1/16'000 des séquences possibles. Sur une séquence de 600 paires de bases typiquement résolue sur gel

d'acrylamide à 6%, on trouve 600 sites possibles de reconnaissance pour tous les 7-mères possibles. Chaque amorce a donc une probabilité de 4% de détecter un ARNm donné. On peut donc déduire le nombre d'amorces HAP à utiliser pour assurer un certain niveau de couverture du transcriptome en DD (Liang & Pardee, 1992).

### **c. Influence de la stratégie expérimentale**

Le taux de faux positifs, considéré comme traditionnellement élevé, dépend en fait de nombreuses petites variations sur les différents matériels (enzymes, tubes, machines, etc...), du type de situations comparées, du design expérimental et de l'expérience de l'expérimentateur. En particulier, si les échantillons comparés sont très différents, comme des échantillons de foie et de cerveau, le taux de faux positifs sera faible. Lorsque les deux situations comparées sont très proches, comme lors de la comparaison de cellules traitées par une drogue avec les mêmes cellules non traitées, le taux de faux positifs est élevé. Peu de gènes sont régulés par la drogue et l'expérimentateur doit souvent découper des bandes plus faiblement différentielles. Effectuer les réactions de DD en duplicat permet de repérer les différences reproductibles, ne dépendant pas de variations infimes lors des réactions de PCR ou des pipetages, et, diminue drastiquement le taux de faux positifs. La reproductibilité des profils obtenus en duplicat est une indication rapide de la qualité de l'expérience.

Un autre facteur important est l'existence de contrôles internes à l'expérimentation. L'induction effective de l'expression de gènes par un traitement doit être effectivement vérifiée indépendamment du DD sur une cible connue du traitement. Dans le profilage du transcriptome des tumeurs, il importe d'avoir accès aux tissus normaux des mêmes patients de manière à s'affranchir des variations individuelles, et, de constater parmi les séquences isolées en DD la présence de gènes précédemment associés au cancer.

Le design expérimental est crucial. Entre autres, il est essentiel de prévoir les méthodes de validation (Northern blot, Hybridation *In Situ*), leur besoin en ARNm (utilisation de techniques d'amplification des ARNms) et leur étendue (Reverse Northern blot en masse ou gène individuel).

## **3. Une comparaison systématique de quelques techniques ouvertes**

Wan et *al.* (Wan et al., 1996) ont effectué une comparaison précise des divers systèmes ouverts d'analyse du transcriptome en étudiant les gènes induits par l'interféron

gamma dans une lignée cellulaire. Ils ont comparé les techniques les plus communément utilisées de Soustraction Electronique (Electronic Substraction, ES), d'Hybridation Soustractive (Subtractive Hybridization, SH) et de Differential Display (DD).

L'ES consiste à constituer des banques d'ADNc, à les séquencer et à comparer l'abondance relative de chaque ADNc au sein des diverses banques par analyse informatique. L'ES possède l'avantage de produire des données informatiques réutilisables. Plusieurs milliers d'ADNcs doivent être séquencés et les transcrits les plus abondants sont préférentiellement isolés. De plus, tous les ADNcs isolés ne correspondent pas à des séquences altérées, entraînant une perte de détection de la majorité de la sous population des ADNcs différenciellement exprimés. Pour détecter à un niveau de probabilité de 95% un ARNm exprimé différenciellement par un facteur 3 fois à une prévalence de 1/20'000, il faut assurer le séquençage d'environ 126'000 ADNcs dans chaque situation expérimentale. Pour assurer une analyse quasi-exhaustive des ARNm (prévalence <1/70'000), un nombre de clones supérieur à 441'000 doit être séquencé pour chaque situation expérimentale. Bien que la méthode SAGE allège le procédé en produisant de petits fragments concaténés, séquençables en grand nombre par plasmide et réduisant le nombre de séquençages à effectuer, les techniques de soustraction électronique requièrent un nombre élevé de réactions de séquençage.

L'hybridation soustractive (Subtractive Hybridization, SH) élimine une grande partie des efforts de séquençage en concentrant l'analyse sur la population des ADNcs spécifiques d'une situation A contre une situation B, en éliminant par hybridation en phase liquide des séquences communes. L'ADNc de la situation testée, appelé « tester », est hybridé en excès (> 20 fois) avec les ARNms du contrôle, appelé « driver ». Les duplex doubles brins ADNc/ARNm sont éliminés après hybridation par rétention sur des colonnes d'hydroxyapatite ou par interaction biotine-streptavidine. La population de molécules simple brin est alors utilisée pour la génération d'une sonde radioactive marquée qui servira à cribler des banques d'ADNc, ou, pour la construction de banques de séquences soustraites. Au cours de ce procédé de SH les séquences de membres de familles de gènes possédant une forte homologie peuvent être soustraites de manière erronée par une séquence apparentée. Certains gènes surexprimés peuvent être perdus, cette technique détectant mieux les différences qualitatives de type absence/présence des transcrits. Wan et ses collègues (Wan et al., 1996) n'ont détecté qu'une sous population très restreinte des séquences régulées par l'interféron

gamma, en accord avec les propriétés de cette technique dans la littérature. Une analyse de la sensibilité de la technique a toutefois démontré sa capacité à isoler des ARNm différenciellement exprimés à une faible prévalence comparée à l'ES. Cette propriété particulière s'est reflétée dans le faible recoupement entre les ADNcs isolés avec ces deux techniques. L'isolement à de nombreuses reprises de mêmes séquences (redondance) est bien corrélé avec la qualité de la séquence en tant que vrai positif. Ce constat permet de cibler les meilleurs clones, à valider par les techniques de validation consommant des quantités importantes d'ARNm (Northern blot). Ainsi est il plus probable d'isoler par cette technique des ADNcs exprimés de manière abondante que des ADNcs exprimés à une faible prévalence, qui restent néanmoins détectables.

Avec la technique du Differential Display (DD), on isole aussi bien les augmentations que les diminutions d'expression et, dans une certaine mesure, on peut isoler des événements d'épissage alternatif. La transition de la bande de DD sur gel à la séquence finale confirmée pour son expression différentielle est assez longue et délicate. Certaines publications, précédant celle de Wan *et al.* (Wan *et al.*, 1996), avaient mis en doute la capacité du DD d'isoler des séquences exprimées à faible niveau (Bertioli *et al.*, 1995; Debouck, 1995). Du fait d'une transcription inverse à partir de la queue poly-A, la séquence codante complète est rarement clonée, et, les séquences peuvent ne pas avoir d'homologie avec les séquences connues, étant dans la partie 3' non traduite. Les principaux obstacles à la mise en oeuvre de la technique de DD sont la longueur et le besoin en ARN des étapes de validation.

Wan (Wan *et al.*, 1996) n'a trouvé que très peu de gènes isolés en commun entre le DD et l'ES, et, entre le DD et la SH. Ces données indiquent que beaucoup de gènes régulés par l'IFN gamma restent à découvrir et que chaque technique a ses biais propres. En totale opposition avec Bertioli *et al.*, qui avaient mené une analyse sur des ARNm introduits de manière artificielle (« spiking ») dans des populations d'ARNm et conclut à une capacité du DD à détecter uniquement les ARNm très abondants (prévalence  $>1/100$ ), et, en accord avec ses principes, le DD est capable d'isoler des séquences correspondant à des transcrits faiblement exprimés (prévalence moyenne  $1/20'000$  mais pouvant aller jusqu'à  $1/200'000$ ). Cependant, certaines séquences abondamment exprimées détectées en SH et ES ne sont pas identifiées en DD. Les clones identifiés en DD dépendent en effet avant tout de l'affinité des amorces arbitraires utilisées pour la séquence, plus que de la prévalence du transcrit. Le DD est donc une méthode de choix pour l'identification de séquences différenciellement exprimées demeurées inconnues à ce jour. Une redondance assez importante a été observée,

en rapport avec l'hybridation éventuelle d'une amorce avec plusieurs régions d'un même transcrit, ou, de plusieurs amorces sur différentes régions d'un même transcrit. Les mésappariements au cours des premiers cycles de PCR peuvent également expliquer cette redondance. Même si les étapes de confirmation sont longues et contraignantes, un des avantages du DD est l'utilisation possible de quelques microgrammes d'ARNs totaux contre la même quantité d'ARNs poly-A pour les autres techniques. L'usage de la PCR permet de mener des études complètes avec des échantillons biologiques tels que les biopsies tumorales. Wan *et al.* (Wan et al., 1996) conseillent l'usage du DD par rapport à l'ES et la SH du fait de son moindre besoin en ARN (ARN totaux contre ARN poly-A), de sa capacité à isoler les transcrits indépendamment de leur prévalence (du fait des propriétés intrinsèques des amorces arbitraires), de sa détection des ARN sur- et sous exprimés en une seule comparaison, du coût limité des matériels nécessaires à sa mise en œuvre, et, de l'évaluation précoce, dès les gels de DD, du succès de la procédure (on doit typiquement observer environ 5% de bandes différenciellement exprimées). De plus, Wan et Al. (Wan et al., 1996), ont démontré un taux de faux positifs moindre qu'avec les deux autres méthodes.

En résumant la littérature, il apparaît que les diverses techniques ouvertes d'analyse du transcriptome possèdent leurs limites, leurs avantages et leurs biais propres. L'expérimentateur doit sélectionner une des approches compatibles avec son étude, en fonction de son design expérimental. S'agissant des prélèvements tumoraux des patients, nous devons faire face à une quantité d'ARN très faible. Nous tentons de plus d'isoler des acteurs encore inconnus de la carcinogenèse avec une couverture vaste des transcrits qu'ils soient rares ou abondants. Nous effectuons des comparaisons multiples entre des tumeurs de stade clinique divers et leurs tissus normaux correspondants. Les micro-arrays Affymetrix n'étaient de plus que peu accessibles (coût élevé et installation des infrastructures à l'IGBMC un an après le début du projet DD) lors du début de notre étude. La technique du DD était donc la plus adaptée à notre étude.

#### **4. Modifications du DD et intégration de plateformes de validation**

Il est possible d'effectuer un séquençage direct des bandes éluées du gel de DD puis ré-amplifiées; les ADNcs étant simples brins dans le gel dénaturant et en faible quantité (Buess et al., 1997; Wang & Feuerstein, 1995). Toutefois, une bande de DD pouvant contenir plusieurs ADNcs (Callard et al., 1994 ; Welsh et al., 1992) migrant à la même position, une

une étape de sous clonage est généralement nécessaire (FigIII.2). Dans l'optique des analyses en masse, les progrès principaux ont porté sur l'utilisation combinée de la technique du Reverse Northern Blot avec le DD (Wang & Feuerstein, 1995), et, le développement de sondes complexes amplifiées

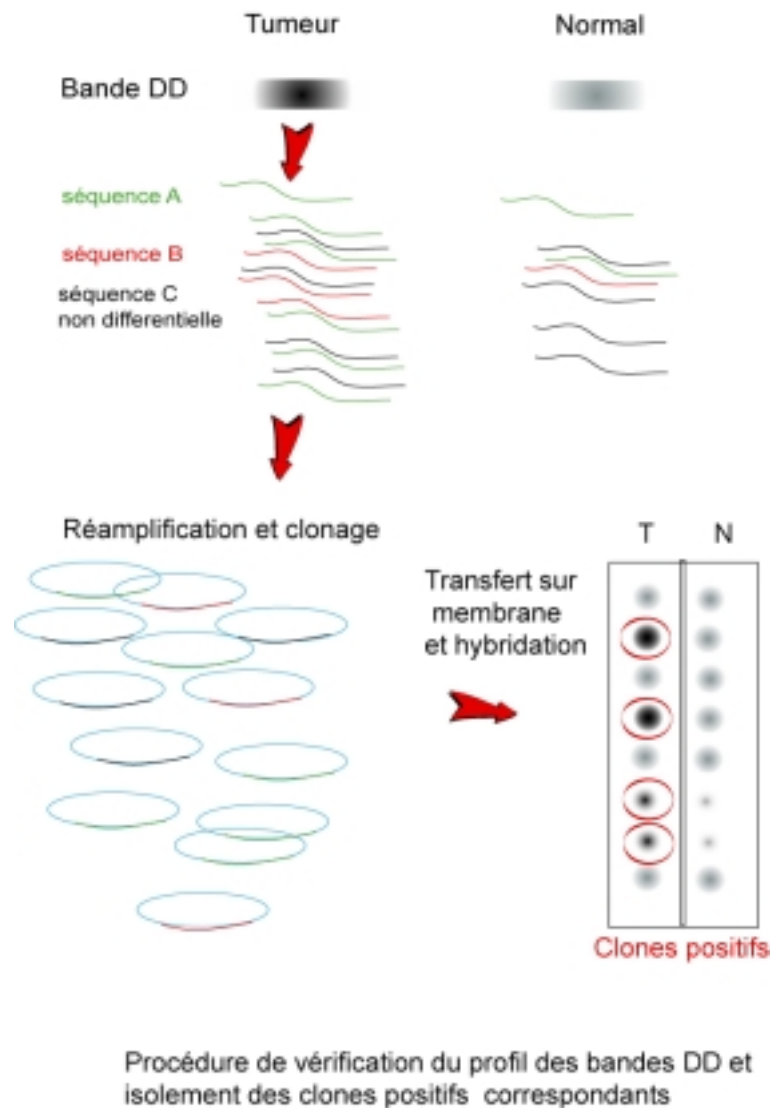


Figure III.2

### a. Analyse en Reverse Northern blot

La construction de macro-arrays (Martin et al., 1998) ou de micro-arrays a été rapidement combinée au DD de manière à effectuer, avec une quantité limitée d'ARN, un second criblage de validation sur l'ensemble des clones, exigé par le taux de faux positif

associé à cette technique,. Cette analyse en masse peut être combinée à la synthèse de sondes complexes amplifiées (Poirier & Erlander, 1998).

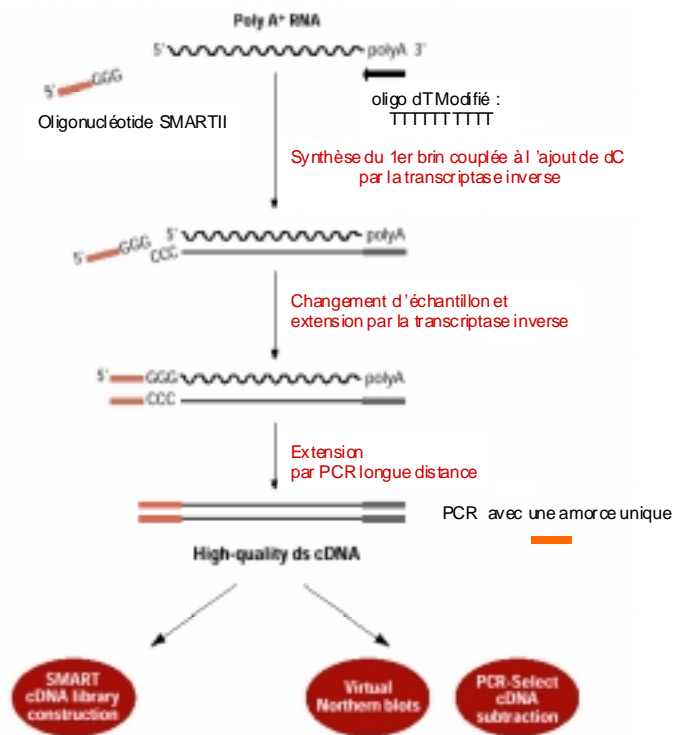
## **b. Synthèse de sondes**

### **α. Sondes linéaires**

La technique des ARNs amplifiés a permis d'obtenir une amplification linéaire des ARNs par utilisation de la transcriptase inverse avec des d'amorces portant le site consensus de l'enzyme T7 ARN polymérase et synthèse *in vitro* d'ARN amplifié (ARNa) à partir des ADNcs (Van Gelder et al., 1990). L'ARNa peut être à nouveau converti en ADNc radioactif, résultant en une meilleure sensibilité, une meilleure stabilité de la sonde et un bruit de fond plus réduit que pour les ARNa radioactifs (Poirier & Erlander, 1998). Dans certaines de ces techniques, lors de la synthèse du premier brin d'ADNc une boucle peut être formée, permettant la poursuite de la transcription inverse, sa propagation vers l'amorce poly-dT et l'intégration de la séquence complémentaire à l'amorce Poly -dT (Poirier et al., 1997). Il est alors possible de réaliser des cycles de PCR après digestion de la boucle, en n'utilisant que cette amorce ou une amorce spécifique correspondant à la partie 5' (non poly-dT) de cette amorce. A chaque cycle de PCR, la matrice simple brin est copiée. L'absence d'une deuxième séquence sélectionnée par un seconde amorce empêche l'amplification exponentielle, traditionnellement obtenue en PCR. Affymetrix a poursuivi le développement de sondes amplifiées linéairement pour l'analyse de ces puces à ADN.

### **β. Sondes linéaires basées sur la technologie SMART**

La technologie SMART (FigIII.3) permet en mettant à profit l'ajout de CCC par la transcriptase inverse à l'extrémité du simple brin d'ADNc de générer un second brin par changement de brin de l'enzyme suite à l'ajout d'une amorce SMARTII portant une extrémité GGG. Après synthèse du deuxième brin une amorce unique, incorporée à l'amorce poly dT et à l'amorce SMARTII, permet de réaliser une amplification qui demeure linéaire dans les limites de la phase exponentielle de PCR.



Principe de la technologie SMART (d'après Clontech)

Figure III.3

### γ. Sondes complexes utilisant les amorces de DD

Un des problèmes principaux du DD, est la vérification de l'expression de séquences exprimées à très faible nombre de copies. En effet, les capacités intrinsèques des amorces permettent d'isoler des clones correspondant à des transcrits rares. Les sondes complexes traditionnellement synthétisées pour les arrays ne reproduisent pas ce phénomène et les transcrits rares représentent donc une très faible fraction du marquage. De manière à reproduire la focalisation du marquage par affinité particulière des amorces de DD, les amorces de DD ont été utilisées pour synthétiser les sondes (Buess et al., 1997 ; Trenkle et al., 1999). De cette manière, même si des phénomènes de compétition existent lors de la PCR, le marquage est focalisé sur les séquences correspondant aux clones d'intérêt isolés en DD.

Certains auteurs synthétisent une sonde complexe avec un mélange d'amorces spécifiques des gènes présents sur la membrane (Martin et al., 1997) de manière similaire aux sondes proposées dans le kit commercial Atlas (Clontech).



Comme exposé plus loin nous avons testé et parfois combiné les divers types de synthèse de sondes existantes, de manière à obtenir divers marquages avec divers biais comme un biais sur les gènes fortement exprimés avec des sondes « random priming » ou à l'inverse, une focalisation du marquage sur les clones isolés en DD, exprimés à faible niveau.

## B. Résultats globaux du DD, face cachée et perspectives

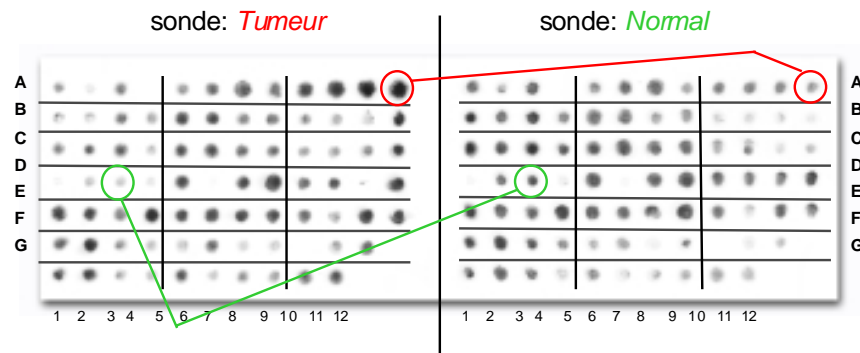
---

Notre première publication établit une liste finale de clones différentiellement exprimés, dans le cadre de l'utilisation d'un protocole de marquage focalisé (mélange d'amorces HAP) sur les transcrits isolés dans notre étude. Cependant, avant de parvenir à cette liste, de nombreuses adaptations techniques aux contraintes expérimentales ont été testées au cours de la longue phase de validation secondaire du DD. Un grand nombre de clones ont démontré une expression différentielle dans l'une ou l'autre de ces conditions. De plus, une analyse de gènes individuels s'est faite par les techniques de Virtual Northern blot (VN) et de Northern blot (NB) en tenant compte de 3 critères : leur profil d'expression en bande DD, leur expression différentielle lors d'au moins une étape de validation préliminaire en Reverse Northern blot (RN) et le caractère de nouveauté de leur séquence. L'étude en DD a généré bien plus de clones positifs que la seule liste présentée dans la première publication. Toutefois tous les clones positifs n'ont pas été isolés avec le même type de sondes ou n'ont pas présenté une reproductibilité suffisante pour être inclus à cette publication ; sans doute du fait existe-t-il des phénomènes de compétition entre les amorces lors de la synthèse de la sonde complexe « DD ». Un nombre important de gènes validés sur les patients individuels est encore en cours de validation en masse avec d'autres types de sondes. Certains gènes favoris sont en cours de caractérisation fonctionnelle (Parties V à VII).

### **1. Premier criblage en Reverse Northern blot sur colonies bactériennes**

La première méthode de validation en masse mise en oeuvre a été une technique de Reverse Northern blot sur colonies bactériennes. Après un essai initial de dépôt manuel sur membrane, clone par clone (1<sup>er</sup> round de DD), les clones obtenus à l'issue du DD ont été disposés dans des plaques de culture de 96 puits, et, ont été transférés à l'aide d'un répliqueur sur une membrane de nylon en contact avec du milieu LB agar ampicilline. Après une nuit de culture, les colonies bactériennes ont été lysées et l'ADN a été fixé à la membrane à l'aide de solutions dénaturantes et fixatrices, à l'issue d'une étape de fixation aux UV. Ces blots d'ADN ont été utilisés pour effectuer des hybridations en Reverse Northern Blot dont le but principal est de vérifier le sous clonage des séquences expliquant le profil des bandes d'intérêt. Toutefois, il existe un risque d'hybridation croisée avec l'ADN bactérien et la forme

des colonies n'est pas toujours identique, rendant toute quantification difficile. L'appréciation des différences doit être visuelle (FigIII.4).



Exemple :  
Reverse Northern blot sur pousses bactériennes

Figure III.4

Sur les membranes constituées par des colonies bactériennes, nous avons effectué un premier criblage avec deux types de sondes. Une sélection de clones positifs au cours de ce criblage a été utilisée pour validation en Northern blot. La vérification des profils différentiels nous a confirmé la validité de notre méthodologie.

#### a. Sondes linéaires

Nous avons réalisé des hybridations avec une sonde linéaire correspondant aux échantillons E et le pool de tissus normaux correspondants NE (n=3). Ces sondes produisent un signal plus faible et permettent le marquage d'une fraction des clones, sans doute correspondant aux ARNms exprimés à un niveau élevé.

L'analyse des divers pools {E (Early), S (Stable), U (Unstable)}, des patients individuels intégrés à ces pools et de patients supplémentaires est en cours d'analyse par une technique de Reverse Northern blot haute densité, avec des sondes linéaires de patients individuels (mises au point chez *ExonHit Therapeutics* d'après le protocole utilisé pour notre synthèse de sondes linéaires). L'analyse du premier round de DD par cette technique durant les phases initiales de l'étude avait démontré une expression différentielle de 17% de nos clones entre tissu normal et tissu tumoral pour chacun des 7 patients testés individuellement.

Si on considère les différences d'expression retrouvées chez au moins 3 patients, 37% de nos clones étaient confirmés par cette méthode.

### **b. Sondes de type DD**

Nous avons également utilisé les sondes SMART-DD, mettant en jeu le marquage de produit de SMART par l'ensemble des amorces de chaque expérimentateur (14 amorces) pour un pool des tissus tumoraux T contre le pool des tissus normaux correspondants N, attendu qu'environ 80% des bandes sur gel de DD présentaient des différences d'expression transversales qu'il s'agisse des groupes T ou N.

### **c. Résultat global du premier criblage**

A partir de 1'750 bandes, 14'000 clones ont été analysés. Environ 2'900 clones, présentant une expression différentielle avec l'un ou l'autre type de sondes, ont été séquencés. Ils correspondent à environ 900 séquences uniques (FigIII.5, premier criblage). Le Dr Young et moi-même avons chacun effectué 3 cycles de DD pour 14 amorces HAP. Christine Wasyluk, Anne Cromer et Chunhua Zhao ont effectué un des 3 cycles de DD sur 14 amorces supplémentaires. 3 cycles de DD ont été effectués par divers expérimentateurs pour 16 amorces au Centre Paul Strauss. J'ai personnellement découpé 597 bandes (environ 1/3 de l'ensemble de l'étude IGBMC et CPS réunis) en DD qui ont produit environ 240 séquences uniques après le premier criblage en RN.

## **2. Seconde série de criblages en RN sur fragments de PCR**

Après avoir effectué un premier criblage complet sur colonies bactériennes, nous avons entrepris un second criblage de validation par Reverse Northern blot sur fragment de PCR. Ces fragments sont déposés sur les membranes de nylon par un système d'aspiration par pompe à vide adaptée à la plaque 96 puits. La surface des puits est toujours identique et le dépôt est uniforme. Le bruit de fond est plus faible, du fait de l'absence de contaminants potentiels d'origine bactérienne et des séquences du vecteur de clonage, ne correspondant pas à l'insert. Une quantification des résultats est alors possible. L'aspect de quelques spots de ces Reverse Northern blots, mieux définis, est présenté dans notre publication.

## a. Choix des sondes

Lors de la mise au point de la production des diverses sondes, j'ai exploré de nombreuses méthodes :

- Sondes par marquage direct de l'ADNc par random priming:

Ce type de marquage ne permet la détection que d'un nombre très restreint de clones, correspondant a priori aux gènes exprimés à un fort niveau. Le même type de marquage avec de l'ADNc amplifié par la méthode SMART n'améliore que peu la détection des clones. Le problème majeur de cette technique est la dilution du marquage sur l'ensemble des transcrits, dont seulement une faible fraction correspond aux clones disposés sur la membrane. De plus, le marquage n'est pas concentré sur l'extrémité 3' des ARNms. Or les séquences isolées en DD présentent généralement une taille située entre 50 et 600 paires de bases à partir de la queue poly-A des ARNms.

- Sondes linéaires :

J'ai tenté, au cours de ce second criblage, de synthétiser des sondes pour les pool NM et M de façon à pouvoir analyser l'ensemble des pools avec les sondes linéaires E, NE, S, et U. La réaction de transcription inverse a été effectuée à partir de 200ng d'ARN. Toutefois je n'ai pas obtenu un « smear » de bonne qualité lors de la première série d'amplifications linéaires. En ajoutant une série de cycles de PCR linéaire, j'ai pu détecter l'amplification de transcrits. A l'issue d'un troisième cycle de ré amplification, j'ai obtenu un « smear » large, correspondant à l'amplification d'une large collection d'ADNcs de taille variable. L'utilisation de ces stocks amplifiés suivis d'un marquage par random priming sur une première série de clones n'a entraîné le marquage que d'un seul clone, non différentiel entre les pools S et U. Ce clone contient peut être une séquence complémentaire de l'amorce utilisée pour les ré amplifications ou correspond peut être à une séquence extrêmement abondante. Dans le cadre de l'étude de la valeur pronostique de nos clones, la synthèse de sondes linéaires pour 50 patients en triplicats a été réalisée par *ExonHit Therapeutics*, tout comme le dépôt robotisé des fragments de PCR sur membrane haute densité. Les hybridations ont été réalisées et sont actuellement en cours d'analyse au Centre Paul Strauss. Au niveau de notre groupe, la priorité a été donnée aux sondes de type DD sur les pools T et N, en vue d'une première publication.

- Sondes de type DD:

Pour toutes ces sondes le marquage est assuré par une PCR mettant en jeu les amorces utilisées en DD, de façon à focaliser le marquage sur les clones d'intérêt isolés par cette technique. Cette sonde biaisée possède un biais « positif », qui nous permet de détecter les clones exprimés à faible niveau. La plupart des tests effectués sont de type tumeur contre normal du fait de la prépondérance de ce type de différences.

Un nombre assez important de clones n'est pas détecté de manière systématique avec ces sondes, dont des clones pourtant validés sur VN ou Northern blot ; ce résultat peut traduire des phénomènes de compétition entre les amorces, introduisant une absence de marquage de certains transcrits exprimés à faible niveau lors de étapes initiales de la PCR. Les différences, lorsqu'elles sont détectables, vont cependant généralement dans le même sens. La compétition entre amorces s'exprime peut être par le fait que les divers expérimentateurs, utilisant des lots d'amorces différents, ont connu un succès variable lors de cette étape de validation. Les clones les plus fortement différentiels comme LPRP sont systématiquement isolés, tandis que les différences plus subtiles ne sont pas aussi reproductibles. Pour le premier criblage, tous les clones ont été hybridés au moins 2 fois avec la sonde DD « standard » décrite dans notre publication.

J'ai testé des sondes T/N de type DD, amplifiées pendant seulement 15 cycles de PCR-DD contre 30 en conditions standards, dans le but d'analyser une éventuelle distorsion introduite par les cycles de PCR-DD. Ces sondes sont en fait beaucoup plus faiblement marquées que les sondes standard, et n'apparaissent pas assez sensibles pour effectuer une bonne analyse.

Après avoir effectué une validation T/N des clones de DD, une tentative d'analyse au niveau des pools, E, S, U et le pool N avec les sondes DD standard a été menée. Quelques clones comme XL3b.6 (secreted frizzled related protein 1), ont produit des signaux différentiels avec ces sondes, mais certaines différences attendues d'après les VN et les Northern blots n'ont pas été systématiquement retrouvées. Les bandes spécifiques d'un pool tumoral correspondent à une fraction plus réduite des gènes, à des différences plus subtiles que les différences T/N. Leur perte par cette technique de validation est importante.

## **b. Utilisation de sondes DD de moindre complexité**

Le taux de validation variable des clones par la sonde DD standard, complexe du fait de l'utilisation de 14 amorces simultanément, nous a amené à tester l'effet de la réduction du nombre d'amorces sur la sensibilité des sondes. Le problème majeur d'une sonde de complexité réduite, synthétisée à partir d'un nombre restreint d'amorces, est la restriction de son utilisation aux clones de DD isolés avec les amorces concernées.

J'ai effectué un test de sondes T/N synthétisées avec chacune des 3 fractions de mes 14 amorces totales sur un seul cycle (1 HT11). Les sondes générées par fractionnement des amorces HAP ont présenté un plus fort marquage et ont permis de détecter des différences concordant avec le profil DD, perdues avec la sonde DD standard. Certains clones ont également été marqués de manière inattendue, sans doute du fait de phénomènes d'amplifications entre amorces générant une amplification de fragments plus complexe que la réaction de DD. Il s'agit d'une focalisation relative. De son côté, Anne Cromer, dont le lot d'amorces semble être victime d'une forte compétition entre les amorces HAP, a constaté que l'analyse amorce par amorce améliorerait la confirmation de ses clones. Sur ces bases, un second criblage avec une sonde DD amorce spécifique a été décidé de façon à assurer la meilleure exhaustivité possible de la collection de clones isolés, dans la perspective de la construction d'un outil prédictif. Dans ces conditions, on reproduit en fait les conditions du DD et on s'assure du clonage des fragments expliquant le profil observé sur gel.

Ceci a impliqué le regroupement de tous les clones issus de bandes isolées avec chacune des amorces HAP, de façon à limiter le nombre de tubes à hybridation et la quantité de sonde à synthétiser. Tous ces tests secondaires ont été réalisés sur des macro-arrays de fragments de PCR. Il a donc fallu réorganiser un nombre très important de plaques bactériennes en plaques de dilution spécifiques de chaque amorce HAP, pour chacun des cycles de DD. Les clones du 1<sup>er</sup> cycle ont dû être réorganisés par amorce HAP, après avoir déjà été réorganisés en plaques 96 puits par leur ordre chronologique de clonage pour les divers tests sur fragments de PCR à l'issue du criblage initial.

Après ce second criblage amorce par amorce, les clones positifs ont été rassemblés sur de nouvelles plaques 96 puits. Les membranes correspondantes ont été analysées de nouveau avec la sonde DD standard plusieurs fois. Le « bruit de fond » des clones n'étant pas détectables avec une sonde complexe a été éliminé. Les clones ayant présenté une expression différentielle avec l'une ou l'autre sonde ont été inclus dans une collection de clones positifs.

A l'issue de ce second criblage, nous avons isolé 1'400 séquences apparemment uniques, quelques clones ne devant être séquencés qu'en cas de la démonstration de leur valeur pronostique.

Outre la confirmation de nombreux clones déjà sélectionnés lors du premier criblage, ce second criblage m'a personnellement amené à isoler 356 clones supplémentaires dont 207 ont présenté un profil différentiel avec les sondes DD standard et ont été inclus dans nos macro-arrays finaux. Pour ma part, le second criblage a apporté un élargissement de ma collection de gènes et l'isolement de gènes contrôles comme EGR1 non détectés lors du premier criblage. Parmi les 207 clones différentiels avec les sondes standard DD, les 100 clones les plus différentiels ont été séquencés et identifiés par Blast. Ont ainsi été isolées 27 séquences inconnues uniques, et 36 gènes connus uniques. Dix sept de ces gènes connus avaient déjà été isolés lors du premier criblage, et, 19 gènes connus supplémentaires ont été isolés. Ce second criblage a donc permis de confirmer de nombreuses séquences du premier criblage (même clone non repiqué des plaques de glycérol lors de cette seconde analyse ou clone différent, contenant une séquence identique), et de compléter notre collection de signatures des cancers VADS.

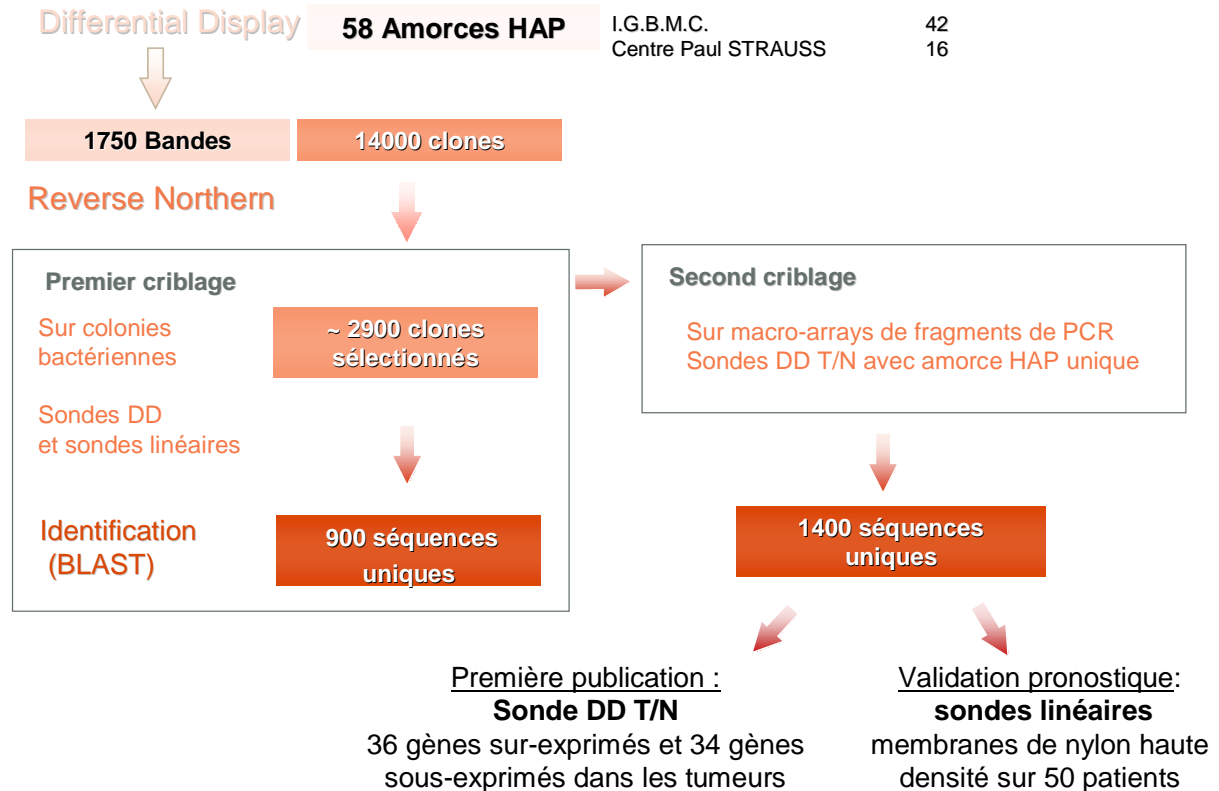
### **3. Etablissement d'arrays finaux pour la première publication**

Seuls les clones les plus fortement différentiels ont été séquencés, de manière à éliminer la redondance sur des macro-arrays finaux. Ces puces rassemblent ultimement les clones positifs uniques de chaque expérimentateur révélés par sondes complexes.

Ces macro-arrays finaux ont été hybridé de multiples fois et quantifiés. Seuls les clones présentant une différence d'expression supérieure à 2 fois, retrouvés de manière reproductible ont été sélectionnés comme différentiels. La liste des 70 gènes différentiellement exprimés est présentée dans notre publication.

J'ai tenté de détecter les clones d'intérêt pronostique sur ces arrays finaux avec des sondes DD issues des pools S et U. Cependant les clones isolés sur ces macro-arrays finaux ont été sélectionnés pour leur différence entre T et N. Les différences plus subtiles entre pools tumoraux sont fortement contre sélectionnées par cette approche. Peu de différences S/U ont été obtenues et l'analyse pronostique sera menée par une autre méthode d'array haute densité avec des sondes linéaires issues de 50 patients en triplicat.





Etapes multiples de validation en Reverse Northern blot

Figure III.5

#### 4. Validation de gènes individuels

Parallèlement à la validation secondaire en masse des clones (RN), les clones positifs avec au moins l'un des types de sondes ont été séquencés. Les séquences-inserts DD correspondant à des séquences inconnues ou à des gènes inconnus dans les cancers VADS ont été analysés individuellement par les techniques de Virtual Northern blot et/ou de Northern blot dont quelques exemples sont présentés (Fig III.6).

##### a. Validation par la technique de Northern blot

La validation par la technique de Northern blot constitue la preuve la plus directe de l'expression différentielle d'un ARNm. En outre, les tissus des patients testés sont différents de ceux utilisés lors du protocole de DD.

Trente quatre candidats « favoris » ont été testés par Northern Blot au Centre Paul Strauss. L'expression différentielle de 22 clones a été confirmée. Huit clones présentent une forte différence d'expression. Quatorze clones ont montré une surexpression dans les tissus tumoraux et 8 clones ont montré une sous expression dans les tissus tumoraux. Pour 11 clones, aucun signal n'a pu être détecté. Ces derniers clones peuvent être des artefacts issus du DD. Des artefacts, faux positifs de DD, devraient cependant générer un signal bien que non différentiel au travers des divers échantillons. Le fait qu'aucun signal ne puisse être détecté peut signifier que nous avons isolé des ARN exprimés à très faible niveau, indétectables en Northern Blot. Un certain nombre de candidats ont été testés au Centre Paul Strauss en RT-PCR quantitative pour notre première publication et les clones ne donnant pas de signal en Northern Blot, représentant une valeur pronostique potentielle seront testés. Quelques clones validés en Northern Blot pourraient être utiles pour l'évaluation du pronostic, en dépit du faible nombre de patients testés. Certains candidats sont des gènes connus, parfois déjà associés au potentiel métastatique.

### **b. Validation par la technique de Virtual Northern Blot**

La technique des « Virtual Northern blot » {VN, (Franz et al., 1999a)} est une technique similaire au Northern blot qui au lieu d'une quantité importante d'ARN déposée sur une membrane, utilise des ADNcs amplifiés synthétisés à partir d'une moindre quantité d'ARN initial. En effet, la quantité d'ARN nécessaire à la fabrication d'un Northern permet la fabrication d'environ 150 membranes de Virtual Northern blot, grâce à l'amplification linéaire en phase exponentielle de PCR des ADNcs par la méthode SMART (Endege et al., 1999 ; Matz et al., 1999 ; Zhu et al., 2001; Zhumabayeva et al., 2001). Après constitution des membranes au Centre Paul Strauss, cette méthode nous a donc permis d'effectuer une validation secondaire pour un grand nombre de gènes, en respectant les contraintes en ARN propres au design expérimental. Nous avons vérifié la validité de cette méthode en comparant le profil en Northern blot et en Virtual Northern blot de plusieurs clones (LPRP, SOD2, CSRP2). Nous avons trouvé une bonne corrélation entre ces deux méthodes comme l'attestent certains clones (DRG1, PIGR, LPRP) présentés dans notre publication. Un autre intérêt de la technique de VN est d'autoriser l'analyse du profil d'expression au niveau des patients utilisés pour la constitution des pools de DD.

Les candidats par lesquels nous sommes particulièrement intéressés sont les protéines membranaires et les protéines sécrétées (application diagnostique et accessibilité), les enzymes (synthèse d'inhibiteurs), les facteurs de transcription, ainsi que les clones de fonction inconnue. Nous avons pu grâce à cette technique VN analyser l'expression d'environ 260 gènes. Cent quarante trois des gènes testés ont généré un signal sur VN, montrant pour 131 d'entre eux un profil différentiel. Quarante vingt clones ont montré une surexpression dans les tissus tumoraux, et, 28 candidats ont potentiellement un profil d'expression spécifique d'un type tumoral particulier. Soixante douze clones ont montré une répression dans les tissus tumoraux.

J'ai personnellement testé 96 clones en VN et obtenu la confirmation d'une expression différentielle pour 73 candidats. J'ai obtenu le plus fréquemment des répressions dans les tissus tumoraux, mais aussi des surexpressions dans tous les tissus tumoraux. J'ai préférentiellement testé des clones surexprimés dans les tumeurs de mauvais pronostic. Cinquante trois clones montrent la présence d'au moins une bande surexprimée dans un tissu tumoral. Parmi ces clones, 30 présentent une surexpression nette dans la majorité des tumeurs. Vingt clones présentent une surexpression dans les tumeurs Unstable (ou « Early et Unstable ») mais pas dans les tumeurs Stable, et fournissent donc un critère potentiel de distinction de ces deux types de tumeurs. L'expression est généralement faible et la validation de la valeur pronostique doit être poursuivie par d'autres techniques, sur une collection plus importante de patients. Trente cinq clones montrent une répression dans les tissus tumoraux uniquement. Dans un nombre significatif de cas (30 % dans mon cas, 56 % des cas de manière globale), aucun signal n'est obtenu. Comme pour la technique de Northern Blot, il peut s'agir de faux positifs de DD, mais également d'un problème de sensibilité de la technique. Il est également possible que lors de la réaction de reverse transcription initiale du procédé SMART, certains ARNs rares aient été perdus. La technique de RT-PCR semi quantitative peut s'appliquer à certains de ces candidats de profil prédictif et de fonction particulièrement intéressante.

Il faut noter que les hybridations sur Virtual Northern Blot présentent fréquemment des bandes multiples ou des profils complexes. On trouve assez souvent plusieurs bandes réprimées dans tous les tissus tumoraux, et, une ou plusieurs bandes de taille différente, correspondant au profil attendu, surexprimées dans les tissus tumoraux (« Unstable » ou « Unstable et Early »). Un autre profil fréquemment obtenu correspond à la détection d'une bande unique réprimée dans tous les tissus tumoraux sauf dans le cas d'au moins un patient Unstable pour lequel l'expression du tissu tumoral est supérieure à celle du tissu normal.

**Article British Journal of Cancer**

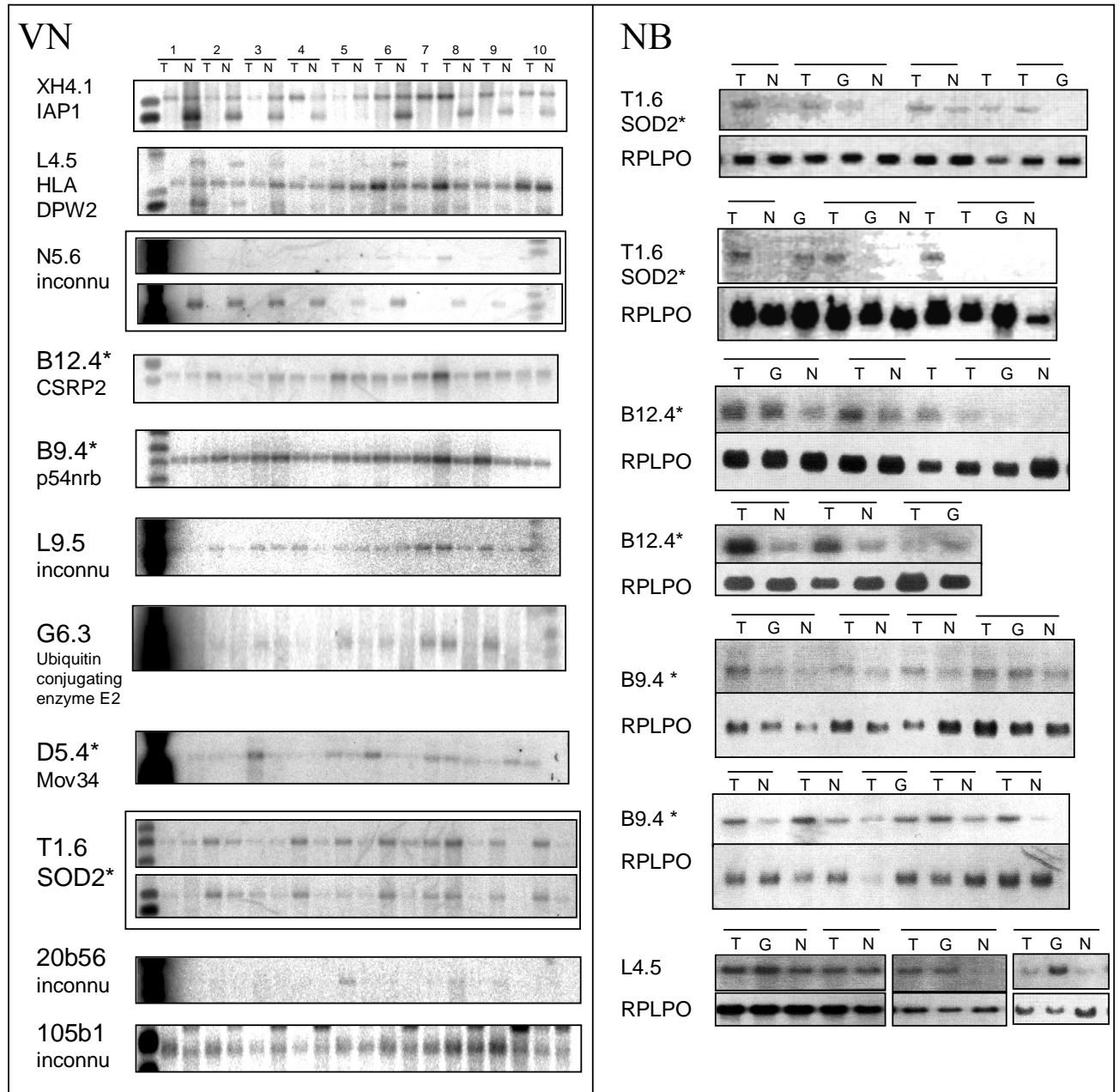
Northern blot: DRG1, APOL2°, PIGR\*, LPRP\* (4/4)

VN : HSPCB°, TRA1, PIGR, LPRP\* (4/6)

RT-PCR : DRG1, APOL2°, HSCP150, RP1-68D18, PIGR\*

**VN/NB positifs présentés dans cette thèse (clones favoris)**

HOP, FLJ10261, P5.1, 99b78°



**Exemples de Virtual Northern/Northern Blots différentiels**

issus de ma liste ou testés par mes soins parmi la liste de Christine Wasyluk (°); anciens clones favoris, clonés et analysés dans des lignées VADS(\*); les échantillons correspondant à un même patient sont regroupés sous une barre (T tumeur, G ganglion lymphatique, N tissu normal).

Figure III.6

Quelques clones ont été séquencés en dépit d'un profil sur Reverse Northern Blot sur colonies inversé par rapport au profil en Differential Display, en raison de l'importance de la différence observée. Ces clones ont donné le même profil en Virtual Northern Blot qu'en Reverse Northern Blot, ou un profil hybride avec bandes multiples, pouvant expliquer les profils détectés auparavant.

Les Virtual Northern blots, produits par le Centre Paul Strauss, ont été calibrés par rapport au gène RPLP0 (anciennement appelé 36b4), et, pour normaliser par rapport à ce gène les tissus normaux ont subi un nombre supérieur de cycles de PCR que les tissus tumoraux correspondants. Il est possible que le gène RPLP0 ne représente pas fidèlement le comportement de l'ensemble des gènes et que ce nombre supérieur de cycles de PCR (en phase exponentielle) dans les tissus normaux explique une surreprésentation de bandes supplémentaires dans les échantillons normaux par rapport aux profils attendus. Ces bandes supplémentaires peuvent également s'expliquer par une plus grande sensibilité du Virtual Northern Blot par rapport au Northern Blot, le matériel ayant été amplifié. Cette plus grande sensibilité pourrait mettre en évidence les transcrits rares issus de l'épissage alternatif ou encore produire des signaux correspondant à des gènes apparentés au gène testé.

Pour les candidats présentant un intérêt pronostic potentiel, je n'ai à ce jour isolé qu'un clone surexprimé dans tous les trois échantillons U présents sur la membrane (B14.4 décrit plus bas). Cela est cohérent avec le fait que même les meilleurs marqueurs de la carcinogenèse actuellement disponibles ne sont pas présents dans 100% des patients. Cela pose un problème pour la validation en Northern blot des marqueurs pronostiques potentiels, puisqu'une membrane ne comprend que rarement plus d'un ou deux patients de type U. Si notre marqueur n'est exprimé que chez 30% des patients, il est probable que sa valeur pronostique intérêt ne sera pas confirmée sur un premier Northern Blot de validation. Il est alors nécessaire de tester un nombre significatif (n=30) de patients par la technique de PCR quantitative en temps réel.

## **5. Analyse de l'expression de la collection de gènes DD dans 15 lignées cellulaires VADS**

Des micro-arrays sur lame de verre des fragments de PCR correspondant aux clones isolés en DD ont été constitués par le service commun de micro-array de l'IGBMC (Bernard Jost). Les contraintes de l'époque, en terme de quantité d'ARN nécessaires à l'utilisation de

ce type de puces, ne nous ont pas permis de réaliser une validation du DD avec des échantillons tumoraux. Le Dr Julia Young a donc assuré une analyse de ces puces avec les ARN extraits de 15 lignées cellulaires VADS.

Après analyse statistique préliminaire, nous avons obtenu un clustering des kératinocytes primaires HEK-a et des lignées spontanément transformées d'une part, et, des lignées tumorales, d'autre part. Cependant seulement un groupe très restreint de 7 gènes est responsable de ce clustering. Pour ma part, ces hybridations ont démontré une expression très forte (51 fois et 31 fois en expérimentation flip/flop par rapport à un pool contrôle de toutes les lignées analysées) du clone SMAP31/HOP, un de mes gènes favoris, dans les kératinocytes primaires HEK-a. Ces cellules, au potentiel prolifératif limité et nécessitant de nombreux facteurs de croissance, ne constituent pas le modèle adapté pour une étude fonctionnelle. J'ai rapidement analysé ces hybridations lorsque le Dr Young a quitté le projet. Le faible nombre de gènes associés au clustering obtenu, l'absence de distinction entre lignées cellulaires d'agressivité différente parmi les lignées tumorales, et, l'absence de distinction entre les lignées d'origine différente dans la sphère ORL ne nous ont pas incité à poursuivre plus loin l'analyse de ces hybridations.

## **6. Analyse de la valeur pronostique : macroarray haute densité**

De façon à analyser la valeur pronostique de notre collection de gènes, nous avons projeté une analyse en triplicat sur 50 patients atteints de cancers de l'hypopharynx. Nous avons ré-amplifié les fragments de PCR correspondant aux 1'200 clones uniques présentant l'expression la plus différentielle en reverse Northern blot et *ExonHit Therapeutics* a généré avec ces fragments 200 copies de 2 membranes de nylon déposées par robot (S1 et S2). Ces membranes ont été hybridées au Centre Paul Strauss en triplicat avec des sondes linéaires issues des 50 patients dans le but de valider la valeur pronostique éventuelle de certains clones issus de DD. L'analyse de ces hybridations est en cours au Centre Paul Strauss.

A terme, la combinaison des clones informatifs issus de DD et de l'analyse Affymetrix pourrait déboucher sur la constitution d'un macro-array ou d'un micro-array spécifique des cancers VADS, et plus particulièrement de l'hypopharynx. Cette membrane pourrait permettre d'évaluer le pronostique.

L'analyse de ces hybridations, réalisées par Régine Millon au Centre Paul Strauss, est en cours. Sur chacune des 2 membranes S1 et S2, 716 clones et 52 contrôles ont été disposés en double sur les 1536 spots disponibles. Un total de 1432 clones a pu être analysé.

Des sondes linéaires ont été synthétisées en triplicats indépendants pour chaque échantillon. Trente échantillons tumoraux (9 E, 10 S, 11 U) et 30 échantillons normaux (6 NE, 6 NS, 5 NU, 3 N macro disséqués) ont à ce jour été examinés.

La reproductibilité des doublets est bonne puisque 85% d'entre eux présentent une variation de moins de 10%, et seulement 2% présentent des variations d'expression supérieures à 25%.

Comme pouvait le laisser craindre l'usage d'une sonde linéaire, le problème principal de cette étude est un faible signal d'hybridation. Trois pourcent des clones présentent un niveau élevé d'expression, 20% présentent un niveau moyen d'expression et 77% des clones présentent un niveau d'expression très faible, délicat à analyser. Cette répartition du signal est quasiment superposable aux proportions relatives décrites dans la littérature des ARNm exprimés à haut, moyen et faible niveau (quelques copies/cellules) dans la cellule. En effet, 86% des transcrits sont exprimés à moins de 5 copies par cellule, et, ne représentent que 25% de la masse totale des ARNm de la cellule (Bishop et al., 1974; Zhang et al., 1997). Ceci semble confirmer que le DD assure bien une analyse du transcriptome de manière relativement indépendante du niveau absolu d'expression des transcrits.

Les résultats sont en cours d'analyse avec la technique SAM (Statistical Analysis of Microarray), le test paramétrique de Wilcoxon-adapté à l'analyse des gènes faiblement exprimés-, ainsi qu'avec un seuil minimum d'expression différentielle de 1,5 fois. Les analyses T/N semblent permettre l'identification de 172 gènes par la méthode SAM, réduits à 107 gènes après application du test de Wilcoxon, puis réduits à 47 gènes (5 surexprimés dans les tumeurs et 42 sous exprimés dans les tumeurs) après application du seuil minimum. Un premier essai de clustering avec les gènes sélectionnés par la méthode SAM indique que ces 172 gènes pourraient contribuer à la distinction des tissus normaux et tumoraux. Les résultats sont très préliminaires mais il semble qu'environ 18 gènes pourraient être prédictifs du type tumoral. Etant donné la faiblesse des signaux observés, l'utilisation de RT-PCR quantitative à large échelle est envisagée sur les clones permettant un clustering des échantillons ou possédant une valeur pronostique potentielle.

La force du Differential Display, sa capacité à détecter les transcrits différentiellement exprimés indépendamment de leur niveau absolu d'expression, est aussi sa grande faiblesse dans le contexte d'une validation en masse des clones, car les sondes complexes disponibles à l'heure actuelle ne sont en général pas assez sensibles. L'utilisation de techniques en cours de

développement, comme les oligodendrimères, pourrait lever cette limite. Une alternative pourrait consister à tester les clones (ou une sélection de clones) par PCR quantitative en temps réel en plaques 96 puits.

## **7. Test fonctionnel en masse :**

A l'issue du premier criblage sur bactéries de nos trois cycles de DD, nous avons ré-amplifié l'ensemble de nos clones et *ExonHit Therapeutics* a cloné l'ensemble de ces fragments dans un vecteur d'expression rétroviral pour effectuer des tests fonctionnels en masse (Shotgun functional test), de type test de résistance à l'apoptose. Malheureusement cette approche n'a pas donné de résultats concluants du fait d'une forte recombinaison du vecteur, impliquant un bruit de fond élevé dans les tests.

A partir de ce constat, nous avons favorisé l'analyse de gènes candidats sur une base individuelle.



## C. Base de données : collection DD dans toute son étendue

---

Notre étude du transcriptome des tumeurs VADS, tant en DD qu'avec la technologie Affymetrix, a généré un grand nombre de clones et de séquences. L'analyse de ces séquences doit être gérée rationnellement et organisée au sein de bases de données informatiques.

Après collecte des séquences isolées par chaque expérimentateur, la constitution d'une base de données a été initiée sous environnement Windows dans Microsoft Access par Carole KNIBBE, étudiante de l'INSA de Lyon. Les données ont ensuite été transférées sous environnement Unix, permettant la mise en oeuvre de recherches automatiques sur des bases de données locales mises à jour sur les bases de données publiques. La base de données et les outils regroupés au sein du programme Gscope, forment l'outil TRANSAT (Transcriptome Analysis Tool), développé par Annaïck CARLES et Raymond RIPP au sein de l'équipe d'Olivier POCH. Cet outil informatique intègre des recherches de localisation sur le génome, d'existence d'ARNm, d'EST et de protéine prédite. Enfin, la classification par geneOntology -permettant une analyse rapide des caractéristiques fonctionnelles potentielles des séquences- a été implémentée.

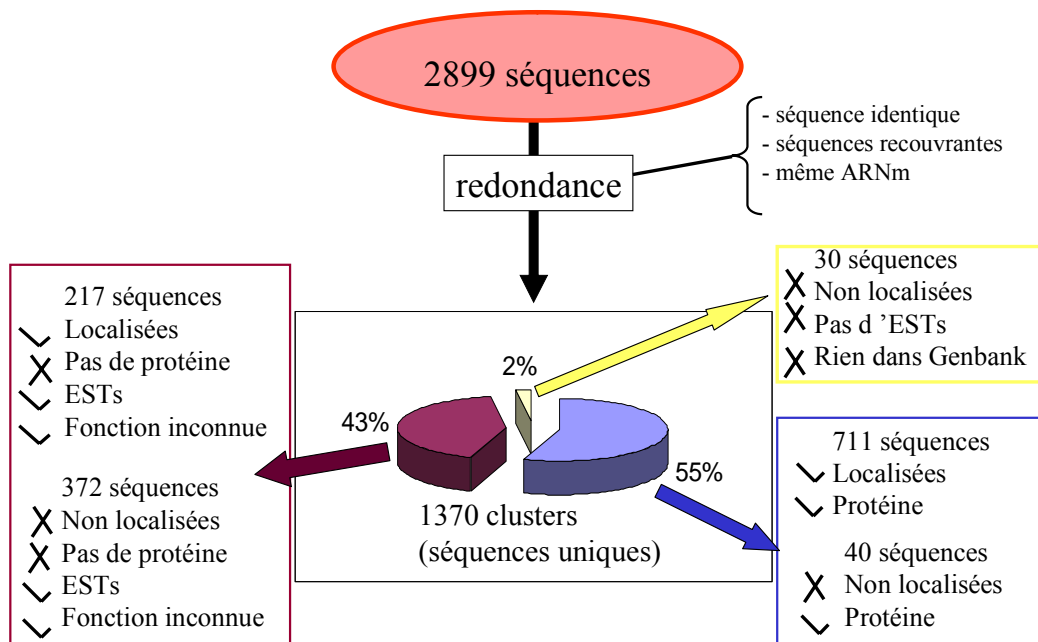
### 1. Identification des séquences

La localisation des séquences sur le génome a été déterminée d'après leur homologie avec les contigs. L'existence de séquences EST homologue, ou même d'ARNm complets, a été analysée. Au niveau fonctionnel, l'existence d'une protéine, et, éventuellement d'une fonction associée, a été recherchée de manière automatique (Fig III.7).

Une absence de localisation chromosomique pour une séquence de la base de données peut traduire plusieurs situations ; l'existence de plusieurs localisations possibles (familles de gènes, duplications géniques, présence de séquences répétées, domaines conservés, etc..) ; un score d'homologie faible (séquence courte) ; l'absence de localisation dans le génome au moment de l'analyse (artefact de recombinaison au sein de E Coli, zone mal séquencée ou ré-assemblée du génome, gène amplifié à partir d'un germe présent dans ou aux environs de la biopsie, séquences répétées masquées, etc...) ; un gène inconnu.

Une faible proportion des gènes isolés, soit 2%, correspond à des séquences totalement inconnues, soit complètement nouvelles, soit artéfactuelles. Pour 43% des gènes isolés, un EST existe confirmant leur expression dans les tissus humains tandis que 55% des gènes

isolés correspondent à une protéine prédite possédant de l'homologie avec une protéine présente dans les bases de données, comme illustré sur le schéma suivant (Figure III.7).



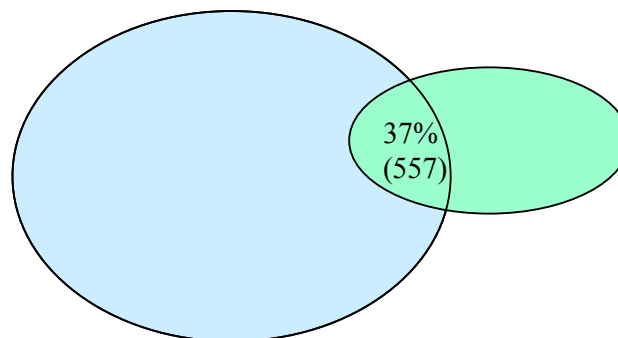
Analyse de la collection DD par Transat

Figure III.7

## 2. Confirmation du taux de couverture du génome

~12600 gènes Affymetrix  
Non sélectionnés

1516 gènes  
sélectionnés en DD



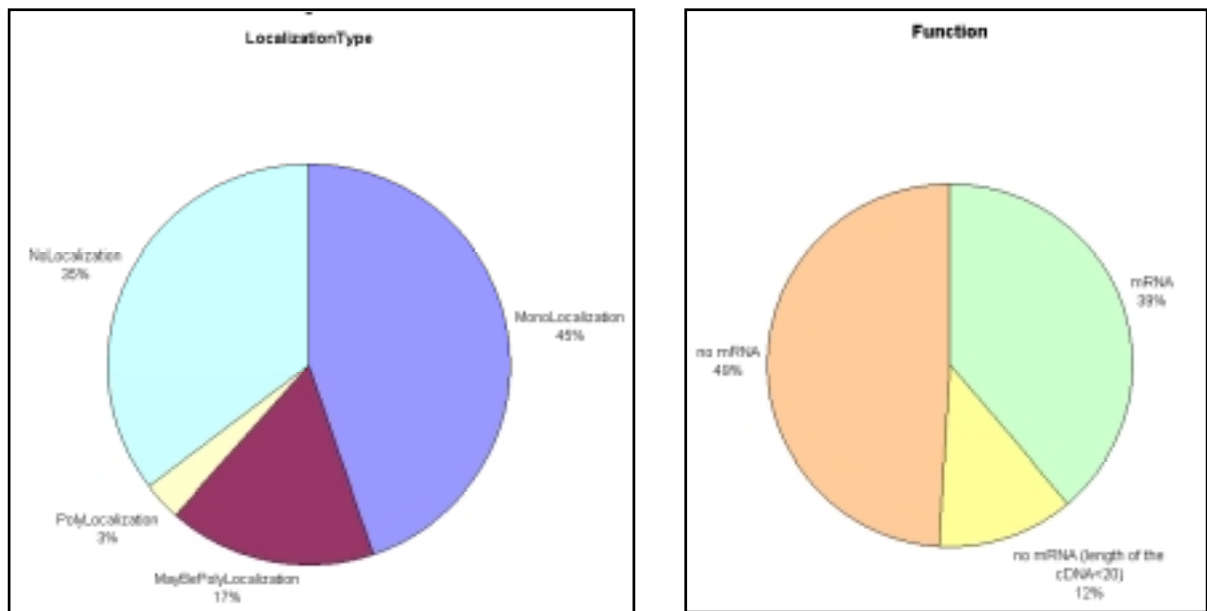
Environ 1/3 des gènes DD sont représentés sur les puces à ADN Affymetrix représentant une couverture d'environ 1/3 du génome.

Relation entre séquences analysées en Affymetrix et en DD (global)

Figure III.8

Notre étude en DD couvrant théoriquement plus de 90% du génome et les puces Affymetrix de 12'000 gènes couvrant environ 1/3 du génome, nous avons analysé la représentation de nos séquences DD sélectionnées comme différentiellement exprimées dans la collection Affymetrix (FigIII.8). Avec un ajout au sein de la base de données générale « DD » d'environ 140 séquences, issues d'une analyse préliminaire en DD de la chimiorésistance dans les tumeurs VADS (Dr Senghua Hao), nous avons constaté que 37% de nos séquences DD étaient représentées sur la puce Affymetrix. Cette proportion semble confirmer le taux de couverture du génome important de notre étude en DD.

### 3. Localisation chromosomique



Localisation sur le génome et existence d'ARNm dans les bases de données publiques

Figure III.9

Quarante cinq pourcent des gènes sont localisés de manière non équivoque sur le génome. Trente cinq pourcent des gènes ne sont pas localisés sur le génome, peut être en relation avec une faible taille pour une certaine proportion des clones (12% des clones ne correspondent pas à un ARNm mais présentent après élimination des amorces HAP et des séquences répétées une séquence spécifique de moins de 20 paires de bases). Trente neuf

pourcent des gènes sont associés de manière non équivoque à des ARNm annotés dans les bases de données. Le reste des clones n'est pas associé à des ARNm mais parfois à des EST ou des protéines dans les bases de données. Le clonage des séquences DD à partir de la queue poly-A peut contribuer à expliquer le manque d'identification d'une fraction des clones (FigIII.9). Bien sur, nous espérons qu'une fraction de ces clones peu identifiés correspond effectivement à des gènes demeurés inconnus.

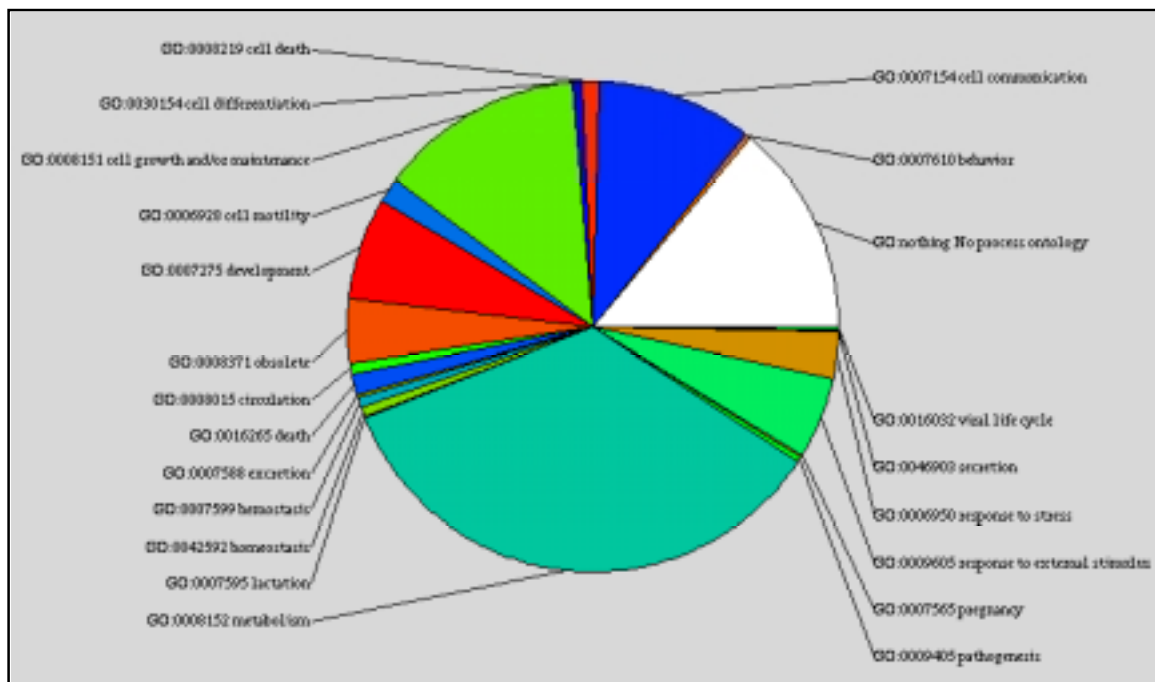
#### **4. Proportion des différents profils d'expression après validation en Reverse Northern blot**

L'analyse du profil d'expression des séquences nous apprend que 48% et 45% d'entre elles correspondent à des bandes respectivement surexprimées et sous exprimées dans l'ensemble des groupes de tumeurs, tandis que seulement 7% des séquences, sélectionnées comme différentielles avec au moins une sonde, correspondent à des bandes de profil d'expression complexe entre les divers groupes tumoraux. On observe donc, entre la représentation des profils complexes au niveau des bandes DD (19% des bandes DD) et la proportion finale des clones sélectionnés, une perte relative de ces séquences de profil complexe. Ce fait peut être expliqué par plusieurs facteurs. Les sondes les plus fréquemment utilisées lors des criblages en RN ont été synthétisées à partir de pools T et de pool N, contenant tous les tissus tumoraux et tous les tissus normaux. Les différences transversales entre tous les types tumoraux ne sont pas altérées tandis que les surexpressions complexes, spécifiques d'un ou deux groupes tumoraux, sont diluées dans le pool T et que leur validation peut être contre sélectionnée. Toutefois, des sondes ont également été synthétisées à partir des divers groupes tumoraux assurant une sélection des événements spécifiques à chaque groupe. Les hybridations avec ces sondes spécifiques ont été moins nombreuses et moins systématiques que pour les différences T/N. Une autre raison, intrinsèque au DD peut expliquer une perte relative des clones spécifiques des groupes tumoraux. Ces groupes, en particulier NM et M, très proches au niveau de l'histopathologie, tous deux constitués de tissus tumoraux, sont *a priori* plus proches entre eux qu'un tissu tumoral (pool T) en regard d'un tissu normal (pool N). Or le taux de faux positifs et la difficulté de détection de clones confirmés en criblage secondaire (notre technique RN) sont d'autant plus élevés que les tissus comparés sont proches (Liang, 1998). Les différences spécifiques de chaque groupe tumoral, subtiles, sont donc en partie contre sélectionnées.

## 5. Prédiction de la fonction à partir des catégories de geneOntology

Dans geneOntology (<http://www.geneontology.org/>), la fonction d'un gène et de sa protéine prédite est analysée sur trois plans : « Process », « Function », et « Component ». Les clones possédant de l'homologie avec une protéine ont été analysés.

### a. « Process »



### Prédiction du «Process» dans geneOntology

Figure III.10

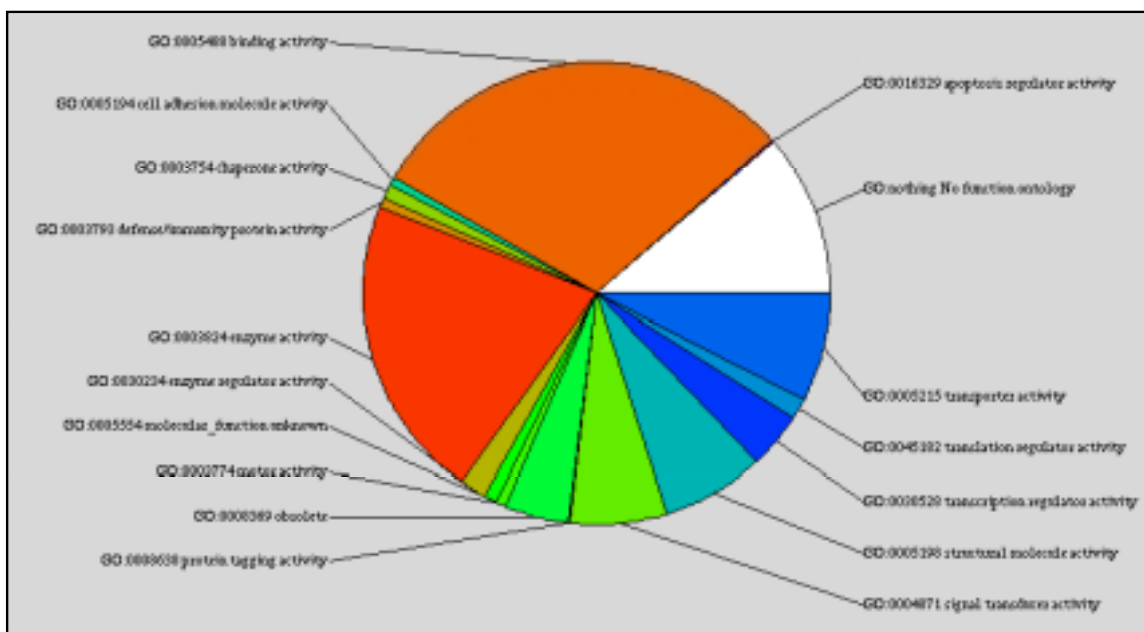
L'implication des protéines correspondantes dans divers grands procédés cellulaires est déterminée par rapport à des catégories « Process » (FigIII.10). On constate une forte proportion de gènes impliqués dans le métabolisme, en relation avec les forts besoins énergétiques des cellules cancéreuses en prolifération. Les gènes impliqués dans la croissance/maintenance cellulaire, les communications intercellulaires, les réponses aux stimuli externes sont également très représentés. Les altérations de ces types de gènes sont par ailleurs fréquemment associées au cancer. L'isolement de gènes associés au développement est cohérent avec les rôles de morphogènes, de régulateurs de la différenciation et de l'identité cellulaire impliqués dans le développement également associés au cancer. Une fraction

importante de gènes associés à des protéines prédites, n'a pas de fonction connue dans geneOntology.

La classification par geneOntology est délicate, certains gènes pouvant être impliqués dans diverses catégories, même parfois inattendues car regroupés dans la catégorie « obsolete ». Bien qu'imparfaite, cette classification constitue une indication globale de la teneur de notre collection.

### b. « Function »

La classification geneOntology définit également une classe « Function », décrivant la fonction de la protéine (FigIII.11).



### Prédictions “function” dans geneOntology

Figure III.11

Les protéines impliquées dans des activités de liaison (à l'ADN comme aux protéines) sont très nombreuses, suivies de près par les enzymes. Ceci est particulièrement intéressant car les enzymes sont des cibles privilégiées du « drug design ». Par rapport à notre publication, on note une faible représentation des gènes associés à la réponse immunitaire, mais cette différence peut être liée aux critères de catégorisation. Une sélection avec les



par geneOntology. Enfin, une fraction significative des gènes correspond à des protéines extracellulaires, possibles marqueurs sécrétés ou circulants.



## D. Première publication

---

Notre première publication présente une liste de 70 gènes différentiellement exprimés entre tissu normal et tissu tumoral, révélés par sondes complexes. Des liens tant physiques sur le génome que fonctionnels sont dégagés de cette étude.

# Differential expression profiling of head and neck squamous cell carcinoma (HNSCC)

F Lemaire<sup>1,4</sup>, R Millon<sup>2,4</sup>, J Young<sup>1,4</sup>, A Cromer<sup>1</sup>, C Wasyluk<sup>1</sup>, I Schultz<sup>2</sup>, D Muller<sup>2</sup>, P Marchal<sup>1</sup>, C Zhao<sup>1</sup>, D Melle<sup>3</sup>, L Bracco<sup>3</sup>, J Abecassis<sup>2</sup> and B Wasyluk<sup>\*,1</sup>

<sup>1</sup>Institut de Génétique et de Biologie Moléculaire et Cellulaire, CNRS/INSERM/ULP, 1 Rue Laurent Fries, BP 10142, 67404 Illkirch cedex, France; <sup>2</sup>UPRES EA 34-30, Centre Paul Strauss, 3 rue de la Porte de l'Hôpital, 67085 Strasbourg, France; <sup>3</sup>Exonhit Therapeutics, 65 Boulevard Masséna, Paris F-75013, France

Head and neck squamous cell carcinoma (HNSCC) is the fifth most common cancer in men with an incidence of about 780 000 new cases per year worldwide and a poor rate of survival. There is a need for a better understanding of HNSCC, for the development of rational targeted interventions and to define new prognostic or diagnostic markers. To address these needs, we performed a large-scale differential display comparison of hypopharyngeal HNSCCs against histologically normal tissue from the same patients. We have identified 70 genes that exhibit a striking difference in expression between tumours and normal tissues. There is only a limited overlap with other HNSCC gene expression studies that have used other techniques and more heterogeneous tumour samples. Our results provide new insights into the understanding of HNSCC. At the genome level, a series of differentially expressed genes cluster at 12p12–13 and 1q21, two hotspots of genome disruption. The known genes share functional relationships in keratinocyte differentiation, angiogenesis, immunology, detoxification, and cell surface receptors. Of particular interest are the 13 'unknown' genes that exist only in EST, theoretical cDNA and protein databases, or as chromosomal locations. The differentially expressed genes that we have identified are potential new markers and therapeutic targets.

*British Journal of Cancer* (2003) **89**, 1940–1949. doi:10.1038/sj.bjc.6601373 www.bjcancer.com  
© 2003 Cancer Research UK

**Keywords:** hypopharynx; 'unknown' genes; functional classes; biomarkers; pharmaceutical targets; virtual Northern

Head and neck squamous cell carcinoma (HNSCC) arises from the surface epithelium of the upper-aerodigestive tract (pharynx, hypopharynx, and larynx) and the oral cavity. Extensive epidemiological studies show that alcohol potentiates tobacco-related carcinogenesis and is also an independent risk factor. Head and neck squamous cell carcinoma is the fifth most common cancer in men with an incidence of about 780 000 new cases per year worldwide (Sankaranarayanan *et al*, 1998). Surgery and radiotherapy are highly effective in the treatment of stage I and II tumours, but over 70% of patients present with locoregionally advanced stage III or IV disease. Locoregional disease recurs in 60% of patients and metastatic disease develops in 15–25% (Genden *et al*, 2003). Furthermore, patients develop second primary tumours at an annual rate of 3–7% (Leon *et al*, 2002). However, less than 30% of HNSCC patients are free of disease after 3 years, and 5-year survival rates have remained largely unchanged in the last three decades (Dimery and Hong, 1993). The characterisation of the molecular determinants of the head and neck carcinogenesis process is essential for the better understanding of this malignancy and the development of rational targeted intervention.

Specific genes have been associated with the development or presentation of HNSCC, but these individual alterations have failed to define prognostic or diagnostic markers (reviewed in Leonard *et al* (1991) and Scully *et al* (2000)). Addressing this issue requires large-scale analysis of gene expression profiles. A number of recent studies have reported gene expression profiles of small numbers of HNSCC patients using commercial or focused microarrays (Leethanakul *et al*, 2000; Xie *et al*, 2000; Alevizos *et al*, 2001; Al Moustafa *et al*, 2002; Belbin *et al*, 2002; El-Naggar *et al*, 2002; Mendez *et al*, 2002). The microarray analysis is limited by the set of genes on the arrays, whereas polymerase chain reaction differential display (PCR-DD) randomly samples the transcriptome. The PCR-DD has been used to discover novel genes that would not have been identified using methodologies that cover a predefined range of genes (Glynn-Jones *et al*, 2001; Sasaki *et al*, 2001; Ying *et al*, 2001). We have performed the first randomised comparative analysis of gene expression of HNSCC patients using PCR-DD. We did not use microdissected tumour or normal components for this analysis since numerous studies have shown that the host tumour microenvironment influences tumour cells (van den Hooff, 1988; Nelson *et al*, 2000; Coussens *et al*, 1999; St Croix *et al*, 2000). We have identified a series of novel genes that exhibit striking differences in expression between HNSCC tumours and histologically normal matched tissues. They should contribute to a better understanding of HNSCC and provide new targets for therapeutics.

\*Correspondence: Dr B Wasyluk; E-mail: boh@igbmc.u-strasbg.fr.

<sup>4</sup>These authors contributed equally to this manuscript and are listed alphabetically.

Received 15 May 2003; revised 9 September 2003; accepted 16 September 2003

## MATERIALS AND METHODS

## Samples

Hypopharyngeal tumours and the corresponding histologically normal tissue, used with consent, were derived from surgical resections of squamous cell carcinoma. The patients had not been treated at the time of surgery, but were subsequently treated with radiotherapy. The samples used were resected near the advancing edge of the tumours avoiding their necrotic centres. They were comprised of 70–80% cancer cells in almost all cases, as assessed on adjacent histological stained sections. Normal samples were collected from the farthest margin of the surgical resections (usually uvula). The tumours were classified according to TNM stages (tumour, node, metastasis) based on the UICC criteria (Sobin and Wittekind, 1997), and grouped into three categories. The early (E) stage corresponds to small-sized tumours (T1/T2), moderately to well differentiated, without lymph node involvement. The two later-stage tumour types were of medium size (T2/T3), homogeneous differentiation and lymph node involvement (N1–N2c). At the time of resection, these later-stage tumours appeared clinically and histologically similar. However, during 3-year follow-up, one group of patients did not develop metastases (no metastatic propensity: NM), whereas the other developed metastases predominantly in the lung, bone, and liver (with metastatic propensity: M).

## PCR-Differential display

Total RNA was isolated with RNeasy (Qiagen, Courtaboeuf, France), DNaseI treated, column purified (Qiagen) and pooled according to the tumour type (3 E, 2 NM and 2 M patients). The corresponding normal RNAs were similarly pooled. The PCR-DD was performed on the pooled samples using 58 5' primers (HAP) in combination with three 3' primers (HT11A/G/C) according to the GenHunter protocol and as described by Liang *et al* (Liang, 1998). All samples were prepared in duplicate from the reverse transcription stage to reduce experimental variability. Differential bands were isolated, reamplified with the corresponding primers, verified by agarose gel electrophoresis, and cloned in the pGEMt-Easy vector (Promega, Charbonnières, France). Eight colonies per band were expanded in liquid culture. A volume of 2  $\mu$ l of the cultures were used for PCR, in the same conditions as the reamplification, with the pGEMe1 and pGEMe2 primers (5'-CGC GGT ACC GGA TCC ATG CAT TGG CGG CCG CGG GAA TTC-3' and 5'-CGC GGT ACC GGA TCC ATG CAT CAT ATG GTC GAC CTG CAG-3', respectively). The fragments (50–800 base pairs) were verified by agarose gel electrophoresis, and subsequently the DNA was spotted directly onto nitrocellulose membranes (Hybond N+, Amersham, Les Ulis, France) using a 96-well vacuum-driven dot blot manifold (Bio-Rad, Marnes-la-Coquette, France). Filters underwent denaturation (1.5 M NaCl, 0.5 M NaOH) and neutralisation (1.5 M NaCl, 0.5 M Tris-HCl pH 7.2, 0.001 M EDTA) followed by UV cross-linking.

**Table 1** Real-time quantitative PCR primers and reaction conditions.

Clone		Primer		PCR		
Name	Acc. no.	Name	Sequence	Location	Size (bp)	T (°C), [MgCl <sub>2</sub> ] (mM), [NTP] ( $\mu$ M)
EMPI	NM_001423.1	EMPI-2F	GACCTCATGCCATGGTCTTT	1393–1412	237	62, 4, 0.5
		EMPI-2R	CTGCATTGAGGGGAATCCTA	1556–1575		
PIGR	AF272149.2	TB6-1F	CCACCGTGGAGATCAAGATT	1532–1551	186	62, 4, 0.5
		TB6-1R	CAGCCCGTGTATTCCACTT	1669–1688		
PON2	NM_000305	AEY125-F	GCCAACAATGGGTCTGTCT	993–1012	198	62, 4, 0.5
		AEY126-R	TGGGTCAATGTTGCTGGTTA	1171–1190		
Apol2	AF305225	AEY131-F	AGGCAGATGAGCTCCGTAAA	363–382	185	62, 4, 1
		AEY131-R	GACCTGCTCAACTCCTCTG	528–547		
DRG1	D87953	AEY139-F	GCTTTGGTCAGAGTGAATTGAA	2717–2736	182	62, 4, 0.5
		AEY140-R	CCGATCCCCGACTTTTCTAC	2879–2898		
PSMD8	NM_004159	AEY147-F	GAAGGAAGATGGTTGGGTA	981–998	189	62, 4, 0.5
		AEY147-R	TCTCTTTGGCTCAGGCTAGG	1150–1169		
RPI-68D18	BM285393	EAW253-F	TGCAAGTCACCACAACAGGT	153 589–153 608	185	62, 4, 0.5
		EAW254-R	AGCCTTGCAATAATGGCTGT	153 753–153 773		
HSPC150	NM_014176	AEY151-F	TGTTCTCAAATTGCCACCAA	390–409	191	62, 4, 1
		AEY152-R	TTGCATGCTTCTGTGCCAC	516–580		
RPLP0	M17885	RPP0-3F	GAAGCTGTGGTCTGATGG	224–243	103	62, 3, 1.5
		RPLP0-R	CCGGATATGAGGCAGCAGTT	307–326		

Listed above are the clone names and accession numbers; the primer name, sequence (5'–3') and location on the sequence associated with the accession number; and the PCR product size and reaction conditions (annealing temperature and concentrations of MgCl<sub>2</sub> and NTP).

**Table 2** Characteristics of the tumours

Patient	Tumour	T	N	M	Diff	Sex	Age (years)	Treatment after surgery	Evolution	Disease-free survival (months)	Overall survival (months)	Actual state
1	E	2	0	0	2	M	53	RX	0	11	15	D
2	E	1	0	0	2	M	61	N	0	15	25	D
3	E	2	0	0	2	M	44	N	0	41	41	A
4	NM	3	2b	0	3	M	66	RX	0	54	58	D
5	NM	3	2b	0	2	M	52	RX	SC	39	44	A
6	M	2	2c	0	3	M	54	RX	M	16	25	D
7	M	3	2b	0	2	M	54	RX	M	5	7	D

All the tumours were localised in the hypopharynx. T, N, and M correspond to the TNM nomenclature for tumour stage (tumour, node, and metastasis). Diff (differentiation): 1 = well, 2 = moderate, 3 = poorly. Treatment after surgery: RX = radiotherapy; N = no treatment. Evolution: 0 = no evolution, M = metastasis, SC = secondary cancer. Actual state: D = dead, A = alive.

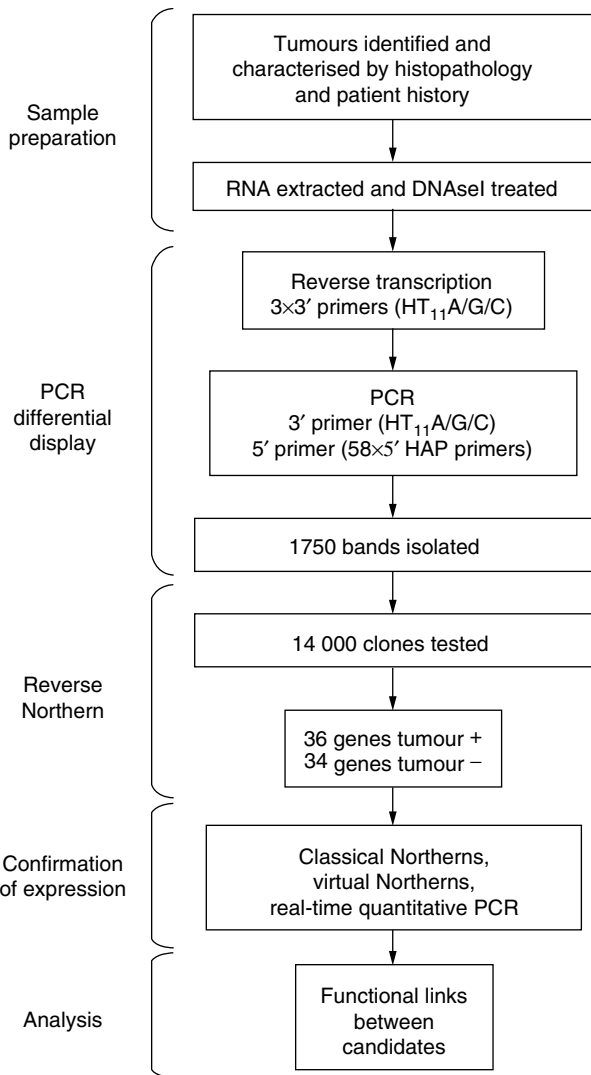
Reverse Northern

Owing to the limiting quantity of patient RNA, the SMART cDNA synthesis system (Clontech, Le Pont de Claix, France) was used to reverse transcribe and amplify total RNA to be used as a probe. The first strand, synthesised from 0.2 µg of total RNA, was amplified for a controlled number of cycles, to ensure linearity, as described by the manufacturer. The labelling was performed with 100 ng of SMART cDNA and a mix of the DD primers that originally generated the clones. The probes were purified through Sephadex G50 columns (Bio-Rad). The filters were hybridised in 10% dextran sulphate/0.1% SDS/10 mM NaCl overnight at 65°C, washed to a stringency of 0.2 × SSC/0.1% SDS at 65°C and exposed on Biomax film for 3–24 h at –80°C, and subsequently on Molecular Dynamics PhosphorImager screens (Orsay, France) for quantification on a Typhoon PhosphorImager analyser (Orsay, France). Positive clones were then expanded from the original liquid cultures and plasmid DNA extracted using standard alkaline lysis followed by purification through Nucleospin miniprep columns (Macherey-Nagel, Hoebdt, France). The sequences of

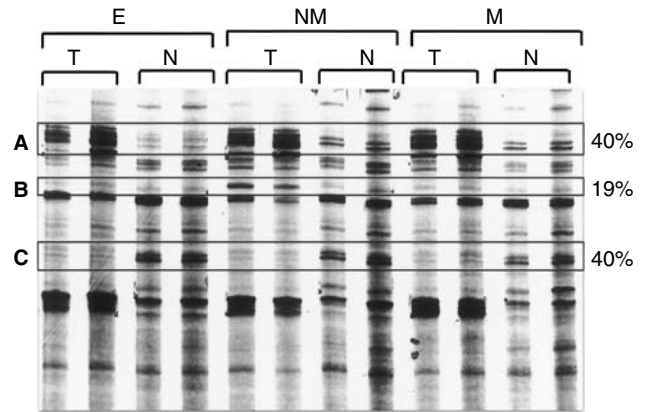
the inserts were analysed with the BLAST algorithm at <http://www.ncbi.nlm.nih.gov/blast/>. Positive clones were then confirmed at least twice with probes generated (as above) from two independent SMART cDNA preparations. The filters included control positive clones that were systematically used for cross comparison.

Classical Northern

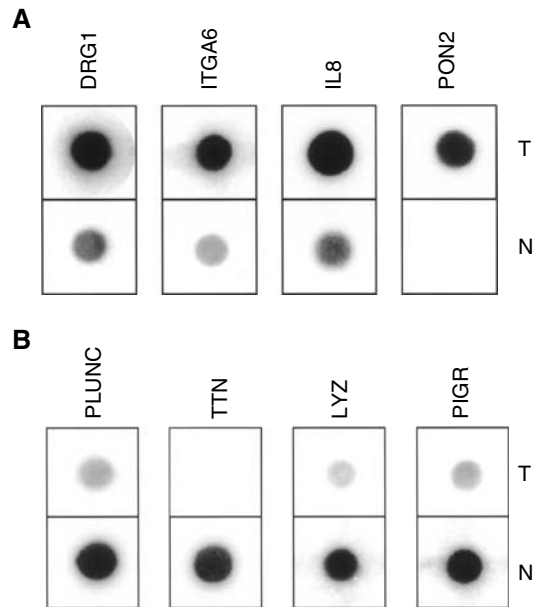
Total RNA was extracted from tissue samples with Trizol (Life Technologies, Cergy Pontoise, France). A measure of 20 µg of RNA was subjected to agarose/6% formaldehyde gel electrophoresis, then transferred to Hybond N+ membranes (Amersham). [<sup>32</sup>P]-labelled probes were generated with the Rediprime system (Amersham). Membranes were prehybridised and hybridised in 50% formamide at 42°C according to the manufacturer's



**Figure 1** Flowchart outlining the study. The flowchart indicates how the tumour samples were selected and processed, the PCR-DD primers that were used and the number of bands isolated, the number of clones tested by reverse Northern, the resulting number of genes identified, the types of confirmation used to validate the results, and the bioinformatics analysis to analyse the results.



**Figure 2** Differential display gel comparing the three stages of tumours (T) with their corresponding normal (N) samples. E = early; NM = no metastatic potential; M = metastatic propensity. Highlighted are the three types of profiles (A, overexpressed in tumour; B, tumour-specific profiles; C, underexpressed in tumour), and the percentages give the overall proportions in these categories.



**Figure 3** Reverse Northern using tumour (T) or normal (N) tissue probes. The genes shown, which are overexpressed (A) or under-expressed (B) in tumours, are the first four in Tables 3A and B, respectively.

**Table 3** Differentially expressed genes identified by PCR differential display (PCR-DD) and reverse Northern analysis in Parts A and B

Order	Clone	Gene	Chromosome	Unigene or acc. no.	Immune
(A) A total of 36 genes were detected with a greater than two-fold higher expression in tumour (T) tissues than in normal (N) samples					
Z V H	B13.4	DRG1, N-myc downstream regulated gene 1	8q24.3	Hs.75789	
	F3.1	ITGA6, alpha 6 integrin	2q31.1	Hs.227730	
	61b	IL8, interleukin 8	4q13-q21	Hs.624	Related
	100b82	PON2, paraoxonase 2	7q21.3	Hs.169857	
	2b11	HPRT, hypoxanthine phosphoribosyltransferase	Xq26-q27.2	Hs.82314	
	XO17.6	Immunoglobulin kappa L chain constant region gene	2p12	NG_000834	Related
	221A	PLAUR, plasminogen activator, urokinase receptor	9q13	Hs.17957	Related
	W4.3	Chr 11, clone RPI1-68D18	11p11-12	ALI33330	
	G6.3	HSPC150, E2 ubiquitin conjugating enzyme	1q31.3	Hs.5199	
	183H	APOBEC3, apolipoprotein B mRNA editing enzyme	22q13.1	AL031856	
	221G	APTX, aprataxin	9p13.3	Hs.14394	
	O1.7	Chr 5, clone CTC-202F10	5q35	AC010216	
	107E	PLA2G7, phospholipase A2, group VII	6p21.2-p12	Hs.93304	Related
	108A	Clone RPI1-367F23, chromosome 9	9q22.31	AL355607	
	120b145	APOL2, apolipoprotein L, 2	22q12	Hs.241412	
	A74j5	RPL27, ribosomal protein L27	17q21.1-q21.2	Hs.111611	
	A80j1	Chr4, clone RPI1-331M18	4	AC108050	
	G6511	A2M, alpha-2-macroglobulin (3' end)	12p13.3-p12.3	Hs.74561	Related
	C65f2	LTB4R, leukotriene B4 receptor	14q11.2-q12	Hs.28408	Related
	275F	BLP1, BBP-like protein 1	8p11.21	Hs.7471	
	126b180	HSPCB, heat shock 90 kDa protein 1, beta	6p12	Hs.74335	
	B14.3	Chr 18, clone RPI1-650P15	18q11	AC021549	
	291D	EFNB2, ephrin-B2	13q33	Hs.30942	
	99ci2	Chr16, clone RPI1-170L3	16p11.1	AC034105	
	P2.8	SPOCK, testican 1	5q31	Hs.93029	
	31_50	PSMB8, proteasome subunit, beta type 8	6p21.3	Hs.180062	
	0954B	IFIT2, interferon-induced protein with tetratricopeptide repeats 2	10q23.33	XM_084477	Related
	XB2v.2	TRAI, tumour rejection antigen (gp96) 1	12q24.2-q24.3	Hs.82689	Related
	31c	LTBP1, latent transforming growth factor beta binding protein 1	2p12-q22	Hs.241257	
	A8015	ND10, nuclear domain 10	17q23.2	Hs.154230	
	A75a1	HMG A2, high mobility group AT-hook 2	12q15	Hs.2726	
	A71d8	MTND4, NADH dehydrogenase 4	Mitochondrion	NC_001807	
	C66p4	DIA1, diaphorase (NADH) (cytochrome b-5 reductase)	22q12.31-qter	Hs.80706	
	C68b1	Stat1, signal transducer and activator of transcription 1	2q32.2-q32.3	Hs.21486	Related
	C78g1	RPL6, ribosomal protein L6	12q23-24.1	Hs.349961	
	A71i7	TDG, thymine-DNA glycosylase	12q24.1	Hs.173824	
(B) A total of 34 genes were detected with a greater than two-fold higher expression in normal (N) compared to tumour (T) tissues					
Z V Z	207D	PLUNC, long palate, lung and nasal epithelium carcinoma associated	20q11.1	Hs.211092	Related
	213A	TTN, titin	2q24.2	Hs.172004	
	G72c7	LYZ, lysozyme	12q13.2	Hs.234734	Related
	XB11b.2	PIGR, polymeric immunoglobulin receptor	1q31-q41	Hs.205126	Related
	XJ19.3	FLJ22671 hypothetical protein	2q37.1	Hs.193745	
	M4.8	GAPDH, glyceraldehyde 3 phosphate dehydrogenase	12p13	Hs.169476	
	XO27.2	LPRP, lacrimal proline-rich protein	12p13	Hs.45033	
	G75e1	SPRR3, small proline-rich protein 3	1q21-q22	Hs.139322	
	XH7.7	Similar to olfactory receptor MOR265-1	12q12	X87825	
	XL6.3	CAPN7, calpain like protease 7	3p24	Hs.7145	
	C78c8	CLECSF2, C-type lectin superfamily member 2	12p13-p12	Hs.85201	Related
	XK1d.3	MTND2, NADH dehydrogenase subunit 2	Mitochondrion	NC_001807	
	XH3.2	CILP, cartilage intermediate layer protein	15q22	Hs.151407	Related
	121a149	Similar to common salivary protein1 mRNA	16p13.3	NM145252	
	137D	PRH, proline-rich protein HaeIII	12p13.2	Hs.73952	
	XJ10.4	KIF13B, kinesin family member 13B (gakin)	8p12	Hs.15711	
	A77s8	KRT13, cytokeratin 13	17q21-q23	Hs.74070	
	XH3.8	ARPP-19, phosphoprotein 19	15q15.3	Hs.7351	
	L1.4	Clone RPI3-253K9 on chromosome X,	X	AK024673	
	B7.4	PER3, period homolog 3	1p36.23	Hs.12592	
	N6.8	Chr11, clone RPI1-630P1 (hypoprotein)	11q	AP002774.3	
	11.7	KIAA0138 gene product	19p13.3	Hs.159384	
	XL9b.8	Chromosome 12 clone RPI1-272B17	12q13	AC06776	
	38b17	CLIC4, chloride intracellular channel 4	1p36.11	Hs.25035	
	XB6.4	KLK-L4, kallikrein-like protein 4 gene	19q19.3-q19.4	Hs.165296	Related
	XL6.2	KIAA1321 gene product	17q11.1	Hs.24336	
	XL9b.1	LTF, lactotransferrin	3q21-q23	Hs.105938	
	I35B	EGLN2, EGL9 homolog	19q13.2	Hs.324277	
	C68a4	GCLC, glutamate-cysteine ligase, catalytic subunit	6p12	Hs.151393	
	A77e2	SPRR2, small proline-rich protein 2	1q21	Hs.355542	
	G80d5	EMPI, epithelial membrane protein 1	12p12.3	Hs.79368	
	G72h1	S100A9, S100 calcium binding protein A9	1q21	Hs.112405	Related
	G74c1	Hypothetical protein DKFZp761C169	5q11.2	Hs.71252	
	G78a5	GJB2, gap junction protein, beta 2, 26 kDa	13q11-q12	Hs.323733	

The genes are listed in the order of the differences in expression levels. Also listed for each PCR-DD clone are the name of the clone, the corresponding gene name, chromosomal localisation, Unigene or accession number, and whether its function is related directly or indirectly with the immune response.

specifications, washed to a stringency of  $0.1 \times$  SSPE/0.1% SDS at  $50^\circ\text{C}$  and exposed to X-ray film (Kodak, Les Ulis, France). The level of expression in tumour samples was analysed in comparison with the matched normal tissues after correction for loading using RPLP0. Ribosomal phosphoprotein P0 (RPLP0, originally called 36B4) is a ubiquitous expressed gene that has been routinely used in different laboratories as an internal control to normalise for the amount of RNA. In a large study (98 cases), we confirmed by real-time quantitative PCR (RT-QPCR) that its expression level remains relatively constant between HNSCC tumours and matched normal tissues (data not shown). RPLP0 gave better results than the commonly used control GAPDH, which was more variable between samples in our experiments.

**Virtual Northern**

A measure of  $0.2 \mu\text{g}$  of total RNA from individual patients was converted into SMART cDNA ((Franz et al, 1999) and the Clontech protocol). The optimal number of cycles for each sample was determined according to the manufacturer's instructions. Aliquots of the PCR products, after different numbers of cycles (15–25), were analysed by agarose gel electrophoresis and Northern blotting with RPLP0 as the probe. The amplification and the fidelity are considered to be optimum when the PCR is in the exponential phase of amplification, one or two cycles before reaching the plateau (range 17–20 cycles). The RPLP0 signal of the optimum PCR was used as an internal standard to equilibrate loading of the virtual Northern. Appropriate amounts of this 'SMART' cDNAs were electrophoresed on agarose gels, transferred to Amersham Hybond N+ nylon filters. Probes were labelled with  $[^{32}\text{P}]\text{dCTP}$  by random priming or PCR with the pGEME1 and pGEME2 primers (see above). Filters were hybridised in dextran sulphate (as above), exposed overnight to PhosphorImager screens and quantified using the Typhoon ImageQuant software. Filters were finally reprobated with RPLP0 to verify equal loading.

**Real-time quantitative PCR**

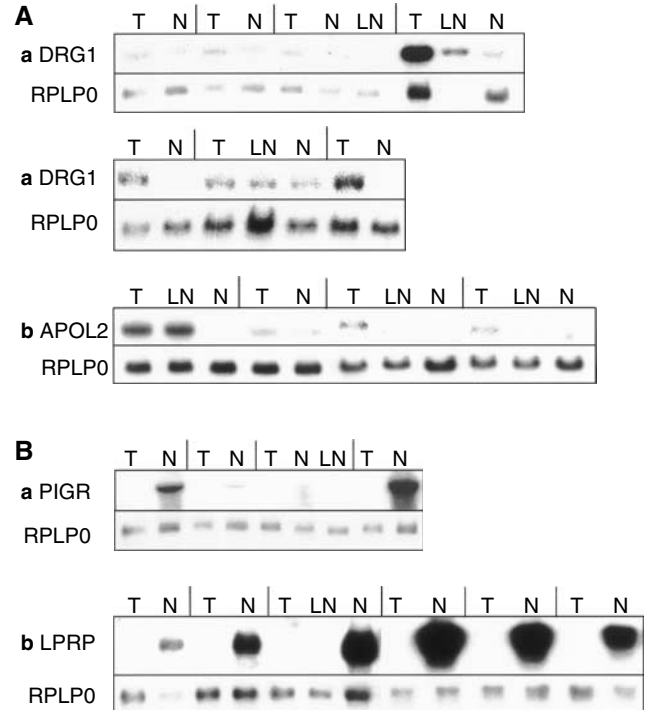
RNA was quantitated with the LightCycler system (Roche Diagnostics, Meylan, France). A measure of  $1 \mu\text{g}$  of total RNA was reverse transcribed with random primers and the Superscript II RT-PCR system (Life Technologies). The PCR reactions were performed with the LC Fast start DNA master SYBR green I reaction mixture according to the manufacturer's instructions. Volumes of  $2 \mu\text{l}$  of 1:50 diluted RT products were used in  $20 \mu\text{l}$  reactions. The nucleotide sequences of the primers and their localisations are shown in Table 1. The primers were chosen with the Primer3 software and their specificity was verified by BLAST analysis on the nr database (non redundant set of GenBank, EMBL, and DDJB databases). For each gene, a standard curve was constructed using serial dilutions of a single standard cDNA (equivalent to 100, 40, 20, 10, 4, 2, and 1 ng of total RNA) derived from a pool of 10 hypopharyngeal tumours. The concentrations of primers,  $\text{MgCl}_2$ , probes, and cDNA were optimised to obtain linear standard curves. Unknown samples were estimated relative to these standard curves. For genes overexpressed in tumours, expression levels were calculated relative to the median values for normal tissue, and *vice versa* for genes expressed at higher levels in normal tissues. PCR reactions were run at least twice for each sample. The mean value was retained whenever the standard deviation did not exceed 15%, and normalised using RPLP0 as an internal control.

**RESULTS**

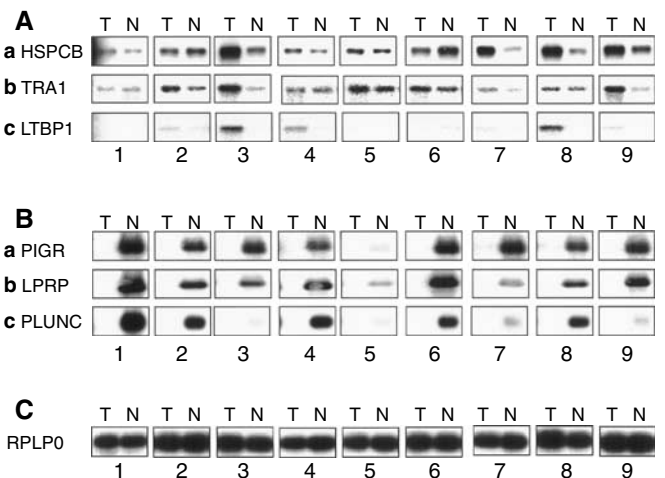
**PCR differential display**

A large-scale PCR-DD was performed on patient RNA derived from three stages of HNSCC (Table 2) and corresponding normal

tissues (see Figure 1 for a methodology outline). Three 3' primers (HT11A/G/C) and 58 5' primers (HAP1-10, 33–80) were combined to cover theoretically over 90% of expressed sequences (Liang, 1998). This experimental design maximises the detection of 'novel' sequences, a strength of PCR-DD compared to DNA arrays. Around 95% of the bands showed no difference in signal intensity across the different samples, as expected. Of the 1750 bands that



**Figure 4** Classical Northern: tumour (T), lymph node (LN) and normal (N) samples from the same patients were analysed. The RPLP0 control is shown under each lane. (A) Genes overexpressed in tumours: (a) DRG1, (b) APOL2. (B) Genes underexpressed in tumours: (a) PIGR, (b) LPRP. The lines separate the samples from particular patients, and comparisons should be made between the samples from each patient.



**Figure 5** Virtual Northern: the lanes 1–3 (E), 6,7 (NM) and 8,9 correspond to individual patients who were pooled for the PCR-DD. Tumour (T) and normal (N) samples from the same patient were compared. (A) Genes overexpressed in tumours: (a) HSPCB, (b) TRA1, and (c) LTBPI. (B) Genes underexpressed in tumours: (a) PIGR, (b) LPRP, and (c) PLUNC. (C) RPLP0 is the internal control.

did show a difference, 40% were increased and another 40% were diminished in all the tumour samples (Figure 2). These two groups are the focus of this initial study. The 19% of the bands that differed between the tumour types will be addressed in future studies. Less than 1% of the bands differed in intensity between the normal samples from different patients, indicating that the differences observed between the normal and tumour samples were due specifically to the development of the tumours and not due to either patient polymorphism or PCR-derived artefacts (Figure 2). The differential bands were isolated and cloned, and eight clones were taken from each band for further analysis.

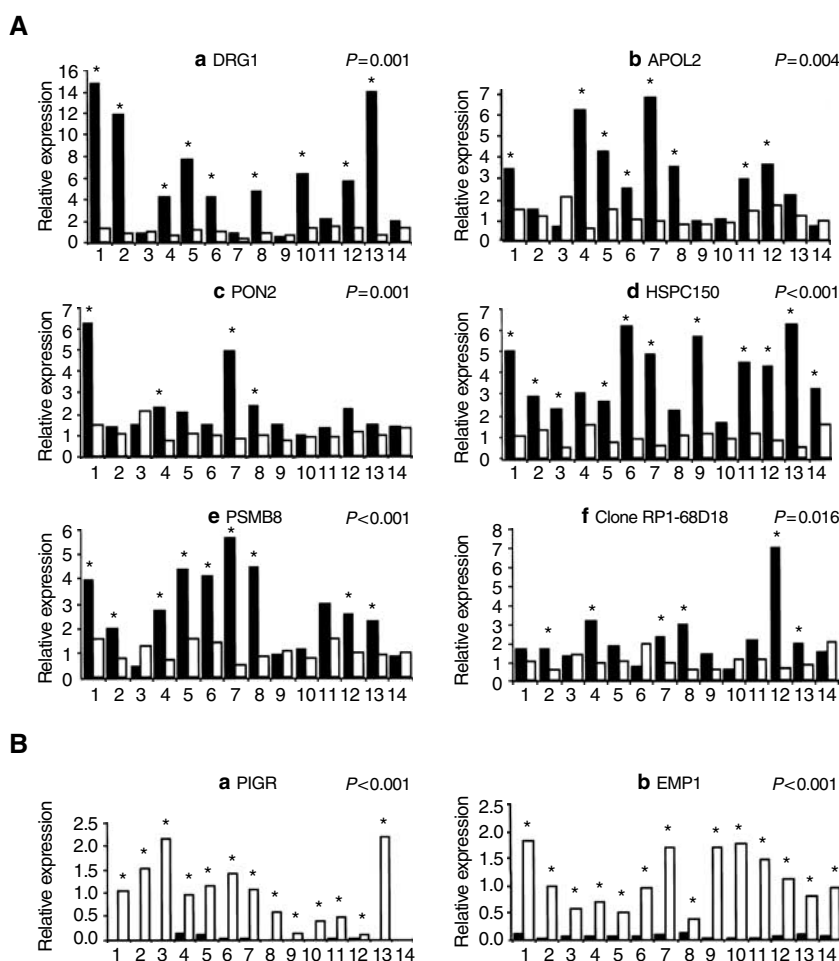
### Identification of the genes

Reverse Northern hybridisation (Zhang *et al*, 1997; Trenkle *et al*, 1999) was performed on the 14 000 clones resulting from the DD to determine which clones among the eight clones derived from each band contained differentially expressed sequences. Macroarrays of the clones were hybridised with probes derived from either pooled tumour or pooled normal RNA, and the resulting signals were quantified. In total, 2500 clones presenting a tumour/normal signal ratio of  $>2.0$  or  $<0.5$  were grouped onto secondary arrays and reprobated twice for confirmation (Figure 3). Clones with consistently differential profiles after multiple hybridisations and tumour/normal ratio of  $>2.0$  (2–5-fold) or  $<0.5$  (0.5–0.07-fold)

were sequenced and identified using the BLAST algorithm. Some of the clones with consistent profiles corresponded to the same gene (1–85 clones per gene). Our final list contains 36 genes that are overexpressed in tumours (Table 3A) and 34 genes that are under expressed (Table 3B). Six of the overexpressed and seven of the underexpressed sequences are novel, in that they do not correspond to known genes.

### Validation of gene expression profiles

To confirm that the large-scale analysis had correctly identified differentially expressed sequences, some up- and downregulated genes were analysed by the classical Northern analysis (Figure 4). As the amount of patient material was too limited to do numerous classical Northern, SMART technology (Clontech) was used to generate virtual Northern (Figure 5). In addition, RT-QPCR was used with a panel of 14 hypopharyngeal carcinomas and matched normal tissues (Figure 6). The results were consistent across these validation techniques (DRG1, Figures 4Aa and 6Aa; APOL2, Figures 4Ab and 6Ab; PIGR, Figures 4Ba, 5Ba, and 6Ba; LPRP, Figures 4Bb and 5Bb; note that the patients were different). We found that PIGR, LPRP, PLUNC, and EMP1 are downregulated in almost all the tumours (Figures 4B, 5B, and 6B). DRG1, APOL2, HSPCB, TRA1, LTBP1, PON2, HSPC150, PSMB8, and clone RP1-68D18 are overexpressed in tumours at various frequencies



**Figure 6** Real-time quantitative PCR. **(A)** Genes overexpressed in tumours (T): (a) DRG1, (b) APOL2, (c) PON2, (d) HSPC150, (e) PSMB8, and (f) RPI-68D18. **(B)** Genes underexpressed in tumours: (a) PIGR and (b) EMP1. The values for the tumours (black columns) and matched normal (N) tissue (white columns) were adjusted according to RPLP0, the internal control. The median of the N values was set to 1. The patients indicated with a star have a matched T/N ratio greater or equal to 2. The *P*-values of *t*-tests between the tumour and normal tissues are indicated.

**Table 4** Comparison with other tumours

Order	Gene	Cancer	Expression	Function	Ref
H N S C C	DRG1 (RTP/ Cap43)	Breast	Overexpression		1
	ITGA6, alpha 6 integrin	Oesophageal carcinomas, breast cancer	Overexpression	Tumorigenicity and metastatic phenotype	2
	IL8, interleukin 8	HNSCC, melanoma	Overexpression	Induction of angiogenesis, metastatic potential	3
	PON2, paraoxonase 2		Mutation	Polymorphism associated to susceptibility to cancer	4
	HPRT	Neuroblastoma cells		Rapid proliferation	5
	PLAUR, plasminogen activator urokinase receptor	HNSCC, skin, breast	Overexpression	Invasion and metastasis	6
	PLA2G7, phospholipase A2 group VII	Lymphoma	Translocation Breakpoint	Pathogenesis	7
	HSPCB, heat shock 90 kDa protein 1 beta (HSP90)	SCC of the tongue, bladder cancer, endometrial cancer	Overexpression	Cell proliferation	8
	TRA1, tumour rejection antigen (gp96) I	Leukaemogenic monocytic sublines, leukemogenesis	Overexpression		9
	LTBP1, latent TGF beta binding protein 1	Ovarian carcinoma	Overexpression	Regulation of TGF beta activity	10
	ND10, nuclear domain 10 protein	Acute promyelocytic leukaemia		Pathogenesis	11
	HMGA2, high mobility group AT-hook 2 (HMGI-C)	Non-small-cell lung carcinoma	Overexpression	DNA binding, neoplastic and undifferentiated phenotypes	12
	MTND4, NADH dehydrogenase subunit 4	Squamous cell cervical carcinoma	Overexpression	Metabolic enzyme	13
N S C C	PLUNC	Lung	Downregulated	Inflammatory response in the upper airways	14
	PIGR, polyimmunoglobulin receptor	Colorectal carcinomas	Diallelic loss	Expressed on secretory epithelial cells	15
	GAPDH, glyceraldehyde 3 phosphate dehydrogenase			Glycolytic function, DNA repair and apoptosis	16
	CAPN7, calpain-like protease 7	Calpain 9, gastric-cancer suppressor		Cysteine proteases; calcium-dependent cell death	17
	KRT13, cytokeratin 13	Oral epithelial dysplasia and squamous cell carcinoma	Downregulated		18
	KLK-L4, kallikrein-like protein 4 gene	Breast cancer tissues and cell lines	Downregulated	Pathogenesis, matrix degradation	19
	GCLC, glutamate-cysteine ligase catalytic subunit			Cellular response to environmental toxicants	20
	SPRR2, SPRR3, small proline-rich proteins 2 & 3	Scirrhou gastric cancer cells with high metastatic potential	Downregulated	Keratinocyte differentiation	21
	GJB2, gap junction protein beta2 26kDa (Connexin 26)	Malignant phenotype, repressed in bladder and breast cancer cells	Downregulated	Keratinocyte differentiation in buccal epithelia; intercellular communication	22

The genes identified in this study that have been described to be implicated other human cancers are listed, with their expression pattern in our study (T > N or N > T), their names, the other cancers concerned, the reported alterations (expression), their functions and references (Ref). References: 1 = Cangul et al (2002); 2 = Mukhopadhyay et al (1999), Tanaka et al (2000); 3 = Bar-Eli (1999), Kawano et al (2001), Liss et al (2001); 4 = Hegele (1999); 5 = Connolly et al (2001); 6 = Mazar (2001), Romer et al (2001), Schmidt and Grunsfelder (2001), van der Pluijm et al (2001); 7 = Lecointe et al (1999); 8 = Cardillo et al (2000), Ito et al (1998), Wataba et al (2001); 9 = Kasukabe et al (1997); 10 = Higashi et al (2001); 11 = Stemsdorf et al (1997); 12 = Manfioletti et al (1991), Rogalla et al (1998); 13 = Cheng et al (2002); 14 = Bingle and Bingle (2000); 15 = Krajci et al (1996); 16 = Sirover (1999); 17 = Yoshikawa et al (2000); 18 = Bloor et al (2001); 19 = Yousef et al (2000); 20 = Walsh et al (2001); 21 = Hippo et al (2001); 22 = Grossman et al (1994), Lucke et al (1999), Singal et al (2000).

(Figures 4A, 5A, and 6A). The differences in expression, measured by RT-QPCR, were at least two-fold in at least half of the tumours for seven of the eight genes analysed. DRG1, APOL2, HSCPC150, and PSMB8, and the novel sequence clone RPI-68D18 are overexpressed in nine, eight, 12, eight, and six patients, respectively. Overall, the expression profiles correlate well with the behaviour observed at the DD band level.

## DISCUSSION

We have compared the expression profiles of hypopharyngeal tumours with matched normal tissues by the PCR-DD. This study of a specific site of HNSCC provides a novel collection of cancer-related genes. Our results are of high quality since the DD sequences were reselected with several rounds of reverse North-



erns, and there was a consistent correlation between the DD profiles and analyses by classical Northern, virtual Northern, and RT-QPCR. The sequences reported here had a consistent DD profile across the tumour samples, whereas other bands (about 20%) with tumour stage-specific profiles need to be studied further with a larger number of tumours. Only eight out of 70 genes overlap between our and other profiles of HNSCC (Leethanakul *et al*, 2000; Alevizos *et al*, 2001; Al Moustafa *et al*, 2002; Belbin *et al*, 2002; El-Naggar *et al*, 2002; Mendez *et al*, 2002; Xie *et al*, 2000), possibly because, in contrast to these other studies, we did not restrict the profiling to particular genes on arrays, since PCR-DD samples the whole transcriptome. Moreover, we restricted our analysis to a very specific site. Six of the common genes are expressed in the same manner (ITGA6, PON2, STAT1, KRT13, SPR2, and EMP1). In contrast to these studies, we found that GJB2 is underexpressed in tumours and DRG1 is overexpressed. In our experiments, DRG1 was shown to be overexpressed by four techniques (DD-PCR, reverse Northern, classical Northern, and RT-QPCR). Furthermore, DRG1 has been shown to be overexpressed in other tumours (see Cangul *et al*, 2002). There is some overlap between our list and profiles of other cancers (see Table 4), which potentially identifies genes with general functions in cancer.

The genes we have identified have a biased chromosomal distribution, with many located at 12p12–13 and 1q21–22 (Table 3). Out of 70 genes, six localise to 12p12–13 (A2M, GAPDH, LPRP, CLECSF2, PRH, and EMP1), and three to 1q21–22 (SPRR3, SPRR2, and S100A9). These are the two most frequently altered regions in nasopharyngeal carcinoma (Marenholz *et al*, 1996; Chen *et al*, 1999; Salomon-Nguyen *et al*, 2000; Sato *et al*, 2001), indicating that transformation has complex effects on epidermoid cell biology.

We identified sequences that might be expressed in non epidermoid cells in the tumours, including endothelial-specific and immune-related genes. EFNB2, which is overexpressed in tumours, is a trans-membrane ligand specifically expressed in arterial endothelial cells (Gale *et al*, 2001). Of the 57 known genes, 16 are immune related (Table 3), and, in particular, the nine that are overexpressed could be considered as potential circulating markers for diagnostic purposes. Certain of the immune-related genes have also been associated with epithelial tissue differentiation and growth control, including PLUNC (Iwao *et al*, 2001), PIGR (Nihei *et al*, 1996), Stat1 (Maziere *et al*, 2000), and HSPCB (Edwards *et al*, 1991), PLAUR (Chapman and Wei, 2001; Ahmed *et al*, 2002) and PLA2G7 (Tao *et al*, 1996). In particular, PLAUR is a pan T cell activating antigen that has also been associated with epithelial-derived tumour development. It interacts with integrins to regulate cell–matrix interactions (Chapman and Wei, 2001; Ahmed *et al*, 2002). PLAUR and PLA2G7 are linked, as PLAUR is activated by PAF, which in turn is a substrate of PLA2G7 (Tao *et al*, 1996).

Some of the differentially expressed genes are involved in detoxification pathways and cellular defences against insults. Physiological response to environmental insult from tobacco and

alcohol is particularly important in HNSCC (Johnson, 2001) and the differential expression of xenobiotic and detoxification enzymes has been reported in other transcriptome level studies (Alevizos *et al*, 2001). We identified two genes involved in antioxidation, GCLC (Talalay, 2000) and PON2 (Ng *et al*, 2001), and another involved in the response to oxidative damage to DNA, TDG (Laval, 1996). Cellular defences against insults could also account for the overexpression of heat shock and stress proteins, such as HSPCB, TRA1 (Maki *et al*, 1990) and DRG1 (Agarwala *et al*, 2000).

Cell-surface receptors, membrane-associated proteins and enzymes that are overexpressed in tumours are potential tumour markers and targets for drug design (Nam and Parang, 2003). We identified four overexpressed cell surface and membrane associated proteins (ITGA6, GJB2, PLAUR, and EFNB2) and nine enzymes (PON2, HPRT, HSPCB150, APOBEC3, PLA2G7, HSPCB, MTND4, DIA1, and TDG). Interestingly, inhibitors of HSPCB are currently being tested in clinical trials (Neckers, 2002).

The major strength of the PCR-DD is to identify unknown genes from limiting amounts of biological material. We identified 13 differentially expressed sequences that exist only in the EST, theoretical cDNA or hypothetical protein databases, or correspond to chromosomal locations. One of these, clone RPI-68D18, was confirmed to be overexpressed in tumours by RT-QPCR. This sequence is homologous to a number of ESTs but otherwise has no significant relationship to cDNAs or proteins in the GENEMBL databases. The differences in expression we report provide insights into the biology of HNSCC and subjects for further study. The gene products that are expressed on the cell surface or have enzymatic activity are particularly noteworthy, since successful therapeutics have been developed against these types of molecules. Finally, the novel sequences may open totally new avenues for further research and development of new therapeutics and markers.

## ACKNOWLEDGEMENTS

We would like to thank: (a) Annaïck Carles, Raymond Ripp and Olivier Poch for help with the bioinformatics; (b) the IGBMC core facilities for help and support; (c) the Fondation pour la Recherche Médicale for a fellowship for JY; (d) the Ligue Régionale (Bas-Rhin/Haut-Rhin) contre le Cancer for funding for an RT-QPCR machine; and (e) ARERS Verre Espoir (No. 138.02), Aventis, the Centre National de la Recherche Scientifique, the Institut National de la Santé et de la Recherche Médicale, the Hôpital Universitaire de Strasbourg, the Association pour la Recherche sur le Cancer, the Fondation pour la Recherche Médicale, the Ligue Nationale Française contre le Cancer (Equipe labellisée), the Ligue Régionale (Haut-Rhin) contre le Cancer, the Ligue Régionale (Bas-Rhin) contre le Cancer, the European Union (FP5 Project QLK6-2000-00159) and the Ministère de la Recherche (Décisions 99H0161 and 98C0372) for financial assistance.

## REFERENCES

- Agarwala KL, Kokame K, Kato H, Miyata T (2000) Phosphorylation of RTP, an ER stress-responsive cytoplasmic protein. *Biochem Biophys Res Commun* 272: 641–647
- Ahmed N, Pansino F, Clyde R, Murthi P, Quinn MA, Rice GE, Agrez MV, Mok S, Baker MS (2002) Overexpression of alpha(v)beta6 integrin in serous epithelial ovarian cancer regulates extracellular matrix degradation via the plasminogen activation cascade. *Carcinogenesis* 23: 237–244
- Al Moustafa AE, Alaoui-Jamali MA, Batist G, Hernandez-Perez M, Serruya C, Alpert L, Black MJ, Sladek R, Foulkes WD (2002) Identification of genes associated with head and neck carcinogenesis by cDNA microarray comparison between matched primary normal epithelial and squamous carcinoma cells. *Oncogene* 21: 2634–2640
- Alevizos I, Mahadevappa M, Zhang X, Ohyama H, Kohno Y, Posner M, Gallagher GT, Varvares M, Cohen D, Kim D, Kent R, Donoff RB, Todd R, Yung CM, Warrington JA, Wong DT (2001) Oral cancer *in vivo* gene expression profiling assisted by laser capture microdissection and microarray analysis. *Oncogene* 20: 6196–6204
- Bar-Eli M (1999) Role of interleukin-8 in tumor growth and metastasis of human melanoma. *Pathobiology* 67: 12–18
- Belbin TJ, SB, Barber I, Socci N, Wenig B, Smith R, Prystowsky MB, Childs G (2002) Molecular classification of head and neck squamous cell carcinoma using cDNA microarrays. *Cancer Res* 62: 1184–1190
- Bingle CD, Bingle L (2000) Characterisation of the human plunc gene, a gene product with an upper airways and nasopharyngeal

- restricted expression pattern. *Biochim Biophys Acta* **1493**: 363–367
- Bloor BK, Seddon SV, Morgan PR (2001) Gene expression of differentiation-specific keratins in oral epithelial dysplasia and squamous cell carcinoma. *Oral Oncol* **37**: 251–261
- Cangul H, Salnikow K, Yee H, Zagzag D, Commes T, Costa M (2002) Enhanced expression of a novel protein in human cancer cells: a potential aid to cancer diagnosis. *Cell Biol Toxicol* **18**: 87–96
- Cardillo MR, Sale P, Di Silverio F (2000) Heat shock protein-90, IL-6 and IL-10 in bladder cancer. *Anticancer Res* **20**: 4579–4583
- Chapman HA, Wei Y (2001) Protease crosstalk with integrins: the urokinase receptor paradigm. *Thromb Haemost* **86**: 124–129
- Chen YJ, Ko JY, Chen PJ, Shu CH, Hsu MT, Tsai SF, Lin CH (1999) Chromosomal aberrations in nasopharyngeal carcinoma analyzed by comparative genomic hybridization. *Genes Chromosomes Cancer* **25**: 169–175
- Cheng Q, Lau WM, Tay SK, Chew SH, Ho TH, Hui KM (2002) Identification and characterization of genes involved in the carcinogenesis of human squamous cell cervical carcinoma. *Int J Cancer* **98**: 419–426
- Connolly GP, Duley JA, Stacey NC (2001) Abnormal development of hypoxanthine–guanine phosphoribosyltransferase-deficient CNS neuroblastoma. *Brain Res* **918**: 20–27
- Coussens LM, Raymond WW, Bergers G, Laig-Webster M, Behrendtsen O, Werb Z, Coughley GH, Hanahan D (1999) Inflammatory mast cells up-regulate angiogenesis during squamous epithelial carcinogenesis. *Genes Dev* **13**: 1382–1397
- Dimery IW, Hong WK (1993) Overview of combined modality therapies for head and neck cancer. *J Natl Cancer Inst* **85**: 95–111
- Edwards MJ, Marks R, Dykes PJ, Merrett VR, Morgan HE, O'Donovan MR (1991) Heat shock proteins in cultured human keratinocytes and fibroblasts. *J Invest Dermatol* **96**: 392–396
- El-Naggar AK, Kim HW, Clayman GL, Coombes MM, Le B, Lai S, Zhan F, Luna MA, Hong WK, Lee JJ (2002) Differential expression profiling of head and neck squamous carcinoma: significance in their phenotypic and biological classification. *Oncogene* **21**: 8206–8219
- Franz O, Bruchhaus II, Roeder T (1999) Verification of differential gene transcription using virtual Northern blotting. *Nucleic Acids Res* **27**: e3
- Gale NW, Baluk P, Pan L, Kwan M, Holash J, DeChiara TM, McDonald DM, Yancopoulos GD (2001) Ephrin-B2 selectively marks arterial vessels and neovascularization sites in the adult, with expression in both endothelial and smooth muscle cells. *Dev Biol* **230**: 151–160
- Genden EM, Ferlito A, Bradley PJ, Rinaldo A, Scully C (2003) Neck disease and distant metastases. *Oral Oncol* **39**: 207–212
- Glynn-Jones E, Harper ME, Seery LT, James R, Anglin I, Morgan HE, Taylor KM, Gee JM, Nicholson RI (2001) TENB2, a proteoglycan identified in prostate cancer that is associated with disease progression and androgen independence. *Int J Cancer* **94**: 178–184
- Grossman HB, Liebert M, Lee IW, Lee SW (1994) Decreased connexin expression and intercellular communication in human bladder cancer cells. *Cancer Res* **54**: 3062–3065
- Hegele RA (1999) Paraoxonase genes and disease. *Ann Med* **31**: 217–224
- Higashi T, Sasagawa T, Inoue M, Oka R, Shuangying L, Saijoh K (2001) Overexpression of latent transforming growth factor-beta 1 (TGF-beta 1) binding protein 1 (LTBP-1) in association with TGF-beta 1 in ovarian carcinoma. *Jpn J Cancer Res* **92**: 506–515
- Hippo Y, Yashiro M, Ishii M, Taniguchi H, Tsutsumi S, Hirakawa K, Kodama T, Aburatani H (2001) Differential gene expression profiles of scirrhous gastric cancer cells with high metastatic potential to peritoneum or lymph nodes. *Cancer Res* **61**: 889–895
- Ito T, Kawabe R, Kurasono Y, Hara M, Kitamura H, Fujita K, Kanisawa M (1998) Expression of heat shock proteins in squamous cell carcinoma of the tongue: an immunohistochemical study. *J Oral Pathol Med* **27**: 18–22
- Iwao K, Watanabe T, Fujiwara Y, Takami K, Kodama K, Higashiyama M, Yokouchi H, Ozaki K, Monden M, Tanigami A (2001) Isolation of a novel human lung-specific gene, LUNX, a potential molecular marker for detection of micrometastasis in non-small-cell lung cancer. *Int J Cancer* **91**: 433–437
- Johnson N (2001) Tobacco use and oral cancer: a global perspective. *J Dent Educ* **65**: 328–339
- Kasukabe T, Okabe-Kado J, Honma Y (1997) TRA1, a novel mRNA highly expressed in leukemogenic mouse monocytic sublines but not in nonleukemogenic sublines. *Blood* **89**: 2975–2985
- Kawano T, Yanoma S, Nishimura G, Tsukuda M (2001) The inhibitory effects of TNP470 on tumour growth of head and neck carcinoma cell producing interleukin-8. *J Laryngol Otol* **115**: 802–807
- Krajci P, Meling GI, Andersen SN, Hofstad B, Vatn MH, Rognum TO, Brandtzaeg P (1996) Secretory component mRNA and protein expression in colorectal adenomas and carcinomas. *Br J Cancer* **73**: 1503–1510
- Laval J (1996) Role of DNA repair enzymes in the cellular resistance to oxidative stress. *Pathol Biol (Paris)* **44**: 14–24
- Lecoite N, Meerabux J, Ebihara M, Hill A, Young BD (1999) Molecular analysis of an unstable genomic region at chromosome band 11q23 reveals a disruption of the gene encoding the alpha2 subunit of platelet-activating factor acetylhydrolase (Pafah1a2) in human lymphoma. *Oncogene* **18**: 2852–2859
- Leethanakul C, Patel V, Gillespie J, Pallente M, Ensley JF, Koontongkaew S, Liotta LA, Emmert-Buck M, Gutkind JS (2000) Distinct pattern of expression of differentiation and growth-related genes in squamous cell carcinomas of the head and neck revealed by the use of laser capture microdissection and cDNA arrays. *Oncogene* **19**: 3220–3224
- Leon X, Ferlito A, Myer III CM, Saffiotti U, Shaha AR, Bradley PJ, Brandwein MS, Anniko M, Elluru RG, Rinaldo A (2002) Second primary tumors in head and neck cancer patients. *Acta Otolaryngol* **122**: 765–778
- Leonard JH, Kearsley JH, Chenevix-Trench G, Hayward NK (1991) Analysis of gene amplification in head-and-neck squamous-cell carcinoma. *Int J Cancer* **48**: 511–515
- Liang P (1998) Factors ensuring successful use of differential display. *Methods* **16**: 361–364
- Liss C, Fekete MJ, Hasina R, Lam CD, Lingen MW (2001) Paracrine angiogenic loop between head-and-neck squamous-cell carcinomas and macrophages. *Int J Cancer* **93**: 781–785
- Lucke T, Choudhry R, Thom R, Selmer IS, Burden AD, Hodgins MB (1999) Upregulation of connexin 26 is a feature of keratinocyte differentiation in hyperproliferative epidermis, vaginal epithelium, and buccal epithelium. *J Invest Dermatol* **112**: 354–361
- Maki RG, Old LJ, Srivastava PK (1990) Human homologue of murine tumor rejection antigen gp96: 5'-regulatory and coding regions and relationship to stress-induced proteins. *Proc Natl Acad Sci USA* **87**: 5658–5662
- Manfioletti G, Giacotti V, Bandiera A, Buratti E, Sautiere P, Cary P, Crane-Robinson C, Coles B, Goodwin GH (1991) cDNA cloning of the HMGI-C phosphoprotein, a nuclear protein associated with neoplastic and undifferentiated phenotypes. *Nucleic Acids Res* **19**: 6793–6797
- Marenholz I, Volz A, Ziegler A, Davies A, Ragoussis I, Korge BP, Mischke D (1996) Genetic analysis of the epidermal differentiation complex (EDC) on human chromosome 1q21: chromosomal orientation, new markers, and a 6-Mb YAC contig. *Genomics* **37**: 295–302
- Mazar AP (2001) The urokinase plasminogen activator receptor (uPAR) as a target for the diagnosis and therapy of cancer. *Anticancer Drugs* **12**: 387–400
- Maziere C, Dantin F, Dubois F, Santus R, Maziere J (2000) Biphasic effect of UVA radiation on STAT1 activity and tyrosine phosphorylation in cultured human keratinocytes. *Free Radic Biol Med* **28**: 1430–1437
- Mendez E, Cheng C, Farwell DG, Ricks S, Agoff SN, Futran ND, Weymuller Jr EA, Maronian NC, Zhao LP, Chen C (2002) Transcriptional expression profiles of oral squamous cell carcinomas. *Cancer* **95**: 1482–1494
- Mukhopadhyay R, Theriault RL, Price JE (1999) Increased levels of alpha6 integrins are associated with the metastatic phenotype of human breast cancer cells. *Clin Exp Metast* **17**: 325–332
- Nam NH, Parang K (2003) Current targets for anticancer drug discovery. *Curr Drug Targets* **4**: 159–179
- Neckers L (2002) Hsp90 inhibitors as novel cancer chemotherapeutic agents. *Trends Mol Med* **8**: S55–61
- Nelson AR, Fingleton B, Rothenberg ML, Matrisian LM (2000) Matrix metalloproteinases: biologic activity and clinical implications. *J Clin Oncol* **18**: 1135–1149
- Ng CJ, Wadleigh DJ, Gangopadhyay A, Hama S, Grijalva VR, Navab M, Fogelman AM, Reddy ST (2001) Paraoxonase-2 is a ubiquitously expressed protein with antioxidant properties and is capable of preventing cell-mediated oxidative modification of low density lipoprotein. *J Biol Chem* **276**: 44444–44449
- Nihei Y, Maruyama K, Endo Y, Sato T, Kobayashi K, Kaneko F (1996) Secretory component (polymeric immunoglobulin receptor) expression on human keratinocytes by stimulation with interferon-gamma and differences in response. *J Dermatol Sci* **11**: 214–222
- Rogalla P, Drechsler K, Schroder-Babo W, Eberhardt K, Bullerdiek J (1998) HMGC expression patterns in non-small lung cancer and surrounding tissue. *Anticancer Res* **18**: 3327–3330

- Romer J, Pyke C, Lund LR, Ralfkiaer E, Dano K (2001) Cancer cell expression of urokinase-type plasminogen activator receptor mRNA in squamous cell carcinomas of the skin. *J Invest Dermatol* **116**: 353–358
- Salomon-Nguyen F, Della-Valle V, Mauchauffe M, Busson-Le Coniat M, Ghysdael J, Berger R, Bernard OA (2000) The t(1;12)(q21;p13) translocation of human acute myeloblastic leukemia results in a TEL-ARNT fusion. *Proc Natl Acad Sci USA* **97**: 6757–6762
- Sankaranarayanan R, Masuyer E, Swaminathan R, Ferlay J, Whelan S (1998) Head and neck cancer: a global perspective on epidemiology and prognosis. *Anticancer Res* **18**: 4779–4786
- Sasaki H, Chen LB, Auclair D, Moriyama S, Kaji M, Fukai I, Kiriyama M, Yamakawa Y, Fujii Y (2001) Overexpression of Hrad17 gene in non-small cell lung cancers correlated with lymph node metastasis. *Lung Cancer* **34**: 47–52
- Sato Y, Kobayashi H, Suto Y, Olney HJ, Davis EM, Super HG, Espinosa III R, Le Beau MM, Rowley JD (2001) Chromosomal instability in chromosome band 12p13: multiple breaks leading to complex rearrangements including cytogenetically undetectable sub-clones. *Leukemia* **15**: 1193–1202
- Schmidt M, Grunsfelder P (2001) Urokinase-type plasminogen activator expression and proliferation stimulation in head and neck squamous cell carcinoma *in vitro* and *in situ*. *Arch Otolaryngol Head Neck Surg* **127**: 679–682
- Scully C, Field JK, Tanzawa H (2000) Genetic aberrations in oral or head and neck squamous cell carcinoma 2: chromosomal aberrations. *Oral Oncol* **36**: 311–327
- Singal R, Tu ZJ, Vanwert JM, Ginder GD, Kiang DT (2000) Modulation of the connexin26 tumor suppressor gene expression through methylation in human mammary epithelial cell lines. *Anticancer Res* **20**: 59–64
- Sirover MA (1999) New insights into an old protein: the functional diversity of mammalian glyceraldehyde-3-phosphate dehydrogenase. *Biochim Biophys Acta* **1432**: 159–184
- Sobin LH, Wittekind C (1997) *TNM Classification of Malignant Tumors*. New York: Wiley-Liss, Inc.
- St Croix B, Rago C, Velculescu V, Traverso G, Romans KE, Montgomery E, Lal A, Riggins GJ, Lengauer C, Vogelstein B, Kinzler KW (2000) Genes expressed in human tumor endothelium. *Science* **289**: 1197–1202
- Sternsdorf T, Jensen K, Zuchner D, Will H (1997) Cellular localization, expression, and structure of the nuclear dot protein 52. *J Cell Biol* **138**: 435–448
- Talalay P (2000) Chemoprotection against cancer by induction of phase 2 enzymes. *Biofactors* **12**: 5–11
- Tanaka Y, Mimori K, Shiraishi T, Ohkura Y, Takubo K, Mafune K, Barnard GF, Mori M (2000) alpha6 Integrin expression in esophageal carcinoma. *Int J Oncol* **16**: 725–729
- Tao Y, Bazan HE, Bazan NG (1996) Platelet-activating factor enhances urokinase-type plasminogen activator gene expression in corneal epithelium. *Invest Ophthalmol Vis Sci* **37**: 2037–2046
- Trenkle T, Welsh J, McClelland M (1999) Differential display probes for cDNA arrays. *Biotechniques* **27**: 554–560, 562, 564.
- van den Hooff A (1988) Stromal involvement in malignant growth. *Adv Cancer Res* **50**: 159–196
- van der Pluijm G, Sijmons B, Vloedgraven H, van der Bent C, Drijfhout JW, Verheijen J, Quax P, Karperien M, Papapoulos S, Lowik C (2001) Urokinase-receptor/integrin complexes are functionally involved in adhesion and progression of human breast cancer *in vivo*. *Am J Pathol* **159**: 971–982
- Walsh AC, Feulner JA, Reilly A (2001) Evidence for functionally significant polymorphism of human glutamate cysteine ligase catalytic subunit: association with glutathione levels and drug resistance in the National Cancer Institute tumor cell line panel. *Toxicol Sci* **61**: 218–223
- Wataba K, Saito T, Fukunaka K, Ashihara K, Nishimura M, Kudo R (2001) Over-expression of heat shock proteins in carcinogenic endometrium. *Int J Cancer* **91**: 448–456
- Xie L, Xu L, He Z, Zhou W, Wang L, Zhang L, Lan K, Ren C, Liu W, Yao K (2000) Identification of differentially expressed genes in nasopharyngeal carcinoma by means of the Atlas human cancer cDNA expression array. *J Cancer Res Clin Oncol* **126**: 400–406
- Ying H, Yu Y, Xu Y (2001) Cloning and characterization of F-LANa, upregulated in human liver cancer. *Biochem Biophys Res Commun* **286**: 394–400
- Yoshikawa Y, Mukai H, Hino F, Asada K, Kato I (2000) Isolation of two novel genes, down-regulated in gastric cancer. *Jpn J Cancer Res* **91**: 459–463
- Yousef GM, Chang A, Diamandis EP (2000) Identification and characterization of KLK-L4, a new kallikrein-like gene that appears to be down-regulated in breast cancer tissues. *J Biol Chem* **275**: 11891–11898
- Zhang H, Zhang R, Liang P (1997) Differential screening of differential display cDNA products by reverse Northern. *Methods Mol Biol* **85**: 87–93

## E. Première publication : commentaires additionnels

---

S'agissant des 70 gènes différentiellement exprimés dans les tumeurs VADS, certains commentaires relatifs aux liens entre les gènes isolés et à l'importance des phénomènes mis en jeu méritent d'être soulignés. Cette collection de gènes correspond à une fraction limitée de notre collection de marqueurs, dont l'expression a été validée de manière robuste avec une sonde focalisée complexe.

### 1. Gènes impliqués dans les phénomènes immunologiques

Seize gènes des 57 gènes connus (28%) sont associés aux phénomènes immunitaires dans la littérature, mettant ainsi en évidence l'importance de l'infiltration de cellules immunitaires au sein de la tumeur, mais aussi de l'expression aberrante d'un répertoire de gènes associés à l'immunologie par les cellules cancéreuses. Ces gènes représentent 9 des 30 gènes connus surexprimés dans les tumeurs (30%) et 7 des 27 gènes connus sous exprimés dans les tumeurs (27%). Les gènes surexprimés dans les tumeurs sont particulièrement intéressants en tant que potentiels marqueurs de la carcinogenèse. Outre les liens entre uPAR (Urokinase Plasminogen Activator Receptor), PLA2G7 (Phospholipase A2 group VII) et les intégrines évoqués dans notre article, la Kallikrein-like L4 (KLK-L4), sous exprimée dans les tumeurs, est reliée à uPAR. En effet, la kallikreine clive les kininogènes de haut poids moléculaires HK en HKa, molécule qui interagit alors avec les intégrines  $\beta 2$  qui, en retour, interagissent avec uPAR pour exprimer ses propriétés anti-adhésives (Chavakis et al., 2000).

Il est difficile d'assigner une seule et unique fonction à un gène associé aux phénomènes immunologiques, tant ces phénomènes sont complexes et impliquent des effets indirects. Les signaux immunologiques échangés au sein du microenvironnement tumoral forment en effet un réseau très complexe. Par exemple, les macrophages associés aux tumeurs (Tumour Associated Macrophages, TAMs) peuvent participer à l'angiogenèse tumorale. Les cellules cancéreuses produisent des protéines chimiotactiques comme MCP1 (Macrophage Chemotactic Protein 1) et TGF  $\beta 1$  (Transforming Growth Factor  $\beta 1$ ) qui est responsable de l'activation des macrophages. En retour, les macrophages activés sécrètent de manière paracrine des molécules de TNF- $\alpha$  (Tumour Necrosis Fator  $\alpha$ ) et d'Interleukine 1 (IL1). Ces

molécules stimulent la sécrétion des facteurs angiogéniques tels IL8 et VEGF (Vascular Endothelial Growth Factor) par les cellules cancéreuses (Liss et al., 2001).

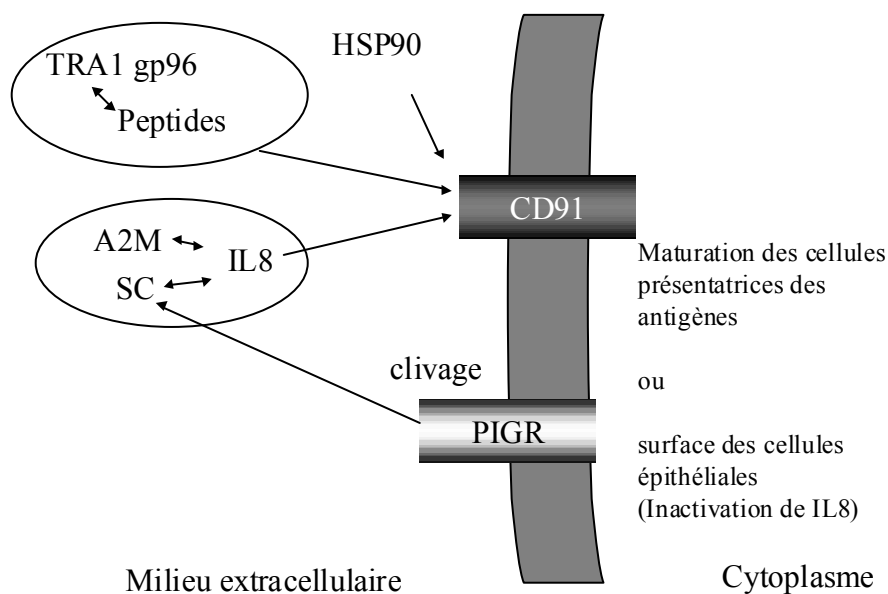
Nous avons isolé des gènes associés au recrutement des monocytes ou à leur activation en macrophages lors des procédés inflammatoires. Plus précisément nous avons isolé des gènes impliqués dans l'absorption des antigènes et associés à la présentation des antigènes par les cellules présentatrices des antigènes APCs (antigen presenting cells), que sont les macrophages activés et les cellules dendritiques. Plusieurs gènes exprimés par les macrophages activés participent au complexe dialogue entre cellules cancéreuses et cellules immunitaires, tels LTBP1 (Latent TGF  $\beta$  Binding Protein 1) et IL8. Les cellules épithéliales de la peau en contact avec le milieu extérieur produisent des cytokines induisant la production de cytokines pro inflammatoires par les fibroblastes du derme.

De manière remarquable, le gène MIF-related Calcium binding protein localisée en 1q21 est associée à la réponse immunitaire, principalement à l'activation des macrophages, ainsi qu'à la différenciation des cellules épithéliales. Plusieurs autres gènes comme BLT1, CAPN7 (Calpain Like Protease 7) et uPA sont impliqués dans la régulation du niveau intracellulaire de  $Ca^{2+}$ , et, possèdent une double fonction dans le contrôle de la différenciation des kératinocytes et dans la réponse immunitaire. LTB4 (Leukotrien B4 receptor) est impliqué dans le chimiotactisme lié à la régulation du niveau de  $Ca^{2+}$  intracellulaire, et, est capable d'induire, comme IL8, l'adhésion des neutrophiles aux molécules d'adhésion ICAM (Intercellular Adhesion Molecules). IL8 et LTB4 peuvent stimuler la migration indépendante des intégrines (CD18) des neutrophiles à travers les cellules endothéliales.

En 12p13, nous avons isolé le gène CLECSF2 (C-type Lectin Superfamily member 2) impliqué dans l'activation des lymphocytes en accord avec la présence d'un cluster de gènes impliqués dans l'immunité, appelé NK gene complex (Brown et al., 1997).

Plusieurs gènes sont des composants d'une même voie (FigIII.13) comme gp96/TRA1, A2M (Alpha 2 Macroglobulin), IL8 (Interleukin 8) et PIGR (calmodulin related poly-immunoglobulin receptor). La protéine de choc thermique gp96, comme la protéine apparentée hsp90, toutes deux surexprimées dans les tumeurs, sont de nouvelles cibles thérapeutiques pour le traitement du cancer selon une stratégie de vaccination (Belli et al., 2002; Maki et al., 1990; Yedavelli et al., 1999). La protéine gp96 est une protéine

chaperonne qui, entre autres, maintient la structure des protéines oncogéniques. La protéine gp96 est impliquée dans l'absorption des peptides exogènes pour la présentation des antigènes par les cellules APC. Cette fonction dépend du récepteur CD91 (Basu et al., 2001) qui lie également l'alpha-2-macroglobuline. L'absorption d'IL8 est régulée par ces mêmes effecteurs. Gp96 induit la maturation des cellules dendritiques et réprime l'expression de son récepteur, de la même manière que d'autres récepteurs associés à l'absorption des antigènes, comme les intégrines  $\alpha 5 \beta 3$  et  $\alpha 5 \beta 5$ . L'alpha 2 macroglobuline est un inhibiteur des protéases qui pourrait protéger les molécules d'adhésion de la dégradation et ainsi jouer un rôle dans la propagation/ l'étalement des cellules. IL8 est une molécule chimiotactique puissante pour les monocytes et un facteur angiogénique distinguant les tumeurs bénignes des tumeurs malignes. La production d'IL8 est induite par divers stimuli comme l'hypoxie, l'hyperthermie et l'oxyde nitrique, impliqués dans les tumeurs et dans les processus inflammatoires. Une hypoxie non létale peut induire une sécrétion du TNF  $\alpha$ , d'IL1 et d'IL8 par les macrophages. Nous avons également isolé une diminution d'expression de PIGR. PIGR est responsable du transport des immunoglobulines IgA complexées aux IgM de la surface basolatérale au pôle apical des cellules épithéliales. Un clivage de PIGR a lieu au cours du transport, entraînant la libération de son domaine de liaison extracellulaire, le Secretory Component (SC). L'alpha-2-macroglobuline est un membre de la superfamille des immunoglobulines (Ig) apparentée à SC. Tout comme l'alpha-2-macroglobuline, SC lie IL8 (Marshall et al., 2001). Dans les voies aériennes, la liaison de PIGR solubilise et inactive IL8. L' IgA sécrétée, SIgA, est l'effecteur le mieux caractérisé du système immunitaire dans les muqueuses et son transport par PIGR requiert localement la coopération des cellules épithéliales et des cellules immunes du plasma.



Exemple d'interactions fonctionnelles (CD91/PIGR)

Figure III.13

## 2. Enzymes

Si les enzymes dans leur ensemble représentent une proportion importante de nos listes de gènes différentiellement exprimés, les enzymes de détoxification sont une cible privilégiée du développement de drogues, d'inhibiteurs. Dans notre étude, nous avons identifié 9 enzymes à partir des 30 gènes connus surexprimés dans les tumeurs (30%) et 7 enzymes à partir des 27 gènes connus sous exprimés dans les tumeurs (26%). HSCP150 (ubiquitin conjugating enzyme E2) et PSMB8 (proteasome subunit beta type 8), sous unité du protéasome possédant une activité catalytique, sont impliquées dans la dégradation des protéines. Dans les tumeurs, si le gène GAPDH est sous exprimé, le gène HPRT est surexprimé. Ce résultat est particulièrement intéressant puisque ces gènes ont longtemps été considéré comme des gènes contrôle invariants (« housekeeping genes »), et, ont donc été utilisés pour la normalisation d'études de l'expression génique, y compris dans les tumeurs. Des variations d'expression de ces gènes de ménage ont pourtant déjà été décrites dans différentes situations biologiques. Par exemple, GAPDH est surexprimé dans les cancers de la prostate (Rondinelli et al., 1997) et les carcinomes du rein (Vila et al., 2000). Nos données posent donc le problème de l'invariabilité de l'expression du gène GAPDH, gène de ménage largement utilisé dans toutes les étapes de validation nécessaires à l'analyse du transcriptome.

En accord avec nos données, certains auteurs ont proposé qu'il n'existe virtuellement aucun gène invariant (Lee et al., 2002). Pour exemple supplémentaire, deux différentes sous unités de l'enzyme NADH déhydrogénase sont régulées différemment, puisque la sous unité 2 (MTDN2) est sous exprimée dans les tumeurs tandis que la sous unité 4 est surexprimée dans les tumeurs. La KLK-L4 (Kallikrein-Like protein 4), apparentée au facteur PSA (Prostate Specific Antigen), également sous-exprimée dans le cancer du sein (Yousef et al., 2000), et la protéase CAPN7 (Franz et al., 1999b) sont toutes deux des protéases possédant des propriétés pro-apoptotiques sous exprimées dans les tumeurs (Yousef & Diamandis, 2002).

### **3. Gènes localisés en 12p13 et 1q21-22**

Nous avons observé une représentation anormalement élevée de gènes situés sur le chromosome 1 (1q21-q22) et le chromosome 12 (12p12-13) dans notre liste finale. Sur un total de 70 gènes, nous trouvons en effet 8 gènes sur le chromosome 12 sous exprimés dans les tumeurs, dont 5 au locus 12p12-13 (GAPDH, LPRP, CLECSF2, PRH, EMP1). De plus, 5 autres clones localisés sur le chromosome 12 sont surexprimés dans les tumeurs, dont 1 localisé en 12p12-13 (A2M). Nous avons isolé 1 gène surexprimé dans les tumeurs localisé en 1q31 ainsi que 4 gènes sous exprimés dans les tumeurs localisés sur le chromosome 1, dont 3 localisés précisément en 1q21 (Small Proline Rich Proteins SPRR2&SPRR3, S100A9). L'expression de la plupart de ces gènes est fortement diminuée dans les tumeurs VADS. Or ces deux régions sont les loci chromosomiques les plus fréquemment amplifiées dans les cancers du nasopharynx (Chen et al., 1999) et possiblement dans d'autres types de cancer (Salomon-Nguyen et al., 2000; Sato et al., 2001 ; Singh et al., 2001 ). L'instabilité génétique de ces régions contribue peut être aux variations de l'expression génique mesurées. Ces deux loci présentent des caractéristiques communes, comme une forte représentation de gènes de la famille « proline rich » (Tesfaigzi & Carlson, 1999), et la concentration de gènes impliqués dans la différenciation terminale des kératinocytes (Chen et al., 1997). La bande cytogénétique 1q21-22 contient en effet le complexe de différenciation épidermique (Epidermal Differentiation Complex, EDC), un groupe de gènes dont l'expression est strictement régulée et dont la fonction est étroitement associée à la différenciation terminale des kératinocytes (Marenholz et al., 1996). Un complexe de gènes associés à l'immunologie, le « NK complex » est localisé en 12p13. La perte de l'expression des gènes en 12p13 et 1q21 pourrait refléter la dédifférenciation des cellules épithéliales au sein de la masse tumorale. Une autre explication



pourrait être une expression particulièrement forte de ces gènes en rapport avec leur fonction (protéines structurales), ou encore, un biais particulièrement fort des amorces correspondant à ces gènes, leur assurant une forte amplification malgré les phénomènes de compétition entre amorces dans les sondes « DD ».

A l'opposé, nous avons identifié une série de gènes surexprimés au sein de la masse tumorale qui sont associés aux kératinocytes comme ITGA6 {Intégrine alpha 6 ; (Tarone et al., 2000)}, DRG1 {Differentiation Related Gene 1 ; (van Belzen et al., 1997)} et APOBEC3, aussi appelée phorboline 1 (Madsen et al., 1999). Certains transcrits spécifiques des kératinocytes seraient exprimés à un fort niveau au sein des tumeurs, peut être en lien avec un besoin continu des cellules cancéreuses de protéines impliquées dans la biologie des cellules épidermoïdes.

#### **4. Comparaison avec une série d'autres études sur les cancers VADS**

Récemment est apparue une série de publications établissant des profils d'expression génique des tumeurs VADS (Al Moustafa et al., 2002; Alevizos et al., 2001 ; Belbin TJ, 2002 ; El-Naggar et al., 2002 ; Leethanakul et al., 2000 ; Mendez et al., 2002 ; Xie et al., 2000 ). Ces travaux font appel à des systèmes fermés comme les membranes Atlas®, et les puces à ADN commerciales Affymetrix®. Apparemment les gènes isolés dans ces diverses études sont très faiblement redondants. Bien que des gènes communs soient isolés de manière systématique, ces différences de résultats peuvent s'expliquer par des différences de matériel tumoral ou de stratégie expérimentale (stade tumoral analysé, localisation précise, degré d'hétérogénéité des tumeurs, technologie employée).

Certaines de ces études ont été complètement (Al Moustafa et al., 2002) ou partiellement réalisées (lignée NHEK-6168 comme contrôle normal pour Belbin et al.) à partir de lignées cellulaires en culture, entraînant une possible déformation des résultats par rapport à l'expression des gènes *in vivo* dans les tumeurs. De plus, beaucoup de ces études ont été menées sur une collection de tumeurs d'origines variées au sein de la sphère ORL incluant des cancers de la cavité buccale, des gencives, des amygdales, du palais, de la mandibule, ou encore de tumeurs laryngées et pharyngées. L'étude de El Naggar inclut également des variants phénotypiques atypiques et rares des tumeurs VADS, des tumeurs dites non conventionnelles. Certaines études se limitent à l'analyse de tumeurs de la cavité buccale (Alevizos et al., Mendez et al.), tandis que notre étude est menée strictement sur une autre

localisation, l'hypopharynx. Les signatures isolées dans ces études sont capables de discriminer les tumeurs des tissus normaux. Toutefois, elles ne sont pas capables de discriminer les tumeurs en fonction de leur stade clinique ou de la durée de survie du patient. El Naggar est capable de discriminer les variants non conventionnels des tumeurs VADS conventionnelles, mais cette distinction ne nécessite pas de diagnostic moléculaire puisque ces variants sont phénotypiquement différents des autres tumeurs VADS. Belbin, bien qu'utilisant des tumeurs de localisation variable, définit deux groupes de tumeurs aux durées moyennes de survie différentes. Cependant, un nombre important de patients regroupés au sein de l'un ou l'autre groupe est décédé d'une cause autre que le cancer (4/8 pour le groupe 1 et 4/9 pour le groupe 2) et, l'âge moyen des patients des 2 groupes est différent. Au sein de chacun des deux groupes, des patients ont présenté une absence de maladie au terme du suivi clinique ou au contraire sont décédés de cancer au cours du suivi clinique. Bien qu'étant un meilleur élément prédictif que les méthodes concurrentes comme les auteurs l'affirment, cette collection prédictive de gènes implique encore de forts risques de classification erronée.

Parmi la liste des 57 gènes connus présentés dans notre travail, 8 gènes (14%) sont également décrits dans ces études concurrentes. Ainsi PON2 (Paraoxanase 2) et ITGA6 (intégrine  $\alpha 6$ ) sont surexprimées dans les tumeurs en accord avec Mendez et al. Xie et al ont, comme nous, rapporté une surexpression de STAT1 (Signal Transducer and Activator of Transcription 1) dans les tumeurs nasopharyngées. La cytokératine 13 (Mendez, Moustafa, El Naggar), SPRR2 (Moustafa, El Naggar), et EMP1 (Alevizos) ont été, en accord avec nos résultats, décrits comme sous exprimés dans les tumeurs. El Naggar a décrit une surexpression dans les tumeurs de GJB2 (Gap Junction Protein Beta 2, aussi appelée connexin26) tandis que nous rapportons sa sous expression. Il est toutefois difficile d'extrapoler la signification de l'expression dans les variants non conventionnels des tumeurs VADS, en regard de la biologie de l'épithélium hypopharyngé normal. En contradiction avec Moustafa et Alevizos, nous avons décrit une surexpression de DRG1 (aussi appelé RTP et Cap43) dans les tumeurs hypopharyngées. A la différence de ces équipes concurrentes, nous avons validé l'expression de ce gène par différentes techniques (Virtual Northern blot, Northern blot et PCR en temps réel). De plus, la protéine Cap43 a déjà été décrite comme étant surexprimée dans les tumeurs, comme cité dans notre article. La différence de profil d'expression observée peut refléter la disparité des types tumoraux analysés, comme discuté précédemment. La faible redondance entre notre étude et les études antérieures sur un ensemble hétérogènes de tumeurs VADS

nous conduit à penser que nous contribuons à mettre en évidence des altérations géniques propres à la carcinogénèse des carcinomes à cellules squameuses.

Cependant, en regroupant les données de toutes ces études, la notre comprise, des voies métaboliques et des voies de signalisation entières apparaissent communément affectées au cours de la carcinogénèse (FigIII.14).

Les voies impliquant les interférons (IFN) et plus particulièrement l'IFN  $\gamma$  sont clairement surexprimées dans les tumeurs. Les interleukines telles que IL1, IL8, IL10, IL6 sont également surexprimées dans les tumeurs. Les gènes de la famille TGF, et TGF  $\beta$  en particulier, sont surexprimés dans les tumeurs. Dans le cas de TGF  $\beta$ , un examen de la littérature indique un rôle critique des altérations de l'expression génique lors la progression tumorale. La surexpression de TGF  $\beta$  correspond à des stades tumoraux relativement avancés, les cellules tumorales ayant acquis une résistance à TGF  $\beta$  et étant alors dotées d'un avantage sélectif lors de la sécrétion de TGF  $\beta$ , tandis que l'hypoexpression de TGF  $\beta$  est associée aux stades précoces du développement tumoral (Derynck et al., 2001; Reiss, 1997). Certains gènes responsables des interactions intercellulaires comme les intégrines et particulièrement les intégrines  $\alpha$ , sont fréquemment surexprimées dans les tumeurs. Le collagène (et ses diverses isoformes), la laminine et la fibronectine, responsables de l'adhésion cellulaire, sont fréquemment surexprimés dans les tumeurs. De même, les enzymes favorisant la croissance des tumeurs par dégradation de la matrice extracellulaire, comme les MMPs (matrix metalloproteinase), les facteurs de croissance VEGF (Vascular Endothelial Growth Factor), PDGF (Platelet Derived Growth factor), FGF (Fibroblast Growth Factor), IGF (Insulin-like Growth Factor) et certaines des protéines associées à leur fonction sont surexprimés dans les tumeurs. La croissance des divers types cellulaires constituant la tumeur semble donc stimulée. Les protéines reflétant le stress de l'environnement tumoral, comme les protéines de choc thermique (heat shock proteins), sont surexprimées. Les voies impliquant le plasminogène, également connu pour son implication dans les procédés de cicatrisation (« wound healing »), sont largement surexprimées dans les tumeurs, comme l'indiquent l'isolement de uPA (Urokinase Plasminogen Activator), uPAR (uPA receptor), PAI1 (Plasminogen Activator Inhibitor 1), et de PAFah (Plasminogen Activating Factor Acetylhydrolase, aussi appelée PLA2G7, phospholipase A2 group VII). Les protéines impliquées dans la dégradation des protéines, l'ubiquitine (et les ubiquitin conjugating enzymes), et le protéasome sont surexprimées dans les tumeurs. La voie des leukotrienes, impliquée dans les processus immunitaires non spécifiques et inflammatoires, est activée dans les tumeurs. On sait qu'en cas d'infection ou de blessure, la prolifération est favorisée, et les

blocages du cycle cellulaire sont parfois levés, ce qui pourrait expliquer la surexpression des acteurs de ces processus dans les tumeurs. Enfin de nombreuses kinases impliquées dans la transduction des facteurs promoteurs de la croissance telles que les MAP Kinases (Mitogen Activated Protein Kinases), JNK (Jun N-terminal kinase) et MEK (MAPK/ERK Kinase) sont elles aussi surexprimées dans les tumeurs.

Familles de gènes isolées/ fonctions isolées dans les diverses études	Exemples	Profil d'expression
Interleukines	IL1, IL8, IL10, IL6	T+
Transforming growth factor beta	TGF beta, Binding proteins comme LTBP1	complexe en fonction de la progression tumorale, T+
Enzymes de dégradation de la matrice extracellulaire	MMPs	T+
Facteurs de croissance	VEGF, PDGF, FGF, IGF	T+
Stress	protéines de choc thermique, chaperones	T+
Inflammation/ cicatrisation	uPA, uPAR, PAI1, PAFah, PLA2G7, leukotrienes	T+
Dégradation des protéines	Ubiquitin conjugating enzymes	T+
Kinases/transduction des signaux extracellulaires	MAP kinases, JNK, MEK	T+
Différenciation des kératinocytes	SPRR, kératines, transglutaminases	T-
Cycle cellulaire	cyclines	T-
TNF alpha	TNF alpha et protéines inductibles par TNF alpha	T-

Figure III.14

A l'opposé, un ensemble de voies sont inactivées dans les tumeurs. Des familles de gènes présentent des profils d'expression différents en fonction de l'effecteur précis envisagé et peuvent avoir un rôle crucial dans la carcinogenèse des carcinomes. Les petites protéines riches en proline (SPRR), à l'exception de SPRR1B surexprimé dans les tumeurs d'après El Naggar, semblent généralement très sous exprimées dans les tumeurs, potentiellement en relation avec leur rôle supposé dans la différenciation terminale des kératinocytes et l'établissement de l'enveloppe cornée (Cornified Enveloppe) de l'épithélium. Egaleme nt liées à la différenciation des kératinocytes, les kératines 2, 4, 15 et 13 sont sous exprimées dans les

tumeurs, tandis qu'une surexpression a été rapportée pour les kératines 5, 6B, 14, 17 et 16 (des résultats contradictoires existent pour les kératines 17 et 16). Les protéines ribosomiques (ribosomal proteins) présentent un profil d'expression hétérogène puisque RPL6 et RPL27 sont surexprimées dans les tumeurs et les protéines ribosomales S6, S19, L14, L32 et L36A sont sous exprimées dans les tumeurs. Les gènes de la famille des cyclines, cycline A1, cycline B1, cyclineD2 et cyclineG2 sont fréquemment rapportées comme étant sous exprimées dans les tumeurs. Les gènes reliés au TNF alpha sont fréquemment sous exprimés dans les tumeurs. Les enzymes comme les transglutaminases 1 et 3 sont sous exprimées dans les tumeurs, peut être en relation avec leur activité dans la différenciation des kératinocytes et la constitution de l'enveloppe cornée. Les enzymes de détoxification Glutathion S transférase (bien que surexprimée lors des phénomènes de résistance aux drogues), ainsi que les aldéhyde déhydrogénases 9 et 10 sont sous exprimées dans les tumeurs.

## **5. Fréquence élevée des pertes d'expression dans les tumeurs**

Tout au long de notre étude, avec toutes les méthodes de validation (Virtual Northern blot, Northern blot et PCR en temps réel), les événements de diminution d'expression dans les tumeurs ont été retrouvés à une très forte fréquence chez les patients, par rapport à une plus grande hétérogénéité des événements de surexpression dans les tumeurs. Les pertes d'expression paraissent plus marquées, fréquentes et homogènes que les événements de surexpression lorsque l'analyse quitte le cadre des pools de patients. Les événements de surexpression, moins fréquemment détectés et plus hétérogènes, semblent correspondre à l'acquisition par la tumeur de propriétés spécifiques à la tumeur menant à une évolution clinique différente. Le maintien d'un état normal est, apparemment, un phénomène régulé de manière extrêmement étroite comme indiqué par le très faible pourcentage de bandes de DD présentant un profil différentiel entre les différents pools de tissus normaux. Les pertes d'expression peuvent être observées dès les stades précoces (groupe E) de la carcinogenèse, et, pourraient être dues à la nécessité d'une perte allélique complète ou une mise sous silence (silencing) complète de gènes impliqués dans des voies de sauvegarde et de surveillance des cellules (gatekeepers et caretakers). On peut également considérer, cet ensemble d'événements communs à tous les types tumoraux analysés, y compris les tumeurs précoces, pourrait refléter une taille assez élevée au moment de l'intervention chirurgicale de ces

tumeurs. Aussi ces tumeurs ne représenteraient pas idéalement les événements les plus précoces de l'initiation tumorale. L'analyse de notre collection de gènes au niveau de lésions pré malignes d'origine VADS, en particulier les tumeurs indiennes étudiées par le Dr Ranju Rahlan, pourrait conduire à l'identification des événements spécifiquement impliqués dans l'initiation tumorale.

De manière analogue, certains gènes surexprimés identifiés par Mendez sont détectés précocement dans les lésions pré malignes. La majorité des surexpressions dans les tumeurs étant transversales, il semble qu'un grand nombre de caractéristiques tumorales soit acquis lors des stades précoces de la carcinogenèse. Dans notre étude, la faible proportion des différences liées à l'évolution tumorale, provient sans doute du fait de la recherche de gènes prédictifs du devenir tumoral au niveau des tumeurs primaires et non de biopsies de métastases pour lesquelles une colonisation et une adaptation au tissu envahi ont dû avoir lieu. Les collections établies par analyse de métastases ne sont, par contre, pas nécessairement prédictives du devenir tumoral à partir d'échantillons tumoraux prélevés lors des stades initiaux du traitement.

**Partie IV**  
**Comparaison des études par puces à ADN**  
**Affymetrix et par Differential Display**

## A. 2<sup>ème</sup> publication

---

Cette publication (Cromer et al., 2003) traite de l'analyse par puces à ADN Affymetrix d'une série de tumeurs hypopharyngées bien caractérisées au plan clinique (stade, traitement) et histologique, comparable à celle utilisée pour l'étude DD. Nous avons ainsi élargi la recherche de marqueurs pronostiques, sur des gènes principalement connus, avec un biais technologique différent. Cette analyse permet d'identifier des marqueurs potentiels parmi des gènes connus, et, *a priori* exprimés à plus fort niveau qu'avec la technique de DD (Lemaire et al., 2003).



## ORIGINAL PAPER

**Identification of genes associated with tumorigenesis and metastatic potential of hypopharyngeal cancer by microarray analysis**

Anne Cromer<sup>1,3</sup>, Annaïck Carles<sup>1,3</sup>, Régine Millon<sup>2,3</sup>, Gitali Ganguli<sup>1</sup>, Frédéric Chalmel<sup>1</sup>, Frédéric Lemaire<sup>1</sup>, Julia Young<sup>1</sup>, Doulaye Dembélé<sup>1</sup>, Christelle Thibault<sup>1</sup>, Danièle Muller<sup>2</sup>, Olivier Poch<sup>1</sup>, Joseph Abecassis<sup>2</sup> and Bohdan Wasylyk<sup>\*,1</sup>

<sup>1</sup>Institut de Génétique et de Biologie Moléculaire et Cellulaire, CNRS/INSERM/IULP, 1 Rue Laurent Fries, BP 10142, 67404 Illkirch cedex, France; <sup>2</sup>UPRES EA 34-30, Centre Paul Strauss, 3 rue de la Porte de l'Hôpital, 67085 Strasbourg, France

Head and neck squamous cell carcinoma (HNSCC) is the sixth most common cancer among men in the developed world. There is a need, for both clinical and scientific reasons, to find markers to identify patients with aggressive disease as early as possible, and to understand the events leading to malignant transformation and susceptibility to metastasis. We report the first large-scale gene expression analysis of a unique HNSCC location, the hypopharynx. Four normal and 34 tumour samples were analysed with 12 600 gene microarrays. Clusters of differentially expressed genes were identified in the chromosomal regions 3q27.3, 17q21.2–q21.31, 7q11.22–q22.1 and 11q13.1–q13.3, which, interestingly, have already been identified by comparative genomic hybridization (CGH) as major regions of gene amplification. We showed that six overexpressed genes (EIF4G1, DVL3, EPHB4, MCM7, BRMS1 and SART1) located in these regions are indeed amplified. We report 119 genes that are highly differentially expressed between 'early' tumours and normal samples. Of these, we validated by quantitative PCR six novel poorly characterized genes. These genes are potential new markers of HNSCC. Comparing patients with relatively nonaggressive and aggressive tumours (without or with clinical evidence of metastasis 3 years after surgery), we identified 164 differentially expressed genes potentially involved in the acquisition of metastatic potential. This study contributes to the understanding of HNSCC, staging patients into prognostic groups and identifying high-risk patients who may benefit from more aggressive treatment.

*Oncogene* advance online publication, 15 December 2003; doi:10.1038/sj.onc.1207345

**Keywords:** HNSCC; chromosomal location; prognosis; gene amplification

**Introduction**

Head and neck squamous cell carcinoma (HNSCC), which affects the oral cavity, the oropharynx, the larynx and the hypopharynx, is the sixth most common cancer among men in the developed world. Well-known risk factors include tobacco and alcohol. Over the last decades, diagnosis and management have improved, but not long-term survival rates. The prognosis of HNSCC is influenced by many factors, such as TNM staging and pathological grading of differentiation. However, since these factors are not sufficient to evaluate outcome, there is a need to identify molecular biomarkers that will help to stage patients in prognostic groups and to identify high-risk patients who may benefit from different treatments. Some genes have been suggested to be potential prognostic markers in HNSCC (Quon *et al.*, 2001). However, molecular markers do not yet contribute to the clinical decision-making process.

Cancer appears to result from the progressive accumulation of genetic aberrations. Amplified chromosomal regions may contain dominant oncogenes, whereas deleted regions may harbour tumour suppressor genes. HNSCC is frequently associated with specific chromosomal aberrations, including amplification of 3q, 8q, 9q, 20q, 7p, 11q13 and 5p, and deletion of 3p, 9p, 21q, 5q, 13q, 18q and 8p (Gollin, 2001). mRNA expression levels often reflect these changes. However, the deregulation of gene expression in cancer is more complex than a simple relationship between genomic aberration and gene expression (Platzer *et al.*, 2002; Pollack *et al.*, 2002).

Expression-array profiling has recently been used to distinguish between cancer subtypes (Golub *et al.*, 1999) and stages of progression (Bittner *et al.*, 2000). The capacity of tumour cells to metastasize could be acquired early in tumorigenesis (Bernards and Weinberg, 2002), and patterns of gene expression can predict the metastatic outcome of patients with breast (van't Veer *et al.*, 2002) and prostate (Singh *et al.*, 2002) cancer. Microarray analysis of gene expression has been reported for HNSCC (Leethanakul *et al.*, 2000; Alevizos *et al.*, 2001; Al Moustafa *et al.*, 2002; El-Naggar *et al.*, 2002; Mendez *et al.*, 2002), but did not concern

\*Correspondence: B Wasylyk; E-mail: boh@igbmc.u-strasbg.fr

<sup>3</sup>These authors contributed equally to this work

Received 22 July 2003; revised 22 October 2003; accepted 11 November 2003

well-defined stages of tumour evolution in a precise head and neck site.

Here we describe the first large-scale gene expression analysis of the hypopharynx, a localization associated with particularly aggressive behaviour (Genden *et al.*, 2003). The aim of our study was to identify genes involved in tumorigenesis, as well as gene expression patterns that will distinguish tumours that will metastasize from those that will not. We have identified by microarray analysis genes whose expression differs between tumours and normal tissues, as well as between

tumours with similar initial stage and histopathological features but with different clinical outcomes.

## Results and discussion

We determined with Affymetrix HG-U95A microarrays the expression profiles of 34 hypopharyngeal cancer samples (including 31 individual tumours, two pools of two and one pool of three additional tumours) and four

**Table 1** Patient and sample characteristics

Class	Patient	Sample	Localiz.	T	N	M	Diff	Sex	Age	Treatment after surgery (4)	Evolution	Actual state	Overall survival (5)	Disease free survival (5)
<i>Normal</i>														
N	pool(1)*	N	U											
N	1047(2)*	N	U											
N (3)	1102*	N	U											
N (3)	1107*	N	U											
<i>'Early' tumours</i>														
Pool_E	PE1*	T	H	2	0	0	2	M	53	RX	0	D	15	11
	PE2*	T	H	1	0	0	2	M	61	N	0	D	25	15
	PE3*	T	H	2	0	0	2	M	44	N	0	A	41	41
E	1047(2)*	T	H	2	0	0	1	M	68	RX	0	A	12	12
E	435*	T	H	1	0	0	1	M	72	RX	SC	A	92	91
E	834*	T	H	1	0	0	2	M	52	N	SC	D	61	42
<i>'Late' tumours</i>														
LR	118	T	H	4	2b	0	1	M	51	RX	LR	D	39	35
LR	429	T	H	3	3	0	2	M	58	RX	LR	D	77	37
LR	684	T	H	3	1	0	3	M	65	RX	LR	D	40	29
LR	702	T	H	3	2c	0	3	M	70	RX	LR	D	5	4
Pool_NM	PNM1*	T	H	3	2b	0	3	M	66	RX	0	D	58	54
	PNM2*	T	H	3	2b	0	2	M	52	RX	SC	A	44	39
NM	065*	T	H	4	2c	0	1	M	67	RX	0	A	118	118
NM	103	T	H	3	1	0	2	M	55	RX	0	A	143	95
NM	167	T	H	4	2c	0	2	M	61	RX	0	A	140	86
NM	190*	T	H	1	2b	0	1	M	70	RX	0	A	63	39
NM	279*	T	H	3	2c	0	1	M	53	RX	0	A	104	104
NM	357	T	H	3	2c	0	2	M	45	RX	0	A	101	101
NM	409	T	H	4	1	0	2	M	43	RX	0	A	116	116
NM	423	T	H	2	2b	0	2	M	64	RX	0	A	83	83
NM	847	T	H	2	1	0	3	F	63	RX	0	A	67	67
NM	856*	T	H	3	2b	0	1	M	59	RX	0	A	66	66
Pool_M	PM1*	T	H	2	2c	0	3	M	54	RX	M	D	25	16
	PM2*	T	H	3	2b	0	2	M	54	RX	M	D	7	5
M	135*	T	H	3	2c	0	2	M	51	RX	M	D	84	14
M	165	T	H	4	2c	0	2	M	71	RX	M	D	22	9
M	203	T	H	3	2b	0	2	M	54	RX	M	D	20	11
M	209*	T	H	3	2c	0	2	M	62	RX	M	D	18	8
M	215	T	H	4	3	0	2	M	43	RX+CT	M	D	12	3
M	218	T	H	3	3	0	2	M	56	RX	M	D	12	8
M	330	T	H	3	2b	0	3	M	63	RX	M	D	7	7
M	408	T	H	1	2c	0	3	M	49	RX+CT	M	A	16	1
M	621	T	H	3	3	0	3	M	41	RX+CT	M	D	34	28
M	629	T	H	3	3	0	2	F	57	RX+CT	M	D	8	8
M	744	T	H	3	2c	0	2	M	51	RX	M	D	9	8
M	829	T	H	4	3	0	2	M	69	RX	M	D	5	1
M	888*	T	H	3	2b	0	2	M	64	RX	M	D	26	21
M	933*	T	H	2	2b	0	2	M	70	RX	M	D	23	22

Sample type: T = tumour, N = normal; localization (localiz.): H = hypopharynx, U = uvula; T, N and M reflect the TNM nomenclature for tumour stage; evolution: 0 = no evolution, LR = local recurrence, M = metastasis, SC = secondary cancer; last follow-up status: D = dead, A = alive; \*samples from batch 1. (1) The N pool contains the normal samples corresponding to the tumours in the TE, TNM and TM pools, the E pool consists of three tumours (PE1, PE2 and PE3), the NM pool two (PNM1 and PNM2) and the M pool two (PM1 and PM2); (2) 1047T and 1047N are tumour and normal samples from the same patient; (3) samples were macrodissected to avoid nonepithelial cells; (4) RX = radiotherapy treated, RX + CT = radiotherapy and chemotherapy treated, N = no treatment; (5) in months

normal samples (including one pool of seven normal tissues from patients whose tumours are in the tumour pools). A full description of the clinical data, including diagnosis and outcome, is shown in Table 1. We used three different approaches to analyse the data, in order to identify gene expression signatures of tumorigenesis and metastatic evolution.

### Global gene expression profile

The data from 34 hypopharyngeal cancer and four normal samples were filtered to exclude genes with low expression, resulting in a working set of 3962 probe sets (see Materials and methods, Preprocessing of the data). We were interested to know if there are patterns of gene expression that cluster samples and single out particular subgroups. Unsupervised hierarchical clustering was used to group samples according to similarity in gene expression, without prior knowledge of sample identity and without any gene selection (Figure 1a). The four normal (N) samples cluster together, indicating a strong similarity in their expression profiles. The tumour (T)

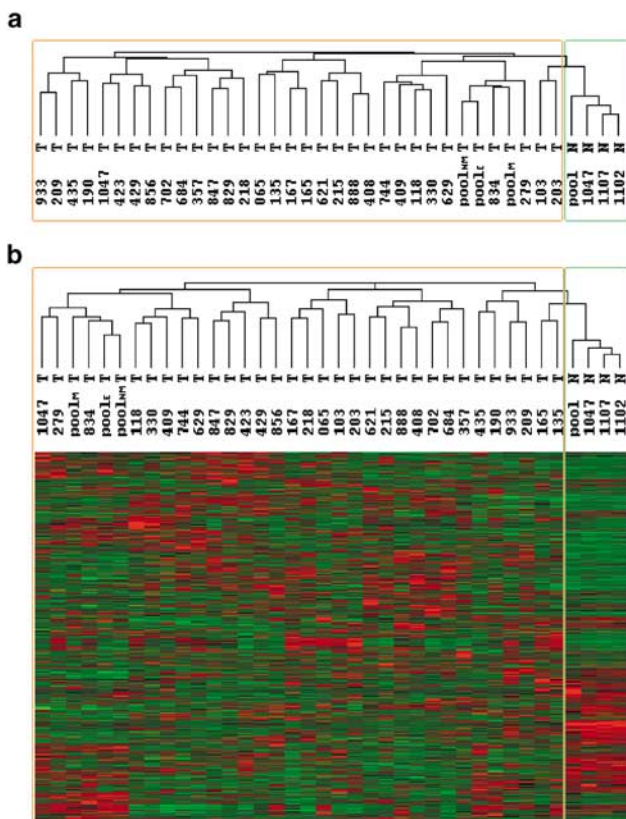
samples are less similar among themselves than the normal samples, as represented by the length of the vertical lines linking them. Cluster analysis can find coherent patterns of gene expression, but provides little information about statistical significance. In order to identify genes that have significant changes in expression, we used significance analysis of microarray (SAM) (Tusher *et al.*, 2001). In a three-step process of paired comparisons (see Materials and methods), we selected 2377 probe sets that are differentially expressed (DE) between tumour and normal tissues. Unsupervised hierarchical clustering with the SAM selected genes resulted in tighter clustering of the normal samples and clearer distinction from the tumour samples (Figure 1b). The expression profiles remained more heterogeneous among tumours than among normal samples. Heterogeneity of expression between tumours can be used for classification (Liu, 2003).

### Chromosomal localization of differentially expressed genes

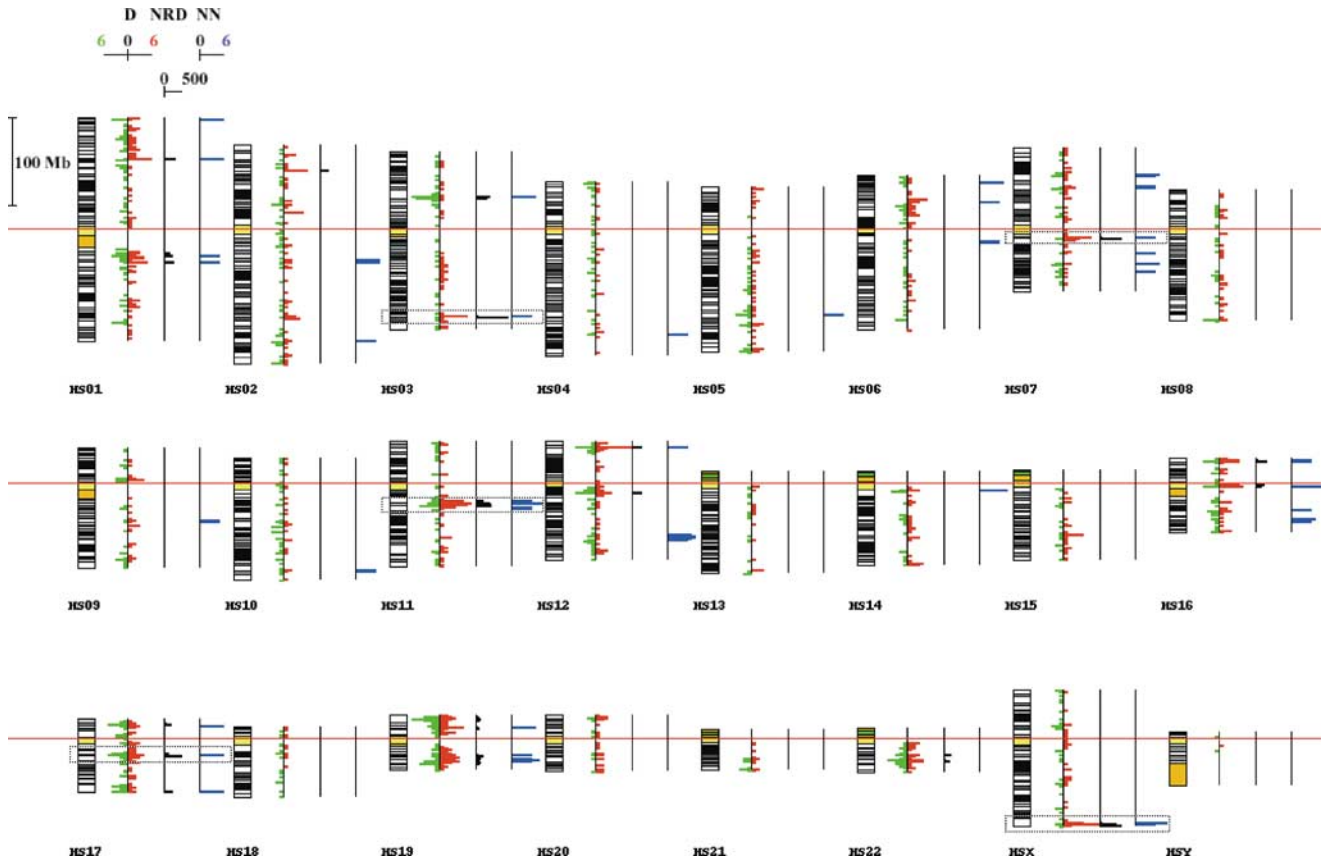
Genomic aberrations associated with malignant transformation can affect gene expression. Gene expression profiles of cell lines show that chromosomal modifications can lead to regional biases in gene expression (Monni *et al.*, 2001; Phillips *et al.*, 2001). We searched for similar regional biases with the SAM selected DE genes using three parameters: local density (LD), the normalized relative density (NRD) and the nearest neighbour (NN) score (Materials and methods; Figure 2). We first showed that the chromosomal distribution of the total gene set on the microarrays is representative of the distribution of known genes (data not shown), which assured us that there is no underlying bias. We found global high densities of DE sequences on chromosomes 19, 17 and 22 and very low densities on chromosomes 13 and Y (Figure 2), which reflects the overall densities of genes on these chromosomes. We can distinguish three groups of chromosomal regions (Figure 2), with different LD, NRD and NN score characteristics.

The first group, which has the characteristics high LD, high NRD and small NN scores, consists of four chromosomal regions (cytobands 2p23.3, 12q13.12, 22q12.1 and 22q13.1). The high density of small genes in these regions could explain their high LD and NRD with low NN scores. A total of 40 genes are located on the 28th Mb of chromosome HS02 (2p23.3) and 39 on the 57th Mb of HS12 (12q13.12; <http://www.ncbi.nlm.nih.gov/mapview/>), which is higher than average (30 genes per Mb). Most of these genes are less than 10 000 base pairs in length. Apparently, the first group distinguishes four regions with numerous small genes.

The second group, which has the characteristics low LD, low NRD and high NN scores, consists of 18 chromosomal regions (only the NN score is visible in Figure 2): cytobands 1p36.33, 2q14.3–q21.2, 2q35, 4q32.3, 5q31.3, 6p25.1, 6p21.32, 6q13, 7p15.3, 7p15.1, 7q22.1, 7q31.1, 7q32.2–q32.3, 9q22.1, 10q26.13–q26.2, 12q23.3–q24.11, 14q11.1–q11.2 and 16q22.1–q23.1. The



**Figure 1** Classification of hypopharyngeal cancer by global gene expression profiling. Unsupervised hierarchical clustering of 38 tumour and normal HNSCC samples using 3962 working probe sets with significant levels of expression (a), and 2377 probe sets selected for being differentially expressed between tumour and normal by SAM (b). The dendrogram indicates the degree of relatedness of the samples, the shorter the vertical branch linking two samples, the closer their expression profiles. The light green box surrounds the normal (N) cluster, and orange box the tumour (T) cluster. Gene expression profiles in (b) are represented with a green to red (lower to higher expression) colour scale



**Figure 2** Chromosomal distribution of sequences differentially expressed (DE) in HNSCC. Three vertical histograms are shown on the right of each chromosome. The first histogram represents the local density (LD, labelled D), that is, the number of DE sequences per million base pairs (Mb; red, overexpressed; green, underexpressed). The second histogram shows the normalized relative density (NRD) calculated for each Mb where more than six DE sequences are located. The third histogram shows in blue the Mb areas where at least two consecutive DE sequences have nearest neighbour (NN) scores greater than four (see Materials and methods). Black dotted boxes surround the strongest clusters. Chromosomes and histograms are drawn to the same scale, and the cytobands are drawn according to [ftp://ftp.ncbi.nih.gov/genomes/H\\_sapiens](ftp://ftp.ncbi.nih.gov/genomes/H_sapiens)

low density of large genes in these regions could explain the low LD, low NRD and high NN scores (LD and NRD are affected by low gene density, but not the NN score). For example, there is only 1 gene per Mb at 14q11.1–q11.2 (<http://www.ncbi.nlm.nih.gov/mapview/>). Apparently, the second group points to 18 chromosomal regions with long genes.

Finally, the third group, which has high LD, high NRD and high NN scores, is particularly interesting because it potentially identifies clusters of over- and underexpressed genes. It consists of 17 chromosomal regions (cytobands 1p34.1, 1q21.3, 1q23.1, 3p21.31, 3q27.3, 7q11.22, 11q13.1–q13.3, 12p13.32, 16p13.3, 16q11.1, 17p13.1, 17q21.2, 17q25.3, 19p13.13, 19q13.2, 19q13.32–q13.33 and Xq28). The strongest clusters of differentially expressed sequences, with the highest LD, NRD and NN scores, are framed with a dotted line in Figure 2 (3q27.3, 17q21.2, Xq28, 7q11.22 and 11q13.1–q13.3). Interestingly, four of these regions (all except Xq28) have been shown to have strong DNA copy number gains by comparative genomic hybridization (CGH) in pharyngeal squamous cell carcinoma (PSCC) (Huang *et al.*, 2002). The correlation is particularly good

on chromosome 3. The cluster in 3p21, consisting of a maximal density of underexpressed sequences combined with a high NRD and a high NN score (Figure 2), corresponds to low DNA copy number by CGH. On the contrary, the cluster of overexpressed sequences located in 3q27.3 corresponds to high DNA copy number by CGH. Interestingly, +3q (gain at 3q) is known to be particularly important for the progression of HNSCC (Redon *et al.*, 2001; Huang *et al.*, 2002) and +7q and +17q for PSCC (Huang *et al.*, 2002). Furthermore, +11q12–13 is one of the smallest recurrent chromosomal regions with a high-level amplification (Huang *et al.*, 2002), particularly in hypopharyngeal tumours (Muller *et al.*, 1997). These results suggest that overexpression is attributable to DNA amplification in these regions. Some of the genes located in these clusters (3q27.3, 17q21.2–q21.31, 7q11.22–22.1 and 11q13.1–q13.3) are known to be involved in tumorigenesis, such as *PAI1* and *PPFIA1*. We were interested in investigating whether ‘poorly known’ genes with potentially interesting functions in these regions are amplified. We investigated *EIF4G1*, *DVL3*, *KRT17* and *KRT16*, *EPHB4*, *MCM7*, *BRMS1* and *SART1* (Table 2), which

**Table 2** Eight selected genes located in the four chromosomal regions with the highest gene-density bias

NRD	Chromosomal location	Genomic location (Mb)	Gene	AC	E/N
1228	3q27.3	180.6	Eukaryotic translation initiation factor 4 gamma (EIF4G1)	Q04637	1.78
	3q27.3	180.7	Segment polarity protein dishevelled homolog DVL-3 (DVL3)	U75651	3.65
950	17q21.31	39.2	Keratin, type 1 cytoskeletal 17 (Cytokeratin 17) (KRT17)	BC011901	7.57
	17q21.31	39.2	Keratin, type I cytoskeletal 16 (Cytokeratin 16) (KRT16)	AF061812	8.45
760	7q22.1	98.9	Ephrin type-B receptor 4 precursor (EC 2.7.1.112) (EPHB4)	U07695	3.12
	7q22.1	98.2	DNA replication licensing factor MCM7	D55716	4.22
556	11q13.1	68.6	Breast cancer metastasis-suppressor 1 BRMS1	AF159141	1.93
	11q13.1	68.3	Squamous cell carcinoma Antigen Recognized by T cells (SART1)	AB006198	1.81

NRD = normalized relative density, Mb = million base pairs, AC = accession number, E/N, 'early' tumour (E)/normal (N) expression ratio; all the information on locations and genes was obtained automatically using the Gscope bioinformatics platform (Ripp *et al*, in preparation; Materials and methods)

are involved in translation, signal transduction, epidermal differentiation, cell growth, DNA replication, tumour suppression and regulation of proliferation, respectively.

Quantitative PCR on genomic DNA showed that EIF4G1 and DVL-3 from 3q27 are amplified (three copies or more) in 5/12 and 6/12 patients, respectively, EPHB4 and MCM7 from 7q22 in 5/12 and 3/12, and BRMS1 and SART1 from 11q13 in 5/12 and 5/12 (Figure 3). The two genes from 11q13 were amplified in the same five patients, indicating that the amplification covers at least the 350 000 base pairs that separate these two genes. In contrast, KRT17 and KRT16 from 17q21 were not amplified in the patients analysed, indicating that gene overexpression is not systematically associated with DNA amplification. There are several large-scale studies that compared DNA and RNA levels. About 10% of overexpressed genes were reported to be amplified (Hyman *et al.*, 2002), whereas about 50% of amplified genes were overexpressed (Hyman *et al.*, 2002; Pollack *et al.*, 2002). However, in another study, only about 4% of amplified genes were overexpressed (Platzer *et al.*, 2002). Other approaches to identify amplified genes based on their genomic distribution have been reported, which however did not normalize for gene density (Crawley and Furge, 2002; Kano *et al.*, 2003). Our approach efficiently highlights genes whose altered expression may result from changes in copy number.

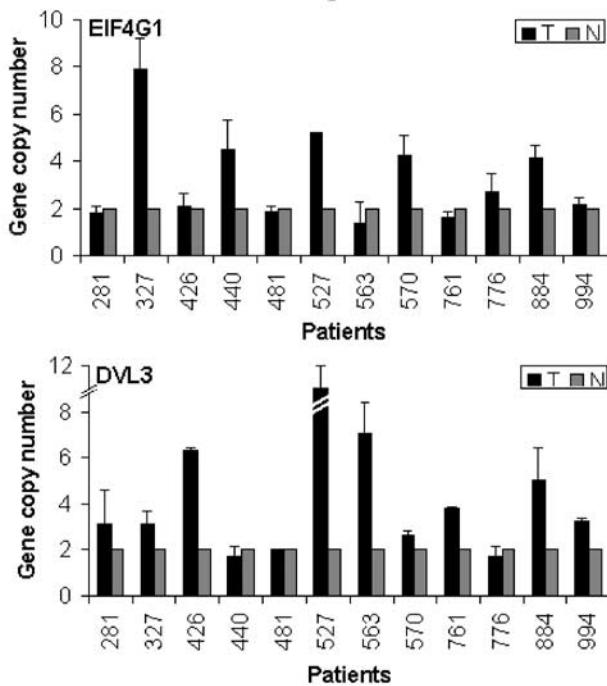
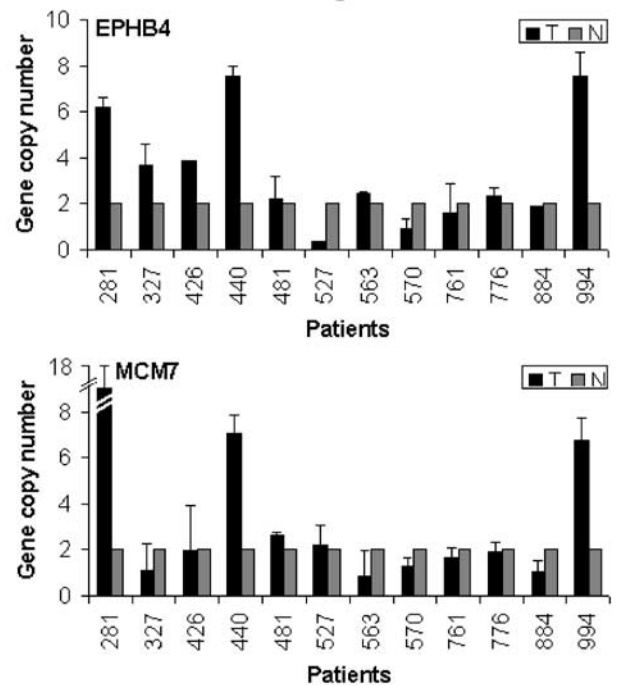
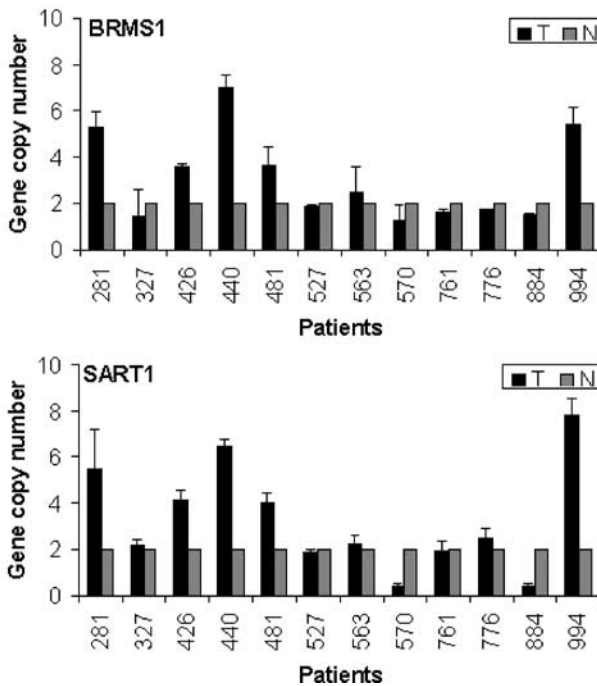
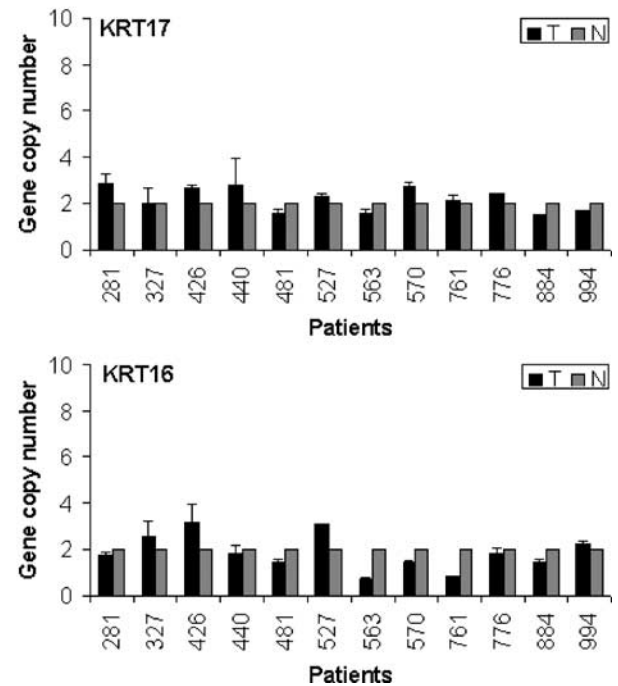
#### Genes differentially expressed in tumorigenesis

We searched for genes involved in tumorigenesis by comparing gene expression profiles in the four normal and four 'early' tumour samples (Table 1). The 'early' tumours are small in size, well or moderately differentiated, with no lymph node involvement. In order to select differentially expressed genes, we used SAM, a method for identifying genes with statistically significant changes in expression (Tusher *et al.*, 2001). In all, 1595 probe sets that are differentially expressed between E and N were selected, of which 136 probe sets (119 unique genes) exhibit a greater than fivefold change (55 genes up in normal and 64 up in tumours; Tables 3 and 4). The duplicate probe sets, for the same gene, gave similar results. These 136 probe sets improve the degree

of relatedness of the samples compared to the working set of 3962 genes (compare the dendrograms in Figure 4a and b).

Among these differentially expressed genes, there are new or uncharacterized sequences. We verified the differential expression of six of these novel genes with quantitative PCR (Q-PCR) on 12 additional hypopharyngeal tumours and their matching normal samples (Figure 5). There was a very good correlation between the microarray and Q-PCR results. For all six genes, 8–10/12 patients (67–83%) had >2-fold differences by Q-PCR, whereas 27–34/34 (79–100%) had >2-fold differences on the arrays. These results validate the gene profiling approach, identify potential new markers for diagnosis, and new genes for the study of the early molecular changes leading to tumour formation. The potential functions of these novel genes were investigated by comparing them with neighbouring co-clustering genes (Figure 4c). AB011112 clusters with two mucins, and a blast analysis defines it as 'transmembrane activator and CAML interactor', but it has not been described in the literature. Y09538 (LIM-domain containing) and AF091087 cluster with unknown sequences (U51712, N74607), a serine protease inhibitor (AJ228139) and an extracellular matrix protein (U68186), raising the possibility that they are involved in extracellular matrix remodelling. Furthermore, the 5' part of AF091087 resembles the *Xenopus laevis* mitotic phosphoprotein 22 (AAM33244). AA418080 clusters with BMP-1, OSF-2, stromelysin 3 and collagen, linking it to the extracellular matrix. M69199, a putative G0/G1 switch protein, clusters with IL-1 beta, GLUT3 and PLOD3, and coincidentally IL-1 upregulates GLUT3 in the ovary (Kol *et al.*, 1997). Finally, the U61836 cluster does not propose a function, but blast analysis suggests that it encodes a polyamine oxidase. The study of these new genes could lead to new diagnosis tools, therapeutic approaches and mechanistic insights.

Some of the genes revealed by our analysis have already been reported to be differentially expressed in HNSCC. We found 12/55 decreased sequences in common with previous studies, and 35/64 increased (Leethanakul *et al.*, 2000; Alevizos *et al.*, 2001; Al Moustafa *et al.*, 2002; El-Naggar *et al.*, 2002; Mendez *et al.*, 2002) (Tables 3 and 4). Our results correlated very well with these other studies, except for the one that

**a Cluster located in 3q27****b Cluster located in 7q22****c Cluster located in 11q13****d Cluster located in 17q12-q21**

**Figure 3** DNA levels measured by quantitative PCR for eight genes with biased chromosomal distributions that are overexpressed in tumours. DNA from 12 patients was analysed for EIF4G1 and DVL3 in 3q27 (a), EPHB4 and MCM7 in 7q22 (b), BRMS1 and SART1 in 11q13 region (c) and KRT17 and KRT16 in 17q12-q21(d). DNA levels were normalized to the GAPDH control. The normal sample DNA levels were adjusted to a copy number of two, and the relative copy numbers of the tumour samples compared to their matching normal samples are shown

compared primary cell cultures of tumours and normal matching samples (Al Moustafa *et al.*, 2002), suggesting that cell culture influences expression. Our results correlate well with a laser capture microdissection study

(Alevizos *et al.*, 2001), in that all of the common genes have the same pattern of expression, suggesting that the strong differences we observe come from epithelial components of the samples. The genes in the 129 gene

**Table 3** Sequences decreased in tumours

Rank	N/E	Unigene	Accession	Description	References
<i>Antioncogene and antiproliferative proteins</i>					
24	5.2	Hs.81134	X52015	Interleukin-1 receptor antagonist	(1); (2)
41	12.4	Hs.103505	X99977	ARS gene, component B	
44	88.3	Hs.76422	M22430	RASF-A PLA2	
53	13.0	Hs.75736	J02611	Apolipoprotein D	
39	6.3	Hs.75106	M25915	Complement cytotoxicity inhibitor (CCI)	(2)
5	11.8	Hs.65424	X64559	Tetranectin	(5)
59	42.2	Hs.183752	AA532495	Clone = IMAGE-996282 (MSMB)	
34	6.8	Hs.2962	AA131149	Clone = IMAGE-587049 (S100P)	
<i>Cancer-associated proteins</i>					
12	71.9	Hs.80395	X76220	MAL gene exon 1	(3)
3	15.9	Hs.79368	Y07909	Progression associated protein (PAP)	(1); (5)
18	6.1	Hs.50964	X16354	Transmembrane carcinoembryonic antigen BGPα	
23	6.2	Hs.220529	M29540	Carcinoembryonic antigen (CEA)	
30	8.0	Hs.20166	AF043498	Prostate stem cell antigen (PSCA)	
22	9.2	Hs.13775	U51712	cDNA/gb = U51712 (SMAP31)	
33	7.5	Hs.73848	M18728	Nonspecific crossreacting antigen	
28	6.3	Hs.44	M57399	Nerve growth factor (HBNF-1)	
46	6.1	Hs.7306	AF056087	Secreted frizzled related protein	
<i>Metabolism</i>					
25	5.4	Hs.75888	U30255	Phosphogluconate dehydrogenase (hPGDH)	
26	5.1	Hs.105435	AF042377	GDP-mannose 4,6 dehydratase	
15	5.3	Hs.2533	U34252	Gamma-aminobutyraldehyde dehydrogenase	(1); (5)
20	12.0	Hs.233441	M57951	Bilirubin UDP-glucuronosyltransferase isozyme 2	
31	8.5	Hs.389	X76342	ADH7	(5)
38	8.4	Hs.575	M74542	Aldehyde dehydrogenase type III (ALDHIII)	
8	16.4	Hs.2022	L10386	Transglutaminase E3 (TGASE3)	(3)
29	7.1	Hs.334841	U29091	Selenium-binding protein (hSBP)	
42	8.3	Hs.5920	AJ238764	N-acetylmannosamine kinase	
16	9.1	Hs.81071	U68186	Extracellular matrix protein 1	(5)
<i>Structural proteins</i>					
4	7.6	Hs.80342	X07696	Cytokeratin 15	(1); (2)
40	5.9	Hs.74070	X14640	Keratin 13	(2); (3); (5)
35	16.7	Hs.3235	X07695	Cytokeratin 4 C-terminal region	(1); (3); (5)
6	7.2	Hs.74304	AF001691	195 kDa cornified envelope precursor (PPL)	
56	8.6	Hs.158295	X05451	Myosin light chain 3 (MLC-3f)	
48	36.3	Hs.931	S73840	Type IIA myosin heavy chain	
51	11.8	Hs.78344	AF013570	Smooth muscle myosin heavy chain SM2	
37	23.0	Hs.115166	AF045941	Sciellin (SCEL)	
1	6.7	Hs.3164	X76732	NEFA protein	
58	8.5	Hs.334629	U96094	Sarcolipin (SLN)	
<i>Mucous-related</i>					
11	6.4	Hs.234642	AB001325	Aquaporine 3 (AQP3)	(2)
52	7.0	Hs.221986	U46569	Aquaporin-5 (AQP5)	
7	5.4	Hs.89603	J05582	Pancreatic mucin	
50	12.6	Hs.198267	AJ010901	MUC4 gene, 3 flanking region	
2	5.9	Hs.89603	HG371-HT26388	Mucin 1, epithelial, alt. splice 9	
60	66.2	Hs.82961	A1985964	Clone = IMAGE-2493903 (TFF3)	
55	107.2	Hs.169224	L08044	Intestinal trefoil factor	
9	23.8	Hs.64867	AJ228139	LETKI precursor	
13	7.4	Hs.8272	AI207842	Clone = IMAGE-1953089 (prostaglandin D2)	
<i>Other</i>					
47	14.8	Hs.73931	M16276	MHC class II HLA-DR2-Dw12 mRNA DQw1-beta	
49	6.1	Hs.204040	AF004230	Monocyte/macrophage Ig-related receptor MIR-7	
45	10.0	Hs.99918	M54994	Bile salt-activated lipase (BAL)	
54	5.1	Hs.250760	F27891	Clone = s4000025D03 (COX6A2)	
27	6.0	Hs.75329	AF063002	LIM protein SLIMMER	
57	14.4	Hs.75329	AF001548	Clone CIT987SK-A-815A9	
17	8.2	Hs.16622	Y09538	ZNF185 gene	
32	5.1	Hs.64742	AB011112	KIAA0540 protein	
36	5.4	Hs.206501	AF091087	Clone 643 unknown	

The rank represents the order of the genes according to SAM score. Numbers refer to: (1) Alevizos *et al.* (2001); (2) Al Moustafa *et al.* (2002); (3) El-Naggar *et al.* (2002); (5) Mendez *et al.* (2002). Genes with an underlined reference have an inverted expression pattern

**Table 4** Sequences increased in tumours

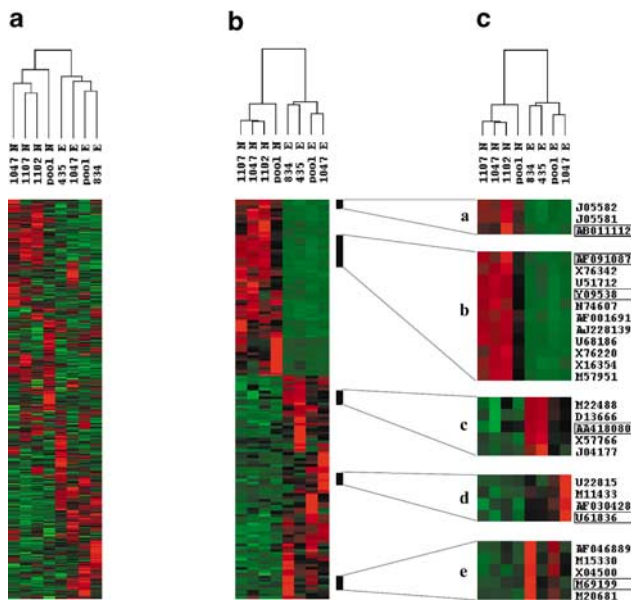
Rank	E/N	Unigene	Accession	Description	References
<i>Tumour invasion and extracellular matrix morphology</i>					
3	87.6	Hs.83169	M13509	Skin collagenase (MMP-1)	(1)
15	15.5	Hs.83326	X05232	Stromelysin (MMP-3)	(5)
12	9.0	Hs.151738	J05070	Type IV collagenase (MMP-9)	
16	25.2	Hs.2258	X07820	Stromelysin-2 (MMP-10)	(4)
37	10.6	Hs.155324	X57766	Stromelysin-3 (MMP-11)	
6	16.3	Hs.1695	L23808	Human metalloproteinase HME (MMP-12)	(5)
20	6.8	Hs.1274	M22488	Bone morphogenetic protein 1 (BMP-1)	
61	5.8	Hs.153357	AF046889	Lysyl hydroxylase isoform 3 (PLOD3)	
41	36.4	Hs.82085	M14083	Beta-migrating plasminogen activator inhibitor 1	(5)
22	9.6	Hs.77274	X02419	uPA	(5)
40	13.1	Hs.179657	U09937	Urokinase-type plasminogen receptor	
<i>Extracellular matrix proteins</i>					
1	10.6	Hs.179573	J03464	Collagen alpha-2 type I	(1); (5)
47	7.3	Hs.82985	Y14690	Procollagen alpha 2(V)	(2); (5)
11	7.1	Hs.75617	X05610	Type IV collagen alpha (2) chain	(5)
5	5.2	Hs.119571	X14420	Pro-alpha-1 type 3 collagen	(5)
9	7.1	Hs.119129	M26576	Alpha-1 collagen type IV	(5)
45	18.3	Hs.82772	J04177	Alpha-1 type XI collagen (COL11A1)	
35	18.9	Hs.172928	Y15915	Collagen (type1 alpha1)/PDGF beta (chimaeric)	
57	5.9	Hs.79914	U21128	Lumican	(5)
17	6.7	Hs.111779	J03040	SPARC/osteonectin	(1); (5)
28	11.6	Hs.204133	X78565	Tenascin-C	(5)
46	36.1	Hs.313	AF052124	Osteopontin	
49	5.4	Hs.83551	U19718	Microfibril-associated glycoprotein (MFAP2)	
26	5.5	Hs.287820	M10905	Cellular fibronectin	(2)
<i>Structural proteins</i>					
2	7.6	Hs.2785	Z19574	Cytokeratin 17	(2); (5)
19	10.0	Hs.115947	AF061812	Keratin 16 (KRT16A)	(2); (5); (6)
36	5.1	Hs.119000	M95178	Non-muscle alpha-actinin	(5)
23	7.8	Hs.75517	U17760	Laminin S B3 chain (LAMB3)	(2); (5)
42	19.7	Hs.54451	Z15008	Laminin (LAMB2)	(2); (6)
66	16.0	Hs.83450	L34155	Laminin-related protein (LamA3)	(2)
4	5.4	Hs.121576	AJ001381	Mutated allele of a myosin class I	
<i>Growth and stress response</i>					
73	39.4	Hs.624	M28130	Interleukin 8 (IL8)	(5)
52	12.7		X04500	Prointerleukin 1 beta	(2); (5)
54	8.4	Hs.126256	M15330	Interleukin 1-beta (IL1B)	(2)
39	5.6	Hs.93913	X04430	Interferon-beta-2 (IL6)	
65	11.3	Hs.789	X54489	Melanoma growth stimulatory activity (MGSA)	(1); (2); (5)
25	5.2	Hs.211600	M59465	Tumour necrosis factor alpha-inducible protein A20	(2)
63	6.0	Hs.265827	U22970	Interferon-inducible peptide (6-16)	(3); (5)
<i>Other</i>					
21	6.2	Hs.136348	D13666	Osteoblast specific factor 2 (OSF-2)	(1); (5)
30	5.0	Hs.90572	U33635	Colon carcinoma kinase-4 (CCK4)	(5)
31	6.5	Hs.418	U09278	Fibroblast activation protein (FAP)	(5)
33	6.7	Hs.135150	AF030428	Lung type-I cell membrane-associated protein (T1A-2)	
51	5.8	Hs.125359	AA704137	Clone = IMAGE-1119984 (similar to THY1)	
59	6.5	Hs.7594	M20681	Glucose transporter-like protein-III (GLUT3)	
74	10.2	Hs.72879	M77481	MAGE-1 antigen	
75	10.8	Hs.36978	U03735	MAGE-3 antigen	
58	5.1	Hs.183648	U22815	LAR-interacting protein 1a gene (PPFIA1)	
53	7.4	Hs.73817	D90144	LD78 alpha precursor (CCL3)	
24	5.5	Hs.118893	D86983	KIAA0230 (peroxidase)	(1)
29	5.7	Hs.373503	M27826	Endogenous retroviral protease	(2)
34	5.0	Hs.91747	AL096719	cDNA DKFZp566N043 (profilin)	
71	10.3	Hs.75212	X16277	Ornithine decarboxylase ODC	
60	9.9	Hs.79389	D83018	nel-related protein 2	
72	5.2	Hs.226307	AL022318	Phorbolin 3	
76	8.6	Hs.80962	U91618	Proneurotensin/proneuromedin N	
44	7.3	Hs.101850	M11433	Cellular retinol-binding protein	(5)
50	11.4	Hs.373503	AA151971	Clone = IMAGE-588365 (similar to S100P)	
55	8.9	Hs.8786	AB014679	N-acetylglucosamine-6-O-sulphotransferase (GlcNAc6ST)	



Table 4 continued

Rank	E/N	Unigene	Accession	Description	References
56	6.5	Hs.118633	AJ225089	2-5 oligoadenylate synthetase 59 kDa isoform	
69	9.6	Hs.105924	AF071216	Beta defensin 2 (HBD2)	
67	18.0	Hs.76118	X04741	Protein gene product (PGP) 9.5	(2)
43	6.0	Hs.288467	AA418080	Clone = IMAGE-767773	
68	7.2	Hs.95910	M69199	G0S2 protein	(2)
48	7.4	Hs.92374	U61836	Putative cyclin G1 interacting protein (C20orf16)	

The rank represents the order of the genes according to the SAM score. Numbers refer to: (1) Alevizos *et al.* (2001); (2) Al Moustafa *et al.* (2002); (3) El-Naggar *et al.* (2002); (4) Leethanakul *et al.* (2000); (5) Mendez *et al.* (2002); (6) Villaret *et al.* (2000). Genes with underlined references have an opposite expression pattern



**Figure 4** Clustering of 'early' tumours by gene expression profiling. Unsupervised hierarchical clustering of the 'early' tumours and normal samples with the 3962 working probe sets, with significant signals (a), and with the 136 probe sets (119 unique genes) selected by SAM that have a fold change greater than 5 (b). The neighbouring genes (accession numbers), with the same expression profiles as the six poorly characterized sequences selected for further study (boxed), are highlighted in (c). Gene expression profiles are represented with a green to red (lower to higher expression) colour scale

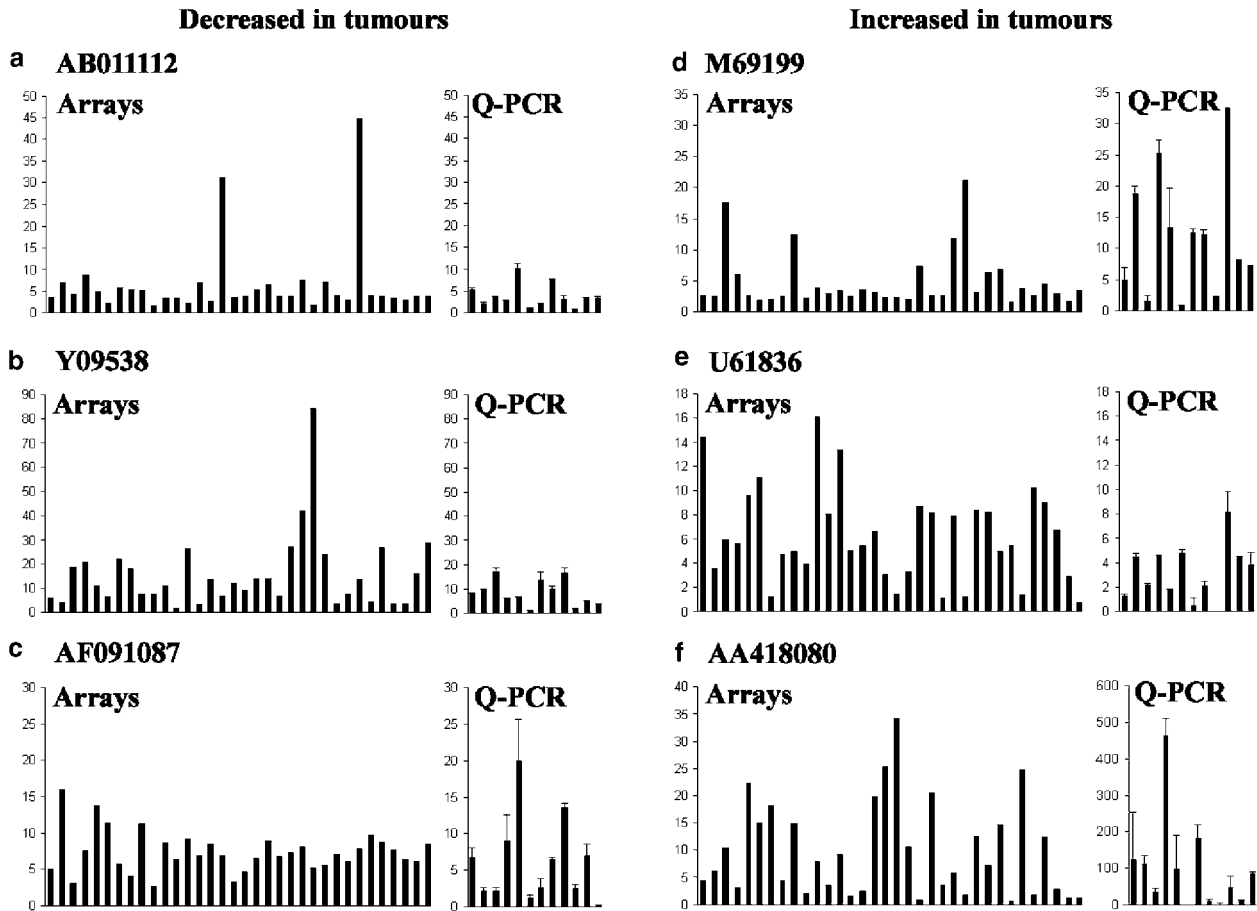
set fall into a number of functional classes. Many of the genes overexpressed in tumour are involved in the formation and remodelling of the extracellular matrix, an important process in tumorigenesis (Kerkela and Saarialho-Kere, 2003). A number of growth and stress response genes are also overexpressed in tumours (Table 4). Genes with decreased expression are involved in inhibition of proliferation, metabolism, structure and mucous-related. Interestingly, IL-1 is increased and the IL-1 receptor antagonist decreased in tumours, a misregulation associated with predisposition to cancer (Arend, 2002) and abnormal cell proliferation in acute myeloblastic leukaemia (Tao *et al.*, 2000).

#### Potential prognostic genes

We selected genes with potential prognostic value by analysing the 30 'late' tumour samples. Since tumour

location is an independent predictor of survival, our study included only hypopharyngeal tumours with similar histological characteristics (T1-T4 tumours with lymph node involvement and no metastasis at the time of surgery). Surgical resection was complete for all patients included in this study (evaluated by a pathologist for absence of tumour cells in the resection margins). All the patients received adjuvant treatment by radiotherapy and four also by chemotherapy. The samples were divided into three subtypes according to clinical behaviour 3 years or more after surgery: patient who did not (NM) or did (M) develop metastases, or had local recurrence (LR) within 3 years (see Table 1). The data were normalized by singular value decomposition (SVD) (Nielsen *et al.*, 2002), which eliminates potential bias due to the use of two batches of experiments (see Materials and methods). Unsupervised clustering using the normalized data for the 3962 working gene set distinguished NM or M from normal (Figure 6Aa and b), but did not distinguish M from NM (Figure 6Ac). In order to identify markers of metastatic propensity, genes with different expression patterns between M and NM were selected by two methods. Firstly, 80 probe sets were identified with the Student's *t*-test ( $P < 0.02$ ). Secondly, 121 probe sets were selected that are differentially expressed at least 1.5-fold in at least half of the samples. The first method selects for homogenous differences, the second for subgroups. Low fold changes could arguably be significant if they arise from relatively few metastatic cells diluted in the tumour mass. Combining the two approaches gives 168 probe sets (164 genes; supplementary table) and 33 common probe sets (32 genes; Tables 5 and 6). Clustering with the 168 probe sets defines two groups of samples and correctly assigns 24/26 (92%) of the M and NM tumours (Figure 6Ba). We noticed that clustering with the 33 common probe sets of the 24 correctly assigned samples defined a set of six M samples that segregate together (on the right in Figure 6Ca). It is perhaps worth noting that these patients have a longer overall survival (Table 1), but the significance of this observation remains to be assessed with additional samples.

The biological significance of the 168 probe sets was assessed by adding other samples to the clustering. When the N group is added, it clusters with the NM group (Figure 6Bb), suggesting that NM tumours could be closer to the normal samples than the M tumours. Similarly, the E samples cluster within NM (Figure 6Bc),



**Figure 5** Differential expression in HNSCC of six poorly characterized sequences. The N/T ratio for underexpressed (a–c) and the T/N ratio for overexpressed (d–f) sequences are shown. The graphs on the left show the fold changes observed on the microarrays (normalized to the average of the normal samples), and the graphs on the right the fold changes observed by Q-PCR on hypopharyngeal samples from additional patients. Expression levels are normalized to ubiquitin, and fold changes are relative to the matching normal samples

suggesting that NM tumours could be closer to the ‘early’ tumours. With LR, 2/4 cluster with NM and 2/4 in a separate subgroup (Figure 6Bd). Similar results are obtained with the 80, 121 and 33 probe subsets (data not shown). These results raise the possibility that the 164 unique genes (that correspond to the 168 probe sets) may be characteristic of the tumours that will metastasize, but do not seem to be associated with local recurrence.

We assigned the 164 unique genes to general functional categories with a gene ontology resource (<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GPL91>). The largest functional group in the poor prognosis M tumours is cell growth and signalling (supplementary table), which includes, for example, cyclin D3 involved in cell cycle control. Genes worthy of note include autotaxin, fusin (CXCR4) and IL8. Autotaxin (ATX) is

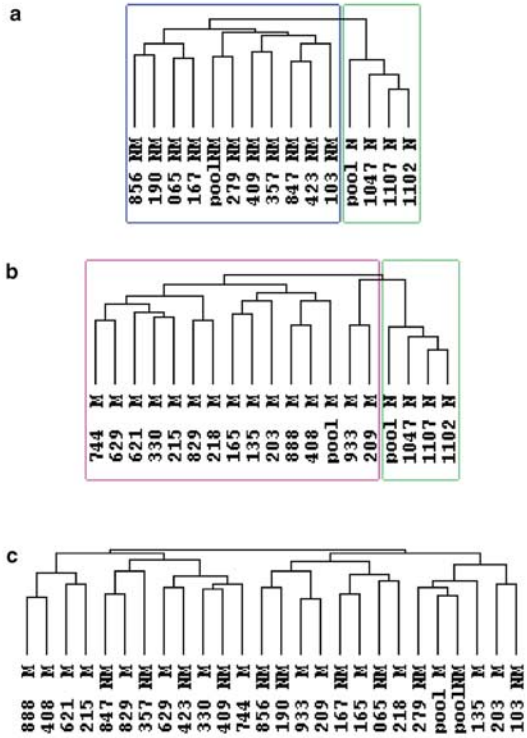
a tumour motility-stimulating protein (Stracke *et al.*, 1992; Umezū-Goto *et al.*, 2002), which is also angiogenic (Nam *et al.*, 2001). Fusin has been implicated in metastasis in prostate cancer, neuroblastoma, ovarian cancer and melanoma (Geminder *et al.*, 2001; Murakami *et al.*, 2002; Taichman *et al.*, 2002). IL-8 contributes to carcinoma cell invasion in the colon, breast and oral cavity (Youngs *et al.*, 1997; Li *et al.*, 2001; Watanabe *et al.*, 2002). NM tumours, with good prognosis, overexpress genes with opposite functions, such as cell adhesion, intercellular junctions and cell shape (supplementary table). Several of the genes selected from the comparison of M and NM tumours have been identified in related studies. Aldehyde dehydrogenase A (van't Veer *et al.*, 2002), hypothetical protein HSPC111 (LaTulippe *et al.*, 2002) and integrin alpha3 (MacDonald *et al.*, 2001)

**Figure 6** Classification of hypopharyngeal samples by gene expression profiling. (A) Dendrograms from unsupervised hierarchical clustering of M tumours with normal (a), NM with normal (b) or NM with M (c) using the 3962 working probe sets with significant expression. (B) Dendrograms from unsupervised hierarchical clustering using the 168 M versus NM selected probe sets (164 unique genes; see Results) and either the 26 M + NM tumour samples (a) or the 24 tumours (excluding those that misclustered in (a)) with the normal (b), the early (c) or the local recurrence (d) samples. (C) Dendrograms from unsupervised hierarchical clustering using the 33 M versus NM common probe set (32 genes) selection (see Results). The profiles of genes overexpressed in M (b) and NM (c) are represented on a green to red (lower to higher expression) colour scale. Blue boxes surround NM clusters, pink M clusters and light green N clusters. The two misplaced samples (829 and 167) in (Ba) are not included in (Bb, Bc, Bd) or C. Arrowheads point to the samples that were added to the clustering: N in (Bb), E in (Bc) and LR in (Bd)

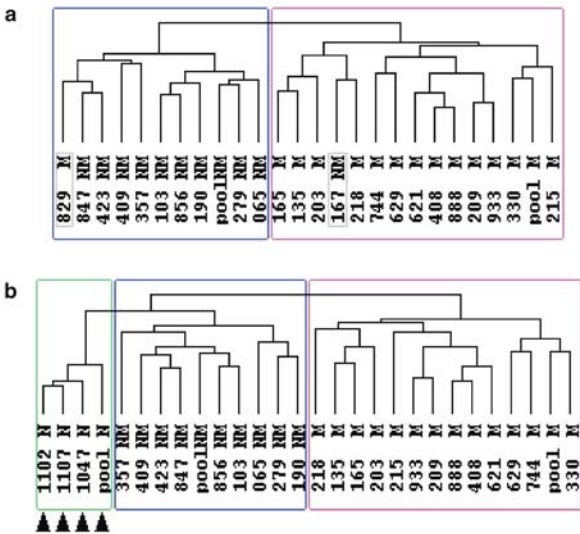
are overexpressed in good prognosis tumours and in NM. However, only one out of three genes that are in common with another HNSCC study falls into a potentially related category (visinin-like 1 is up

in NM, and both ribophorin II and hypothetical protein FLJ10097 are up in M, whereas all three are in the 'better predictor of outcome' group; Belbin *et al.*, 2002).

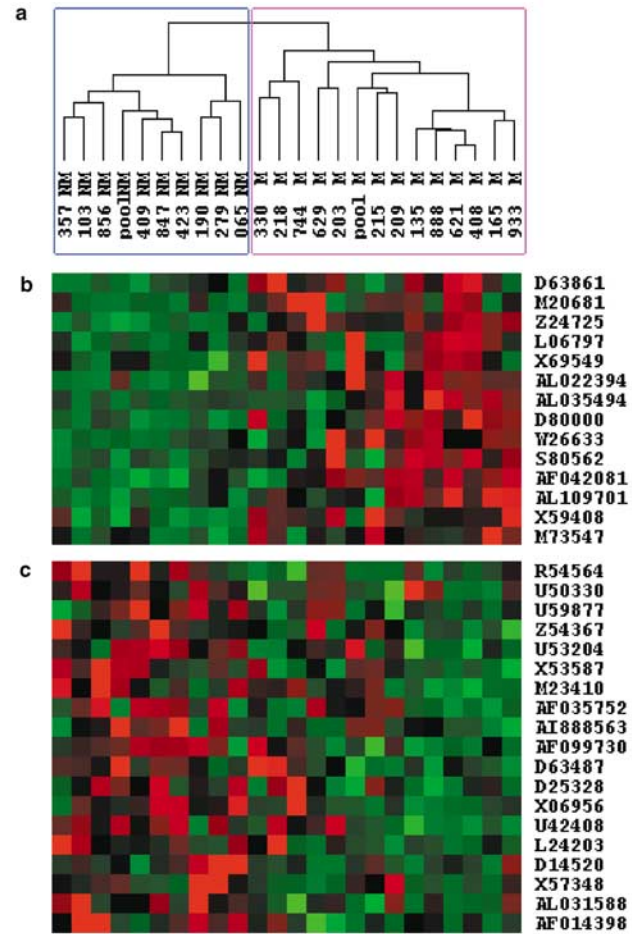
**A**



**B**



**C**



**Table 5** Genes increased in M tumours

<i>M/NM ratio</i>	<i>Unigene</i>	<i>Accession</i>	<i>Description</i>
2.46	Hs.184736	AL035494	Hypothetical protein FLJ10097
1.94	Hs.89414	L06797	Chemokine (C-X-C motif), receptor 4 (fusin)
1.73	Hs.211602	D80000	SMC1 (structural maintenance of chromosomes 1, yeast)-like 1
1.65	Hs.83656	X69549	Rho GDP dissociation inhibitor (GDI) beta
1.62	Hs.143482	D63861	Peptidylprolyl isomerase D (cyclophilin D)
1.60	Hs.177556	W26633	Melanoma antigen, family D, 1
1.57	Hs.75260	Z24725	Mitogen inducible 2
1.55	Hs.178112	M73547	DNA segment, single copy probe LNS-CAI/LNS-CAII
1.52	Hs.14368	AF042081	SH3 domain binding glutamic acid-rich protein like
1.51	Hs. 83532	X59408	MCP: membrane cofactor protein
1.51	Hs.75847	AL109701	CREBBP/EP300 inhibitory protein 1
1.50	Hs.194662	S80562	Calponin 3, acidic
1.50	Hs.7594	M20681	Solute carrier family 2 (facilitated glucose transporter), member 3
1.48		AL022394	Sequence from 20q

**Table 6** Genes increased in NM tumours

<i>NM/M ratio</i>	<i>Unigene</i>	<i>Accession</i>	<i>Description</i>
2.33	Hs.84728	D14520	Kruppel-like factor 5 (intestinal)
2.25	Hs.85266	X53587	Integrin, beta 4
2.13	Hs.5753	AF014398	Inositol(myo)-1(or 4)-monophosphatase 2
2.03	Hs. 98485	AF099730	GJB3: gap junction protein, beta 3, 31 kDa (connexin 31)
1.88	Hs.75318	X06956	Tubulin, alpha 1 (testis specific)
1.83	Hs.139851	AF035752	Caveolin 2
1.81	Hs.82237	L24203	Tripartite motif-containing 29
1.81	Hs.184510	X57348	Stratifin
1.77	Hs.79706	Z54367	Plectin 1, intermediate filament binding protein, 500 kDa
1.76	Hs.149098	AI888563	Smoothelin
1.67	Hs.2340	M23410	Junction plakoglobin
1.66	Hs.18141	U42408	Ladinin 1
1.63	Hs.223025	U59877	RAB31, member RAS oncogene family
1.63	Hs.122552	AL031588	G-2 and S-phase expressed 1
1.59	Hs.99910	D25328	Phosphofructokinase, platelet
1.51	Hs.112028	R54564	Misshapen/NIK-related kinase
1.50	Hs.1274	U50330	Bone morphogenetic protein 1
1.35	Hs.82563	D63487	KIAA0153 protein

We investigated whether we could predict if a tumour is M or NM, using a leave-one-out strategy. We were unable to predict correctly the subgroup of all the samples (data not shown). We found that the number of samples analysed is important for prediction accuracy. Using five M and five NM samples (microarray batch 1) for training and selecting genes, and using the remaining six NM and 10 M (microarray batch 2) for blind validation, we correctly predicted 62% of blind samples. In the converse experiment, with more samples for training, the prediction was improved to 80% (supplementary figure). A total of 98 breast tumours were needed to generate a predictive signature (van't Veer *et al.*, 2002). An analysis of a larger collection of tumours will be needed to establish whether or not a useful predictive signature can be identified for hypopharyngeal carcinoma.

## Conclusion

We have found a pattern of gene expression that distinguishes hypopharyngeal tumours from related

normal tissue. In addition, our study has revealed new uncharacterized sequences implicated in tumour formation. In a parallel large-scale differential display analysis of hypopharyngeal cancer, 1200 sequences were identified that differ in expression between tumour and normal (data not shown; Lemaire *et al.*, 2003). In all, 223 are common between the two studies, indicating that the two approaches are not redundant. Chromosomal aberrations have been shown to occur in HNSCC. We have identified six new genes that are amplified as well as overexpressed in hypopharyngeal carcinoma, using a new analysis method based on genomic clustering of overexpressed genes. Our study has defined a set of 164 genes that classify similar histological and clinical tumours with different outcome on the basis of expression levels. This signature is an 'indicator' but not a 'predictor' of HNSCC outcome, since the number of samples is not sufficient to crossvalidate the data set. Many of these sequences are still uncharacterized and could be major determinants of the capacity to metastasize. They could contribute to the prediction of whether a patient with hypopharyngeal cancer will develop metastases.

## Materials and methods

### Tissue samples

Primary tumour samples were obtained, with informed consent, from 38 patients undergoing surgery for hypopharyngeal tumours as a primary treatment without previous radiation or chemotherapy. A tumour fragment was taken near the advancing edge of the primary tumour (avoiding its necrotic centre), immediately frozen and stored in liquid nitrogen. The rest of the tumour was fixed in 6% buffered formaldehyde and embedded in paraffin for histopathological analysis. Tumour fragments were composed of at least 70% cancer cells, as assessed on adjacent histological stained sections. The TNM system of the UICC was used for tumour-node-metastasis staging (Sobin and Fleming, 1997). Tumours were classified into two groups: 'early' stage tumours (T1–T2), without lymph node involvement following histological examination of cervical lymph node sections, and 'late' stage tumours (mainly T2–T4), with lymph node involvement. None of the patients presented distant metastasis at the time of surgery. After surgery, all of the patients received adjuvant radiotherapy (RX) and four of them radiotherapy combined with chemotherapy. 'Late' tumours were subdivided into no metastatic (NM), metastatic (M) and local recurrence (LR) 'propensity' according to clinical outcome during a 3-year follow-up (Table 1). Normal samples were collected from the farthest margin of resection (usually uvula) from eight patients (seven were pooled). Normal uvula was also obtained from two nonsmokers, without cancer history, treated by uvulo-palato-pharyngo-plasty for obstructive deep apnoea syndrome. The surface epithelium of the latter samples was macrodissected, to enrich for epithelial tissue. Total RNAs were extracted using the RNeasy\* kit (Qiagen, France) with DNaseI treatment. The quality of the RNA preparation was examined by agarose gel electrophoresis.

### Hybridization

Gene expression profiles were analysed using Affymetrix HG-U95A microarrays containing probe sets representing ~12 650 distinct transcription features. cDNA synthesis, cRNA synthesis and labelling, as well as array hybridization, were performed as described in the Affymetrix user's manual (Affymetrix, Santa Clara, CA, USA) using 5 µg of total RNA. The T7 RNA polymerase promoter was incorporated in the first round of double-stranded cDNA synthesis. This cDNA was used for *in vitro* transcription with the ENZO BioArray High Yield IVT kit, to amplify the RNA and incorporate the biotinylated ribonucleotides required for staining after hybridization. The yield and quality of the cRNA was checked by spectrophotometry and capillary electrophoresis using the Agilent 2100 Bioanalyser and RNA 6000 LabChip kit (Agilent technology). A measure of 10 µg of fragmented, biotinylated cRNA was hybridized to the Affymetrix arrays at 45°C for 16 h, as described in the Affymetrix user's manual. Washing and staining of arrays were performed using the GeneChip Fluidics Station 400, and scanning with the Affymetrix GeneArray Scanner.

### Preprocessing of microarray data

Acquisition and quantification of array images as well as primary data analysis were performed using the Microarray Suite v5.0 and Data Mining Tool v2.0 Affymetrix software packages. Microsoft Excel was used for further statistical analyses. All arrays were globally scaled to a target value of

800, using the average signal from all gene features and Microarray Suite v5.0. The data set was filtered for low expression values prior to statistical analysis. Genes exhibiting mean values under an arbitrary threshold value of 650 (half of the mean value of all genes and all chips) among all 38 experiments were eliminated, leaving 3962 working probe sets, which can be considered to be expressed genes.

The samples were analysed in two hybridization batches, which are shown in Table 1. For the NM versus M comparison (Results, Potential prognostic genes section), the results from two experiments with different batches of microarrays were normalized by SVD (Alter *et al.*, 2000), using the Matrix and Linear Algebra package for Excel v1.1 (available at <http://digilander.libero.it/foxes/index.htm>). SVD linearly transforms the expression data from the gene and array matrix to a condensed 'eigengene' and 'eigenarray' representative matrix. A unique eigengene was identified that correlated the best with the two array batches. The influence of this eigengene and its corresponding eigenarray was subtracted from all the data (Nielsen *et al.*, 2002).

### Hierarchical clustering

We applied a hierarchical clustering algorithm to the samples and genes. The algorithm organizes all the data elements into a single tree. We mean-centred genes and arrays and used complete linkage clustering (cluster and tree-view at <http://rana.lbl.gov/EisenSoftware.htm>; Eisen *et al.*, 1998). In the dendrograms, shorter branches connect more similar samples. The red to green colour scale represents the mean-adjusted expression values, where red corresponds to higher and green to lower expression.

### Supervised gene selection

To compare tumour (T) and normal (N) patient samples, we used the Significance Analysis of Microarrays add-in to Microsoft Excel (Tusher *et al.*, 2001; <http://www-stat.stanford.edu/~tibs/SAM/index.html>). The genes were selected with SAM's default parameters and no minimum fold change to englobe the largest significant set of differentially expressed genes. Comparison of E against N results in 1595 significant probe sets, M against N in 1587 and NM against N in 1590, resulting in a total of 2377 unique probe sets. To refine the E versus N comparison, we selected 136 probe sets (119 unique genes) that have an E/N or N/E ratio greater than 5. Genes differentially expressed between NM and M patients were selected by two methods. Firstly, the *t*-test ( $P < 0.02$ ) identified 80 probe sets (79 genes). Secondly, 121 probe sets (118 genes) were selected that, for at least half of the samples of the subtype, exhibit a signal ratio of more than 1.5 between the individual samples and the average of the opposite tumour subtype, but never in the opposite subtype. This gave a combined working set of 168 probe sets (164 unique genes).

### Bioinformatics

Bioinformatics analysis was performed using the Gscope bioinformatics platform (Ripp *et al.*, in preparation) dedicated to managing large collections of protein or nucleotide sequences. A five-step protocol was used to localize the Affymetrix sequences on the human genome and to analyse their chromosomal distribution. Firstly, 12 448 Affymetrix sequences were automatically located on the human genome using BLAST (Altschul *et al.*, 1997). Secondly, the local densities (LD) of Affymetrix sequences per million base pairs (Mb) were calculated. Thirdly, these densities were related to

the number of Affymetrix sequences per chromosome. These relative densities were normalized in relation to relative density in known genes in the human genome of reference ([ftp://ftp.ncbi.nih.gov/genomes/H\\_sapiens](ftp://ftp.ncbi.nih.gov/genomes/H_sapiens)) as following:  $NRD = [(NAffySeq \text{ per Mb}) / (NAffySeq \text{ per chromosome})] / [(NgenesNCBI \text{ per Mb}) / (NgenesNCBI \text{ per chromosome})] \times 100$ , where NRD is the normalized relative density, NAffySeq number of Affymetrix sequences, Mb million base pairs and NgenesNCBI = number of genes known at NCBI. Fourthly, our analysis focused on the 2377 differentially expressed (DE) Affymetrix sequences selected by SAM (see above). Redundancy introduced by multiple probes mapping to the same gene was eliminated, leaving 2037 unique sequences. In addition to LD and NRD, we studied the neighbourhood of the DE sequences. For each of the 2037 DE sequences, the six neighbouring Affymetrix sequences on either side were identified. Out of these 12 sequences, the number of DE sequences was counted to obtain a nearest neighbours (NN) score. Fifthly, the functions of the 2037 DE sequences were established by automatically searching protein data banks.

#### Quantitative PCR on DNA

DNA was isolated from tissue samples by proteinase K digestion, phenol extraction and ethanol precipitation. A measure of 25 ng was used for PCR amplification with Sybr Green and the Roche Lightcycler (Roche Molecular Biochemicals). Oligonucleotide primers are designed to cross intron/exon junctions where possible with primer3 ([http://www-genome.wi.mit.edu/cgi-bin/primer/primer3\\_www.cgi](http://www-genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi)). Their sequences are: DVL3, 5'-TCCATTTTCTAATGGGCTGG-3' and 5'-ACAATGGAGATGCCCAAGAA-3'; EIF4G1, 5'-GCTGGATGGATTGGGGAGGAG-3' and 5'-TGGCCGCAGTGGTGTTTTATT-3'; KRT16, 5'-CCAGAGACCTG-GAGGAACAG-3' and 5'-CGTCTTCACATCCAGCAAGA-3'; KRT17, 5'-CTGGCCCTACCCCACTTTA-3' and 5'-GAGATGACCCTTGCCATCCTG-3'; EPHB4, 5'-TCCTGC-AAGGAGACCTTCAC-3' and 5'-CAGAGGCCTCGCAAC-TACAT-3'; MCM7, 5'-CCAGGCAACATCAACATCTG-3' and 5'-ATTACAGGCGTGAGCAAACA-3'; BRMS1, 5'-ATTGCCAAGCTGGAGGTG-3' and 5'-CTTTCTCTGGG-CTCCTTCCT-3'; SART1, 5'-TGTCCCGTAGGCCAAGT-TAC-3' and 5'-AGAATCGGCGAGTCAGGAAC-3'; GAPDH, 5'-GGAGCAAAAAGGTCATCAT-3' and 5'-GGCATTGTGCAAGAAAGAG-3'; RLPO, 5'-AATGT-TGCCAGTGTCTGTCTG-3' and 5'-AAGGTAGAAGGC-CACATCACC-3'. DNA levels were measured in two independent experiments, normalized to GAPDH DNA levels, and the matching normal samples were normalized to represent two DNA copies.

#### References

- Al Moustafa AE, Alaoui-Jamali MA, Batist G, Hernandez-Perez M, Serruya C, Alpert L, Black MJ, Sladek R and Foulkes WD. (2002). *Oncogene*, **21**, 2634–2640.
- Alevizos I, Mahadevappa M, Zhang X, Ohyama H, Kohno Y, Posner M, Gallagher GT, Varvares M, Cohen D, Kim D, Kent R, Donoff RB, Todd R, Yung CM, Warrington JA and Wong DT. (2001). *Oncogene*, **20**, 6196–6204.
- Alter O, Brown PO and Botstein D. (2000). *Proc. Natl. Acad. Sci. USA*, **97**, 10101–10106.
- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W and Lipman DJ. (1997). *Nucleic Acids Res.*, **25**, 3389–3402.

#### Quantitative PCR on cDNA

Total RNAs from matching tumour and normal samples from individual patients were extracted using the RNeasy kit (Qiagen). First-strand cDNA was synthesized with 1 µg of RNA in 20 µl of a reaction mixture containing 0.3 µg of hexanucleotide primers (Boehringer-Mannheim), 200 U of Superscript II RNase H reverse transcriptase (Invitrogen) and 40 U of RNasin (Promega), dNTP and buffer. The mixture was incubated at 42°C for 50 min and then heated to 95°C for 10 min. A measure of 2 µl of a 1/25 dilution of the reverse transcriptase reaction was used for PCR amplification with Sybr Green and the Lightcycler (Roche Molecular Biochemicals). Oligonucleotide primers were designed to cross exon/exon junctions with primer3 ([http://www-genome.wi.mit.edu/cgi-bin/primer/primer3\\_www.cgi](http://www-genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi)). Their sequences are: AB011112, 5'-CACGGAAGGAGTATTTGACCA-3' and 5'-CGATACTGCAGGAGGAGAAAG-3'; Y09538, 5'-CGTG AAGGAGTACGTGAATGC-3' and 5'-ATGGCAGCAGAT ACCAAGATG-3'; U61836, 5'-CACTTCTTGAGCAGGGT TTCA-3' and 5'-ACGCCATTCTTGGGAATA-GAGG-3'; AA418080, 5'-GTAGCCATGACATTTGGAGCAC-3' and 5'-GACAACATGGTGCAGAGAGGT-3'; AF091087, 5'-CAG GGAGAAGCATTGATTGAT-3' and 5'-TTCTCTCCCTTC AACCTGTGA-3'; M69199, 5'-TCAGAGAAACCGCTGA-CATCT-3' and 5'-ATGCAAAATGGTG-GTCATTGT-3'; ubiquitin B, 5'-GCTTTGTTGGGTGA-GCTTGT-3' and 5'-CGAAGATCTGCATTTTGACCT-3'. Gene expression levels were measured in two independent experiments and normalized to ubiquitin expression levels.

#### Acknowledgements

We thank (a) Diemunsch F. for technical assistance; (b) the IGBMC core facilities; (c) the Ligue Regionale contre le Cancer, the Association pour la Recherche sur le Cancer, the Ministère de la Recherche et de la Technologie, and the Fondation pour la Recherche Médicale for fellowships to A Cromer, A Carles, G Ganguli, F Lemaire and J Young; (c) the Ligue Régionale (Bas-Rhin/Haut-Rhin) contre le Cancer for funding for an RT-QPCR machine; (d) The Ministère de la Recherche for the purchase of the Affymetrix arrays; and (e) ARERS Verre Espoir (no. 138.02), Aventis, the Centre National de la Recherche Scientifique, the Institut National de la Santé et de la Recherche Médicale, the Hôpital Universitaire de Strasbourg, the Association pour la Recherche sur le Cancer, the Fondation pour la Recherche Médicale, the Ligue Nationale Française contre le Cancer (Equipe labellisée), the Ligue Régionale (Haut-Rhin) contre le Cancer, the Ligue Régionale (Bas-Rhin) contre le Cancer, the European Union (FP5 project QLK6-2000-00159) and the Ministère de la Recherche (Décisions 99H0161 and 98C0372) for financial assistance.

- Arend WP. (2002). *Cytokine Growth Factor Rev.*, **13**, 323–340.
- Bellin TJ, Singh B, Barber I, Socci N, Wenig B, Smith R, Prystowsky MB and Childs G. (2002). *Cancer Res.*, **62**, 1184–1190.
- Bernards R and Weinberg RA. (2002). *Nature*, **418**, 823.
- Bittner M, Meltzer P, Chen Y, Jiang Y, Seftor E, Hendrix M, Radmacher M, Simon R, Yakhini Z, Ben-Dor A, Sampas N, Dougherty E, Wang E, Marincola F, Gooden C, Lueders J, Glatfelter A, Pollock P, Carpten J, Gillanders E, Leja D, Dietrich K, Beaudry C, Berens M, Alberts D and Sondak V. (2000). *Nature*, **406**, 536–540.
- Crawley JJ and Furge KA. (2002). *Genome Biol.*, **3**, 751–758.

- Eisen MB, Spellman PT, Brown PO and Botstein D. (1998). *Proc. Natl. Acad. Sci. USA*, **95**, 14863–14868.
- El-Naggar AK, Kim HW, Clayman GL, Coombes MM, Le B, Lai S, Zhan F, Luna MA, Hong WK and Lee JJ. (2002). *Oncogene*, **21**, 8206–8219.
- Geminder H, Sagi-Assif O, Goldberg L, Meshel T, Rechavi G, Witz IP and Ben-Baruch A. (2001). *J. Immunol.*, **167**, 4747–4757.
- Genden EM, Ferlito A, Bradley PJ, Rinaldo A and Scully C. (2003). *Oral. Oncol.*, **39**, 207–212.
- Gollin SM. (2001). *Head Neck*, **23**, 238–253.
- Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri MA, Bloomfield CD and Lander ES. (1999). *Science*, **286**, 531–537.
- Huang Q, Yu GP, McCormick SA, Mo J, Datta B, Mahimkar M, Lazarus P, Schaffer AA, Desper R and Schantz SP. (2002). *Genes Chromosomes Cancer*, **34**, 224–233.
- Hyman E, Kauraniemi P, Hautaniemi S, Wolf M, Mousses S, Rozenblum E, Ringner M, Sauter G, Monni O, Elkhouloun A, Kallioniemi OP and Kallioniemi A. (2002). *Cancer Res.*, **62**, 6240–6245.
- Kano M, Nishimura K, Ishikawa S, Tsutsumi S, Hirota K, Hirose M and Aburatani H. (2003). *Physiol Genom.*, **13**, 31–46.
- Kerkela E and Saarialho-Kere U. (2003). *Exp. Dermatol.*, **12**, 109–125.
- Kol S, Ben-Shlomo I, Ruutuainen K, Ando M, Davies-Hill TM, Rohan RM, Simpson IA and Adashi EY. (1997). *J. Clin. Invest.*, **99**, 2274–2283.
- LaTulippe E, Satagopan J, Smith A, Scher H, Scardino P, Reuter V and Gerald WL. (2002). *Cancer Res.*, **62**, 4499–4506.
- Leethanakul C, Patel V, Gillespie J, Pallente M, Ensley JF, Koontongkaew S, Liotta LA, Emmert-Buck M and Gutkind JS. (2000). *Oncogene*, **19**, 3220–3224.
- Lemaire F, Millon R, Young J, Cromer A, Wasylyk C, Schultz I, Muller D, Marchal P, Zhao C, Melle D, Bracco L, Abecassis J and Wasylyk B. (2003). *Br. J. Cancer*, **89**, 1940–1949.
- Li A, Varney ML and Singh RK. (2001). *Clin. Cancer Res.*, **7**, 3298–3304.
- Liu ET. (2003). *Curr. Opin. Genet. Dev.*, **13**, 97–103.
- MacDonald TJ, Brown KM, LaFleur B, Peterson K, Lawlor C, Chen Y, Packer RJ, Cogen P and Stephan DA. (2001). *Nat. Genet.*, **29**, 143–152.
- Mendez E, Cheng C, Farwell DG, Ricks S, Agoff SN, Futran ND, Weymuller Jr EA, Maronian NC, Zhao LP and Chen C. (2002). *Cancer*, **95**, 1482–1494.
- Monni O, Barlund M, Mousses S, Kononen J, Sauter G, Heiskanen M, Paavola P, Avela K, Chen Y, Bittner ML and Kallioniemi A. (2001). *Proc. Natl. Acad. Sci. USA*, **98**, 5711–5716.
- Muller D, Millon R, Velten M, Bronner G, Jung G, Engelmann A, Flesch H, Eber M, Methlin G and Abecassis J. (1997). *Eur. J. Cancer*, **33**, 2203–2210.
- Murakami T, Maki W, Cardones AR, Fang H, Tun Kyi A, Nestle FO and Hwang ST. (2002). *Cancer Res.*, **62**, 7328–7334.
- Nam SW, Clair T, Kim YS, McMarlin A, Schiffmann E, Liotta LA and Stracke ML. (2001). *Cancer Res.*, **61**, 6938–6944.
- Nielsen TO, West RB, Linn SC, Alter O, Knowling MA, O'Connell JX, Zhu S, Fero M, Sherlock G, Pollack JR, Brown PO, Botstein D and van de Rijn M. (2002). *Lancet*, **359**, 1301–1307.
- Phillips JL, Hayward SW, Wang Y, Vasselli J, Pavlovich C, Padilla-Nash H, Pezullo JR, Ghadimi BM, Grossfeld GD, Rivera A, Linehan WM, Cunha GR and Ried T. (2001). *Cancer Res.*, **61**, 8143–8149.
- Platzer P, Uppender MB, Wilson K, Willis J, Lutterbaugh J, Nosrati A, Willson JK, Mack D, Ried T and Markowitz S. (2002). *Cancer Res.*, **62**, 1134–1138.
- Pollack JR, Sorlie T, Perou CM, Rees CA, Jeffrey SS, Lonning PE, Tibshirani R, Botstein D, Borresen-Dale AL and Brown PO. (2002). *Proc. Natl. Acad. Sci. USA*, **99**, 12963–12968.
- Quon H, Liu FF and Cummings BJ. (2001). *Head Neck*, **23**, 147–159.
- Redon R, Muller D, Caulee K, Wanherdrick K, Abecassis J and du Manoir S. (2001). *Cancer Res.*, **61**, 4122–4129.
- Singh D, Febbo PG, Ross K, Jackson DG, Manola J, Ladd C, Tamayo P, Renshaw AA, D'Amico AV, Richie JP, Lander ES, Loda M, Kantoff PW, Golub TR and Sellers WR. (2002). *Cancer Cell*, **1**, 203–209.
- Sobin LH and Fleming ID. (1997). *Cancer*, **80**, 1803–1804.
- Stracke ML, Krutzsch HC, Unsworth EJ, Arestad A, Cioce V, Schiffmann E and Liotta LA. (1992). *J. Biol. Chem.*, **267**, 2524–2529.
- Taichman RS, Cooper C, Keller ET, Pienta KJ, Taichman NS and McCauley LK. (2002). *Cancer Res.*, **62**, 1832–1837.
- Tao M, Li B, Nayini J, Andrews CB, Huang RW, Devemy E, Song S, Venugopal P and Preisler HD. (2000). *Cytokine*, **12**, 699–707.
- Tibshirani R, Hastie T, Narasimhan B and Chu G. (2002). *Proc. Natl. Acad. Sci. USA*, **99**, 6567–6572.
- Tusher VG, Tibshirani R and Chu G. (2001). *Proc. Natl. Acad. Sci. USA*, **98**, 5116–5121.
- Umezū-Goto M, Kishi Y, Taira A, Hama K, Dohmae N, Takio K, Yamori T, Mills GB, Inoue K, Aoki J and Arai H. (2002). *J. Cell Biol.*, **158**, 227–233.
- van't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, Peterse HL, van der Kooy K, Marton MJ, Witteveen AT, Schreiber GJ, Kerkhoven RM, Roberts C, Linsley PS, Bernards R and Friend SH. (2002). *Nature*, **415**, 530–536.
- Villaret DB, Wang T, Dillon D, Xu J, Sivam D, Cheever MA and Reed SG. (2000). *Laryngoscope*, **110**, 374–381.
- Watanabe H, Iwase M, Ohashi M and Nagumo M. (2002). *Oral Oncol.*, **38**, 670–679.
- Youngs SJ, Ali SA, Taub DD and Rees RC. (1997). *Int. J. Cancer*, **71**, 257–266.

Supplementary Information accompanies the paper on Oncogene website (<http://www.nature.com/onc>).

## B. Commentaires : comparaison DD/Affymetrix

---

### 1. Contribution personnelle et cadre de l'étude

Je n'ai participé que très épisodiquement, à titre consultatif (participation initiale au choix des échantillons, au choix des échantillons S et U comme prioritaires, correction du manuscrit), à cette étude au titre de mon appartenance au groupe « Tête et Cou » et de mon implication dans les résultats obtenus en DD par les divers expérimentateurs. Cromer et al. (Cromer et al., 2003) ont mené une analyse par puces à ADN Affymetrix de 12600 gènes sur 38 échantillons (34 T, 4 N). Anne Cromer a assuré cette étude, assistée par le Dr Régine Millon du Centre Paul Strauss pour les étapes de validation, entre autres, et d'Annaïck Carles, bio-informaticienne.

### 2. Résultats résumés

Cette étude a révélé après élimination du bruit de fond une expression différentielle de 2'377 gènes différents entre les tumeurs et les tissus normaux. Un groupe de 119 gènes apparaît qui peut discriminer les tumeurs de stades précoces (E) des tissus normaux (N). Un autre groupe de 164 gènes est potentiellement capable de distinguer les tumeurs présentant ou non des capacités métastatiques (M et NM ; anciennement U et S), dès le traitement par chirurgie. Sur les 1'200 séquences uniques isolées en DD, 223 (18,5%) sont également isolées comme différentielles dans l'étude Affymetrix, soulignant la complémentarité des ces deux approches pour aboutir à une meilleure compréhension de la carcinogenèse des tumeurs VADS. Toutefois peu de ces séquences font partie des groupes prédictifs issus du clustering.

### 3. Comparaison avec le DD

Sur une collection d'un peu plus de 1'200 séquences uniques identifiées en DD (hors étude de chimiorésistance), 428 des séquences sont représentées sur la puce à ADN Affymetrix. Un tel taux de 36% est cohérent avec le taux de couverture du génome en DD, estimé par le nombre d'amorces utilisées, puisque la puce Affymetrix examine environ 1/3



des 30'000-40'000 gènes estimés du génome (Lander et al., 2001). Parmi les séquences validés en DD dans notre publication initiale, certains gènes ont également été détectés comme différentiellement exprimés dans l'étude Affymetrix. Vingt sept des 70 gènes (39%) isolés dans la publication DD sont présents sur la puce Affymetrix, ce qui demeure cohérent avec le taux de couverture théorique de notre étude DD. Cependant, au niveau des listes de gènes sélectionnés et confirmés par clustering dans l'étude Affymetrix, peu de gènes sont communs (8 sur 70). Ces deux approches faiblement redondantes présentent donc des biais et des capacités différentes. Si les gènes identifiés ne sont pas toujours identiques, ils appartiennent souvent à une même famille de gènes, voire ils sont impliqués dans des processus biologiques de même nature. Les protéines structurales (kératines), les interleukines, les ephrines, les protéines des jonctions serrées (Gap Junction proteins, GJB, connexin), les protéines dégradant la matrice (MMPs, Kallikrein-like protein 4), les apolipoprotéines, ou les ubiquitin conjugating enzymes, sont, entre autres, isolées dans ces deux études. Dans les deux études, il semble qu'une partie de l'expression différentielle au niveau de l'ARNm soit potentiellement corrélée à des altérations au niveau génomique. Les deux points chauds isolés en DD, 12p12-13 et 1q21-22 sont également retrouvés dans l'étude Affymetrix, puisqu'elles font partie d'un groupe de gènes de localisations surreprésentées, potentiellement distinctes dans les tumeurs et dans les tissus normaux (groupe 3 : 1q21.3 et 12p13.32).

Du point de vue des gènes candidats, SMAP31 (partie V) fait partie des gènes sélectionnés comme sous exprimés dans les tumeurs, confirmant le profil d'expression de ce gène obtenu en DD, RN, VN, Northern blot et PCR quantitative (Partie V).

Si on compare la liste totale des gènes potentiellement prédictifs de la nature tumorale des échantillons en Affymetrix, et la collection des 70 gènes identifiés par DD, on s'aperçoit que la protéine ribosomale L6, la protéine ribosomale L27a et l'apolipoprotéine L2 sont décrites comme sous exprimées dans les tumeurs en Affymetrix, en opposition au résultat obtenu en DD. En ce qui concerne la kératine 13, RTP/DRG1, PON2, PLAUR et l'intégrine  $\alpha 6$  les résultats sont concordants. La majorité des séquences (62/70) ne sont pas décrites comme différentielles en Affymetrix, peut être en lien avec la capacité de la technique DD de détection de séquences exprimées à faible niveau (exclusion des gènes proches du bruit de fond du « working set » Affymetrix), ou, avec l'absence éventuelle de ces gènes de la puce Affymetrix.

J'ai effectué une comparaison supplémentaire, à partir d'une sélection informatique, de l'ensemble des séquences DD associées à des différences T/N, présentant un profil d'expression documenté (profil E/N) sur Affymetrix. Trois cent trente neuf (339) séquences annotées pour les deux techniques sont isolées (28% de la collection DD hors de l'étude de chimiorésistance, confirmant le taux de couverture de notre étude DD). Deux cent trente neuf (239) séquences ont un profil cohérent en DD et en Affymetrix (70,5%) et 100 séquences présentent un profil inversé (29,5%). Cette apparente divergence des résultats doit être relativisée. Tout d'abord, cette comparaison ne correspond pas à une sélection des séquences les mieux validées, publiées dans les deux études. Ensuite, le fold change en Affymetrix des séquences apparemment divergentes est faible. En effet, le fold change moyen des gènes surexprimés dans les tumeurs (T+) en Affymetrix inversés par rapport au DD est de 1,4 fois tandis que le fold change moyen des gènes sous exprimés dans les tumeurs (T-) en Affymetrix inversés par rapport au DD est -1,48 fois. Ces différences sont donc peu significatives (variation de moins d'un facteur 2). En fait, seuls 32 gènes (9,4%) ont un profil contradictoire entre les deux études avec un fold change supérieur à 2 fois, et, pourraient correspondre à des divergences significatives. Il est probable que, dans cette comparaison, les apparentes divergences correspondent aux bruits de fond combinés des deux études.

Les différences observées entre les résultats obtenus en DD et en Affymetrix peuvent avoir plusieurs causes. Tout d'abord, les deux technologies présentent des biais différents. La technologie Affymetrix utilise des sondes de type linéaire. En revanche, la technologie du DD, utilisée pour identifier des séquences faiblement exprimées, nous a conduit à utiliser des sondes focalisées par PCR avec les amorces HAP. Nous avons déjà constaté que les phénomènes de compétition lors de la PCR ou les limites d'extension des fragments lors de la PCR (optimum d'environ 800bp) pouvaient expliquer une variation légère des profils obtenus. Le filtrage des gènes exprimés à faible niveau en Affymetrix, aboutissant au working set de 3'962 gènes, explique sans doute les différences entre les gènes isolés par les deux approches. Il est également possible que les régions analysées d'un même transcrit soient différentes entre le fragment DD et le « probe set » correspondant Affymetrix, bien que ce dernier soit censé correspondre à des séquences disposées tout du long du transcrit. L'épissage alternatif, comme les propriétés des sondes, pourraient expliquer certains résultats éventuellement divergents. Enfin, les gènes sélectionnés dans l'étude Affymetrix comme prédictifs de la nature tumorale du tissu, ont été sélectionnés sur une collection de patients en grande partie différente de celle utilisée pour l'étude DD.

**Partie V:**  
**Caractérisation du candidat F1.4 : SMAP31**

## A. Matériel et méthodes non présentés dans les articles

---

Je vais présenter, en préambule à la caractérisation des gènes favoris, les protocoles utilisés n'ayant pas déjà été décrits dans les articles ou le manuscrit joint précédemment.

### 1. Immunohistochimie

Les sections ont été déparaffinées par 2 lavages de 10 minutes dans la solution Histolemon. Un cycle de réhydratation a été entrepris par une série de 2 lavages de 5 minutes dans de l'éthanol 100%, 80% et 50%. Après lavage 5 minutes dans de l'eau Millipore (MQ), les sections ont été bouillies 5 minutes au four micro ondes dans de l'acide citrique pH5.2 et laissées à refroidir pendant 30 minutes. Après rinçage dans l'eau MQ pendant 5 minutes, les lames ont été lavées dans un mélange (1 volume H<sub>2</sub>O<sub>2</sub> 30%/9 volumes methanol). Les lames ont ensuite été lavées trois fois 5 minutes dans du PBS. Les sections histologiques ont été délimitées avec un stylo spécial (Dako pen). Les sections ont été bloquées 20 minutes dans une solution de PBS 0,05% Tween 20 (PBST) contenant 4% de sérum albumine bovine (BSA), avant mise en contact avec l'anticorps primaire (sérum purifié ou pré immun) dilué au 1/200<sup>ème</sup> dans du PBST / 1%BSA pendant une heure à température ambiante.

Les sections ont été lavées 3 fois 8 minutes dans du PBST avant mise en contact avec l'anticorps secondaire de chèvre reconnaissant les protéines de lapin (Goat Anti Rabbit Biotynilated, GARB) couplé à la biotine dilué au 1/200<sup>ème</sup> pendant 30 minutes dans du PBST/1%BSA à température ambiante.

Les sections ont été lavées 3 fois pendant 8 minutes dans du PBST pour éliminer l'anticorps secondaire libre. La streptavidine-peroxydase diluée au 1/200<sup>ème</sup> dans du PBS a été mise en contact des sections pendant 30 minutes. L'excès a été éliminé par 3 lavages de 8 minutes dans du PBS. La section a ensuite été révélée avec le substrat DAB (DAB substrate Kit, Vektor SK4100) pendant 10 minutes avant lavage à l'eau MQ. Les sections ont ensuite été traitées au methyl green pour reconnaître les structures histologiques et déshydratées par lavage à des concentrations croissantes d'éthanol, puis dans la solution histosol. Après déshydratation complète, les lames ont été montées dans la solution de montage Eukitt.

Comme pour les sections analysées en Hybridation *In Situ*, une barre correspondant à 50µm est représentée sur les images.

## 2. Immunocytochimie

Les cellules ont étéensemencées sur des boîtes 6 puits, au fond desquelles des lames couvre objets avaient été disposées. Les cellules ont adhéré à la surface de la lame, pendant 3h environ ou 12h suivant la méthode de transfection, puis ont été transfectées. Après expression des plasmides 24h, les puits ont été lavés 3 fois avec du PBS. Après aspiration complète du liquide, 500µl d'acétone/méthanol (50%/50%), conservé à -20°C a été ajouté dans les puits. Ce liquide a été aspiré immédiatement et remplacé par 1,5ml d'acétone/méthanol (50%/50%). La boîte a été incubée pendant 10 minutes (ou 12h) à -20°C. Le liquide a été éliminé et on a attendu un séchage complet avant stockage éventuel à -20°C. Lors de la reprise du protocole, les cellules ont été réhydratées dans 1ml de PBS pendant 2 minutes. Dans toutes étapes ultérieures, il est critique d'aspirer le milieu et de le remplacer sans dessèchement des plaques. On a disposé sur les lames 300µl de PBS/3% BSA (Bovine Serum Albumine), et la réaction de blocage a été poursuivie pendant 30 minutes à température ambiante. On a ajouté ensuite 1ml de PBS, qui a ensuite été remplacé par 200µl d'une solution de PBS/0,5% BSA contenant l'anticorps primaire anti-flag (2FL 1B11) dilué au 1/300<sup>ème</sup> ou les sérums de lapin dilués au 1/200<sup>ème</sup> ou au 1/1000<sup>ème</sup>. Des pièces de gaze imbibées de PBS ont été disposées pour éviter la déshydratation avant incubation 2 heures à 37°C ou alternativement 12h à 4°C (moins de bruit de fond). La gaze a été récupérée et, sans laisser les plaques se déshydrater, 1ml de PBS a été ajouté dans chaque puits et laissé en contact avec agitation modérée pendant 10 à 30 minutes. Ce lavage a été répété 2 fois avant ajout de 200µl de l'anticorps secondaire (chèvre anti-souris conjugué au fluorochrome Cy3 ou âne anti-lapin conjugué au fluorochrome Cy3 pour les sérums) dilué au 1/800<sup>ème</sup> dans du PBS/0,5% BSA. Des pièces de gaze imbibées de PBS ont été disposées pour éviter la déshydratation. Les plaques ont été incubées pendant 1h à 1h30 à 37°C. Après récupération des pièces de gaze, 1ml de PBS a été ajouté à chaque puits, et laissé en contact pendant 10 à 30 minutes avec agitation modérée. Ce lavage a été répété 2 fois avant élimination du liquide et mise en contact pendant 1 minute maximum avec une solution de Hoechst/PBS (25µl/5ml) pour obtenir le marquage des noyaux. Les puits ont été lavés 3 fois avec 1ml de PBS. Les lamelles couvre objets ont été récupérées et ont été montées sur des lames porte objets dans une solution de montage (PBS, glycérol, propyl gallate). Les lames ont été observées sous un microscope à fluorescence.

## 3. FACS

Après avoir laissé les plasmides s'exprimer dans les cellules pendant 48h, le milieu de culture est récupéré et 3 lavages sont effectués avec 10ml de PBS. Un lavage supplémentaire est effectué dans 10ml de PBS-EDTA 0.1%. Pour pouvoir analyser les cellules en apoptose en suspension dans le milieu, tous les lavages sont récupérés et seront centrifugés avec les cellules trypsinées. Les cellules sont trypsinées avec 1ml de trypsine diluée à température ambiante, jusqu'à décollement de la boîte de culture. La suspension de cellules dans la trypsine est neutralisée par 3ml de milieu complet DMEM (Dulbecco's Minimum Essentiel Medium) +10% de sérum de veau foetal (FCS). Les lavages et la suspension de cellules sont regroupés et centrifugés pendant 5 minutes à 800 rpm. Le surnageant est éliminé. Dans le cadre de l'utilisation de l'anticorps anti CD20, les cellules sont reprises dans 50µl de milieu DMEM 10%FCS + 10µl d'anticorps anti CD20 FITC et incubé pendant 45 minutes à 4°C. Cette étape n'est pas nécessaire lors de l'utilisation de la GFP-spectrine (Green Fluorescent Protein-spectrin). Les culots de cellules sont lavés 3 fois avec du PBS, avant d'être repris dans 200µl de PBS et fixées par ajout goutte à goutte en agitation de 400µl d'éthanol 95% conservé à -20°C pour une concentration finale de 70% d'éthanol. La fixation a été entreprise 12h à 4°C. Les cellules sont ensuite centrifugées à 1200 rpm (cellules plus petites), et l'éthanol éliminé. Les cellules sont réhydratées dans 3ml de PBS pendant 30 minutes. Un lavage supplémentaire dans 3ml de PBS est effectué avant resuspension dans 250µl de RNase A (1mg/ml dans du PBS) et ajout de 250µl d'iodure de propidium (50µg/ml dans du PBS). Les cellules sont incubées 12h à 4°C avant analyse dans le cytomètre de flux FACSCalibur (Becton Dickinson) et traitement des données par le programme ModFit LT 3.1.

#### **4. Cultures cellulaires :**

La lignée de kératinocytes de peau spontanément immortalisés HaCaT a été cultivée dans le milieu DMEM + 10% FCS. Les cellules Fadu d'origine pharyngées (après adaptation) ainsi que les cellules Det562, isolées d'une tumeur métastatique d'origine pharyngée, ont été cultivées dans le même milieu.

Les cellules COS ont été cultivées dans le milieu DMEM + 5% FCS. Les cellules RPMI2650, originaires de la cavité buccale, ont été cultivées dans le milieu MEM +10%FCS + 1mM Sodium pyruvate+ AANE (acides aminés non essentiels) + Glutamine.

#### **5. Transfection**

#### **a. Méthode « JetPEI »**

Les transfections ont été menées suivant les spécifications du fabricant sur des cellules ayant été divisées la veille de l'expérimentation, s'étant attachées durant 12h au support pour atteindre environ 70% de confluence (JetPEI easy protocol ; PolyTransfection).

#### **b. Méthode « BBS » (Phosphate de Calcium)**

Les transfections avec la technique BBS (BES buffered solution) ont été réalisées sur des cellules trypsinées le matin de l'expérimentation, ayant adhéré environ 3h, à une densité d'environ 30% de confluence. Le milieu de transfection a été laissé en contact avec les cellules 12h, puis, les boîtes ont été lavées trois fois au PBS avant reconstitution du milieu normal. Les cellules ont ensuite été incubées au minimum 12h pour permettre l'expression des plasmides.

### **6. Test de clonogénicité et clones stables**

Les cellules ont été transfectées avec les plasmides pSG5 puro Flag, contenant un gène de résistance à la puromycine, et, l'expression a été maintenue pendant 24h sans sélection. La sélection pour la résistance à la puromycine a été introduite par ajout au milieu de 25µl de puromycine 1mg/ml à 50ml de milieu de culture. Le milieu de culture a été changé tous les deux jours de façon à éliminer les cellules mortes en suspension, et, à maintenir la sélection. Les premières colonies ont pu être observées entre 5 jours et 7 jours de sélection. Toutefois la sélection a été maintenue pendant au moins 3 semaines au cours des étapes de repiquage et d'amplification des clones avant congélation.

### **7. Western Blotting**

Les protéines ont été séparées sur un gel de polyacrylamide constitué d'un gel de stacking : (5% acrylamide/bisacrylamide 29/1, TrisHCl-SDS pH 6.8) et d'un gel de séparation (10 à 15% acrylamide/bisacrylamide 29/1, TrisHCl-SDS pH 8.8).

Les protéines ont été soumises à une électrophorèse pendant 1 heure à 150 volts dans le tampon « Tris Glycine SDS PAGE », puis ont été transférées sur membrane de nitrocellulose, pendant 1h30 à 150 Volts dans le tampon de transfert. La qualité du transfert a été évaluée par

coloration au rouge ponceau. Les membranes ont été saturées pendant 2h avec la solution PTMS (PBS Tween Milk solution, contenant, 0,05% de Tween 20, 5% de lait déshydraté). L'anticorps primaire (ou le sérum) a été mis en contact avec la membrane 12h à 4°C à des dilutions variant de 1/1000 à 1/200 suivant la sensibilité de l'anticorps dans la solution PTMS. Le lendemain, les membranes ont été lavées 3 fois 5 minutes avec du PBST. L'anticorps secondaire adéquat (Chèvre anti lapin pour les sérums, Chèvre anti souris pour l'anticorps flag) couplé à la peroxidase (HRP Horse Radish Peroxydase) a été mis en contact pendant 1h dans du PBST à une dilution 1/10000. L'excès d'anticorps a ensuite été éliminé par 3 lavages de 5 minutes avec du PBST. Les membranes ont été révélées par exposition sur film après ajout des substrats du kit PIERCE SupersignalSubstrat ECL.

Dans le cas des anticorps anti p-ERK1/2 (Phospho-p44/42 MAP kinase Antibody), le TBST a été utilisé à la place du PBST sur des extraits de cellules transfectées puis cultivées 12h en absence de sérum de manière à limiter l'activation basale des MAP kinases. Cet anticorps est détecté avec un anticorps secondaire de chèvre anti lapin (Goat anti rabbit). Une révélation ultérieure avec l'anticorps anti-TBP (souris) permet de vérifier la quantité relative de protéines pour chaque échantillon, après transfert sur la membrane.



## B. Communication courte sur F1.4/HOP

---

Nous rapportons dans une communication courte l'implication de HOP dans la carcinogénèse des cancers VADS, suggérant un rôle éventuellement plus général de suppresseur de tumeurs de HOP que dans le seul choriocarcinome.

## Implication of the HOP tumour suppressor in HNSCC carcinogenesis.

Frédéric Lemaire<sup>1</sup>, Régine Millon<sup>2</sup>, Danielle Muller<sup>2</sup>, Yannick Rabouel<sup>2</sup>, Laurent Bracco<sup>3</sup>,  
Joseph Abecassis<sup>2</sup>, Bohdan Wasyluk<sup>1#</sup>

**Running title:** HOP tumour suppressor in HNSCC

**Keywords:** Hypopharynx, homeodomain, development, differentiation.

<sup>1</sup>Institut de Génétique et de Biologie Moléculaire et Cellulaire, CNRS/INSERM/ULP, 1 Rue Laurent Fries, BP 10142, 67404 Illkirch cedex, France. <sup>2</sup>UPRES EA 34-30, Centre Paul Strauss, 3 rue de la Porte de l'Hôpital, 67085 Strasbourg, France. <sup>3</sup>Exonhit Therapeutics S.A., 65 Boulevard Masséna, Paris F-75013, France.

# To whom correspondence should be addressed. Tel: 33 (0) 3 88 65 34 11. Fax: 33 (0) 3 88 65 32 01. Email: boh@igbmc.u-strasbg.fr.

## **Summary**

We report that HOP is expressed in the suprabasal layer of normal uvula and expression strongly decreases in hypopharyngeal carcinoma. Interestingly, HOP has very recently been shown to be a tumour suppressor involved in differentiation, suggesting that HOP has a similar role in head and neck squamous cell carcinoma (HNSC).

## **Introduction**

HNSCC occurs through malignant conversion of basal layer keratinocytes of the epithelium of the upper aerodigestive tract (pharynx, hypopharynx, and larynx) and the oral cavity. HNSCC is the fifth most common cancer worldwide (Sankaranarayanan et al., 1998) and the 5-year survival rate is only 18% for hypopharyngeal carcinoma (Dimery & Hong, 1993). We have performed Differential Display (DD) (Lemaire et al., 2003) and DNA microarray (Cromer et al., 2003) analysis of hypopharyngeal tumours to search for new biomarkers and targets for drug design. In this complementary study, we report that expression of HOP is strongly decreased in tumours. Interestingly, in a series of recent independent reports, this protein under different names (mOB1/HOP/LAGY/NECC1/SMAP31) has been shown to be a potent tumour suppressor involved in development (Adu et al., 2002; Asanoma et al., 2003; Chen et al., 2002; Chen et al., 2003; Shin et al., 2002). We show here that HOP is expressed at a precise stage of epidermal differentiation and its expression is lost in HNSCC, suggesting that HOP may be a general tumour suppressor gene involved in keratinocyte differentiation.

## **Materials & Methods**

Tumour samples, Northern and Real-Time Quantitative PCR (RT-QPCR) were as described previously (Lemaire et al., 2003). The primers sequences were: HOP, 5'-TCAACAAGGTCGACAAGCAC-3' and 5'-TCTGTGACGGATCTGCACTC-3'; the ubiquitous gene RPLPO, 5'-GAAGGCTGTGGTGCTGATGG-3' and 5'-CCGGATATGAGGCAGCAGTT-3'. F1.4-pGEMT-Easy: pGEM-T Easy with a 572bp insert corresponding to SMAP31 (now HOP transcript variant 1, Accession Number

BC014225, nt 675-1246). F1.4 is a DD clone that did not appear to be differentially expressed by the sensitive Reverse Northern approach we described previously (Lemaire et al., 2003), but scored differential by other approaches (see Results). In situ hybridisation (ISH) was performed by a standard technique (Cau et al., 1997; Strahle et al., 1994). Non-radioactive digoxigenine labelled RNA probes were synthesized from 15µg of linearised F1.4-pGEMT-Easy (antisense: Aat2 and SP6 RNA polymerase; sense: Spe1 and T7 polymerase).

## Results

We isolated a differentially expressed sequence (F1.4) in a DD analysis of RNA from HNSCC and matching normal tissue. The sequence corresponded to SMAP31/LAGY, which at that time was annotated in the NCBI database as a clone under-expressed in choriocarcinoma versus normal placenta villi (SMAP31) and lung cancer versus normal lung (LAGY). More recently, it has been called HOP (Chen et al., 2002; Shin et al., 2002 ) since it is a homodomain-only protein.

HOP (F1.4) expression in the patients used for the DD was analysed on Virtual Northern blots, that contain cDNA synthesised by SMART (Lemaire et al., 2003) (Figure 1A). In all 9 matched tumour-normal pairs there was a striking decrease in expression in the tumours. The major transcript migrates around 1,000 bp, which corresponds to the sizes reported for HOP [1200 bp, (Chen et al., 2002; Shin et al., 2002)], LAGY [HOP/LAGY transcript variants: 1013bp, NM\_139212; 989bp, NM\_139211, 1265bp, NM\_032495, (Chen et al., 2003)], and NECC1/SMAP31-22, SMAP31-R and SMAP31-12 [980bp, AB059409, 989bp, AB059410 and 1097bp, AB059408 respectively, (Asanoma et al., 2003)]. HOP expression in other samples was analyzed by Northern blotting (Figure 1B) and slot blotting (Figure 1C), which do not involve PCR amplification. As expected, the predominant band on

Northern blots migrate around 1,000 bp compared to the ribosomal RNA markers (Figure 1B). HOP expression is decreased in 4 out of 5 matched tumour-normal pairs (Patients 1-5) and in several individual tumours (lanes 14-17) compared to a pool of normal tissues (lane 18) and the other normal samples (lanes 3, 5, 8, 11, 13 and 19). On the slot blots, all 10 tumour-normal pairs have decreased expression in the tumours (Figure 1C, compare samples with equivalent signals in the RPL0 loading controls). Expression of HOP was also decreased in involved lymph nodes (L; Figure 1B, lanes 2, 7, 10; Figure 1C, columns 14, 17, 20 and 23).

In Situ Hybridization (ISH) of frozen sections was used to localise HOP mRNA expression (Figure 2). HOP expression was detected in the suprabasal layer of the epithelium of normal tissue (2A, antisense; 2B, Hematoxylin/Eosin). There was no significant signal with the sense probe and normal tissue, or with the antisense probe and several tumour samples (data not shown). The expression pattern is compatible with a role for HOP in epithelial differentiation.

These results suggested that HOP expression might be associated with the degree of differentiation of the tumours. To test this possibility, Real-Time Quantitative PCR (RT-QPCR) was used to measure HOP expression in a panel of 33 hypopharyngeal carcinomas of different histopathological differentiation status, for which there were 23 matched normal tissues (Figure 3 and data not shown). HOP expression was found to be decreased in 22/23 primary tumours compared to normal matched samples, with N/T fold changes ranging from 3.2 to 128 (average = 27 fold). There was no statistically significant difference in HOP expression between tumours with different features, suggesting that there is no association between HOP expression and differentiation status (ANOVA,  $p=0.16$ ).

## Discussion

We have shown that HOP expression is strikingly decreased in hypopharyngeal HNSCC, in 44 out of 46 patients analysed. There was no statistically significant correlation between residual expression and differentiation grade of the tumours. We also observed decreased HOP expression in tumours by micro-array analysis (Cromer et al., 2003). HOP was found to be expressed in the suprabasal layer of the epithelium, pointing to a role in keratinocytes differentiation.

There is a striking recent convergence of results from several laboratories that indicate that HOP is an important new tumour suppressor. At the end of 2002, several studies showed that HOP (mOB1) is expressed during mouse development (Adu et al., 2002) and is involved in cardiac development (Chen et al., 2002; Shin et al., 2002). More recently, HOP was shown to be a choriocarcinoma suppressor gene involved in cytotrophoblast differentiation (Asanoma et al., 2003), and to be under-expressed in primary lung tumours compared with normal lung (Chen et al., 2003). HOP is located in a chromosomal region (4q11-q12) that is frequently deleted in solid tumours, including lung tumours, hepatocellular carcinoma and bladder cancer (Koo et al., 1999; Petersen et al., 1997; Sakakura et al., 1999; Wong et al., 1999). Our study provides evidence for a role of HOP in HNSCC.

HOP consists of 73 amino-acids with an unusual 60 amino-acid homeodomain, and is the smallest homeodomain protein to date. Homeodomain proteins are important for embryogenesis and development [reviewed in (Chi & Epstein, 2002; Duboule, 1994)]. HOP does not possess conserved homeodomain residues needed for DNA binding, but modulates cardiac genes expression by direct interaction with SRF and inhibition of SRF binding to DNA (Chen et al., 2002; Shin et al., 2002).

HOP is widely expressed in mouse and human tissues, suggesting it may have an important role in many cell types. Using adult human tissue dot blots (Clontech ref PT3307-1), we found expression in oesophagus, lung, placenta, thyroid gland, foetal lung, and brain (in decreasing order, data not shown). A splicing variant of mOB1, comprising an additional protein coding exon 5, can be found in cDNA libraries and in nucleotide sequences databases. We did not detect this variant by RT-PCR (data not shown). The cellular localisation of HOP is predicted to be mainly nuclear. We found by immunocytochemistry with our specific antibodies and HaCat spontaneously immortalized keratinocytes endogenous HOP is nuclear (data not shown) consistent with other studies (Chen et al., 2002; Shin et al., 2002).

HOP expression has been detected in primary cultures of normal lung but not in lung tumour derived cell lines (Chen et al., 2003). Similarly, we detected, by semi-quantitative RT-PCR, expression in primary keratinocytes and HaCat cells, but not in a panel of 15 HNSCC derived cell lines (data not shown). Reintroduction of HOP in choriocarcinoma cell lines inhibits cell proliferation and tumour formation in nude mice (Asanoma et al., 2003). We report that HOP expression is decreased in HNSCC tumours compared to their normal matched samples, consistent with studies in other cancer types (Asanoma et al., 2003; Chen et al., 2003). The reduction of HOP/LAGY expression in lung squamous cell carcinoma is correlated with increasing TNM staging. A complete loss of expression of HOP in two poorly differentiated lung tumour samples has also been reported (Chen et al., 2003). In our study, there is no statistically significant association with tumour differentiation, suggesting that loss of HOP expression is important in all tumours despite the degree of differentiation. We have shown that HOP is expressed in the suprabasal layer of the upper aerodigestive tract epithelium, suggesting that it is involved in keratinocyte differentiation. HOP is potentially responsible for the balance between proliferation and differentiation of cardiomyocytes in



developing heart (Shin et al., 2002). HOP over-expression in choriocarcinoma cell lines induces the expression of CSH1, a marker of differentiated syncytiotrophoblasts (Asanoma et al., 2003). Taken together, the results in HNSCC and in other cancer types indicate that HOP could be an important tumour suppressor gene in a wide range of solid tumours.

## **Acknowledgements**

We thank: Annaïck Carles, Raymond Ripp and Olivier Poch for help with the bioinformatics, Valérie Fraulob and Pascal Dollé for help with ISH, and the IGBMC core facilities. We acknowledge for financial assistance: the Ligue Nationale Française contre le Cancer (Equipe labellisée), the Ligue Régionale (Haut-Rhin) contre le Cancer, the Ligue Régionale (Bas-Rhin) contre le Cancer, ARERS Verre Espoir (n° 138.02), the Centre National de la Recherche Scientifique, the Institut National de la Santé et de la Recherche Médicale, the Hôpital Universitaire de Strasbourg, the Association pour la Recherche sur le Cancer, the Fondation pour la Recherche Médicale, the European Union (FP5 project QLK6-2000-00159) and the Ministère de la Recherche (Décisions 99H0161 and 98C0372).

## Figure Legends

**Figure 1.** Analysis of Hop RNA expression by Virtual Northern blots (A), Northern blots (B) and Northern slot blots (C). Bars indicate Tumour (T) and normal (N) samples from the same patients [A, Patients 1-10; B, Patients 1-5; C, Patients (Pat.) 1-10]. Comparisons should be made between matched samples or globally between unmatched samples [B, tumours from Patients 6-10 and the pool of normal samples (P)]. For the slot blots (C), mRNA was spotted at 4 concentrations, increasing from the bottom to the top of membranes. RPLP0 is the loading control.

**Figure 2.** ISH of normal tissue. The uvula was hybridized with an antisense probe (A) or stained with Hematoxylin/Eosin (B). The black arrowhead indicates the specific stain and the white arrowhead the basal layer. The bars represent 50µm.

**Figure 3.** Real time quantitative RT-PCR.

HOP expression was analysed in matched tumour and normal samples (1-23) of different differentiation (diff) status. The values for the tumours (open columns) and matched normal tissue (black columns) were adjusted according to the RPLP0 internal controls.

## References

- Adu, J., Leong, F.T., Smith, N.R., Leek, J.P., Markham, A.F., Robinson, P.A. & Mighell, A.J. (2002). Expression of mOb1, a novel atypical 73 amino acid K50-homeodomain protein, during mouse development. *Mech Dev*, **119 Suppl 1**, S43-7.
- Asanoma, K., Matsuda, T., Kondo, H., Kato, K., Kishino, T., Niikawa, N., Wake, N. & Kato, H. (2003). NECC1, a candidate choriocarcinoma suppressor gene that encodes a homeodomain consensus motif. *Genomics*, **81**, 15-25.
- Cau, E., Gradwohl, G., Fode, C. & Guillemot, F. (1997). Mash1 activates a cascade of bHLH regulators in olfactory neuron progenitors. *Development*, **124**, 1611-21.
- Chen, F., Kook, H., Milewski, R., Gitler, A.D., Lu, M.M., Li, J., Nazarian, R., Schnepf, R., Jen, K., Biben, C., Runke, G., Mackay, J.P., Novotny, J., Schwartz, R.J., Harvey, R.P., Mullins, M.C. & Epstein, J.A. (2002). Hop is an unusual homeobox gene that modulates cardiac development. *Cell*, **110**, 713-23.
- Chen, Y., Petersen, S., Pacyna-Gengelbach, M., Pietas, A. & Petersen, I. (2003). Identification of a novel homeobox-containing gene, LAGY, which is downregulated in lung cancer. *Oncology*, **64**, 450-8.
- Chi, N. & Epstein, J.A. (2002). Getting your Pax straight: Pax proteins in development and disease. *Trends Genet*, **18**, 41-7.
- Cromer, A., Carles, A., Millon, R., Ganguli, G., Chalmel, F., Lemaire, F., Young, J., Dembele, D., Thibault, C., Muller, D., Poch, O., Abecassis, J. & Wasylyk, B. (2003). Identification of genes associated with tumorigenesis and metastatic potential of hypopharyngeal cancer by microarray analysis. *Oncogene*.

- Dimery, I.W. & Hong, W.K. (1993). Overview of combined modality therapies for head and neck cancer. *J Natl Cancer Inst*, **85**, 95-111.
- Duboule, D. (1994). How to make a limb? *Science*, **266**, 575-6.
- Koo, S.H., Kwon, K.C., Ihm, C.H., Jeon, Y.M., Park, J.W. & Sul, C.K. (1999). Detection of genetic alterations in bladder tumors by comparative genomic hybridization and cytogenetic analysis. *Cancer Genet Cytogenet*, **110**, 87-93.
- Kornberg, T.B. (1993). Understanding the homeodomain. *J Biol Chem*, **268**, 26813-6.
- Lemaire, F., Millon, R., Young, J., Cromer, A., Wasylyk, C., Schultz, I., Muller, D., Marchal, P., Zhao, C., Melle, D., Bracco, L., Abecassis, J. & Wasylyk, B. (2003). Differential expression profiling of head and neck squamous cell carcinoma (HNSCC). *Br J Cancer*, **89**, 1940-9.
- Petersen, I., Langreck, H., Wolf, G., Schwendel, A., Psille, R., Vogt, P., Reichel, M.B., Ried, T. & Dietel, M. (1997). Small-cell lung cancer is characterized by a high incidence of deletions on chromosomes 3p, 4q, 5q, 10q, 13q and 17p. *Br J Cancer*, **75**, 79-86.
- Sakakura, C., Hagiwara, A., Taniguchi, H., Yamaguchi, T., Yamagishi, H., Takahashi, T., Koyama, K., Nakamura, Y., Abe, T. & Inazawa, J. (1999). Chromosomal aberrations in human hepatocellular carcinomas associated with hepatitis C virus infection detected by comparative genomic hybridization. *Br J Cancer*, **80**, 2034-9.
- Sankaranarayanan, R., Masuyer, E., Swaminathan, R., Ferlay, J. & Whelan, S. (1998). Head and neck cancer: a global perspective on epidemiology and prognosis. *Anticancer Res*, **18**, 4779-86.
- Shin, C.H., Liu, Z.P., Passier, R., Zhang, C.L., Wang, D.Z., Harris, T.M., Yamagishi, H., Richardson, J.A., Childs, G. & Olson, E.N. (2002). Modulation of cardiac growth and development by HOP, an unusual homeodomain protein. *Cell*, **110**, 725-35.

Strahle, U., Blader, P., Adam, J. & Ingham, P.W. (1994). A simple and efficient procedure for non-isotopic in situ hybridization to sectioned material. *Trends Genet*, **10**, 75-6.

Wong, N., Lai, P., Lee, S.W., Fan, S., Pang, E., Liew, C.T., Sheng, Z., Lau, J.W. & Johnson, P.J. (1999). Assessment of genetic changes in hepatocellular carcinoma by comparative genomic hybridization analysis: relationship to disease stage, tumor size, and cirrhosis. *Am J Pathol*, **154**, 37-43.

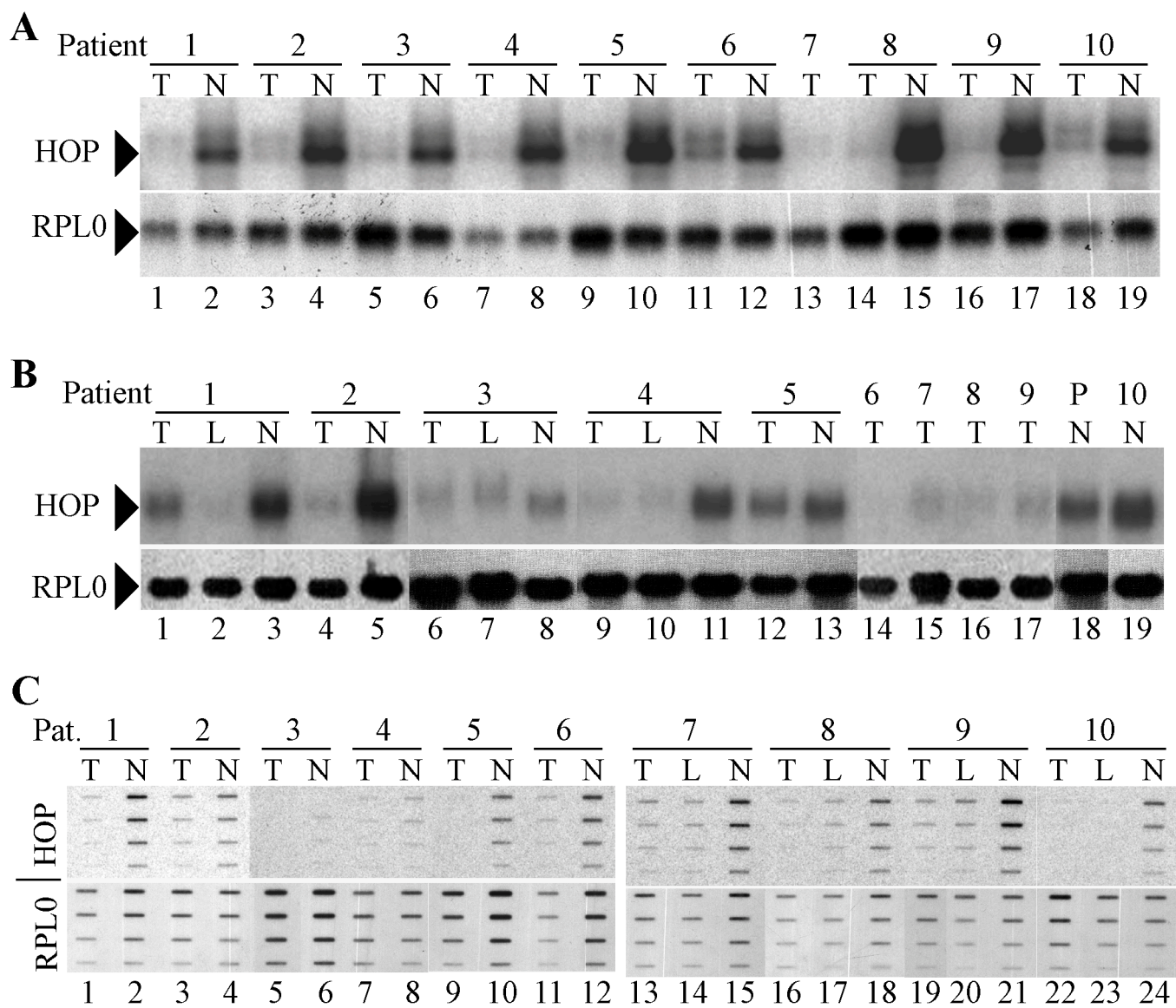


Figure 1 Lemaire et al.

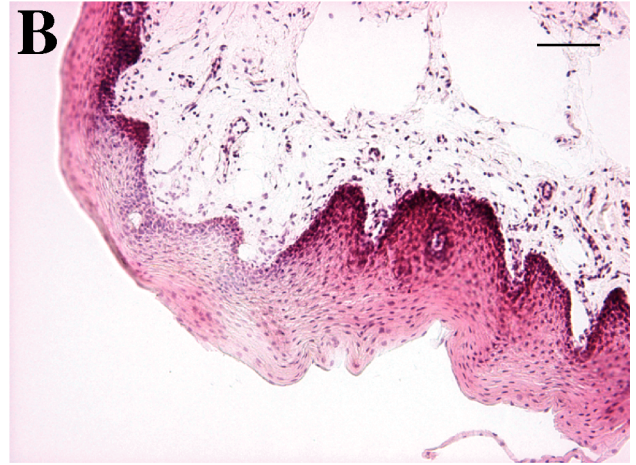
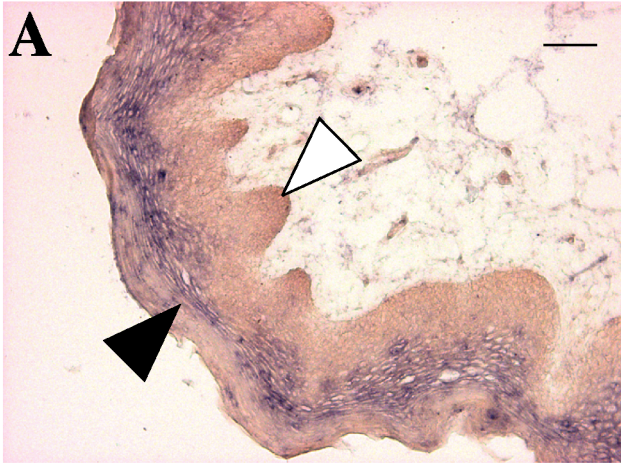


Figure 2 Lemaire et al.



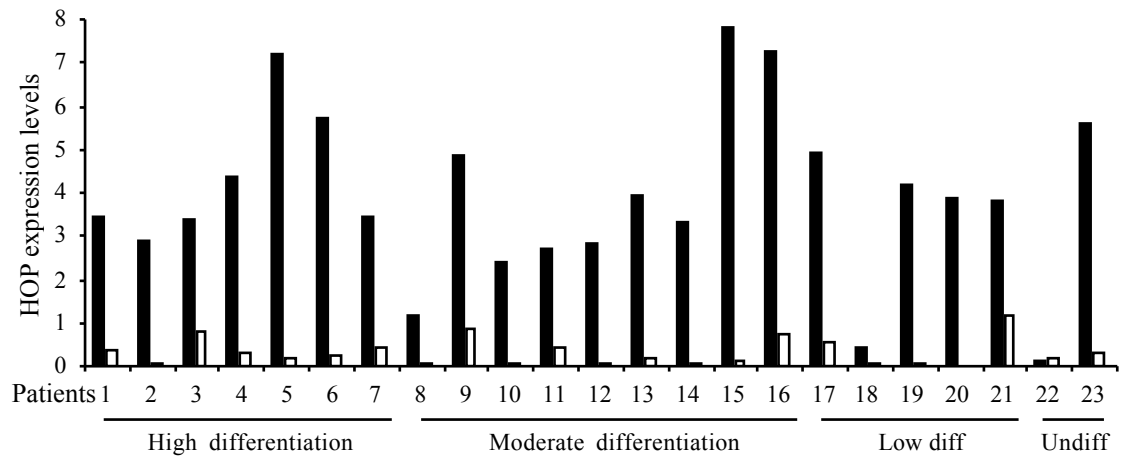


Figure 3 Lemaire et al.

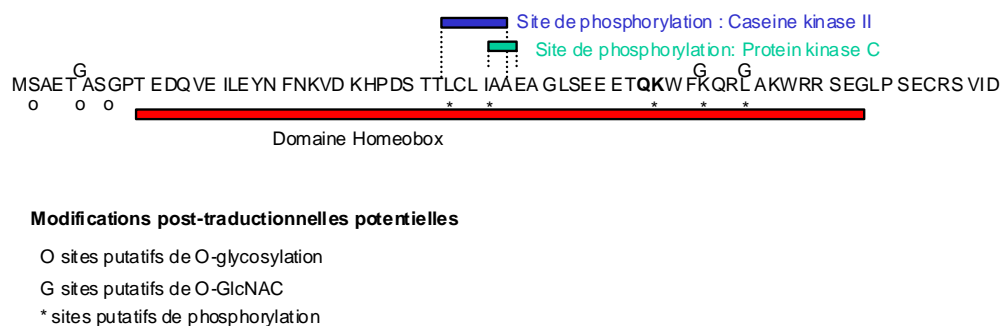
## C. Caractérisation de F1.4/HOP/SMAP31/NECC1/mOB1

---

### 1. Analyse informatique de SMAP31

Une analyse informatique des sites putatifs de modifications post-traductionnelles a révélé la présence de sites potentiels de glycosylation, ainsi que de phosphorylation (non décrits dans le manuscrit précédent, FigV.1).

Une prédiction informatique de la localisation subcellulaire de cette protéine par PSORT (<http://psort.nibb.ac.jp/>), d'après sa composition en acides aminés, a indiqué une présence prédominante de la protéine au niveau nucléaire (43,5%), une localisation cytoplasmique partielle (30,4%), une présence partielle au niveau des mitochondries (17,6%) et du cytosquelette (8,7%). La protéine alors inconnue SMAP31 possédait donc une localisation nucléaire prédite compatible avec une interférence putative de cette protéine avec le facteur Pax6, et, un rôle éventuel de liaison à l'ADN.



### Structure de la protéine de SMAP31

Figure V.1

### 2. Caractérisation de l'expression dans les lignées cellulaires VADS

L'expression de HOP n'est détectable avec les amorces ayant servi au clonage de la séquence codante dans aucune des lignées VADS analysées. Il semble, d'après l'analyse par

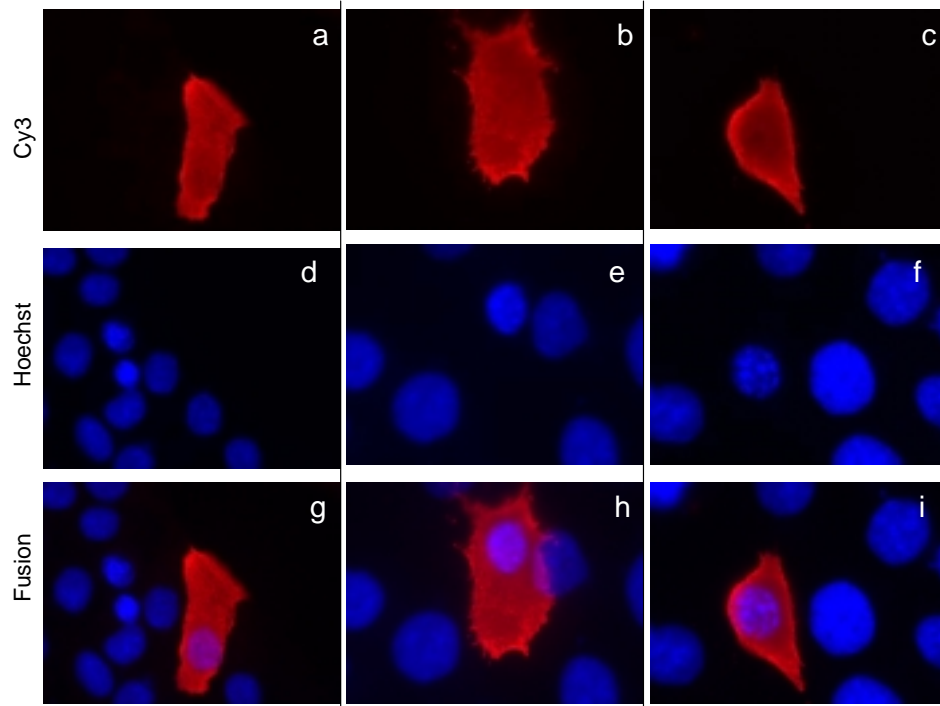
puces à ADN à façon construite à l'IGBMC, que F1.4 est exprimé dans les kératinocytes primaires HEK-a et faiblement dans la lignée de kératinocytes spontanément immortalisés HaCaT, qui sera notre modèle d'étude de ce clone. Les lignées dérivées de tumeurs d'origine pharyngée Fadu et Det562, n'expriment pas ou très faiblement ce clone. L'effet de la restauration de l'expression de SMAP31 dans ces lignées a été étudié. L'ensemble de ces trois lignées permet également l'étude des autres clones analysés à titre individuel.

### **3. Clonage de la séquence codante**

Lors de la caractérisation initiale de SMAP31/HOP, l'ORF (Open Reading Frame, phase ouverte de lecture) prédite contenait deux codons ATG proches. Le rôle relatif de ces deux codons n'étant pas connu et les étapes ultérieures de l'étude mettant de toute façon en jeu l'utilisation d'une étiquette Flag de 18aa, j'ai préféré cloner la version la plus longue de 91 aa de la séquence codante, possédant 18 aa supplémentaire par rapport à la séquence codante longue de 73aa maintenant publiée (accession number NP\_631958). Cette séquence a été clonée en phase de lecture dans le vecteur d'expression pSG5 puro Flag, après amplification par PCR haute fidélité à partir des échantillons normaux SMART N. Aucune amplification n'a été obtenue lors de l'amplification en parallèle de l'échantillon de Pool Tumoral T.

### **4. Etude de la localisation subcellulaire de la protéine transfectée par immunocytochimie avec l'anticorps Anti-Flag**

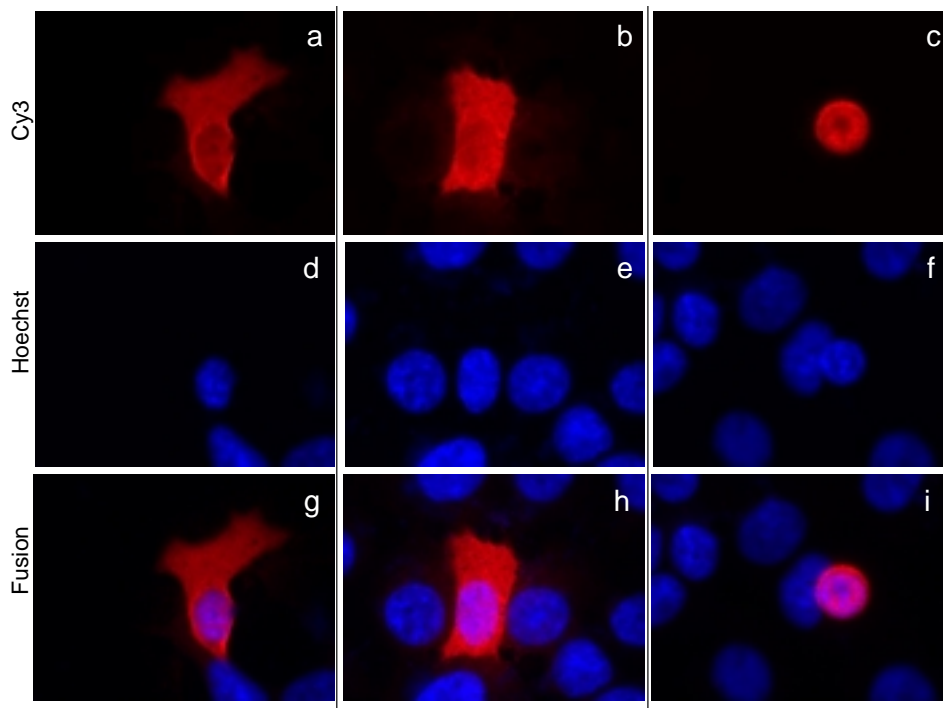
Lors de premières expériences (FigV.4), la localisation avait été observée comme majoritairement nucléaire ou périnucléaire. Au cours d'expérimentations plus récentes avec la méthode de transfection JetPEI, contrairement à la localisation attendue comme majoritairement nucléaire, la localisation a été détectée comme largement cytoplasmique dans les lignées HaCaT (Fig IV.2) et Fadu (FigV.3). Cette différence peut être due à une surexpression intense de la protéine, entraînant une localisation cytoplasmique par saturation du noyau. Je ne peux exclure un effet des séquences supplémentaires comme du peptide Flag.



**Immunocytochimie pour HOP-Flag; lignée HaCaT**

Trois exemples de localisation avec l'anticorps anti-flag sont montrés (a,b,c) ainsi que les marquage des noyaux correspondants (c,d,f) et la fusion de ces images (g,h,i).

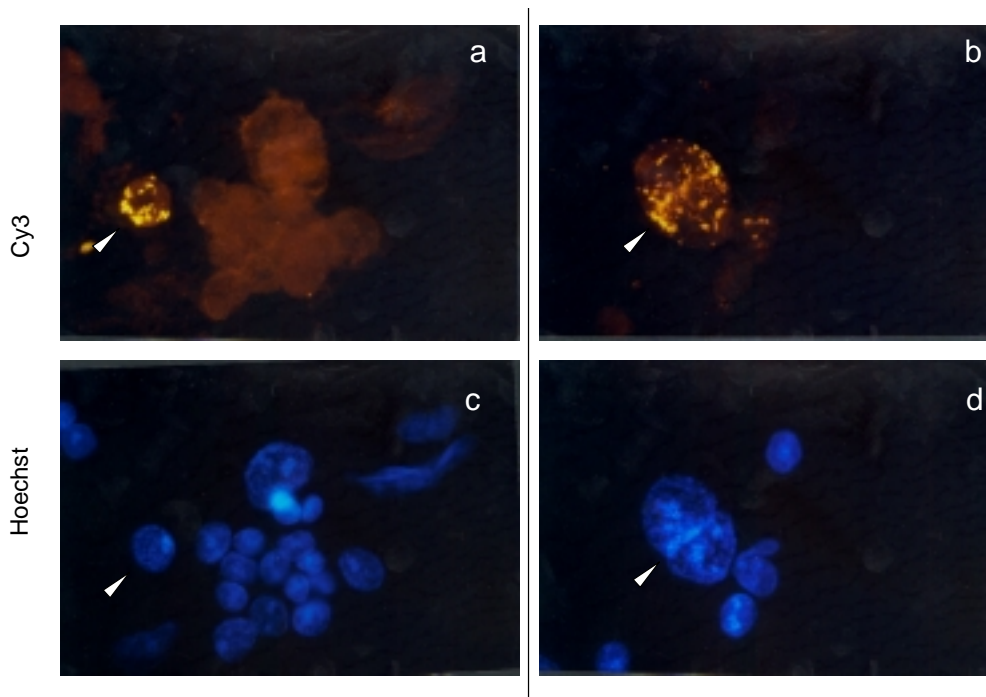
Figure V.2



**Immunocytochimie pour HOP-Flag; lignée FadU**

Trois exemples de localisation avec l'anticorps anti-flag sont montrés (a,b,c) ainsi que le marquage des noyaux correspondants (c,d,f) et la fusion de ces images (g,h,i).

Figure V.3



#### Immunocytochimie pour HOP-Flag; lignée Det562

(Expérience ancienne avec la technique de transfection au BBS)

Deux exemples de localisation avec l'anticorps anti-flag sont montrés (a,b) ainsi que les marquages des noyaux correspondants (c,d).

Figure V.4

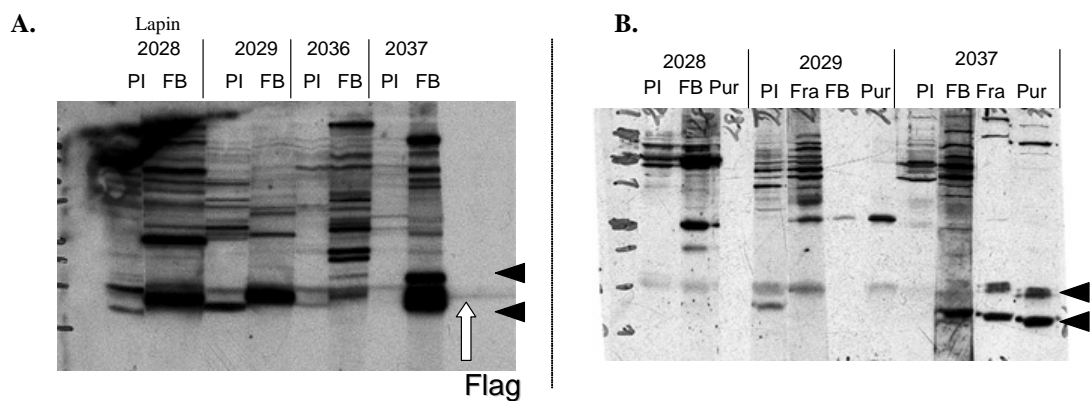
## 5. Obtention d'anticorps spécifiques

Une fois confirmée l'expression différentielle de HOP, par diverses méthodes mettant en jeu ou non des amplifications par PCR, et, compte tenu du rôle potentiel de HOP dans la biologie de l'épithélium normal indiqué par Hybridation *In Situ*, des peptides ont été synthétisés, couplés à l'ovalbumine et injectés à des lapins dans le but d'obtenir des anticorps polyclonaux spécifiques. De tels anticorps, permettent l'analyse du gène endogène, et de la régulation de l'activité de son promoteur dans diverses conditions physiologiques. Chaque peptide a été injecté dans deux lapins indépendants. Le sérum pré immun et des prélèvements réguliers de la 4<sup>ème</sup> à la 8<sup>ème</sup> semaine après injection ont été collectés. A l'issue de la 8<sup>ème</sup> semaine, une réinduction finale par une seconde injection du peptide immunogène a été menée, et, une récolte finale du sérum a été effectuée après 4 semaines supplémentaires (final bleed, FB).

J'ai analysé le profil des bandes détectées pour chaque lapin, pour chaque prélèvement, sur des extraits de cellules COS transfectées par HOP en orientation sens dans PSG5-Flag (FigV.5A). En plus de bandes surnuméraires de grande taille, une bande très forte

de taille proche de 10kD a été détectée, avec notamment le sérum 2037. Cette taille est cohérente avec la taille attendue de 10,1 kDa (8,1kDa + 18 aa supplémentaires). Cette bande est de taille identique à la bande détectée par l'anticorps AntiFlag. L'anticorps AntiFlag ne produit (entre mes mains, et celles d'une collègue) qu'un signal relativement faible en Western Blot, bien que le signal soit intense pour d'autres utilisations comme l'immunocytochimie. Une deuxième bande d'environ 15kD est détectée, et pourrait être un doublet. Il pourrait s'agir d'un variant d'épissage alternatif d'une séquence homologue chez le singe (cellules COS), comme d'un homologue de l'isoforme longue de HOP (figure IV.6) nouvellement intégrée aux bases de données, ou, de protéines possédant une homologie de séquence tels des facteurs contenant des homéodomaines (famille Pax par exemple). La faible taille de HOP n'a permis de synthétiser que peu de peptides différents potentiellement immunogènes. HOP ne contenant presque uniquement qu'un homéodomaine, structure conservée, il existe un risque de réactivité croisée.

Les échantillons ont été poolés avant purification pour chaque lapin, lorsque cela était possible. La réactivité des anticorps purifiés a été testée sur la protéine endogène dans la lignée cellulaire HaCaT, et, un profil identique a été détecté avec le sérum 2037 (FigV.5B) Une grande partie des bandes surnuméraires de haut PM a été éliminée ou fortement atténuée par la purification. Quelques bandes de grande taille restent cependant présentes dans le sérum PI.



**Sérum lapin 2037 purifié:** détection de deux bandes (10 et 15kDa).  
Taille attendue 10.1kDa. Même taille que la bande détectée par Flag

#### Caractérisation et purification de sérums Anti-HOP.

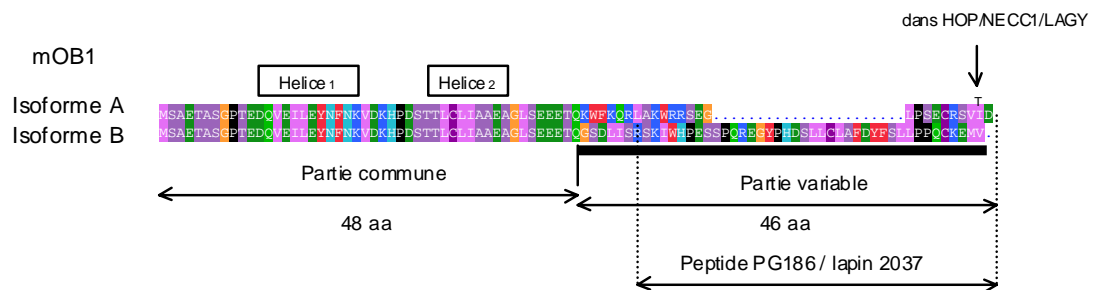
**A.** Cellules COS transfectées pour HOP en orientation sense (clone 31S)

**B.** Protéine endogène dans la lignée HaCaT

Les sérums pré immuns (PI), final bleed (FB), une fraction intermédiaire de purification (Fra) et le sérum purifié (Pur) sont présentés.

Figure V.5

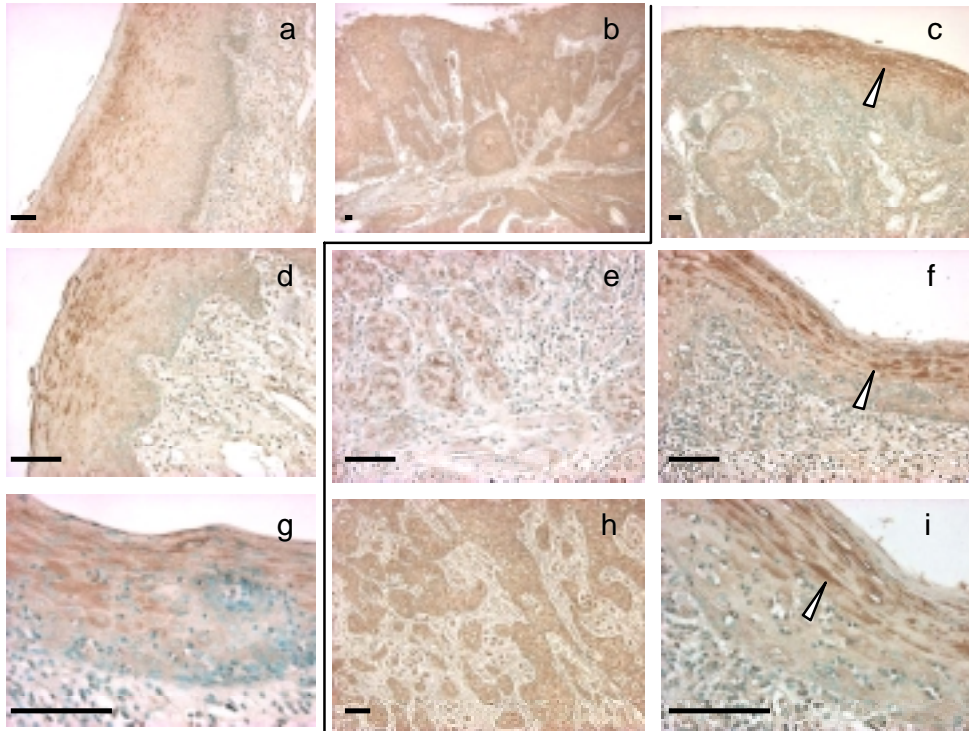
Au cours de l'étude un variant supplémentaire de HOP a été décrit (FigV.6). Compte tenu de l'existence potentielle de deux isoformes de la protéine HOP (mOB1 isoformes A et B), il est important de noter que le sérum 2037 n'est censé détecter que la forme courte, habituellement décrite de HOP, et, ne détecte pas la version longue, à ce jour mal caractérisée. La détection sur la protéine endogène d'un doublet de taille supérieure à la taille attendue semble donc correspondre à des modifications post-traductionnelles de la protéine, qui contient des sites potentiels de glycosylation et de phosphorylation, ou, à la détection de protéines de type Pax ou « goosecoïd » contenant un homéodomaine homologue à HOP (FigV5B). Cependant, le peptide 2037 ne reconnaît qu'une fraction de l'homéodomaine, hors des hélices 1 et 2, dont l'hélice 3/4 non fonctionnelle.



Variants putatifs de HOP  
Figure V.6

## 6. Caractérisation de l'expression par Immunohistochimie

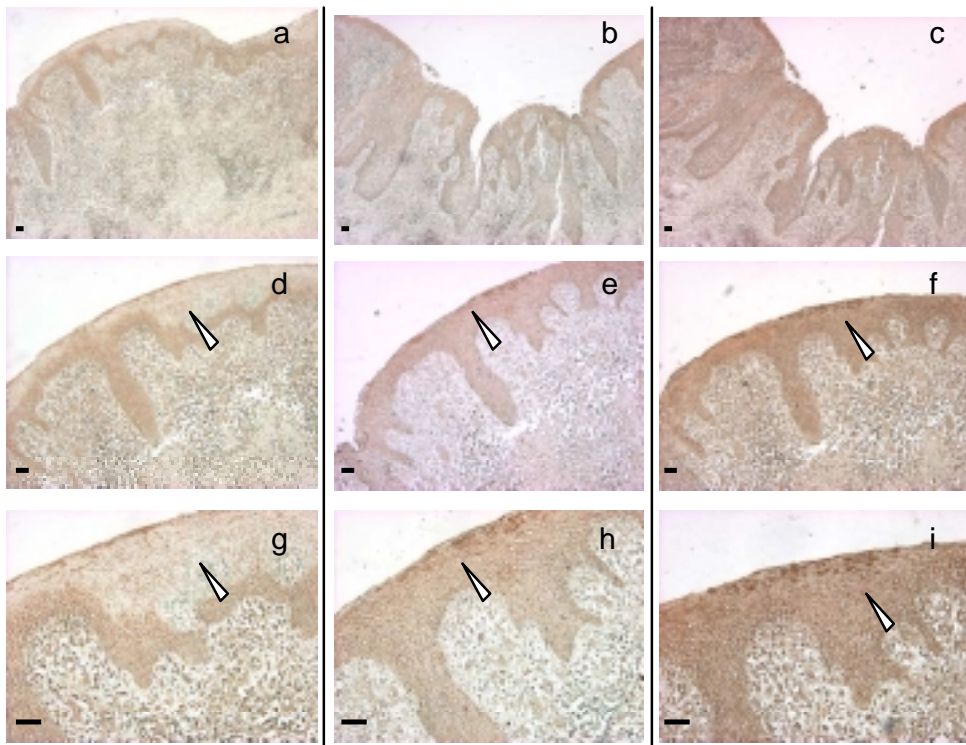
J'ai voulu confirmer au niveau protéique le profil d'expression de l'ARNm de HOP obtenu en hybridation *In Situ*, par l'analyse de tumeurs et de tissus normaux VADS par la technique d'immuno-histochimie. Le sérum pré immun (PI) est utilisé comme contrôle négatif. Il contient quelques bandes de grandes tailles qui peuvent générer un bruit de fond, compliquant l'analyse visuelle des lames. Nous avons utilisé les sérums non purifiés (FB), plus affins, au prix d'un bruit de fond potentiellement élevé, ainsi que le sérum purifié (Pur) issu du lapin 2037. Les lames utilisées sont des lames paraffinées de tumeurs, portant des sections de grande taille dans lesquelles j'ai recherché des zones d'épithélium histologiquement normal. J'ai également examiné l'expression de HOP au sein de la masse tumorale.



**Immunohistochimie pour HOP; tumeur 24656T**

Des exemples de marquage par le sérum pré immun (a, b, d, g), le sérum purifié 2037 (c, e, f, h, i) sont présentés.

Figure V.7



**Immunohistochimie pour HOP; tumeur 26312T**

Des exemples de marquage par le sérum pré immun (a, d, g), le sérum purifié 2037 (b, e, h) et le FB 2037 (c, f, i) sont présentés.

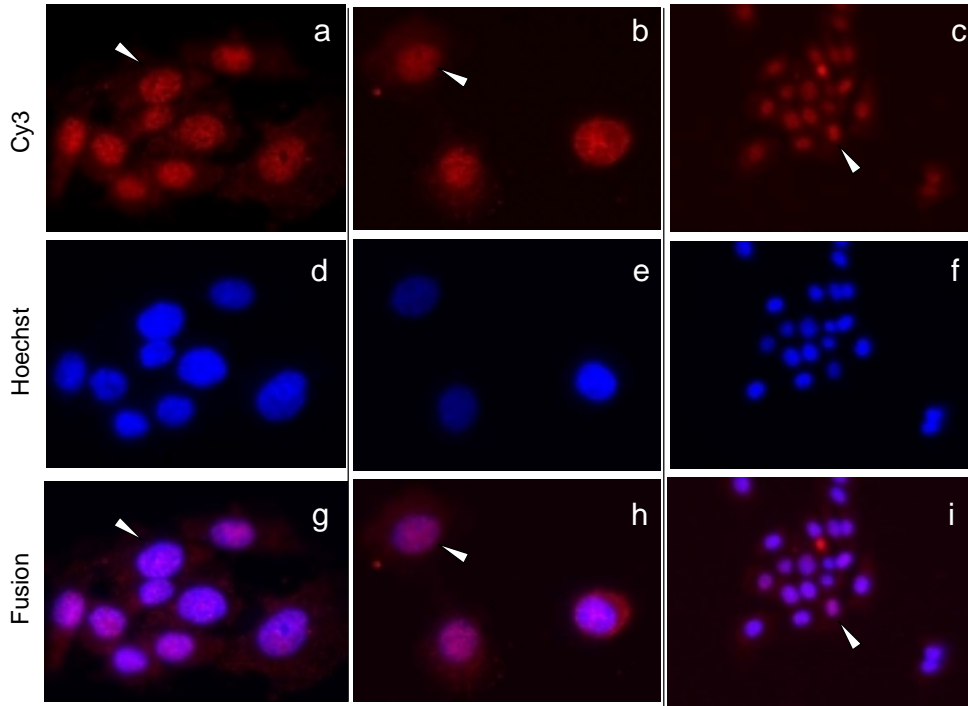
Figure V.8



Le bruit de fond avec le sérum pré immun PI est fort. Ce bruit de fond fort peut être issu des bandes détectées en WB avec le sérum pré immun, être liée à une concentration trop élevée du sérum utilisé, ou, être lié aux conditions de saturation et de blocage de la lame. L'utilisation de sérum de chèvre (Goat serum) plutôt que de d'albumine bovine (BSA) pourrait contribuer à diminuer le bruit de fond. Un marquage moindre dans certains épithéliums de surface des biopsies nous permet de constater une coloration plus marquée pour les sérums purifiés et FB dans une zone qui semble, en accord avec les résultats d'ISH, correspondre aux kératinocytes en cours de différenciation, certainement de la couche granuleuse (FigV.7c,f,i et FigV.8 e,h,f,i par rapport aux images a,d,g de ces figures). Au sein des masses tumorales, un marquage parfois fort est détecté, sans qu'il soit de manière évidente spécifique, puisqu'il est détecté dans un certain nombre de cas avec le PI. Il semble cependant qu'une expression clairsemée, en amas de quelques cellules isolées, similaire à celle observée en ISH, puisse être spécifique.

## **7. Etude de la localisation subcellulaire endogène par immunocytochimie avec l'anticorps spécifique**

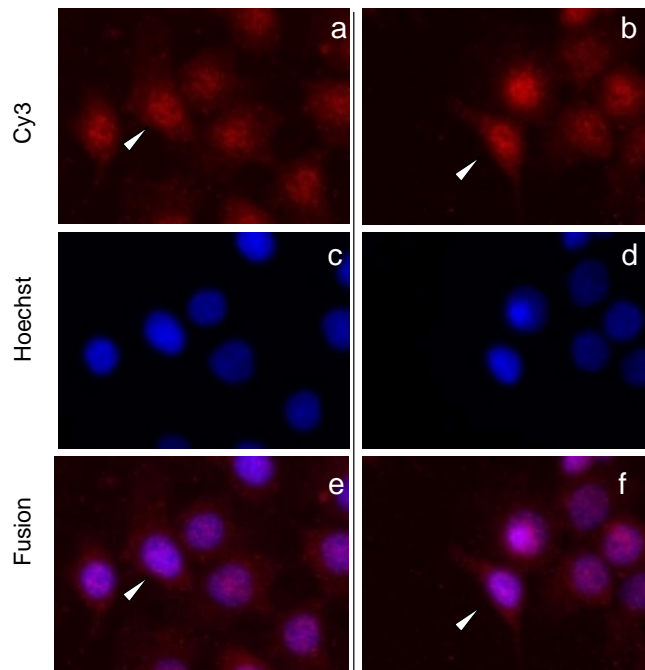
La localisation de HOP avec l'anticorps purifié a été recherchée dans les lignées HaCat, Fadu et Det562. L'anticorps a été utilisé à une dilution 1/1000 et à une dilution 1/200. La détection de HOP est meilleure à la dilution 1/200, la dilution 1/1000 produisant un marquage faible, de l'ensemble de la cellule qui pourrait correspondre à du bruit de fond. A la dilution 1/200, un marquage majoritairement nucléaire est détecté pour les cellules HaCaT (FigV.9). De manière surprenante un marquage est détectable, de façon toutefois plus faible, pour les lignées tumorales VADS Fadu (FigV.10) et Det562 (FigV.11c). Une fraction du marquage demeure cytoplasmique. Ce marquage cytoplasmique est plus élevé dans la lignée tumorale Det562, métastatique (FigV.11a et b). L'anticorps secondaire seul (non présenté) produit un bruit de fond caractéristique et clairement différent, puisque les noyaux sont généralement exclus du marquage du bruit de fond, sans doute pour des problèmes d'accessibilité, tandis que le cytoplasme est marqué faiblement. La localisation nucléaire de HOP dans les cellules HaCaT est cohérente avec les résultats obtenus pour HOP dans le développement cardiaque (Chen et al., 2002; Shin et al., 2002) et avec les prédictions informatiques.



#### Immunocytochimie pour HOP; lignée HaCaT

Trois exemples de localisation avec le sérum 2037 purifié sont montrés à la dilution 1/1000 (a et b), et à la dilution 1/200 (c) ainsi que les marquages des noyaux correspondants (d,e,f) et la fusion de ces images (g,h,i).

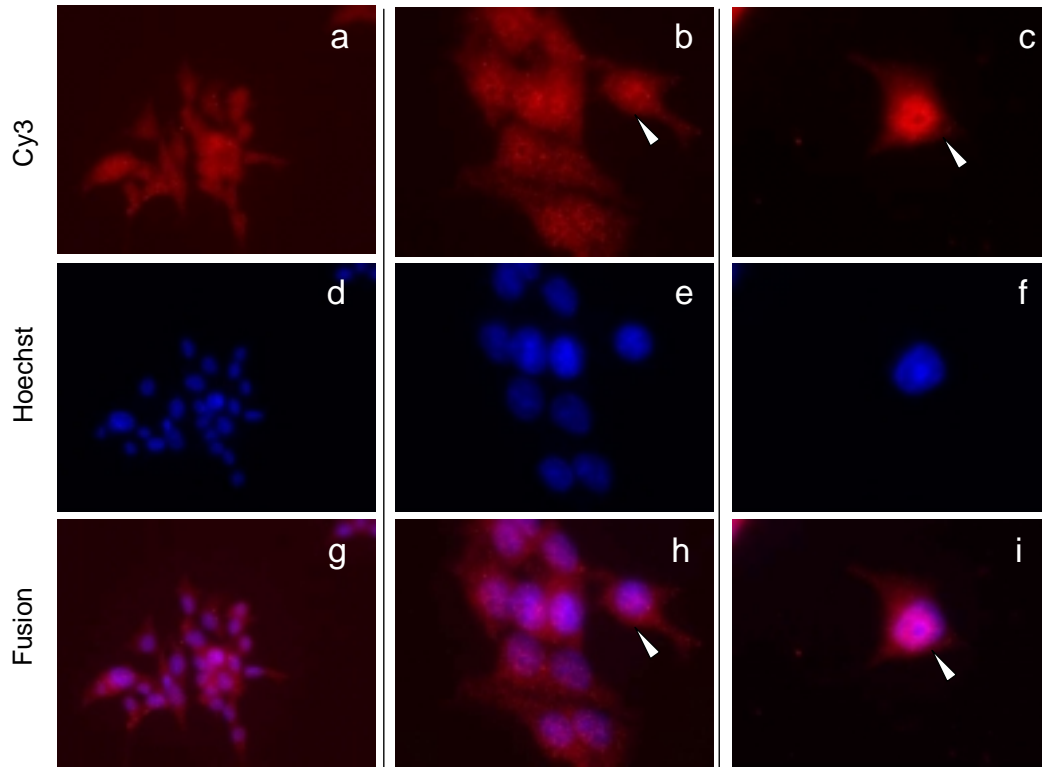
Figure V.9



#### Immunocytochimie pour HOP; lignée Fadu

Deux exemples de localisation avec le sérum 2037 purifié sont montrés à la dilution 1/1000 (a et b) ainsi que les marquages des noyaux correspondants (c,d) et la fusion de ces images (e,f).

Figure V.10



### Immunocytochimie pour HOP; lignée Det562

Trois exemples de localisation avec le sérum 2037 purifié sont montrés à la dilution 1/1000 (a), et à la dilution 1/200 (b et c) ainsi que les marquages des noyaux correspondants (d,e,f) et la fusion de ces images (g,h,i).

Figure V.11

Du fait de la conservation des domaines homéodomaines, il est possible, particulièrement dans les lignées tumorales dans lesquelles HOP n'est que peu ou pas exprimé, que nous détectons également d'autres protéines à homéodomaines (famille Pax dont certains membres sont exprimés dans les tumeurs).

## E. Etude fonctionnelle préliminaire

---

### 1. Etude par FACS

#### a. Considérations générales sur mes études par FACS

De manière à déterminer la fonction potentielle de mes gènes d'intérêt (F1.4, B14.4, 99b78, P5.1-non présenté dans cette thèse-), validés en VN et NB, j'ai entrepris d'étudier l'effet de leur surexpression et de l'expression de leur ARN antisens dans des lignées cellulaires VADS. J'ai utilisé la technique de cytométrie de flux FACS (Fluorescence Activated Cell Sorting) pour trier les cellules transfectées et analyser la répartition dans le cycle cellulaire et la population des cellules en apoptose.

De nombreuses conditions ont été testées pour cette analyse :

- Lignées cellulaires : les lignées cellulaires COS, Fadu, HaCaT, Det562 et RPMI2650 ont été analysées. La lignée RPMI2650, contrairement aux autres lignées VADS analysées, possède une version sauvage de p53, et, est bien caractérisée pour cet usage dans notre laboratoire.
- Choix du marqueur de sélection des cellules transfectées : J'ai testé la co-transfection du marqueur membranaire CD20 qui a entraîné une très faible détection de cellules positives avec un anticorps spécifique couplé au FITC) par rapport au contrôle de transfection GFP (green fluorescent protein). J'ai également testé la GFP-spectrine, protéine de fusion associée à la membrane et préservée lors des étapes de fixation du protocole. La détection a été améliorée mais demeure assez faible (10% des cellules détectées par rapport à la GFP).
- Technique de transfection : J'ai testé la méthode BBS ainsi que la méthode JetPEI, qui fournit un meilleur taux de transfection-évalué par co-transfection de la GFP dans les boîtes contrôles-, avec une meilleure reproductibilité.

Malgré de nombreuses tentatives, ces expériences se sont révélées peu informatives, tantôt du fait de la détection d'une faible proportion des cellules transfectées (amélioration avec la GFP-spectrine), tantôt du fait d'un effet non spécifique d'induction de l'apoptose et de perturbation de la répartition des cellules dans le cycle cellulaire par la seule transfection (effet observé avec les contrôles pSG5, pSG5 puro Flag et pBSK). La transfection des clones contrôles positifs comme p53 et MDM2 n'a pas systématiquement entraîné les effets attendus,

jetant un doute sur les rares effets mesurés avec les clones d'intérêt. Les résultats obtenus (analyse par le programme ModFit LT 3.1) ne sont donc, pour le moment, que des indications, faute d'avoir été répétés et améliorés par usage de vecteurs assurant une plus forte expression, comme pCMV (reclonage effectué dans ce vecteur) ou pCDNA3 par exemple.

## **b. Résultats pour le clone HOP**

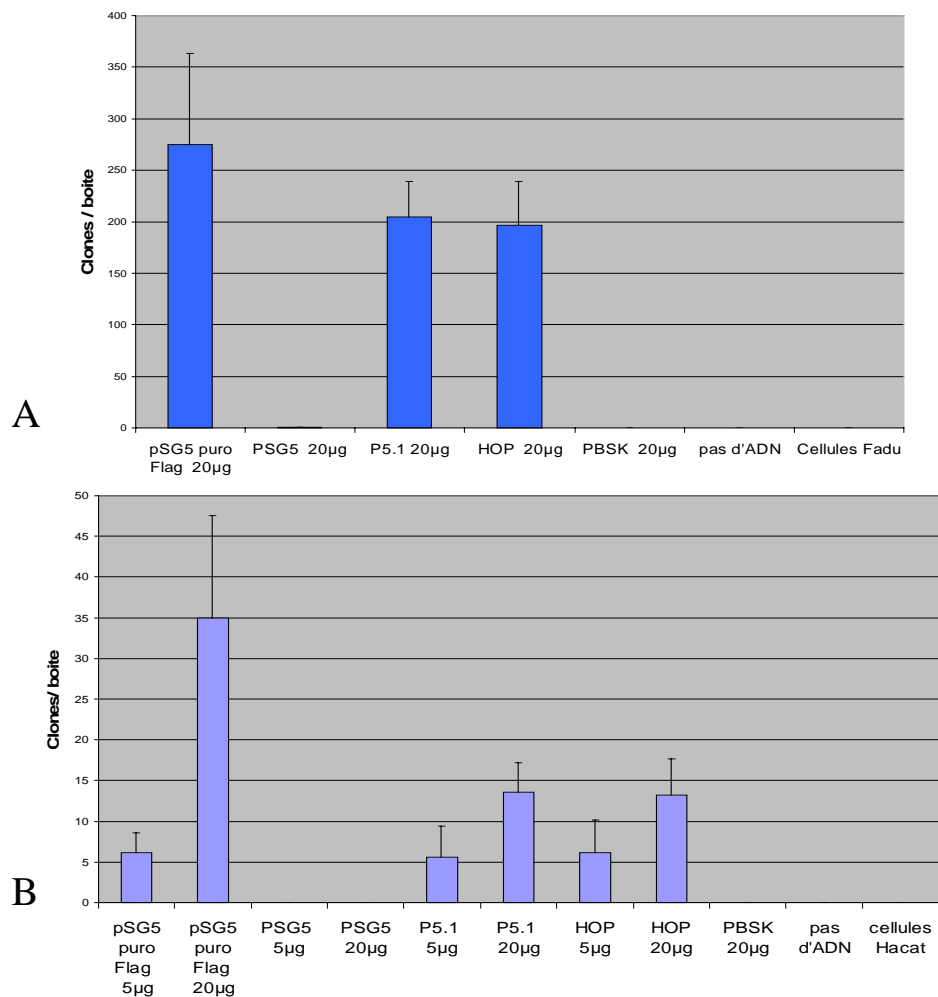
Les expériences par FACS n'ont pu apporter que peu d'indications supplémentaires. La transfection du clone sens SMAP8S semble entraîner une baisse de la prolifération tandis que l'expression du clone antisens SMAP6AS pourrait provoquer une augmentation modérée de la prolifération. Si elles étaient confirmées, ces données seraient compatibles avec les fonctions associées à HOP comme gène suppresseur de tumeurs dans la différenciation et dans la limitation de la prolifération. En effet, l'allongement du temps de doublement des clones stables sur exprimant NECC1/HOP dans les lignées cellulaires dérivées de choriocarcinomes (Asanoma et al., 2003) semble appuyer les tendances observées. Toutefois ces résultats doivent être confirmés. L'utilisation du vecteur pCMV pourrait permettre d'augmenter les effets observés. La technique d'ARN interférence (RNAi) à l'aide du vecteur pSUPER pourrait permettre de mieux observer les effets de l'abolition de l'expression du gène HOP.

## **2. Clones stables et test de clonogénicité**

Une première série de tests de clonogénicité a été effectuée dans les cellules HaCaT (FigV.12A) et Fadu (FigV.12B). Les clones correspondant aux premières séquences codantes clonées dans pSG5 puro flag ont été, outre d'autres anciens clones favoris abandonnés, les clones F1.4 (HOP) et P5.1. Ce dernier clone, initialement inclus à cette thèse, correspond à un gène surexprimé dans les tumeurs, validé en VN et en NB -avec un profil moins robuste que mes autres favoris- possédant une faible homologie avec les cullines, protéines interagissant avec le gène suppresseur de tumeurs Von Hippel Lindau VHL. Cette analyse réalisée après transfection au BBS a généré un nombre important de clones par boîte de culture lorsque le marqueur de résistance à la puromycine était introduit (pSG5 puro flag, et gènes clonés dans ce vecteur d'expression). Toutefois, dans les cellules Fadu, les différences de nombre de clones obtenus sur les boîtes correspondant au contrôle pSG5 puro flag, et, les clones HOP et P5.1 ne sont pas significatives après comptage sur une série de 5 boîtes (FigV.12A). Dans la lignée HaCaT, le nombre de clones obtenus est significativement différent entre le contrôle,

et, les clones HOP et P5.1. Cependant, il semble que le comportement des 2 plasmides HOP et P5.1 suive la même évolution. Le nombre de clones obtenus augmente avec la quantité de plasmide transfectée et donc, la quantité du marqueur de résistance. Le fait que les plasmides codant pour HOP et P51 se comportent de manière similaire, bien que ces deux gènes aient des fonctions *a priori* opposées, laisse supposer un éventuel effet non spécifique.

Les cellules HaCaT ne sont pas strictement des cellules normales puisqu'elles sont immortalisées et possèdent une version mutée de p53. Si l'effet observé s'avérait spécifique, l'expression de la protéine HOP entraînerait une moindre croissance que le vecteur vide, en accord avec un rôle modérateur de la prolifération d'un gène suppresseur de tumeurs potentiel.



## Tests de clonogénicité pour HOP

Clones sens et contrôles dans les lignées Fadu (A) et HaCaT (B).

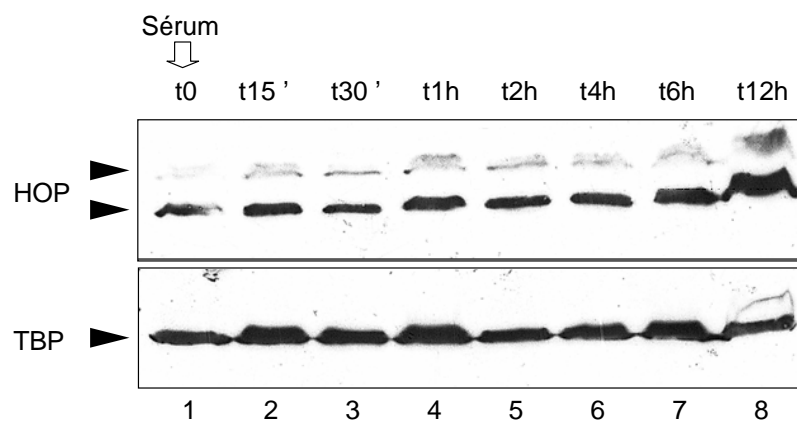
Figure V.12

J'ai tenté à deux reprises de répéter ces tests de clonogénicité dans les cellules HaCaT par transfection du clone sens 8S et du clone antisens 6AS avec la méthode de transfection JetPEI, permettant l'obtention d'un nombre plus élevé de cellules transfectées (contrôle GFP). Malheureusement ces expériences se sont révélées infructueuses.

### 3. Etude de l'induction par le sérum

L'obtention d'anticorps contre HOP nous permet d'analyser la régulation de la protéine endogène. Après avoir privé des cellules HaCaT de sérum pendant 24h, et, avoir ainsi bloqué leur prolifération, l'expression de HOP a été analysée après réinduction par le sérum (FigV.13). L'expression de HOP est faiblement induite (t15 minutes pour le doublet supérieur, t12h pour la bande inférieure) après induction par le sérum dans les cellules HaCaT, suggérant un rôle potentiel dans le contrôle de la prolifération. Cette information fonctionnelle est très préliminaire car il est reconnu que plus de 1000 gènes sont induits par le sérum.

La fonction de HOP est dépendante d'une interaction avec le SRF (Serum Response Factor) dans le développement embryonnaire cardiaque, puisqu'elle forme une boucle de régulation négative des gènes induits par Nkx2.5 sur les gènes régulés par le SRF (Chen et al., 2002; Shin et al., 2002 ). Cette induction de HOP par le sérum pourrait aussi être liée à un contrôle négatif des gènes induits par le SRF promouvant la prolifération.



Induction de HOP par le sérum dans les cellules HaCaT  
Cinétique sur 12h avec le sérum 2037.

Figure V.13

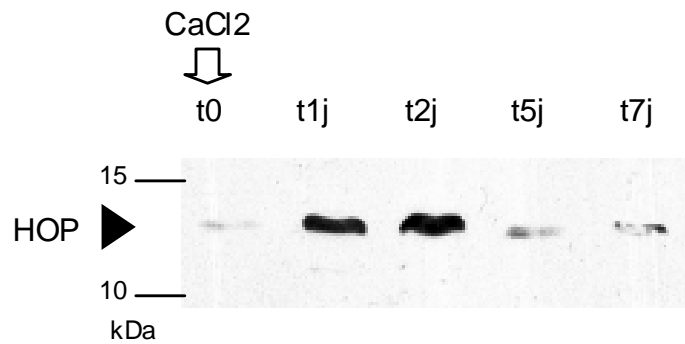
#### 4. Etude de l'implication dans la différenciation des kératinocytes

Du fait de l'expression de HOP dans les kératinocytes en cours de différenciation des couches supra basales de l'épithélium normal, et, des données publiées sur les protéines identiques à HOP, nous avons décidé d'étudier le rôle de cette protéine dans la différenciation des cellules HaCaT (Fusenig & Boukamp, 1998). Les cellules HaCaT sont des kératinocytes spontanément immortalisés capables de se différencier lors de l'élévation du niveau de calcium ( $\text{CaCl}_2$ ) dans le milieu de culture (Breitkreutz et al., 1993). *In vivo* un gradient de calcium est présent dans l'épiderme, s'exerçant de façon croissante de la couche basale aux couches superficielles (Menon et al., 1985). *In Vitro*, le calcium induit la différenciation des kératinocytes (Hennings et al., 1980 ; Yuspa et al., 1989). Les cellules HaCaT peuvent reconstituer une structure d'épithélium stratifié en co-culture avec des fibroblastes du derme (Schoop et al., 1999). Bien que non tumorigénique, la lignée HaCaT peut subir une conversion maligne par transfection de H-Ras (Breitkreutz et al., 1991). La différenciation peut être suivie en mesurant l'expression des marqueurs de la différenciation des kératinocytes que sont les kératines 1/10, correspondant à des cellules en cours de différenciation, et, la kératine 14, correspondant à des cellules peu différenciées (Breitkreutz et al., 1991, Schoop et al., 1999). De manière préliminaire, nous avons donc étudié l'effet de l'augmentation du  $\text{CaCl}_2$  sur l'expression de la protéine HOP (Fig IV.14). Nous avons repris les conditions et les temps d'induction utilisés dans diverses études visant à mettre en évidence par micro arrays, puis à confirmer, par RT-PCR et Western blot, l'induction effective de gènes exprimés lors de la différenciation des kératinocytes (Abiko et al., 2003; Seo et al., 2002 ). Il semble que l'expression de HOP soit nettement induite dès les 2 premiers jours après induction de la différenciation, puis, diminue dramatiquement entre le 2<sup>ème</sup> et le 5<sup>ème</sup> jour. Une seule bande peut être détectée, le doublet d'environ 15kDa disparaissant dans ces conditions de culture, sans explication à ce jour. Cette expression transitoire, en pic, d'un facteur impliqué dans la différenciation, induisant l'expression d'effecteurs secondaires, est similaire à l'expression dans la différenciation des myofibroblastes d'un autre de mes anciens gènes favoris validé en NB, CSRP2 (Weiskirchen et al., 2001). Il convient d'évaluer le rôle de la densité cellulaire puisque l'expression de récepteur KGFR (keratinocyte growth factor receptor), capable de moduler l'équilibre prolifération/différenciation, est sensible à la densité cellulaire dans les cellules HaCaT (Capone et al., 2000).

Cette expérience reste à compléter en révélant le contrôle TBP (la coloration au rouge Ponceau était homogène après transfert, la bande TBP devrait être équilibrée). J'ai dû arrêter



cette expérimentation pour entreprendre l'écriture de ma thèse. Nous pensons présenter la caractérisation de l'expression de HOP dans les cancers VADS, en tant que short report, pour valoriser rapidement le travail déjà effectué, et, signaler un rôle peut être plus général de HOP comme gène suppresseur de tumeur, mis en évidence jusqu'alors dans les seuls choriocarcinomes et carcinomes du poumon



Expression de HOP lors de la différenciation induite par le  $\text{CaCl}_2$  des cellules HaCaT  
Cinétique avec le sérum 2037.

Figure V.14

## E. Matériel biologique

Clone F1.4 Résumé	Séquence	Localisation
Gène SMAP31/HOP		
fragment DD dans pGEM-T EASY (Clontech)	CTAAATGTACATAA TTAGAAAAGAAAA TAACAA TAGGAAGCTA TGTG TATCTTCTGTGTAAGCAGTGGCTTCACTGGAAAAATGGTGTGGCTA GCATTTCCCTTTGAGTCATGATGACAGATGGTGTGAAAACCATCTAA GTTTGCTTTTGACCATCACCTCCCAGTAGCAATTTGCTTTCATAATCC ATTTAGCAATCCAGGCCTCTGTGAAAAGATAATATGAGGGAGAAG GGAACACATTTCTCTGAA	572bp. BC014225 (HOP transcript variant 1) 675-1246 (572/572) //AF454763 Lagy complete cds 426-997
Protéine correspondante	MSAETASGPTEDQVEILEYNFNKVDKHPDSTTLCLIAEAGLSE EETQKWFQRLAKWRRSEGLPSECRSVTD	HOP , 73 aa, 219bp
Fragment séquence codante (CDS) cloné	GGATCCCCTATGCTCATTTTCTGGGCTGTACAGAGAAGACTGGA AGAGCGCGCAGGGACCATGTCGGCGGAGACCGCGAGCGGCCCA CAGAGGACCAGGTGGAATCCTGGAGTACAACCTCAACAAGGTG ACAAGCACCCGGATTCCACCACGCTGTGCCTCATCGCGCCGAGG CAGGCCCTTCCGAGGAGGAGCCAGAAATGGTTTAAGCAGCGCC TGGCAAAGTGGCGGCGCTCAGAAGCGCTG	91 aa, 276bp ATG alternatif
Orientation Sens SMAP 8S	Sens	
Orientation Antisens SMAP 6AS	Antisens	
<b>Protéine correspondante</b>		
	<u>M</u> LIFLGCYRRRLEERAGT <u>M</u> SAETASGPTEDQV EILEYNFNKVDKHPDSTTLCLIAEAGLSEET QKWFQRLAKWRRSEGLPSECRSVTD	
<b>Peptides injectés</b>		
PG186	AKWRRSEGLPSECRSVTDC	Lapins 2029 & 2037, aa 56 à 73
PG185	CLEYNFNKVDKHPDSTT	Lapins 2028 & 2036, aa 17 à 32

Matériel biologique obtenu pour l'étude de HOP

Figure V.15

**Partie VI**  
**Caractérisation du candidat B14.4 :**  
**hypothetical protein FLJ10261**

## A. Caractérisation du candidat B14.4 : Hypothetical protein FLJ10261

---

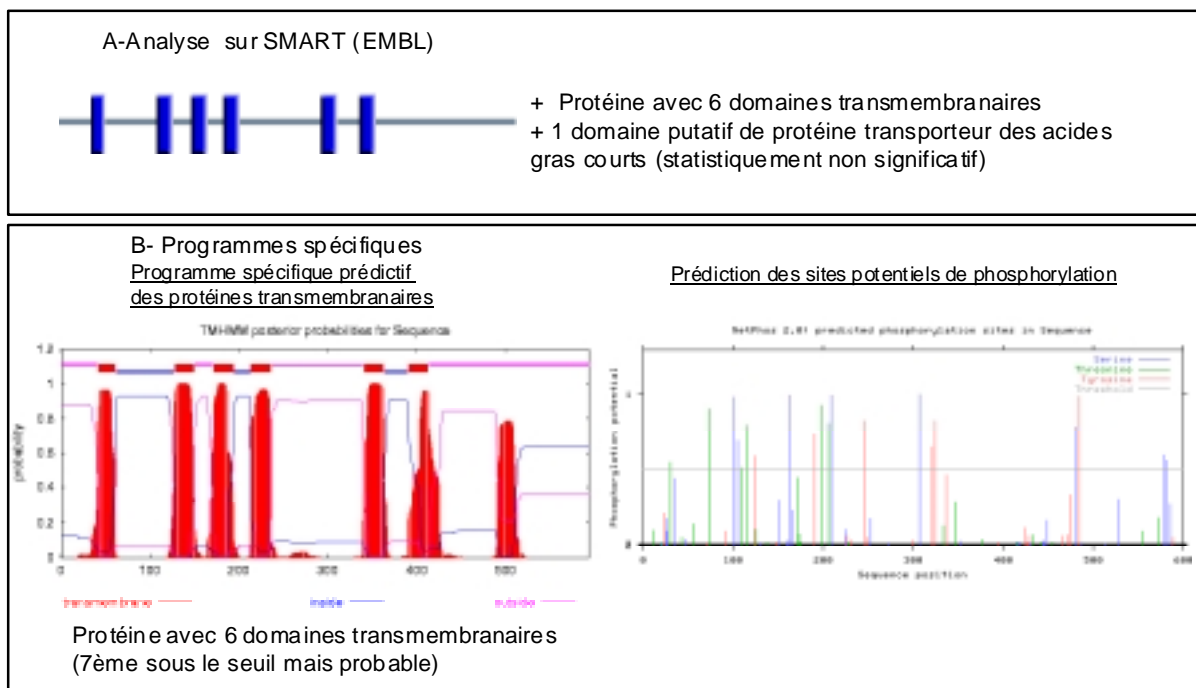
### 1. Profil DD

Le clone B14.4 contient une séquence surexprimée dans les tumeurs en DD. L'analyse en Reverse Northern blot a révélé visuellement sur certaines séries d'hybridation de type T/N DD, une forte surexpression dans le « pool » tumoral. La séquence de ce clone a donc été déterminée.

### 2. Identification

La séquence particulièrement courte, de 80bp, a été identifiée par BLAST comme le gène inconnu FLJ10261, isolé en tant qu'ADNc complet de 3'052bp à partir d'une banque d'embryon complet, localisé sur le chromosome 11. Aucune fonction n'était attachée à ce gène. Une recherche informatique de la séquence codante a indiqué l'existence d'une séquence codante de 1'800bp correspondant à une protéine contenant 6 domaines transmembranaires indiqué lors de l'analyse (FigVI 1A) par BlastP et par SMART (<http://smart.embl-heidelberg.de/>). Une analyse complémentaire avec le programme TMHMM (<http://www.cbs.dtu.dk/services/TMHMM/>), dédié à la prédiction des domaines transmembranaires, a révélé l'existence d'un 7<sup>ème</sup> domaine transmembranaire, juste sous le seuil de détection du programme (FigVI.1B). Cette protéine est donc, potentiellement, un récepteur impliqué dans la transduction des signaux extracellulaires. Ce type de protéines sont des cibles privilégiées du « drug design » du fait de leur accessibilité et de la synthèse possible de ligands synthétiques ou d'anticorps thérapeutiques, comme le démontrent les exemples de l'anticorps modifié Herceptin (Trastuzumab) ciblant EGFR, du Rituxan ciblant l'antigène CD20 exprimé par les lymphomes des cellules B ainsi que du Gleevec, un inhibiteur de la tyrosine kinases Abl participant à la carcinogenèse des leucémies chroniques myéloïdes (CML) et des tumeurs gastrointestinales (Kim, 2003 ; Shawver et al., 2002). Récemment une version plus longue de l'ARNm de ce gène a été décrite (Kato, 2003). Ce variant code pour une protéine contenant un huitième domaine transmembranaire, impliquant

un rôle de transporteur transmembranaire. Cependant, comme pour le variant 7 transmembranaire, une analyse par PipeAlign (<http://igbmc.u-strasbg.fr/PipeAlign/>) des protéines apparentées n'indique une homologie qu'avec trois groupes de protéines. Ces protéines sont des protéines hypothétiques conservées chez l'homme, la souris et la drosophile, pour lesquelles aucune information fonctionnelle n'est disponible. Aucune protéine de fonction connue ne présente d'homologie avec cette protéine, nous fournissant de ce fait peu d'informations sur son ligand ou ses substrats.



Analyse bio-informatique initiale de FLJ10261

Figure VI.1

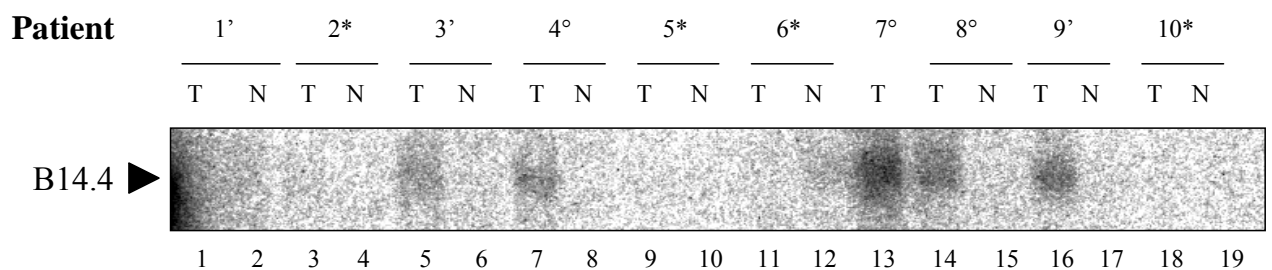
En outre, des sites de phosphorylation potentiels ont été détectés correspondant sensiblement aux extrémités des domaines transmembranaires. Ces sites pourraient moduler l'activité de ce récepteur/transporteur membranaire.

### 3. Validation en Virtual Northern blot

Une étude en Virtual Northern blot (FigVI.2) a été tentée malgré la faible taille de l'insert DD (80bp). De façon surprenante, un signal faible mais clairement détectable a été obtenu. Une bande diffuse a été détectée dans les échantillons tumoraux de tous les patients

de type U (patients 4, 7, 8). Deux patients de type E (patients 3, 9), pour lesquels l'évolution tumorale n'est pas connue, ont également présenté une surexpression dans les tissus tumoraux. Ce gène ne s'exprime donc pas chez les patients présentant une absence bien caractérisée de potentiel métastatique. Le caractère possiblement prédictif de la progression tumorale de ce gène nous a conduit à cloner sa séquence codante puis à effectuer une validation complémentaire en Northern blot. La séquence codante de 1800bp (variant 7 transmembranaire, seul connu à l'époque) a été clonée dans le cadre de lecture dans le vecteur d'expression pSG5-puro-Flag, de manière à pouvoir exprimer et détecter cette protéine étiquetée par Flag.

Le Virtual Northern blot a été répété avec la séquence codante complète et n'a pas présenté le profil attendu. Une bande de taille inférieure au transcrit attendu a été détectée comme sous exprimée dans les tumeurs. Un « smear » a été détecté dans les tumeurs suggérant une synthèse incomplète de l'ADNc. Le Northern blot effectué simultanément avec la même sonde a confirmé la surexpression de B14.4 dans les tumeurs, mettant en évidence les limites éventuelles du Virtual Northern blot, lorsque la taille du transcrit est supérieure à 2,5-3kB.

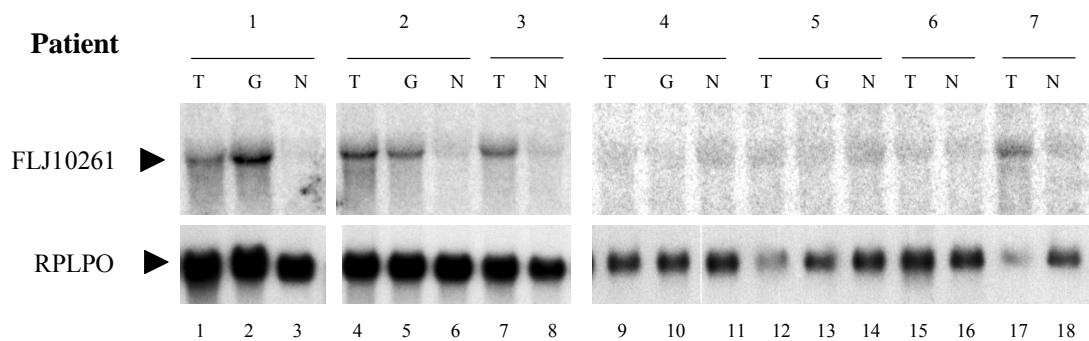


### Virtual Northern Blot pour B14.4

Une bande diffuse (d'environ 1200bp est détectée dans les tumeurs de type Early (°) et Unstable (°). L'expression n'est pas détectée dans les tumeurs de type Stable (\*)

Figure VI.2

## 4. Validation en Northern Blot



### Northern Blot pour FLJ10261

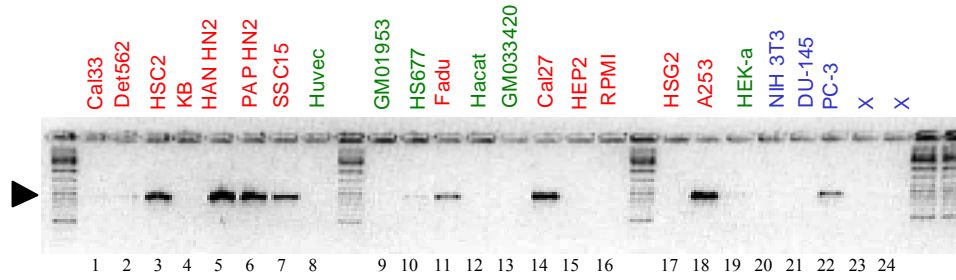
Les tumeurs (T), les ganglions (G) et les tissus normaux (N) des mêmes patients ont été analysés. Une bande d'environ 4kb est détectée comme surexprimée dans les tumeurs. Le contrôle RPLPO est présent.

Figure VI.3

L'analyse en Northern blot avec la séquence codante de 1'800bp (FigVI.3) a révélé l'expression différentielle de FLJ10261 dans 5/7 patients (1, 2, 3, 7 et 5 marginalement). Une bande unique a été détectée, d'une taille estimée entre 4 et 4.5 kB, significativement supérieure à la taille attendue de 3kB dans les bases de données, même compte tenu du manque de précision de l'estimation de la taille sur Northern blot. Ce résultat s'explique sans doute par l'existence récemment rapportée de variants de taille plus importante de la protéine FLJ10261 que celui répertorié dans les bases de données lors de notre caractérisation (Kato, 2003). La version 8 transmembranaire de ce gène pourrait donc être la forme effectivement altérée dans les tumeurs, et, posséder une fonction dans la carcinogénèse. Une nouvelle entrée dans les bases de données, récente (été 2003) confirme l'existence de ce transcrit plus long d'environ 1000bp, comportant un domaine transmembranaire supplémentaire (Voir : Avancée des connaissances plus bas). Une des deux membranes (patients 4 à 7), plus usagée ou comportant moins de patients altérés pour ce gène, n'a permis d'hybridation que pour 2 des cas analysables.

## 5. Expression dans les lignées cellulaires VADS

Une analyse de l'expression de FLJ10261 a été menée par RT-PCR semi-quantitative dans 22 lignées cellulaires (FigVI.4), principalement d'origine VADS, de manière à déterminer les systèmes cellulaires adéquats pour les études fonctionnelles.



Analyse de l'expression de FLJ10261 dans des lignées cellulaires par PCR semi-quantitative

Une bande unique de taille attendue est détectée dans des lignées cellulaires

Figure VI.4

Plus particulièrement, FLJ10261 est exprimé dans les cellules Fadu (colonne11), faiblement exprimé dans les cellules Det562 (colonne2), d'origine hypopharyngée, et, n'est pas détectable dans les cellules « normales », kératinocytes de la peau spontanément immortalisés HaCaT (colonne12). Ces trois lignées sont nos modèles d'études, couvrant l'expression de tous mes clones d'intérêt.

La lignée RPMI2650, dérivée d'une tumeur de la cavité buccale, et, utilisée pour des expériences d'analyse du cycle cellulaire par FACS, n'exprime pas ou peu FLJ10261 (colonne16).

## 6. Caractérisation de l'expression par hybridation *In Situ*

La sonde a été synthétisée après clonage de la séquence codante de 1'800bp. Suivant les tumeurs analysées, 4 types de profils d'expression ont été mis en lumière :

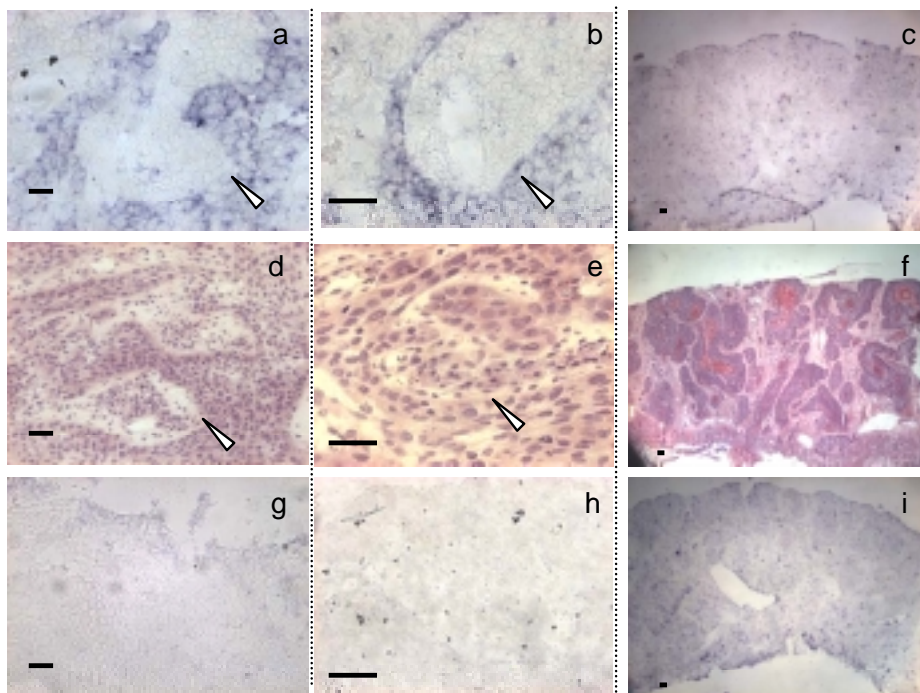
-Aucune expression détectable. La tumeur ne présente apparemment pas d'altération pour ce gène (FigVI.5c).

- Une détection dans des sous régions de nature inconnue au sein des masses tumorales, avec exclusion de marquage « en champignon » pour d'autres cellules cancéreuses adjacentes, pour la tumeur 840T (FigVI.5 a et b).

- Une détection dans les cellules de la couche basale des épithéliums hyperplasiques de structure encore bien défini pour la tumeur 771T (FigVI.7 a,b,c). Cette expression par des cellules, à potentiel prolifératif illimité, ayant des capacités prolifératives importantes, peu différenciées, est cohérente avec un rôle d'oncogène potentiel.

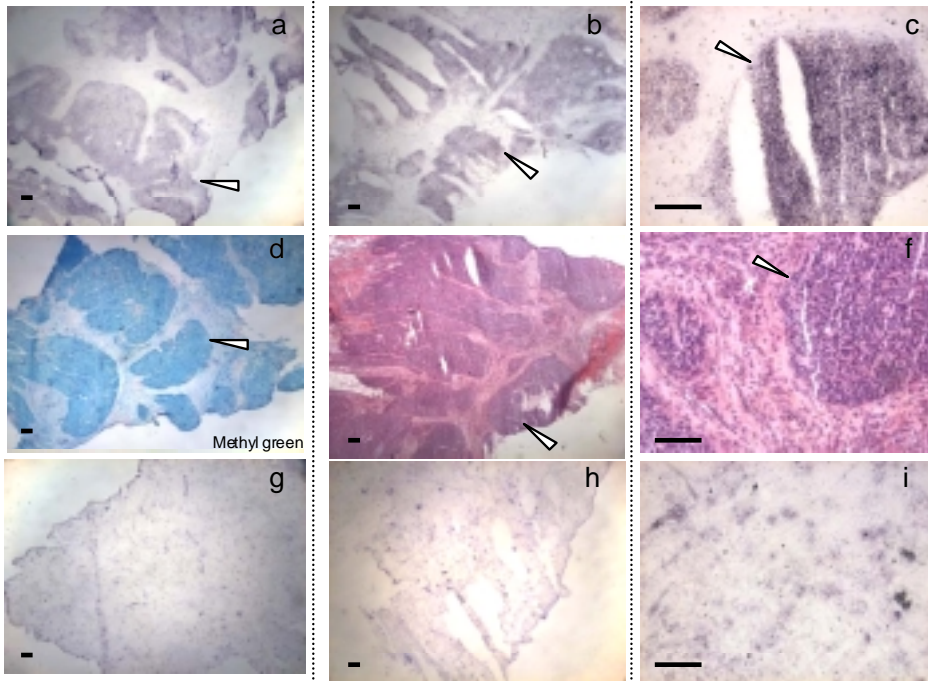


- Une détection dans toutes les cellules des masses tumorales pour les tumeurs 245T et 413T (FigVI.6 a,b,c). Les tumeurs correspondantes doivent avoir acquis une altération pour ce gène à un stade précoce de la carcinogenèse, lors de l'initiation tumorale. Cependant la séquence d'activation de ce gène variable sur le petit nombre de tumeurs analysées ne semble pas étayer un rôle dans l'initiation tumorale en général.



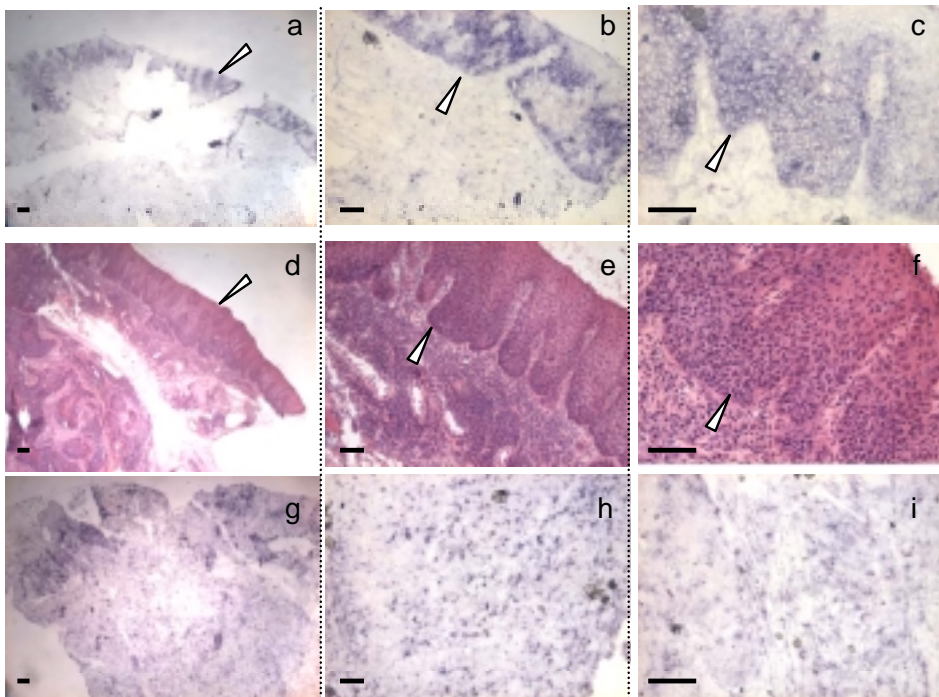
Hybridation In Situ pour FLJ10261; tumeur 840T et 347T  
 3 exemples de marquage pour la sonde antisens SpeI/T7 (a, b, c) sont présentés ainsi que les structures histologiques correspondantes (d, e, f) et le marquage pour la sonde sens Aat2/SP6 correspondant (g, h, i). Les sections (a, b, d, e, g, h) correspondent à la tumeur 840T.

Figure VI.5



Hybridation In Situ pour FLJ10261; tumeur 413T  
 3 exemples de marquage pour la sonde antisens SpeI/T7 (a, b, c) sont présentés ainsi que les structures histologiques correspondantes (d, e, f) et le marquage pour la sonde sens Aat2/SP6 correspondant (g, h, i).

Figure VI.6

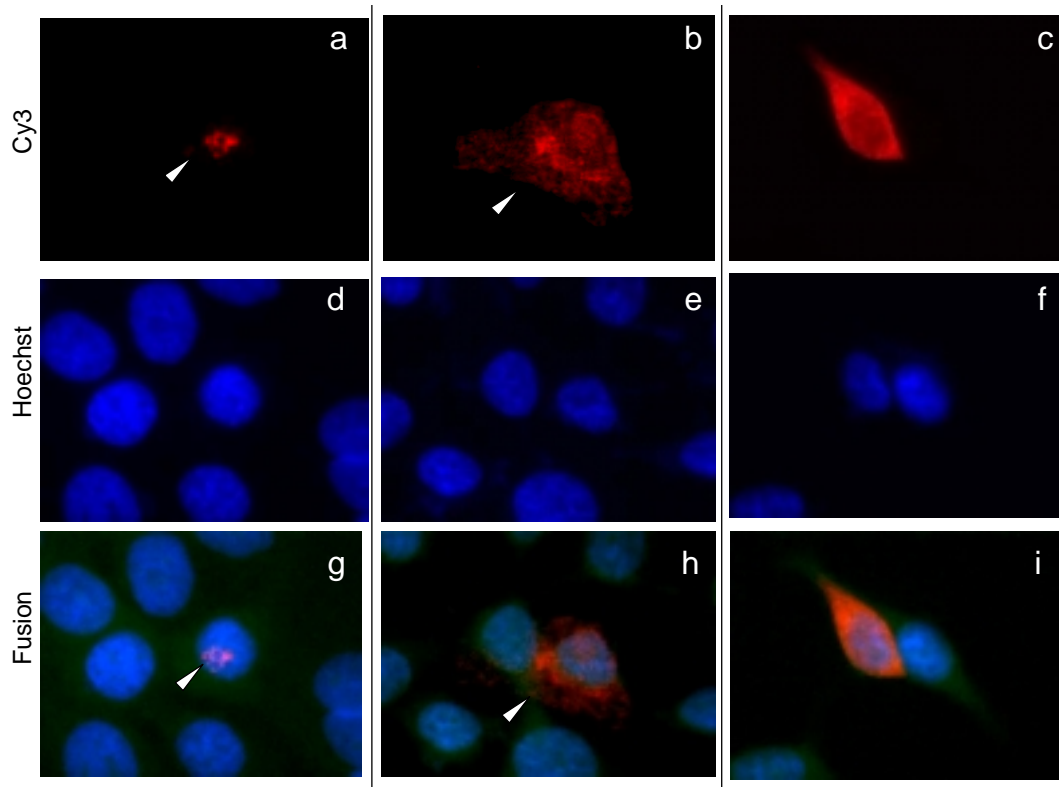


Hybridation In Situ pour FLJ10261; tumeur 771T  
 3 exemples de marquage pour la sonde antisens SpeI/T7 (a, b, c) sont présentés ainsi que les structures histologiques correspondantes (d, e, f) et le marquage pour la sonde sens Aat2/SP6 correspondant (g, h, i).

Figure VI.7

Lorsqu'un signal est détectable, FLJ10261 est donc exprimé par les cellules peu différenciées à potentiel prolifératif élevé, comme les cellules de la couche basale de l'épithélium et les cellules cancéreuses.

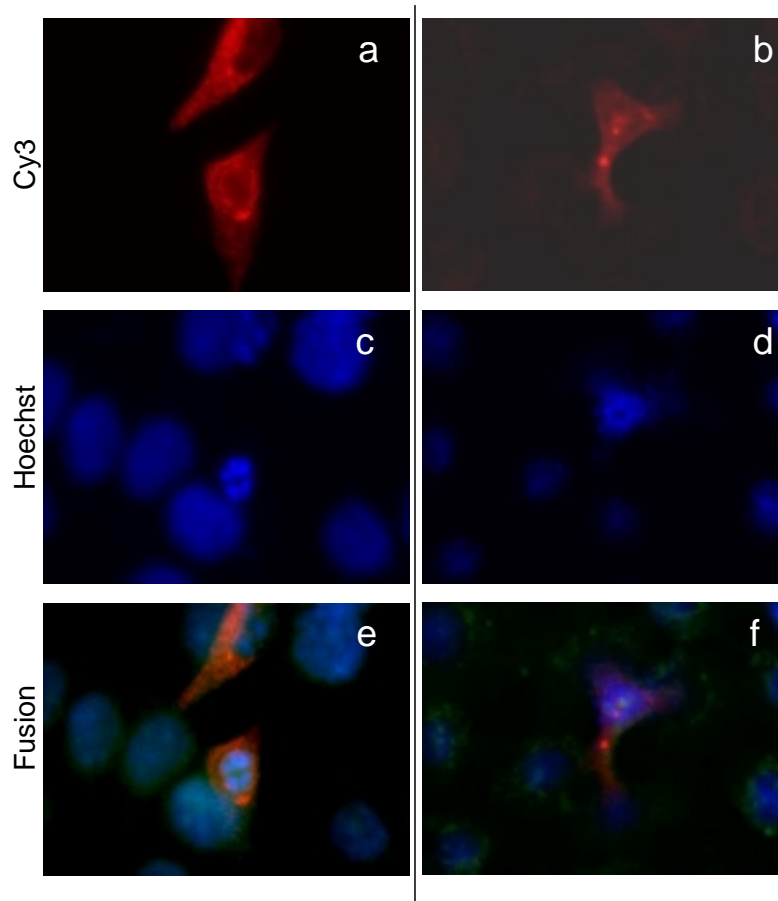
### 7. Etude de la localisation subcellulaire de la protéine transfectée par immunocytochimie avec l'anticorps Anti-Flag



#### Immunocytochimie pour FLJ10261-Flag; lignée HaCaT

Trois exemples de localisation avec l'anticorps anti-flag sont montrés (a,b,c) ainsi que les marquages des noyaux correspondants (c,d,f) et la fusion de ces images (g,h,i).

Figure VI.8



### Immunocytochimie pour FLJ10261-Flag; lignée Fadu

Deux exemples de localisation avec l'anticorps anti-flag sont montrés (a,b) ainsi que les marquage des noyaux correspondants (c,d) et la fusion de ces images (e,f).

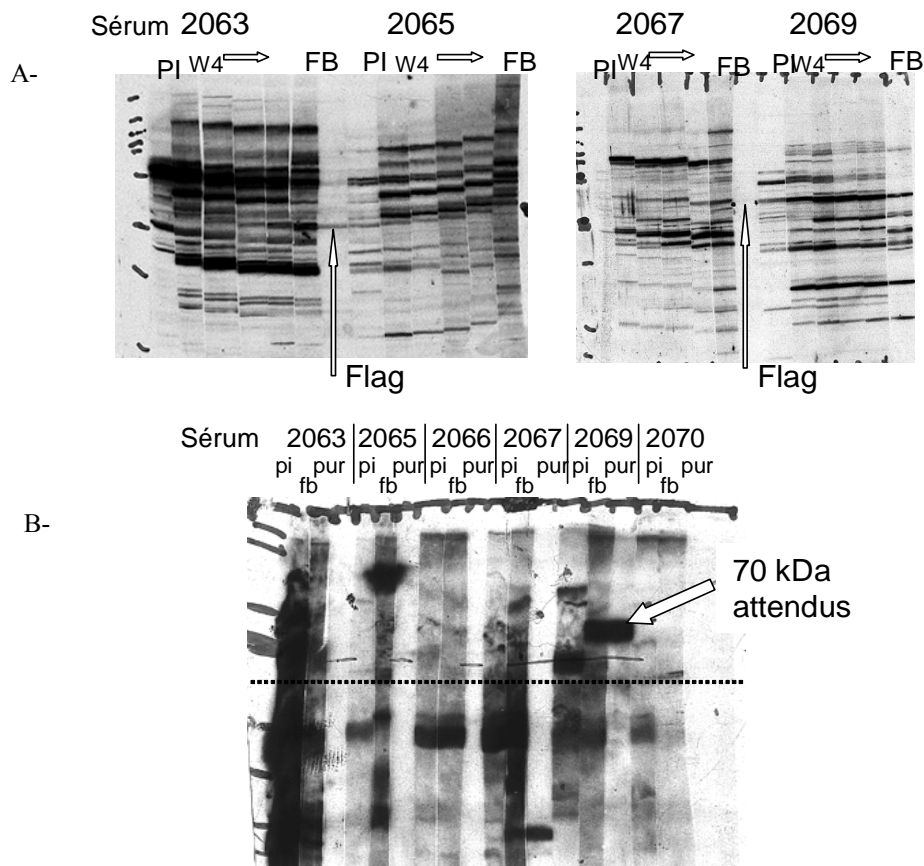
Figure VI.9

De manière surprenante, dans quelques cellules HaCaT transfectées, un signal est parfois détectable au sein du noyau (FigVI.8a). Cependant, les cellules transfectées présentent en général un marquage de toute la cellule qui pourrait correspondre à un marquage membranaire (FigVI.8 b et c). Dans les cellules Fadu (FigVI.9 a et b), il semble que les prolongements membranaires sont encore d'avantage marqués, indiquant une localisation membranaire potentielle. On ne peut exclure que le marquage soit cytoplasmique en l'absence d'une analyse complémentaire par microscopie confocale.

## 8. Obtention d'anticorps spécifiques

Les sérums correspondant à 4 peptides ont été testés. La plupart des sérums pré immuns détectent de nombreuses bandes non spécifiques sur des extraits de cellules COS

transfectées par le clone pSG5 puro-flag 12S, exprimant la protéine FLJ10261 transmembranaire étiquetée par Flag (FigVI.10A). Parmi de nombreuses bandes, une bande de taille attendue (70kDa attendus) a été détectée de manière croissante avec les sérums des semaines successives après immunisation. Une purification a été nécessaire, compte tenu du nombre élevé de bandes non spécifiques détectées. Au cours de cette purification, la plupart des bandes non spécifiques ont été éliminées. Seul le sérum 2069 permet la détection d'une bande de taille attendue sur la protéine endogène dans les cellules Fadu (FigVI.10B).



### Caractérisation et purification de sérums Anti-FLJ10261

A- Cellules COS transfectées pour B14 12S (sens)

B-Protéine endogène dans la lignée Fadu

Les sérums pré immuns (PI), final bleed (FB), ainsi que les différentes semaines d'immunisation (W) et le sérum purifié (Pur) sont présentés.

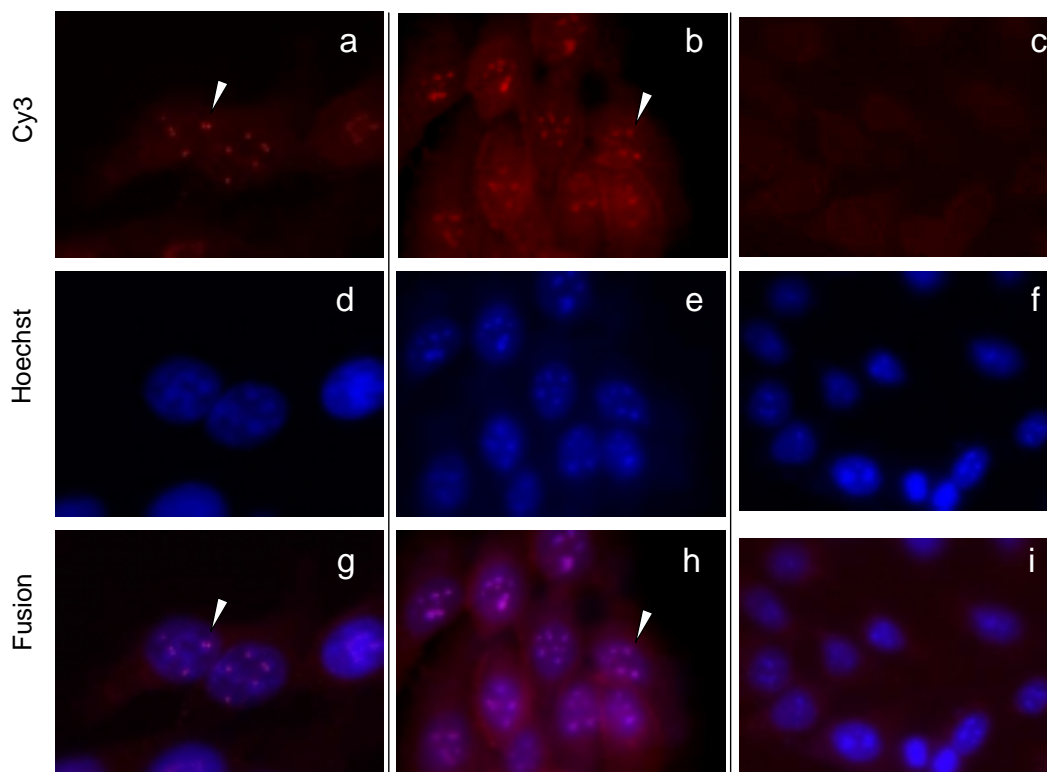
Figure VI.10

Dans une moindre mesure, le sérum 2067 permet la détection d'une bande unique, de taille nettement inférieure à la taille attendue. Cette taille pourrait éventuellement résulter d'un

clivage ou d'une modification de la protéine endogène. Le sérum pré immun correspondant ne détecte que peu de bandes non spécifiques.

## 9. Etude de la localisation subcellulaire endogène par immunocytochimie avec l'anticorps spécifique

### a. Lignée Hacat



#### Immunocytochimie pour FLJ10261; lignée HaCaT

Trois exemples de localisation avec le sérum 2069 purifié sont montrés à la dilution 1/1000 (a), à la dilution 1/200 (b) et sans anticorps primaire (c) sont présentés ainsi que les marquages des noyaux correspondants (d,e,f) et la fusion de ces images (g,h,i).

Figure VI.11

La lignée HaCaT n'exprime que peu ou pas le gène FLJ10261 (avec les amorces utilisées en RT-PCR semiquantitative). Cependant une localisation très particulière et très délimitée est observée, la protéine apparaissant localisée sous forme de « corps nucléaires » finement délimités (FigVI.11 a et b). Il pourrait s'agir de grands complexes au sein desquels adviennent les phénomènes d'épissage, appelés « speckles ».

Une localisation nucléaire de certains récepteurs possédant une activité tyrosine kinase (RTK) a déjà été rapportée, et, les fonctions de ces protéines au niveau nucléaire sont de plus en plus étudiées. Les complexes entre les facteurs de croissance et leurs récepteurs ne demeurent en effet pas à la surface de la cellule au niveau des membranes, comme synthétisé par Carpenter {Revue, (Carpenter, 2003)}. Avant d'être rapporté pour des récepteurs à activité tyrosine kinase, ce type de mécanisme a été rapporté pour d'autres récepteurs et d'autres protéines transmembranaires. Le récepteur Notch est clivé en deux fragments après liaison de son ligand, par l'action séquentielle de deux protéases localisées dans la membrane. De manière comparable, la protéine précurseur de la maladie d'Alzheimer (APP) est clivée par protéolyse et module l'expression de gènes cibles (Gao & Pimplikar, 2001). Le ligand Delta de Notch est également une protéine transmembranaire, dont la protéolyse, et une fonction nucléaire subséquente, est régulée par Notch (Bland et al., 2003). La cycline D1 est un gène cible direct de Notch (Jeffries et al., 2002). Notch inhibe la transactivation par AP-1 au niveau nucléaire (Chu et al., 2002).

Les récepteurs à activité tyrosine kinases existent sous deux formes : la forme intacte membranaire et la forme soluble de leur fragment cytoplasmique après clivage. Le mécanisme d'acheminement de la forme cytoplasmique jusqu'au noyau n'est pas connu. En particulier, les récepteurs de la famille ErbB sont exprimés dans les cellules épithéliales et sont activés par liaison de leur ligand EGF (Epidermal Growth Factor). Le récepteur à activité tyrosine kinase ErbB-4 est activé par clivage, après liaison à son ligand et internalisation de la membrane, ou, après activation de la protéine kinase C par le TPA sans internalisation. Le clivage résulte en la libération de la forme soluble dans le milieu extracellulaire, ou, dans les vésicules intracellulaires en cas d'internalisation des membranes. Le clivage des protéines RTK est effectué par deux types d'enzymes protéolytiques, TACE (Tumor Necrosis Factor  $\alpha$  Converting Enzyme) et PS-1 ( $\gamma$  secretase). TACE clive le domaine extracellulaire du récepteur tandis que PS-1 clive le domaine intracellulaire, qui va pouvoir être redirigé vers le noyau et s'y accumuler. Il est intéressant de noter que TACE est également une protéine possédant un domaine transmembranaire, possédant une activité métalloprotéase dans son domaine extracellulaire.

Une isoforme non clivable de ErbB-4, ne possédant pas de domaine extracellulaire a été isolée dans des tumeurs humaines. Dans le cas de ErbB-1, la liaison du ligand au récepteur est nécessaire au clivage et à la localisation nucléaire. ErbB-3 a été rapporté comme constitutivement nucléaire sous forme non clivée. L'addition de son ligand influence les proportions de ErbB-3 entre le noyau et le cytoplasme. FGFR1 (Fibroblast Growth Factor

Receptor I) est localisé dans le noyau en présence de FGF. La localisation nucléaire impliquerait la protéine importin  $\beta$  (Reilly & Maher, 2001).

Les données concernant une fonction nucléaire des RTK ont principalement été rapportées pour ErbB-1. En effet, ErbB-1 est capable de lier une séquence spécifique (ATRS, adenosine/thymidine –rich sequence) dans les promoteurs de gènes cibles (Lin et al., 2001). ErbB-1 se lie au promoteur de la cycline D1. L'inhibition de l'activité de PS-1 bloque l'inhibition de croissance par le ligand heregulin. FGFRI agirait comme un facteur de transcription sur le promoteur de FGF2. Certains auteurs indiquent que la constitution d'un stock cytosolique de FGFRI résulte d'une association instable de ce récepteur avec la membrane (Myers et al., 2003). Un autre rôle proposé des RTK est l'adressage de facteurs associés à leur domaine de récepteur jusqu'au noyau.

Outre la famille ErbB, d'autres protéines de type RTK comme CSF-1, c-Kit, MGF, Ax1, TrkA, Met, et Tie-1 sont clivées dans leur domaine extracellulaire. Cependant le clivage du domaine intracellulaire par PS-1 n'a pas été démontré. L'ensemble de ces RTK sont présentes dans le nucléoplasme et ne sont apparemment pas associées à l'enveloppe nucléaire

Des protéines transmembranaires sont retenues à la membrane du réticulum endoplasmique, et, sont solubilisés puis dirigés vers le noyau après clivage. En particulier, des facteurs de transcription transmembranaires sont piégés dans le réticulum endoplasmique. Les Sterol regulatory element binding proteins SREBP-1 et SREBP-2 ainsi que les facteurs de transcription ATF6 et beta sont clivés par la sérine protéase Site1 protease (SP-1), puis, par la métalloprotéase Site-2 protéase (SP-2).

Un clivage par le système ERAD (Endoplasmic Reticulum Associated Degradation), est capable d'extraire des protéines transmembranaires complètes de la membrane du réticulum endoplasmique. Le système ERAD permet, principalement, l'élimination des protéines mal conformées du réticulum et leur translocation dans le cytoplasme où elles sont dégradées. Un modèle pour la localisation nucléaire des protéines membranaires par importation rétrograde du réticulum endoplasmique au golgi puis au noyau a été proposé.

Des protéines transmembranaires liées à l'adhésion cellulaire comme CD44 et l'E-cadherin sont également clivées par protéolyse.

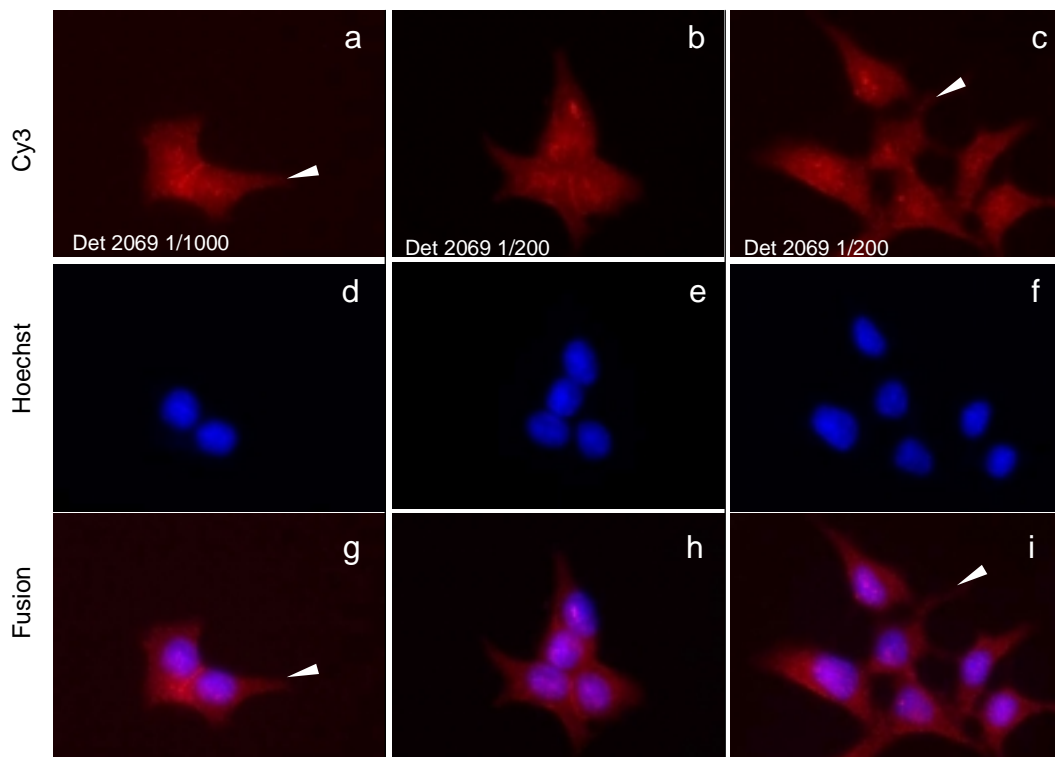
De manière particulièrement intéressante pour le clone B14.4, la protéine transmembranaire ZO-2 est un constituant des jonctions serrées, formant une barrière imperméable entre les cellules épithéliales et endothéliales, co-localisée dans des sous régions du noyau (speckles) avec le facteur d'épissage SC35 (Islas et al., 2002). Cette localisation est observée dans des conditions de faible densité de culture (aussi en cas de cicatrisation). Le marquage nucléaire



diminue lorsque la culture approche de la confluence. L'adressage de cette protéine dans le noyau est dépendant du cytosquelette. Il pourrait être intéressant de comparer l'expression de B14.4 avec des facteurs d'épissage et d'évaluer son rôle dans les jonctions serrées. Toutefois, B14.4 ne contient pas de domaines conservés de ce type de protéines comme les domaines PDZ, et GK.

Une séquence responsable de l'adressage des protéines transmembranaires dans le noyau a été isolée (Meyer & Radsak, 2000). Cette séquence DRLRHR n'est pas présente dans la séquence des versions 7 et 8 transmembranaires de FLJ10261, ni dans la protéine transmembranaire 99b78 (partie VII).

### b. Lignée Det562

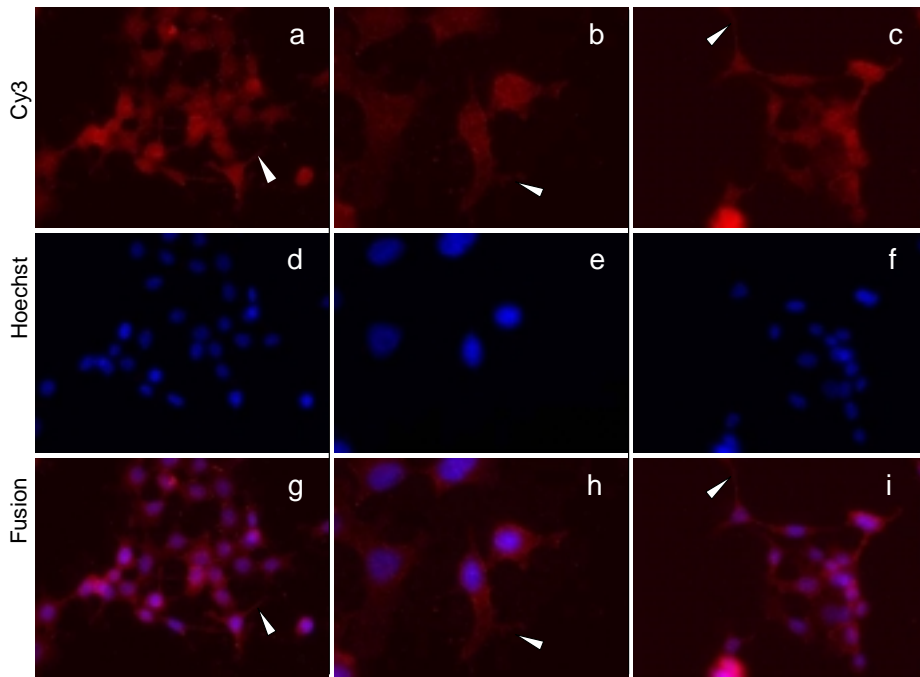


#### Immunocytochimie pour FLJ10261; lignée Det562

Trois exemples de localisation avec le sérum 2069 purifié sont montrés à la dilution 1/1000 (a), et à la dilution 1/200 (b et c) sont présentés ainsi que les marquages des noyaux correspondants (d,e,f) et la fusion de ces images (g,h,i).

Figure VI.12

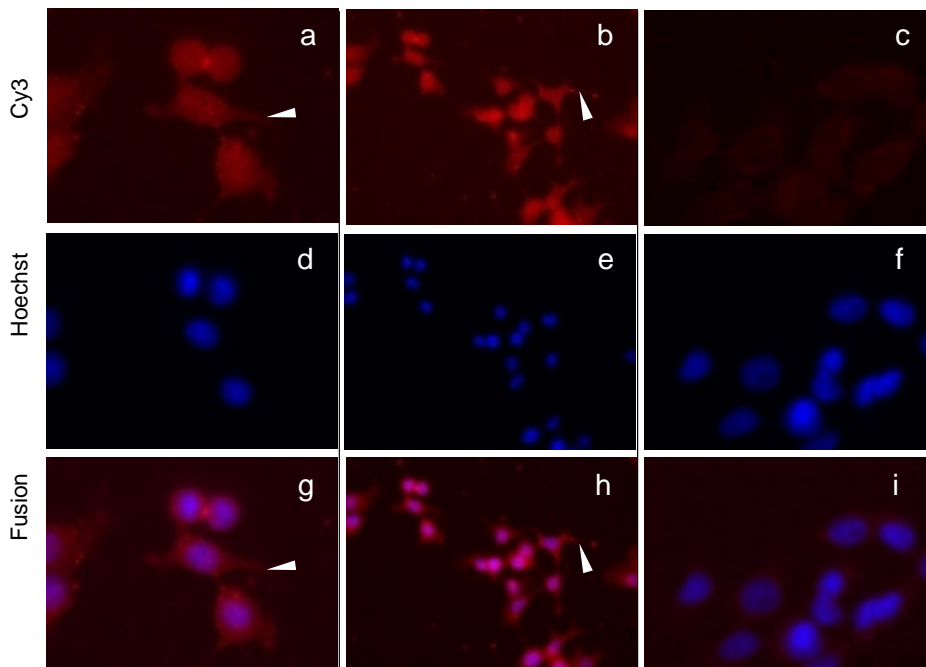
### c. Lignée Fadu



**Immunocytochimie pour FLJ10261; lignée Fadu**

Trois exemples de localisation avec le sérum 2069 purifié sont montrés à la dilution 1/200 (a), à la dilution 1/1000 (b et c) sont présentés ainsi que les marquages des noyaux correspondants (d,e,f) et la fusion de ces images (g,h,i).

Figure VI.13



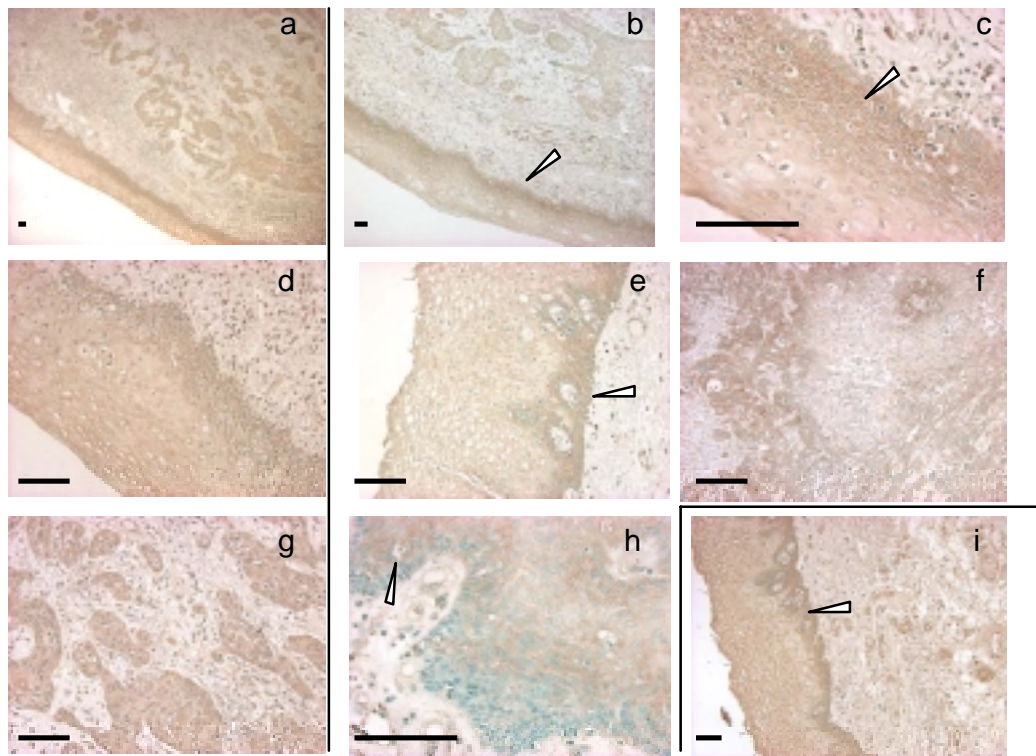
**Immunocytochimie pour FLJ10261; lignée Fadu**

Deux exemples de localisation avec le sérum 2069 purifié sont montrés à la dilution 1/200 (a, b) et un contrôle sans anticorps primaire (c) sont montrés ainsi que les marquages des noyaux correspondants (d,e,f) et la fusion de ces images (g,h,i).

Figure VI.14

Pour les deux lignées Det562 (FigVI.12) et surtout Fadu (FigVI.13&14), exprimant plus fortement FLJ10261, la localisation semble être, comme attendu, membranaire puisque les prolongements membranaires sont nettement visibles. Une analyse supplémentaire par microscopie confocale doit être réalisée pour confirmer la nature membranaire de FLJ10261 et exclure une localisation cytoplasmique.

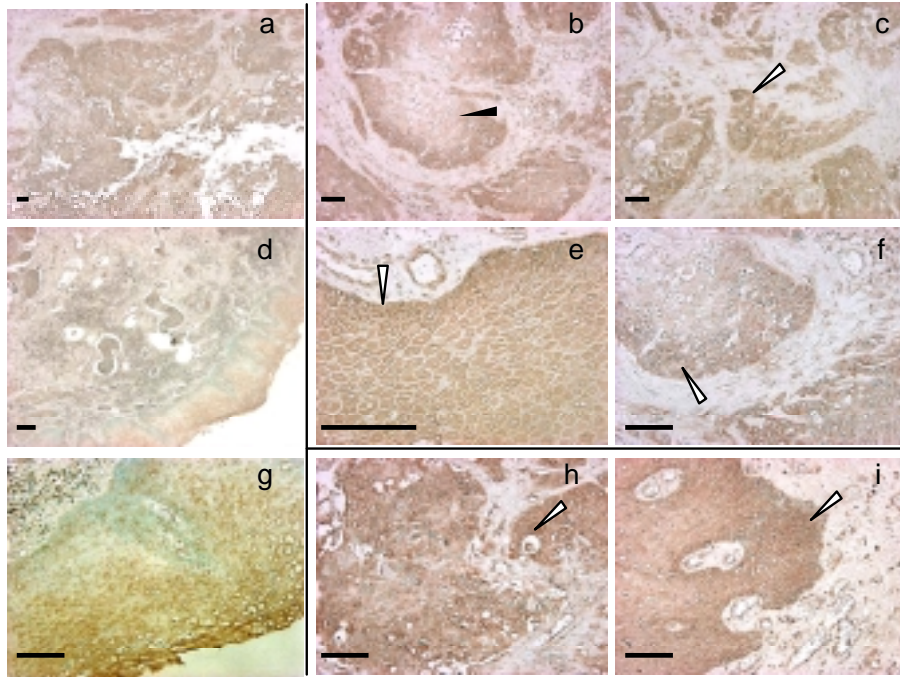
## 10. Caractérisation de l'expression par Immunohistochimie



Immunohistochimie pour FLJ10261; tumeur 24189T

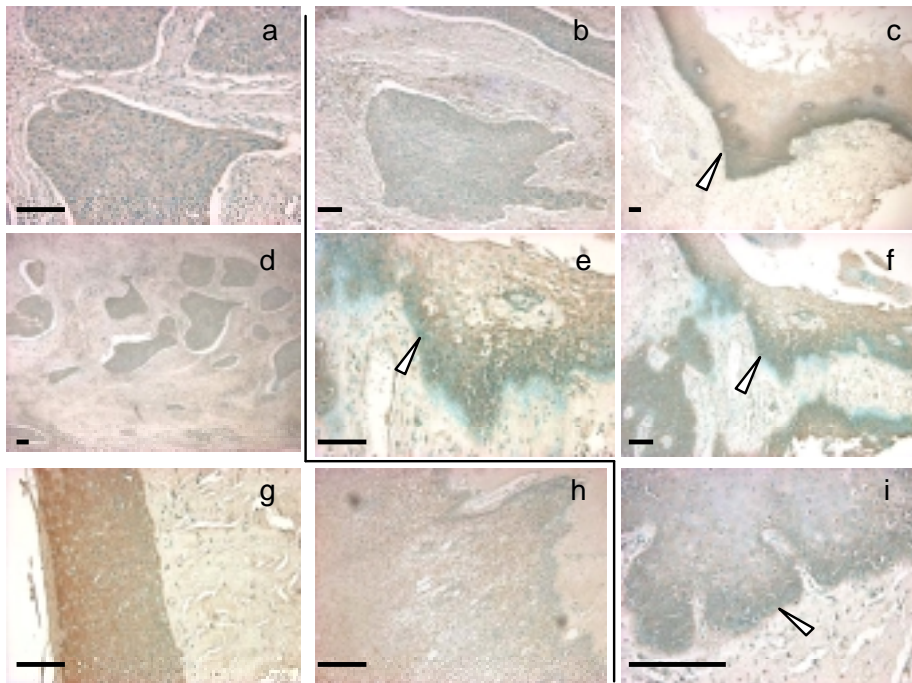
Des exemples de marquage par le sérum pré immun (a, d, g), le sérum purifié 2069 (b, c, e, f, h) et le FB 2069 (i) sont présentés.

Figure VI.15



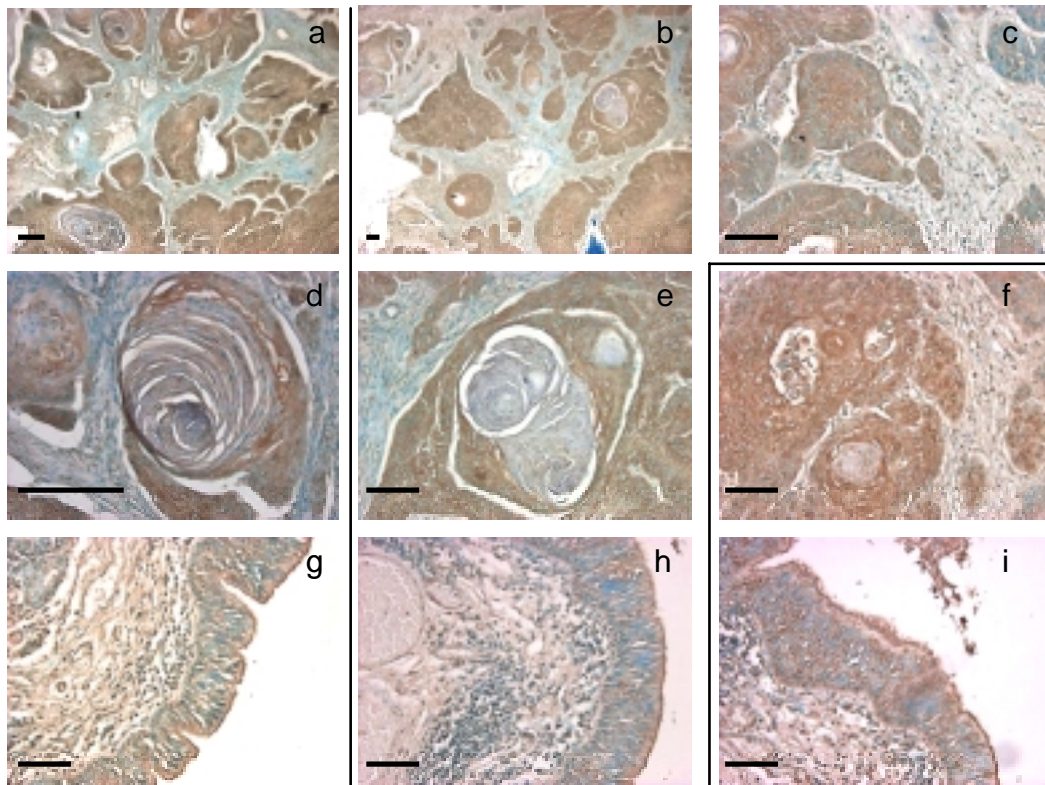
Immunohistochimie pour FLJ10261; tumeur 31073T  
Des exemples de marquage par le sérum pré immun (a, d, g), le sérum purifié 2069 (b, c, e, f) et le FB 2069 (h, i) sont présentés.

Figure VI.16



Immunohistochimie pour FLJ10261; tumeur 35574T  
Des exemples de marquage par le sérum pré immun (a, d, g, h), le sérum purifié 2069 (b, c, e, f, i) sont présentés.

Figure VI.17



#### Immunohistochimie pour FLJ10261; tumeur 29335T

Des exemples de marquage par le sérum pré immun (a, d, g), le sérum purifié 2067 (b, c, e, h) et le FB 2067 (f, i) sont présentés.

Figure VI.18

Le sérum pré immun produit un signal fort qui rend l'analyse difficile. Nous envisageons de le purifier de manière analogue au sérum purifié. Une amélioration du blocage des interactions antigène/anticorps par utilisation de sérum de chèvre pourrait également contribuer à diminuer le bruit de fond. Une optimisation de la dilution de sérum utilisée pourrait améliorer la qualité de l'analyse.

Il semble pourtant que les cellules basales des zones comportant un épithélium structuré soient légèrement plus marquées avec le sérum purifié qu'avec le sérum pré immun (FigVI.15 et V.17e). De même, dans la tumeur 24189T (FigVI.15f), ainsi que dans une moindre mesure dans la tumeur 31073T (FigVI.16), le marquage semble être plus faible, exclu de certaines zones des masses tumorales (FigVI.16b), de manière comparable aux figures « en champignon » observées en Hybridation *In Situ* sur la tumeur 840T.

L'épitope reconnu par le sérum 2069 est situé au niveau intracellulaire, à l'extrémité C terminale de la protéine. Il est possible que ce sérum détecte un signal dans le noyau des

cellules HaCaT correspondant à une forme clivée soluble, co-localisée avec des facteurs d'épissage dans des speckles.

### Position des peptides (épitopes) sur FLJ10261 7 transmembranaire cloné

MEMCDQRHNIIMCPLCDKTCYWKMSACATARASHLFDNPATVFFSVFMALWAATFMEHWKRRQMRLNRYRWDLTGFEEEEAVKDHPRAEYEARVLEKSLKESRNKETD  
 KVCLTWRDRFPAYLTNLVSIIFMIAVTFIAVLGVIYIRISMAAALAMNSPSVRSNIRVTATAVIINLVVILLDEVYGCIAARWLTKIEVPKTEKSFEEERLIFKAFLLKF  
 VNSYTPIFYVAFKGRFVGRPGDYVYIFRSFRMEECAPGGCLMELCISLIMLGKQLIQNNLFEIGIPKMKKLIRYLKQKQSPDHEECVKRQRYEVDYNLEPPFAGLT  
 PEYMEMIIQFGFVTLFVASFPLAPLALLNIIIEIRLDAKKFVTELRFPVAVRAKDIGIWIYNILRGIGKLAVIINAFVISFTSDFIPRLVLYMYSKNGTMHGFVNHTLSS  
 FNVSDFQNGTAPNDPLDLGYEVQICRYKDYREPPWSENKYDISKDFWAVLAARLAFVIVFQNLVFMMSDFVDWVIPDKDISQIHKKEKVLVLELPMREEQDKQQLLETW  
 MEKERQKDEPPCNHNTKACPDSLGSPPASHAYHGGVL

PG216 lapins 2063 & 2065 PG219 lapins 2067 & 2070  
 PG217 lapins 2066 & 2069 PG218 lapins 2068 & 2071

### Prédiction de la position dans la cellule (versions 8 et 7 trans membranaires)

Extérieur	Transmembrane	Intérieur
<p>WVYEILKRTTCTKAKYSMGITSLLANGVYAAAYPLHDGDYNGENVEFNDKLLYEENARYGVFYKYQPIDLVRKYFGEKIGLYFAWLGVYQMLIPASIVGIIIVFLYGCATM            DENIPSEMCDQRHNIIMCPLCDKTCYWKMSACATARASHLFDNPATVFFSVFMALWAATFMEHWKRRQMRLNRYRWDLTGFEEEEAVKDHPRAEYEARVLEKSLKESRN            NKETDKVKLTWRDRFPAYLTNLVSIIFMIAVTFIAVLGVIYIRISMAAALAMNSPSVRSNIRVTATAVIINLVVILLDEVYGCIAARWLTKIEVPKTEKSFEEERLIFKAF            LLKLVNSYTPIFYVAFKGRFVGRPGDYVYIFRSFRMEECAPGGCLMELCISLIMLGKQLIQNNLFEIGIPKMKKLIRYLKQKQSPDHEECVKRQRYEVDYNLEPPFA            GLTPEYMEMIIQFGFVTLFVASFPLAPLALLNIIIEIRLDAKKFVTELRFPVAVRAKDIGIWIYNILRGIGKLAVIINAFVISFTSDFIPRLVLYMYSKNGTMHGFVNHTL            SSVNVDQNGTAPNDPLDLGYEVQICRYKDYREPPWSENKYDISKDFWAVLAARLAFVIVFQNLVFMMSDFVDWVIPDKDISQIHKKEKVLVLELPMREEQDKQQLLET            WMEKERQKDEPPCNHNTKACPDSLGSPPASHAYHGGVL</p>	<p>VLAARLAFVIVFQNLVFMMSDFVDWVIPDKDISQIHKKEKVLVLELPMREEQDKQQLLETW</p>	<p>WVYEILKRTTCTKAKYSMGITSLLANGVYAAAYPLHDGDYNGENVEFNDKLLYEENARYGVFYKYQPIDLVRKYFGEKIGLYFAWLGVYQMLIPASIVGIIIVFLYGCATM            DENIPSEMCDQRHNIIMCPLCDKTCYWKMSACATARASHLFDNPATVFFSVFMALWAATFMEHWKRRQMRLNRYRWDLTGFEEEEAVKDHPRAEYEARVLEKSLKESRN            NKETDKVKLTWRDRFPAYLTNLVSIIFMIAVTFIAVLGVIYIRISMAAALAMNSPSVRSNIRVTATAVIINLVVILLDEVYGCIAARWLTKIEVPKTEKSFEEERLIFKAF            LLKLVNSYTPIFYVAFKGRFVGRPGDYVYIFRSFRMEECAPGGCLMELCISLIMLGKQLIQNNLFEIGIPKMKKLIRYLKQKQSPDHEECVKRQRYEVDYNLEPPFA            GLTPEYMEMIIQFGFVTLFVASFPLAPLALLNIIIEIRLDAKKFVTELRFPVAVRAKDIGIWIYNILRGIGKLAVIINAFVISFTSDFIPRLVLYMYSKNGTMHGFVNHTL            SSVNVDQNGTAPNDPLDLGYEVQICRYKDYREPPWSENKYDISKDFWAVLAARLAFVIVFQNLVFMMSDFVDWVIPDKDISQIHKKEKVLVLELPMREEQDKQQLLET            WMEKERQKDEPPCNHNTKACPDSLGSPPASHAYHGGVL</p>

MEMCDQRHNIIMCPLCDKTCYWKMSACATARASHLFDNPATVFFSVFMALWAATFMEHWKRRQMRLNRYRWDLTGFEEEEAVKDHPRAEYEARVLEKSLKESRNKETD  
 KVCLTWRDRFPAYLTNLVSIIFMIAVTFIAVLGVIYIRISMAAALAMNSPSVRSNIRVTATAVIINLVVILLDEVYGCIAARWLTKIEVPKTEKSFEEERLIFKAFLLKF  
 VNSYTPIFYVAFKGRFVGRPGDYVYIFRSFRMEECAPGGCLMELCISLIMLGKQLIQNNLFEIGIPKMKKLIRYLKQKQSPDHEECVKRQRYEVDYNLEPPFAGLT  
 PEYMEMIIQFGFVTLFVASFPLAPLALLNIIIEIRLDAKKFVTELRFPVAVRAKDIGIWIYNILRGIGKLAVIINAFVISFTSDFIPRLVLYMYSKNGTMHGFVNHTLSS  
 FNVSDFQNGTAPNDPLDLGYEVQICRYKDYREPPWSENKYDISKDFWAVLAARLAFVIVFQNLVFMMSDFVDWVIPDKDISQIHKKEKVLVLELPMREEQDKQQLLETW  
 MEKERQKDEPPCNHNTKACPDSLGSPPASHAYHGGVL

7/8ème domaine sous le  
 seuil de détection

### Position des épitopes et orientation dans la cellule

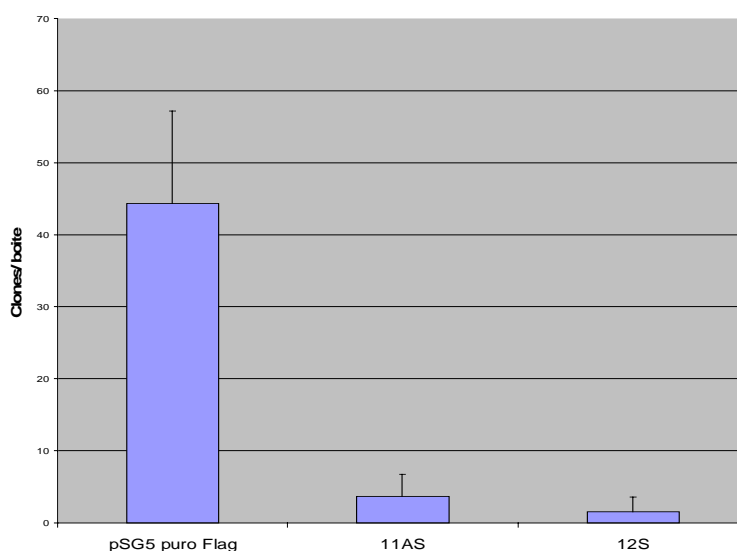
Figure VI.19

## B. Etude fonctionnelle préliminaire:

---

### 1. Test de clonogénicité

J'ai, à deux reprises, tenté d'effectuer un test de clonogénicité dans la lignée Fadu (FigVI.20) avec les plasmides sens (12S) et antisens (11AS) correspondant à la séquence codante de la protéine FLJ10261-7 transmembranaire. Des clones stables ont été obtenus et la transfection des plasmides 11AS et 12S semblent entraîner une altération du nombre de clones obtenu. Il paraît cohérent que la transfection du clone antisens 11AS provoque une moindre formation de clones que le contrôle, en rapport avec l'induction éventuelle de l'apoptose ou une inhibition possible de la prolifération. Toutefois, la transfection du clone sens 12S entraîne le même effet, ce qui semble indiquer que les variations observées pourraient ne pas être spécifiques. Il est possible que, pour des raisons inconnues, la transfection de ce plasmide n'aie pas été aussi efficace que pour le plasmide contrôle pSG5 puro flag et pour l'antisens. Un autre explication pourrait être un effet dominant négatif de la version 7 transmembranaire en orientation sens sur la forme 8 transmembranaire endogène, produisant un effet similaire à l'antisens.



Tests de clonogénicité pour FLJ10261  
Clones sens (12S), Antisens (11AS) et contrôle (pSG5 puro Flag)  
dans la lignée Fadu

Figure VI.20

## **2. Analyse par FACS**

Au cours d'une expérience la transfection du clone B14 11AS (antisens) a entraîné une diminution de la proportion des cellules en phase S du cycle cellulaire (-12.9% contre -5% avec pSG5 puro flag) dans les cellules transfectées en même temps qu'une augmentation de la fraction des cellules en apoptose supérieure à celle observée avec le vecteur pSG5 puro flag vide (+7.45% contre +3% avec pSG5 puro flag). Ce résultat n'a pas pu être reproduit de manière régulière et la transfection du vecteur pBSK entraîne une variation assez proche de celle causée par B14 11AS dans les cellules transfectées. Ce résultat est donc peu convaincant mais s'il pouvait être reproduit d'une manière plus claire (clonage dans pCMV), il pourrait être cohérent avec le rôle d'oncogène putatif de FLJ10261, promouvant la prolifération et autorisant une résistance à l'apoptose.

## **3. Etude de l'induction par le sérum**

Le gène FLJ10261 ne semble pas régulé par le sérum après analyse en western blot.



## C. Avancée des connaissances durant l'étude

---

Courant 2003, Katoh a publié une caractérisation *In Silico* de ce clone (Katoh, 2003). Parallèlement à cette publication, une version plus longue de l'ORF codant pour une protéine contenant 8 domaines transmembranaires a été ajoutée aux bases de données. Cette combinaison de 8 domaines transmembranaires permet la formation de pores dans les membranes cellulaires et est donc associée à des transporteurs membranaires.

La région 11q13 est une des régions les plus fréquemment amplifiées dans les cancers humains. Sur la base de l'amplification du locus CCND1-EMS1 localisé en 11q13 dans les cancers de l'oesophage {36-50% ; (Williams et al., 1993 ; Yoshida et al., 1993)}, de la vessie {21% ; (Proctor et al., 1991)}, du sein {15% ; (Dickson et al., 1995)} et du foie {13% ; (Zhang et al., 1993)}, Katoh a tenté de déterminer par analyse informatique la fonction du gène FLJ10261 un de seuls gènes non caractérisés dans ce locus à ce jour. La fréquence des amplifications du locus 11q13 dans les tumeurs hypopharyngées a déjà été démontrée au Centre Paul Strauss et mise en relation avec une invasion ganglionnaire. Elle ne possède toutefois pas de valeur pronostique significative (Muller et al., 1994; Muller et al., 1997).

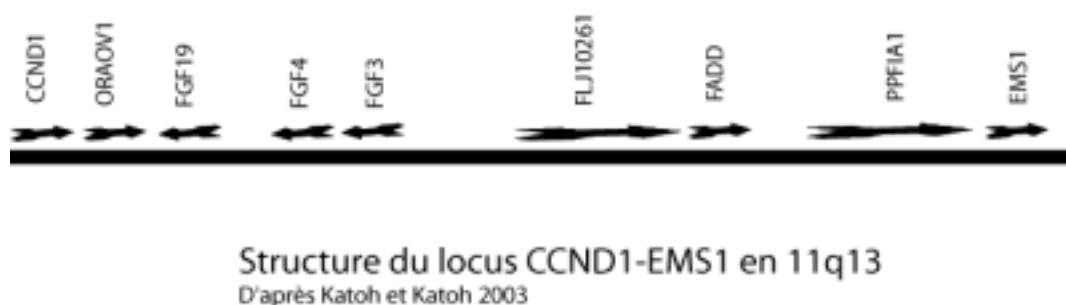


Figure VI.21

Tous les gènes de ce locus sont caractérisés sauf FLJ10261. CCND1 est impliqué dans le contrôle du cycle cellulaire (Han et al., 1999). ORAOV1 (Overexpressed in Oral Cancer 1) est surexprimé dans les cancers buccaux (Huang et al., 2002). Plusieurs gènes sont de la famille des « fibroblast growth factors » (FGF19, FGF4, FGF4) impliqués dans le contrôle de la croissance des fibroblastes et dans le cancer (Ornitz & Itoh, 2001). FADD est un médiateur connu de l'apoptose {apoptotic signaling adaptor protein, (Sheikh & Huang, 2003)}. PPFIA1

est impliqué dans les interactions protéines/protéines (PTRF interacting coiled-coil protein). EMS1 est impliqué dans la régulation des jonctions adhérentes (Schuurin et al., 1993).

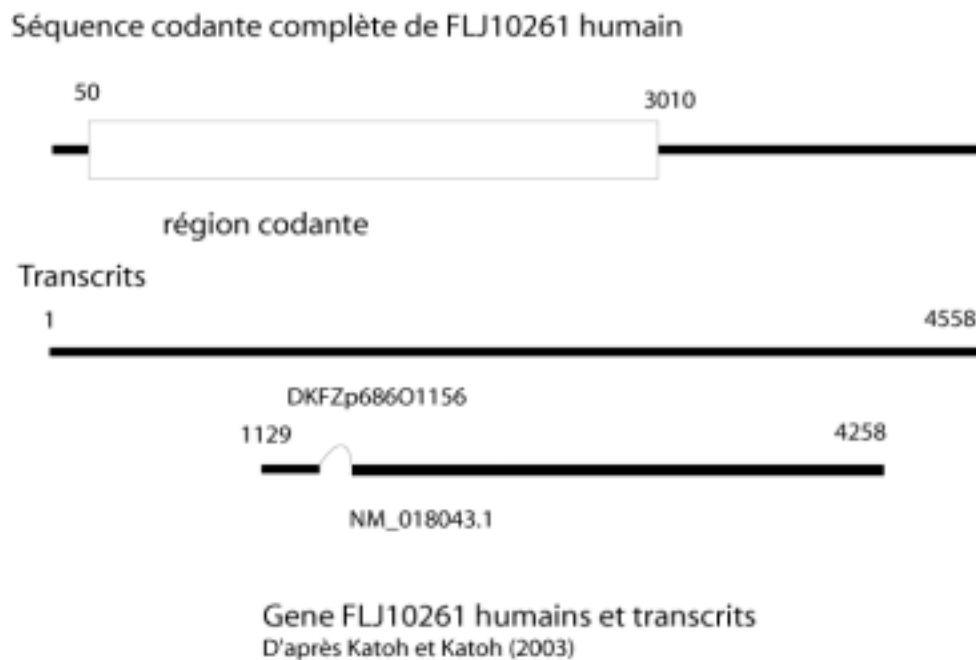
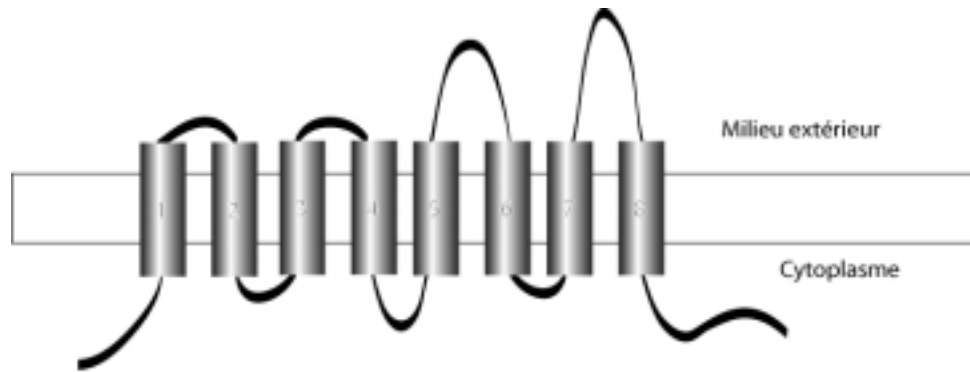


Figure VI.22

Le gène FLJ10261, composé de 26 exons, est situé entre le gène FGF3 et FADD au sein du locus CCND1-EMS1. Il existe 2 isoformes de FLJ10261 présentant ou non l'exon 15. Le gène FLJ10261 est homologue des gènes C12orf3 localisé en 12p13, C11orf25 localisé en 11p14 et FLJ34272 localisé en 12q23.

La protéine FLJ10261 humaine possède une homologie de 89,8% d'identité avec la protéine chez la souris et respectivement 58,4%, 38,3% et 38,6% d'identité avec les protéines C12orf13, C11orf25 et FLJ34272. D'après Katoh (Katoh, 2003), la présence de sites de glycosylation communs indique l'appartenance à une famille de protéines contenant 8 domaines transmembranaires. Des prédictions informatiques indiquent que toutes ces protéines contiennent 8 domaines transmembranaires et que leurs extrémités N- et C-terminales sont orientées vers le cytoplasme, suggérant un rôle comme transporteurs de substrats non identifiés (figVI.23). L'analyse fonctionnelle n'est pas poussée plus loin, cet article étant entièrement *In Silico*.



Structure putative des protéines de la famille FJL10261  
D'après Katoh et Katoh (2003)

Figure VI.23

Au niveau des bases de données, une analyse *In Silico* indique que l'ARNm de FLJ10261 est exprimé dans les tumeurs VADS, parathyroïdes, du sein, du pancréas et gastriques. Cette expression dans les tumeurs VADS signifie que cet ARNm a été trouvé présent dans une banque d'ADNc d'origine VADS. L'étude de Katoh ne présente aucune preuve physique de l'expression de cet ARNm dans les cancers VADS et encore moins d'une expression différentielle.

Il est intéressant de noter l'identification par ces auteurs de clusters de gènes homologues situés sur les régions 11q13 mais aussi 12p13, région identifiée comme point chaud des remaniements associés à la carcinogenèse dans notre étude DD et au niveau génomique (Chen et al., 1999 ; Lemaire et al., 2003). La présence de ces gènes homologues, membres d'une même famille potentielle de protéines transporteurs, relie potentiellement ces deux régions.

## D. Matériel biologique

---

Clone B14.4 Résumé	Séquence	Localisation
fragment DD dans pGEM-T EASY (Clontech)	AAGCTTCTCAACGATATTTGGTAACTAATAAACTCCCTCCACCTTTGAGCTA CAGAAAAAAAAATCCTCAATCTACCATATAATTGATATTTGAAAAAAAAAAG CTT	Homo sapiens cDNA FLJ10261 fis, clone HEMBB1000975, Accession Nb AK001123.1 , match 88bp, nt 2784-2697
Fragment séquence codante (CDS) cloné	ATGGAGATGTGTGACCAGAGACACAATATCACCATGTGCCCGCTTTGCGAC AAGACCTGCAGCTACTGGAAGATGAGCTCAGCCTGCGCCACGGCCCGCG CCAGCCACCTCTTCGACAACCCGCGCCACGGTCTTCTCTGTCTTCATGGC CCTCTGGGCTGCCACCTTCATGGAGCACTGGAAGCGGAAACAGATGCGAC TCAACTACCGCTGGGACCTCACGGGCTTTGAAGAGGAAGAGGAGGCTGTC AAG	Homo sapiens hypothetical protein FLJ10261 (FLJ10261), mRNA, Accession Number NM_018043.1, nt (43-1827)
Orientation Sens 12S		
Orientation Antisens 11AS		
<b>Protéine correspondante</b>	MEMCDQRHNITMCP LCKTCSYWK MSSACATARASHLFDNPAIVFFSVFM ALWAATFMEHWKRKQMLNRYRDLTGFE EEEAVKDHPRAEYEARVLEK SLKKE SRNKETDKVKLTWRDRFPAYLTNLVSIIFMIAVTFAIVLGVIIYRISMAAA LAMNSSPSVRSNIRVTVTATAVINLVIIILLDEVYGC IARWLTKIEVPKTEKSFE ERLIFKAFLLKFVNSYTPIFYVAFFKGRFVGRPGDYVYIFRSFR	
<b>Peptides injectés</b>		
PG218	KRKQMLNRYRDLTGFE E C	Lapins 2068 & 2071
PG219	YKDYREPPWSENKYDISC	Lapins 2067 & 2070
PG216	KLKQQSPPDHEECCKKRKC	Lapins 2063 & 2065
PG217	EKERQKDEPPCNHHTC	Lapins 2066 & 2069

Matériel biologique obtenu pour l'étude de FLJ10261

Figure VI.24

**Partie VII**  
**Caractérisation du candidat 99b78**

## A. Caractérisation du candidat 99b78 :

---

### 1. Profil en DD

Ce clone a été isolé en DD comme surexprimé dans les tumeurs par Christine Wasyluk. Cette dernière ne poursuivant pas ce projet, j'ai poursuivi la validation de ce clone en l'incluant à mes membranes de macro arrays pour validation en Reverse Northern blot, après compilation des résultats des différents expérimentateurs. En particulier, la différence d'expression de ce clone observée avec divers types de sondes, et, son caractère inconnu ont guidé ce choix.

### 2. Profil en Reverse Northern blot

Ce clone a présenté une surexpression dans le pool tumoral en Reverse Northern blots avec des sondes DD, et, une surexpression dans des tumeurs de type E et U avec des sondes linéaires sur macro arrays de nylon (clones du 1<sup>er</sup> cycle de DD analysés avec des sondes linéaires par ExonHit Therapeutics).

### 3. Identification du gène

La séquence-insert de DD de 318 bp n'a pas initialement présenté d'homologie significative pour un gène ou un ARNm (homologie sur 21bp avec le clone Homo sapiens genomic DNA, chromosome 21q, section 75/105).

Un allongement par assemblage d'ESTs (TIGR ; <http://tigrblast.tigr.org/tgi/>) a permis de générer un clone TIGR de 1600bp contenant un ORF, une séquence codante de 579 bp. L'analyse de la protéine codée par ce gène a révélé la présence d'un domaine transmembranaire, confortant la validité de l'assemblage. Plus tard, notre clone (comme la séquence codante prédite et le clone TIGR correspondant) a présenté de l'homologie pour un clone situé en 1p33 sur le génome (Human DNA sequence from clone RP4-705P8 on chromosome 1p33-34.2 d'une taille de 32kb). Récemment plusieurs clones IMAGE ont été décrits comme présentant de l'homologie pour 99b78. Toutefois aucun de ces clones ne contient d'ORF significative annotée dans la séquence nucléotidique. Sans assemblage,

aucune séquence codante ne peut être prédite pour ce clone. Une analyse de la structure putative du gène (Spidey) a révélé que ce gène est composé de 3 exons, de 40, 220 et 319 bp respectivement (coordonnées génomiques 3'990'531-3'990'570, 3'920'989-3'921'208, 3'920'245-3'920'563 sur le contig NT\_032977.6), séparés par de très larges introns sur le chromosome 1.

Le clone 99b78 est localisé en 1p33-34.2. En 1p33, des remaniements génétiques indiquent la présence de gènes suppresseurs de tumeurs potentiels comme d'oncogènes potentiels. Dans les lymphomes des cellules B, des délétions ont été détectées en 1p33-34 (Cigudosa et al., 1999). Dans les carcinomes hépatocellulaires, on retrouve 1p34.3-35 et 1p33-34.1 parmi les régions les plus fréquemment amplifiées en CGH (Zimonjic et al., 1999). Des amplifications en 1p33-35 sont détectées dans les chondrosarcomes (Larramendy et al., 1997) et des amplifications récurrentes sont isolées en 1p33-p32 dans les sarcomes (Menghi-Sartorio et al., 2001). Un gène suppresseur de tumeurs potentiel, PTCH2 (Novel Human Patched-like gene), est localisé en 1p33-34, une région délétée dans divers types de tumeurs mais présente des variants d'épissage et une surexpression dans les carcinomes des cellules basales de la peau (Zaphiropoulos et al., 1999).

La protéine prédite ne possède pas d'homologie significative avec d'autres protéines, bien que possédant un domaine transmembranaire. Une analyse des protéines apparentées par PipeAlign n'a fourni aucune protéine homologue, de la même famille que 99b78.

Récemment, un blastp de cette protéine (FigVII.1) sur l'ensemble des bases de données de protéines ne produit qu'une faible homologie de 14 aa identiques sur une région de 43aa (32%) avec une enzyme alcool dehydrogénase procaryote (dehydrogenase related to short-chain alcohol dehydrogenase de Magnetococcus sp MC-1). Etant donné l'étiologie des cancers de l'hypopharynx, cette homologie partielle pourrait tout de même constituer un indice sur la fonction de 99b78, en tant qu'enzyme dégradant les alcools, associée aux membranes. Une autre zone d'homologie faible (31% d'identité sur une zone de 77 aa) est trouvée pour la protéine fibronectin, type III and protein kinase family member/ Hypothetical protein C16D9.2a de *Caenorhabditis elegans*. La région d'homologie ne correspond pas au domaine responsable de l'activité kinase proprement dite, ni à aucun autre domaine fonctionnel annoté dans la séquence.

Récemment les clones IMAGE 3916394 (1'014bp, mRNA with retained intron, Tissue : skin, melanotic melanoma), IMAGE 5178133 (1'775bp, mRNA, Tissue : adult brain, lung, testis pooled), IMAGE 5180231 (1'773bp) ont été trouvés homologues au fragment initial de DD de façon identique au clone situé sur le chromosome 1. Ces clones ne

correspondent à aucune protéine prédite d'après les données annotées. Cependant une recherche de la présence d'ORFs dans ces séquences révèle qu'elles codent potentiellement pour des protéines en partie identiques à notre clone 99b78 prédit (FigVII.1). En effet, lorsqu'une ORF significative peut être trouvée, les 119 premiers aa sont strictement identiques à la séquence que nous avons isolée. Le clone IMAGE 3916394 ne code pas pour une protéine, peut être en lien avec la présence signalée d'un intron anormalement retenu dans le clone IMAGE. Ce clone est très anormal attendu qu'il contient apparemment 230 bp ayant une faible homologie (31% d'identité en tblastX du clone image sur les bases de données de protéines) avec des protéines structurales de virus. Le clone IMAGE 5178133 ne contient pas la séquence codante complète d'une protéine de 256 aa, identique à la protéine 99b78 prédite pour les 119 premiers aa. Cette protéine présente un domaine supplémentaire ayant une faible homologie (38% d'identité) avec une enzyme de réparation de l'ADN, DNA mismatch repair enzyme (predicted ATPase) de *Pseudomonas fluorescens* PfO-1. Le clone IMAGE 5180231 code pour une protéine de 220aa, identique à la protéine précédente sur ses 217 premiers aa. Ce clone IMAGE, contrairement au précédent, contient un codon stop. La séquence codante correspond donc soit à un variant d'épissage alternatif contenant un exon final différent de la séquence prédite, soit à la rétention d'un intron contenant un codon stop.

La recherche forcée d'une ORF autour de notre insert DD dans le clone génomique RP4-705P8 permet d'obtenir une protéine identique à la séquence prédite pour les 105 premiers aa. La divergence des séquences indique la présence probable d'un intron après l'aa 105. Il est donc possible que la structure prédite par Spidey soit incomplète et que la limite d'un autre exon se situe vers l'acide aminé 105.

La présence de grands introns peut être également associée à l'existence d'exons supplémentaires non caractérisés, dans des variants d'épissage alternatif. La séquence DD que nous avons isolé se trouve sur une partie commune à tous ces ADNcs.

La séquence codante initialement prédite par TIGR semble pertinente, attendu que nous avons réussi à cloner une séquence strictement identique à la prévision à partir d'ADNc tumoral, et, que la structure prédite du gène sur le chromosome semble cohérente. Les enzymes utilisées pour le clonage étaient des enzymes adaptées à l'amplification haute fidélité et longue distance des fragments de PCR (PCR Advantage Mix du protocole SMART), également utilisés pour les techniques de 5' RACE. De plus, cette séquence code pour un domaine transmembranaire, un domaine fonctionnel (ainsi que d'autres domaines putatifs), dont la présence n'est probablement pas liée au hasard. Des variants d'épissage alternatif existent peut être pour les clones IMAGE n'ayant pas retenu d'intron (IMAGE 5178133 & IMAGE



5180231). Mise à part notre analyse « forcée » des clones IMAGE, produisant des ORF potentielles, aucune protéine apparentée à 99b78 n'est détectée dans les bases de données protéiques par une analyse avec PipeAlign (<http://igbmc.u-strasbg.fr/PipeAlign/>).



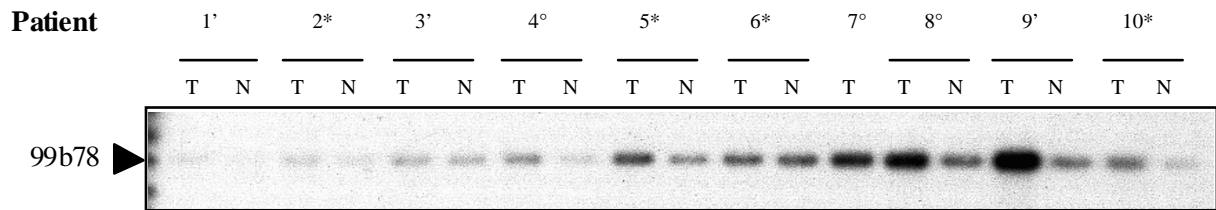
### Isoformes possibles et structure de 99b78

Figure VII.1

## 4. Validation en Virtual Northern blot

Un transcrit d'environ 3'000bp est détecté en VN avec une sonde synthétisée à partir de l'insert DD (FigVII.2). Cette taille est supérieure à celle du clone TIGR attendu. Le clone TIGR, assemblage d'ESTs, pouvait ne pas contenir l'ARNm complet, cependant l'ORF prédite contenue dans le clone TIGR que nous avons amplifié et cloné à partir d'ARN tumoral paraissait complète, du codon start au codon stop. Outre cette possibilité, on ne peut exclure soit un artefact lors du procédé SMART (dimère de 1'600bp d'environ 3'000bp), soit l'existence d'un variant de plus grande taille, contenant ou non des exons codants pour des domaines supplémentaires de 99b78. Il est important de garder en perspective qu'aucune

information sur l'existence d'un transcrit n'était disponible lorsque toutes les premières validations (VN, NB, slot blots) ont été réalisées. La détection d'une bande unique et surexprimée dans les tumeurs (patients 2, 3, 4, 5, 8, 9, 10) a donc constitué une information cruciale.

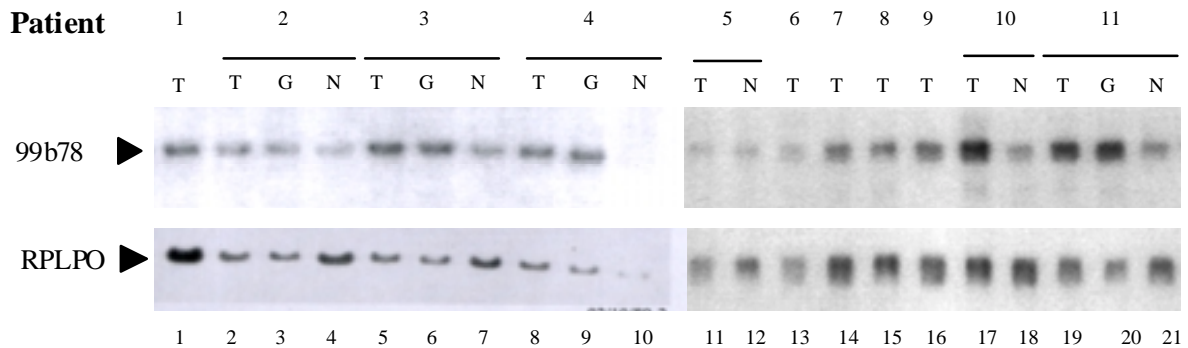


#### Virtual Northern Blot pour 99b78

Une bande d'environ 3kb (a) est détectée comme surexprimée dans les tumeurs indépendamment du type tumoral E(°), S(\*) ou U(°).

Figure VII.2

## 5. Validation en Northern blot



#### Northern Blot pour 99b78

Les tumeurs (T), les ganglions (G) et les tissus normaux (N) des mêmes patients ont été analysés. Une bande unique est détectée comme surexprimée dans les tumeurs. Le contrôle RPLPO est présenté.

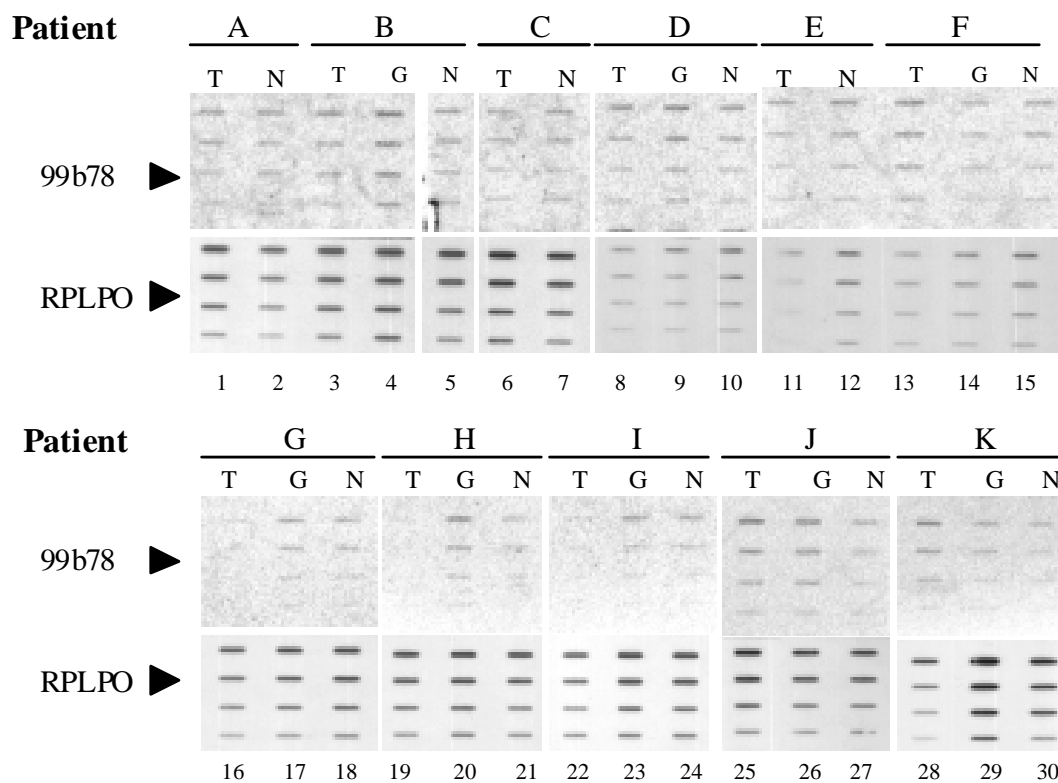
Figure VII.3

En utilisant l'insert DD comme sonde (FigVII.3), une bande unique, identique dans tous les échantillons, a été détectée comme surexprimée dans les tumeurs de 5/6 couples d'échantillons tumoraux et normaux issus d'un même patient (patients 2, 3, 10, 11

en tenant compte du contrôle RPLP0 pour le patient 5). L'expression est maintenue à un niveau au moins égal dans les ganglions lymphatiques envahis (colonnes 3, 6, 9 et 20).

## 6. Validation en Northern Slot blots

L'analyse d'une collection indépendante de patients en Northern slot blot avec la séquence codante clonée (FigVII.4), nous a permis de confirmer, après analyse visuelle, l'expression différentielle de 99b78 dans 6 (voir 7) tumeurs ou ganglions lymphatiques sur 12 couples d'échantillons tumoraux et normaux issus d'un même patient (patients D, F, H, J, K et E compte tenu de RPLP0, l'expression dans le ganglion étant peut être plus fort que dans le tissu normal pour le patient B).



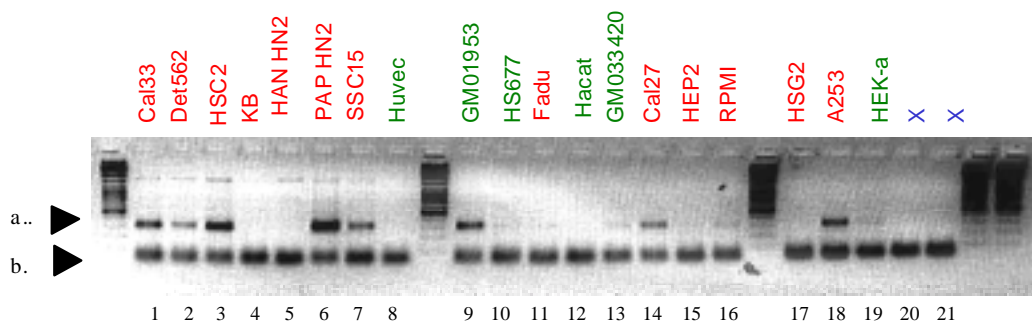
### Northern Slot Blot pour 99b78

Les tumeurs (T), les ganglions (G) et les tissus normaux (N) des mêmes patients ont été analysés. Le contrôle RPLP0 est présenté.

Figure VII.4

L'extension du nombre de patients analysés va être poursuivie au Centre Paul Strauss par Realtime Quantitative PCR sur des séries d'une dizaine de patients dans un premier temps. L'existence de variants potentiels (si les clones IMAGE correspondent bien à des variants d'épissage alternatif) pourrait amener à dessiner des amorces spécifiques de chaque variant, en plus d'amorces au sein de la séquence codante que nous avons cloné.

## 7. Etude de l'expression dans les lignées cellulaires VADS



Analyse de l'expression de 99b78 dans des lignées cellulaires par PCR semi-quantitative

Les amorces utilisées ne sont pas les amorces utilisées pour le clonage de la séquence codante et permettent l'amplification d'un petit fragment. Une bande représentant 99b78 de taille attendue (a) ainsi que des dimères d'amorces peuvent être détectés (b) dans des lignées cellulaires.

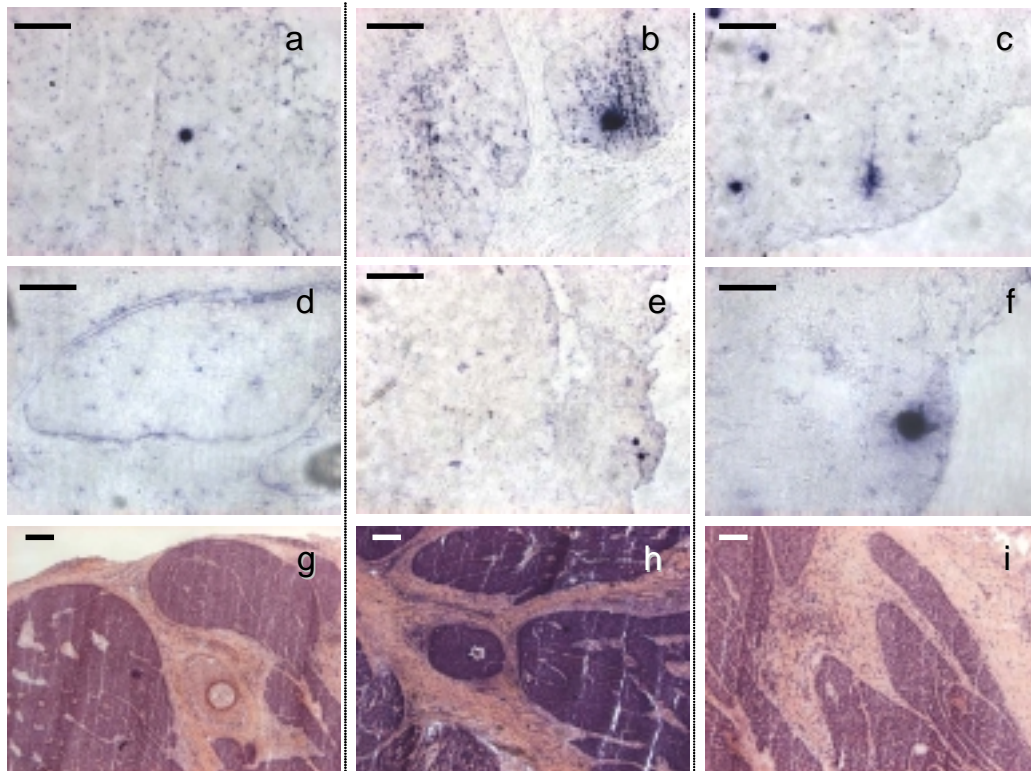
Figure VII.5

L'expression de 99b78 est détectée dans diverses lignées cellulaires VADS (FigVII.5), et, plus particulièrement dans les cellules Det562 (colonne 2). L'expression de 99b78 n'est pas détectée dans les cellules HaCaT (colonne 12), et dans les cellules Fadu (colonne 11). Les cellules RPMI2650 utilisées pour l'analyse par FACS semblent exprimer très faiblement 99b78 (colonne 16). L'effet de la transfection des clones antisens de 99b78 a donc été évalué dans les cellules Det562.

## 8. Etude par Hybridation *In Situ*

De manière à déterminer de façon plus précise le compartiment tumoral responsable de l'expression de 99b78, j'ai effectué une Hybridation *In Situ* sur lames congelées de

tumeurs de l'hypopharynx, après synthèse de sonde à partir de l'insert DD (FigVII.6). Le signal obtenu ne semble pas être spécifique mais correspondre à des agrégations artéfactuelles, puisque les mêmes structures sont observées avec les sondes correspondant tant au sens qu'à l'antisens. Un léger surlignage des bordures des masses tumorales peut être observé avec la sonde SpeI (FigVII.6d), mais la surface concernée est bien inférieure à la dimension d'une cellule. Ce marquage est donc probablement artéfactuel.



#### Hybridation In Situ pour 99b78; tumeur 245T

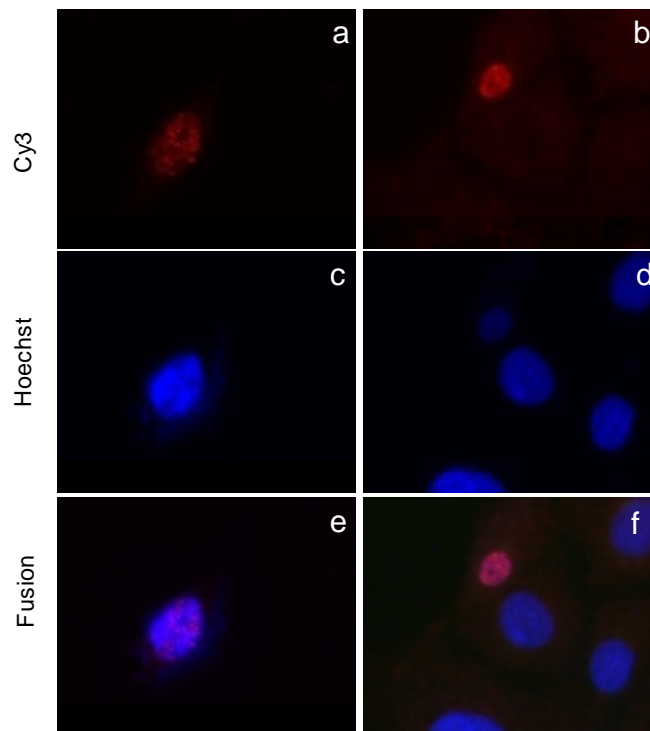
Trois exemples de marquage pour la sonde SpeI/T7 (a, d), sont présentés ainsi que les structures histologiques correspondantes (g, h, i), le marquage pour la sonde NaeI/Sp6 (b, e) et pour la sonde Aat2/SP6 (c, f). Aucun marquage spécifique n'est détecté.

Figure VII.6

### 9. Immunocytochimie pour le clone 99b78-Flag

Bien que je n'ai pas reproduit à de nombreuses reprises l'analyse de la localisation subcellulaire de la version étiquetée par Flag en fusion avec la partie N terminale de 99b78, il semble que cette localisation soit nucléaire (FigVII.7 a et b). Comme présenté plus bas, cette localisation, initialement considérée comme aberrante, est peut être cohérente *a posteriori*, du

fait, d'un potentiel clivage de 99b78, détecté avec le sérum 2026 dont l'épitope est également situé sur la partie N terminale de 99b78.



Immunocytochimie pour 99b78 (31S)-flag; lignée HaCaT  
2 exemples de localisation avec l'anticorps anti-flag (a et b) sont présentés, ainsi que les marquages des noyaux correspondants (c,d) et la fusion de ces images (e,f).

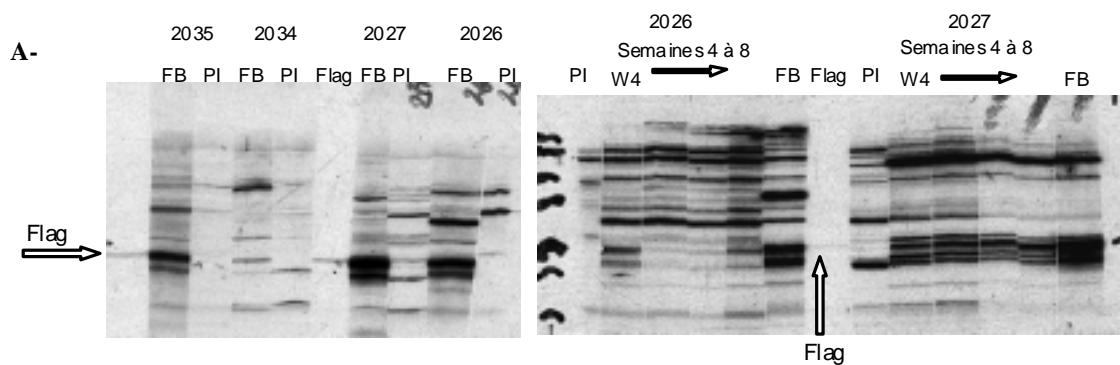
Figure VII.7

## 10. Obtention et caractérisation d'anticorps polyclonaux

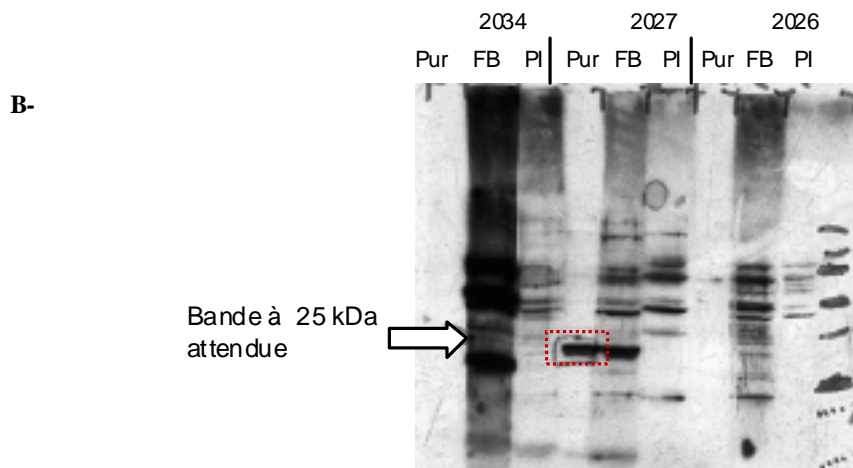
Les anticorps correspondant à 4 peptides ont été testés sur des extraits de cellules COS transfectées avec le clone pSG5-puro-flag 31S, contenant la séquence codante de 99b78 dans le cadre de lecture (FigVII.8A). Les sérums pré immuns, les sérums collectés aux semaines 4 à 8 après injection ainsi que la saignée finale (Final Bleed FB) ont été analysés pour leur capacité à reconnaître une bande à la taille attendue. L'anticorps Anti-flag (signal assez faible dans mes mains) a été utilisé de manière à confirmer la taille de la bande, en plus de l'estimation de la taille par rapport au marqueur. Les sérums permettant la détection d'une bande de taille cohérente avec la taille attendue ont été purifiés. La capacité de ses sérums purifiés à détecter la protéine spécifique endogène a été testée sur des extraits protéiques de

cellules Det562 (FigVII.8B), lignée exprimant 99b78. Une bande de taille cohérente avec la taille attendue (25kDa pour une taille théorique de 23,8kDa et migrant au même niveau que le Flag) a pu être détectée avec les sérums correspondant aux lapins 2026 et 2027.

Il faut noter que les sérums pré immuns détectent des bandes non spécifiques. Ces bandes se sont avérées très pénalisantes pour l'étude de l'expression de 99b78 dans les tumeurs par immunohistochimie (contrôle négatif avec les sérums pré immuns). Certains sérums pourtant prometteurs (2035) ont perdu toute capacité de détection après purification. Seuls les sérums 2026 et 2027 ont conservé des capacités de détection de la protéine endogène (visible sur un autre gel pour 2026).



2027, 2026, 2035: bande détectée vers 25kDa (attendue 23.8 kDa)  
même taille que pour l'anticorps anti-Flag



### Caractérisation et purification de sérums Anti-99b78.

A- Cellules COS transfectées pour 99b78 31S (sens)

B-Protéine endogène dans la lignée Det562

Les sérums pré immuns (PI), final bleed (FB), ainsi que les différentes semaines d'immunisation (W) et le sérum purifié (Pur) sont présentés.

Figure VII.8

Les peptides 2026 et 2027, utilisés pour les analyses en immunocytochimie et en immunohistochimie, correspondent à des épitopes disposés de part et d'autre de la membrane. Ces épitopes sont présents dans toutes les isoformes putatives de 99b78. La prédiction de l'orientation des domaines extramembranaires est étonnante -le domaine d'homologie avec une kinase est indiqué comme potentiellement dirigé vers l'extérieur, extracellulaire-. Il est possible que cette protéine transmembranaire ne soit pas attachée à la membrane plasmique mais plutôt à des membranes comme celles du réticulum endoplasmique ou d'autres organites. Si tel était le cas, le domaine possédant une faible homologie avec une kinase serait orienté vers le cytoplasme de la cellule.

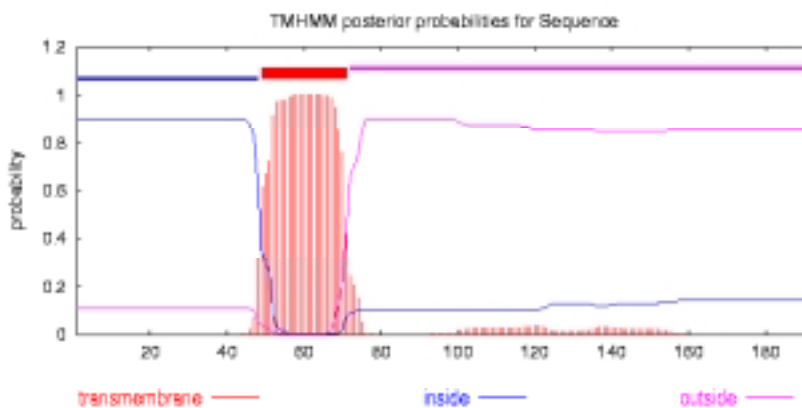
Protéine 99b78 par TIGR ORF 579bp 192 aa /

**inside** **Transmembrane domain**

---

MGITCLKVDMRWPKSHEEWPPGMMMKSSAGGKVVSSHSIQLQNWSKFNHFLRVILKAVGLVLLWLLWLLFQLLQLQNEKDLAVLSS  
 IQWVLDGLPDDVKGSWSWLVLLPFVGRMPPAPALLGAPECASFPSSTSVLNLWHGIGICAFRTSFLPILQDVCCPSTGQDPAGVGNTGNTNP  
RHQTSALGMANWSCVL\*

Peptide PG 183 (Iapins 2026 & 2034)      Peptide PG 184 (Iapins 2027 & 2035)



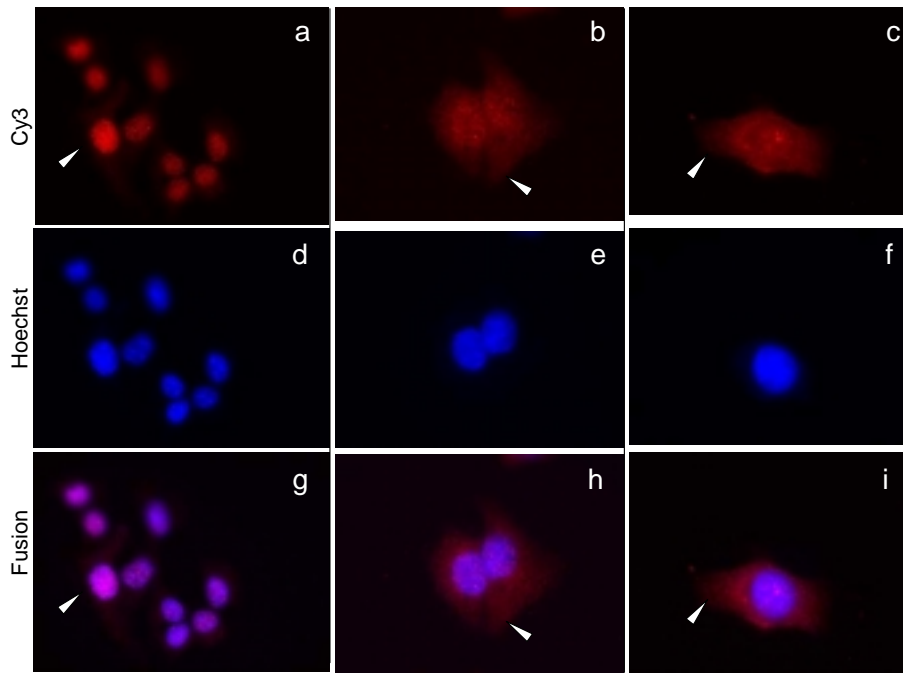
Position des épitopes et structure de 99b78

Figure VII.9

## 11. Immunocytochimie pour le clone 99b78

L'étude par immunocytochimie avec les sérums purifiés nous indique la localisation cellulaire de la protéine endogène, et, nous permet de mieux cerner les fonctions potentielles de la protéine. Les sérums purifiés 2026 et 2027 ont été utilisés à deux dilutions différentes (1/1000 et 1/200).

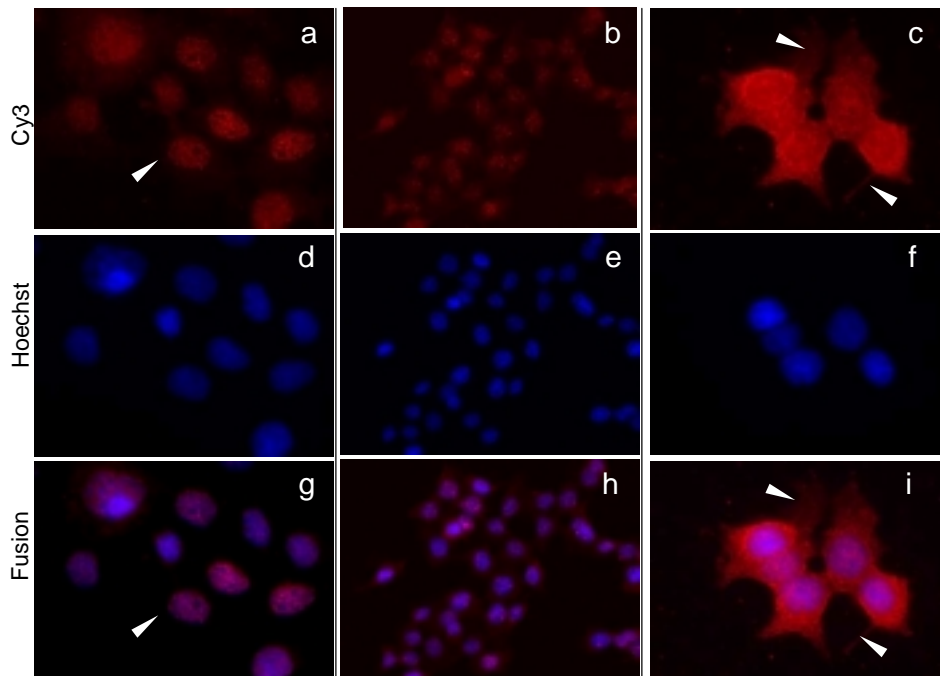




**Immunocytochimie pour 99b78; lignée HaCaT**

Trois exemples de localisation sont montrés avec le sérum 2026 purifié à la dilution 1/1000 (a) et avec le sérum 2027 purifié à la dilution 1/1000 (b) et 1/200 (c) ainsi que les marquages des noyaux correspondants (d,e,f) et la fusion de ces images (g,h,i).

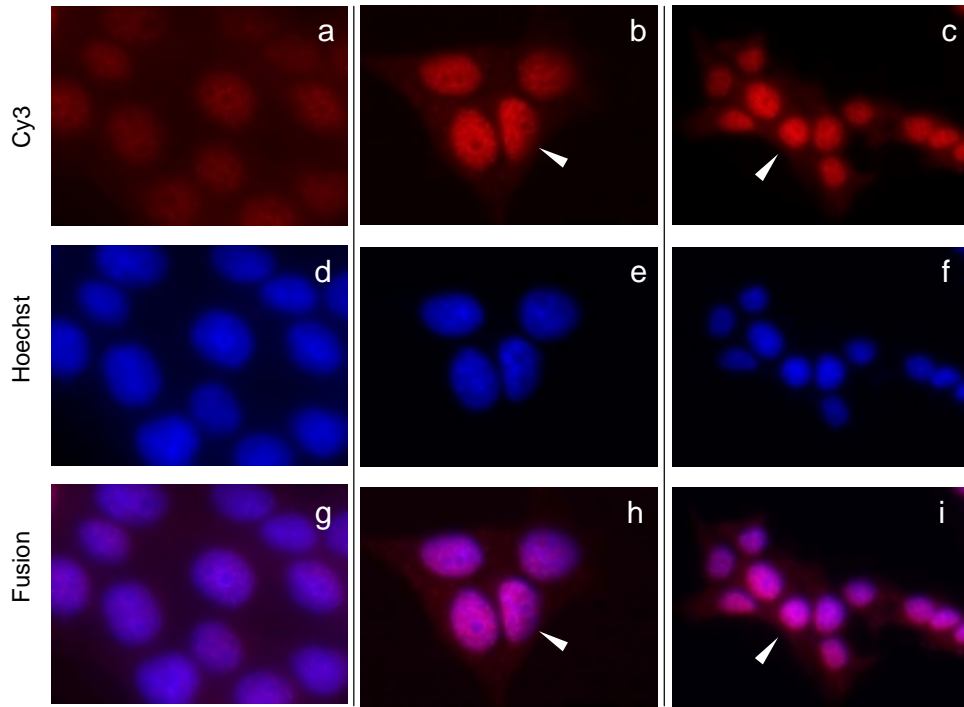
Figure VII.10



**Immunocytochimie pour 99b78; lignée FadU**

Trois exemples de localisation sont montrés avec le sérum 2026 purifié à la dilution 1/200 (a) et avec le sérum 2027 à la dilution 1/1000 (b) et 1/200 (c) ainsi que les marquages des noyaux correspondants (d,e,f) et la fusion de ces images (g,h,i).

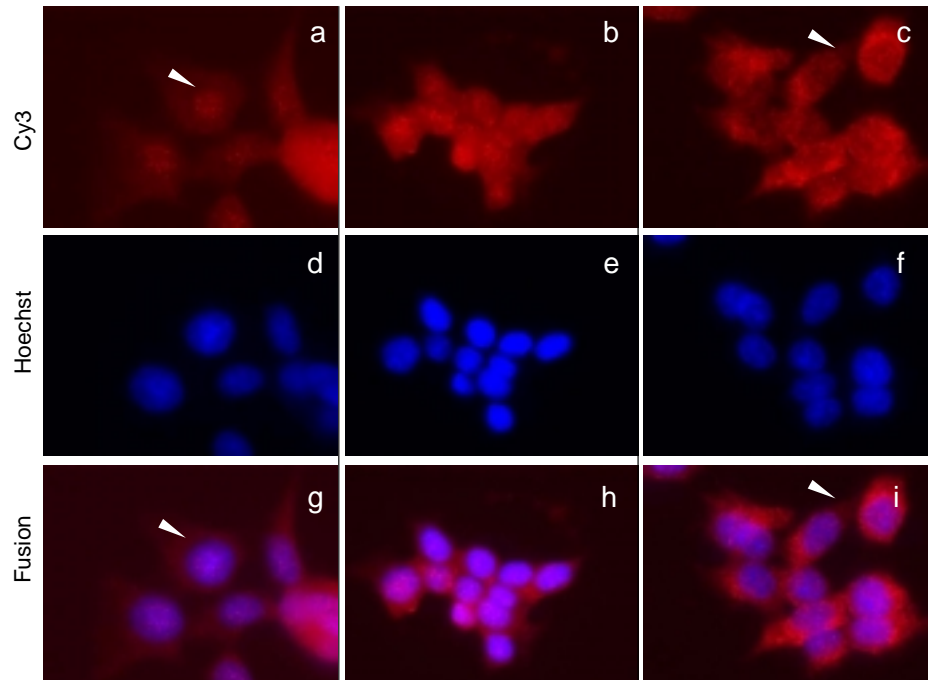
Figure VII.11



**Immunocytochimie pour 99b78; lignée Det562**

Trois exemples de localisation avec le sérum 2026 purifié sont montrés à la dilution 1/1000 (a) et à la dilution 1/200 (b, c) ainsi que les marquages des noyaux correspondants (d,e,f) et la fusion de ces images (g,h,i).

Figure VII.12



**Immunocytochimie pour 99b78; lignée Det562**

Trois exemples de localisation avec le sérum 2027 purifié sont montrés à la dilution 1/1000 (a, b) et à la dilution 1/200 (c) ainsi que les marquages des noyaux correspondants (d,e,f) et la fusion de ces images (g,h,i).

Figure VII.13

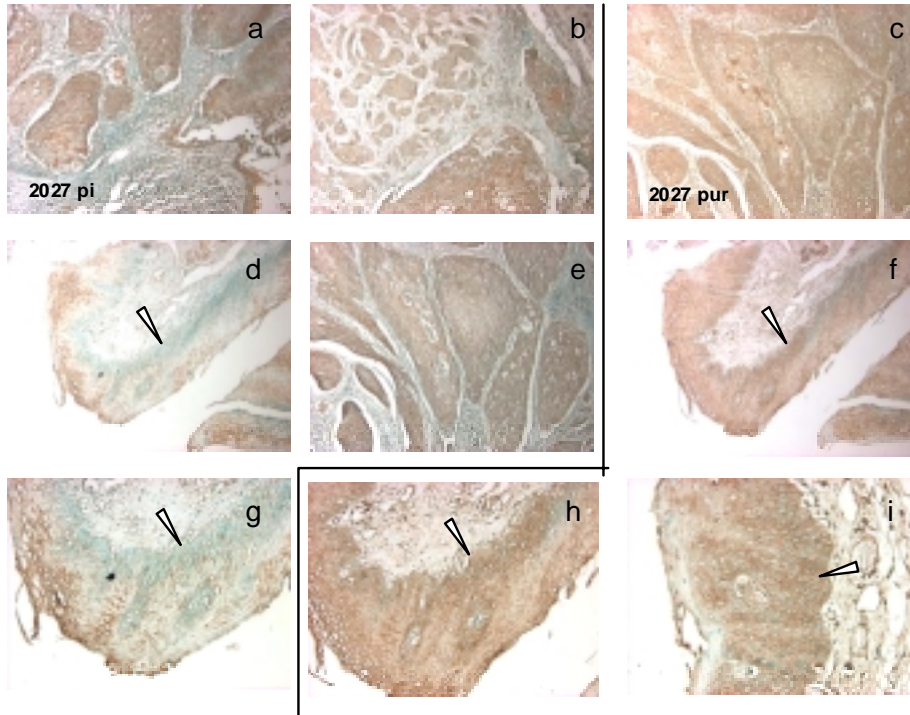
Avec le sérum purifié 2026, une expression presque exclusivement nucléaire est détectée dans les lignées HaCaT (FigVII.10a), Fadu (FigVII.11a) et Det562 (FigVII.12). Un tel marquage nucléaire avait également été repéré avec l'anticorps anti-flag, qui est fusionné à l'extrémité N, dans laquelle se situe l'épitope correspondant au sérum 2026 par rapport au domaine transmembranaire.

Avec le sérum 2027, une localisation partiellement (HaCaT, FigVII.10 b) ou majoritairement nucléaire est détectée à une dilution 1/1000 dans les cellules Det562 (FigVII.13 a et b), et Fadu (FigVII.11b). Lorsque l'anticorps est utilisé à une concentration supérieure de 1/200, la localisation apparaît comme partiellement nucléaire, mais surtout cytoplasmique ou membranaire (FigVII.10c, FigVII.13c et FigVII.12c). Comme attendue, la détection est nette dans les cellules Det562. Des prolongements cytoplasmiques sont très visibles sur certaines cellules Fadu et Det562, ce qui pourrait indiquer une localisation membranaire. Seule une analyse complémentaire en microscopie confocale (ou une co-localisation avec un marqueur membranaire) permettrait d'être catégorique sur la nature membranaire de la localisation de 99b78.

## 12. Immunohistochimie

Les deux sérums 2026 et 2027 ont été utilisés de manière à étudier l'expression de la protéine 99b78 dans les tumeurs, information que nous n'avons pas obtenue au niveau de l'ARNm en hybridation *In Situ*.

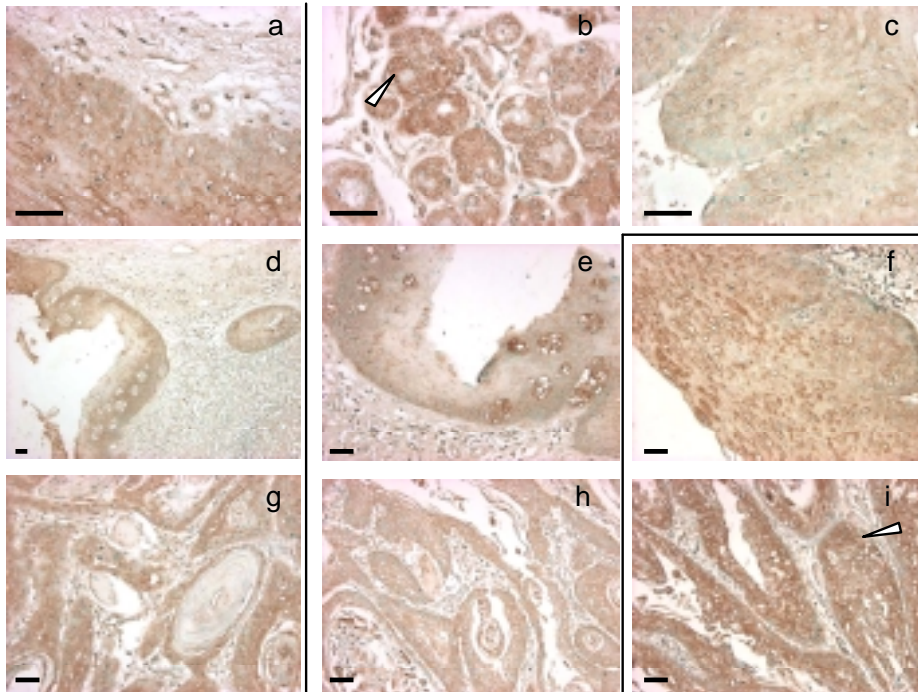
Sur les sections de la tumeur 31680T (FigVII.14), pour laquelle le sérum pré immun ne produit que peu de signal sur la bordure de l'épithélium, il semble qu'un léger marquage des cellules de la couche basale puisse être observée avec le sérum 2027 (FigVII.14 f, h, i). Dans une moindre mesure, un marquage similaire est observable sur la tumeur 25004T (FigVII.16h). Ce type de cellules sont pluripotentes et en prolifération, ce qui est cohérent avec un rôle d'oncogène potentiel de 99b78. Sur les autres sections, la détection d'un signal plus fort que celui détecté avec le sérum pré immun n'est pas évidente.



Immunohistochimie pour 99b78; tumeur 31680T

Des exemples de marquage par le sérum pré immun (a, b, d, e, g), le sérum purifié 2027 (c, f, h, i) sont présentés.

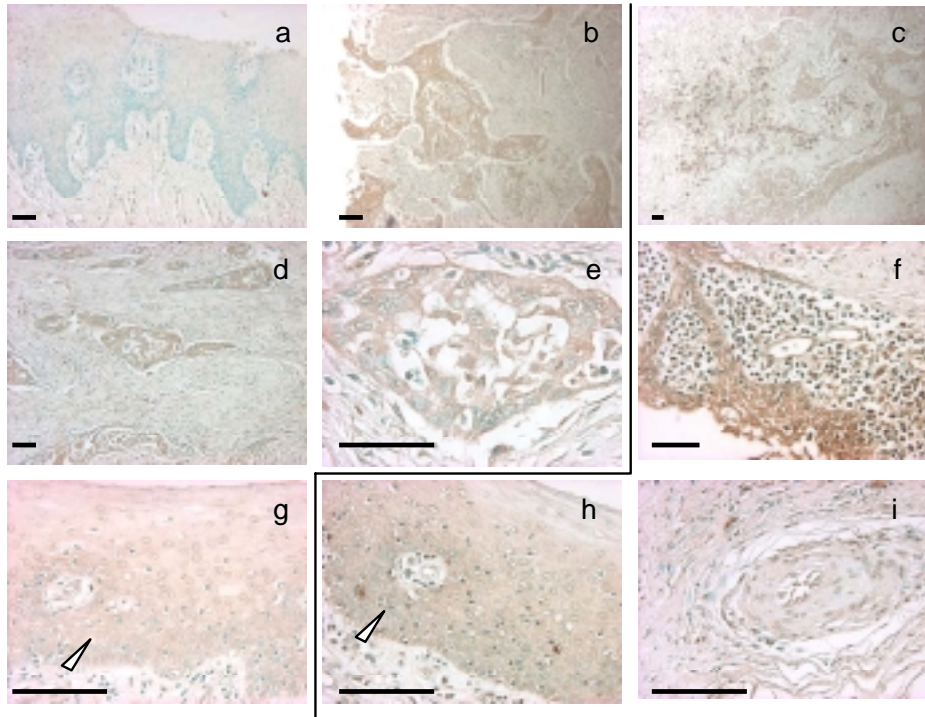
Figure VII.14



Immunohistochimie pour 99b78; tumeur 27847T

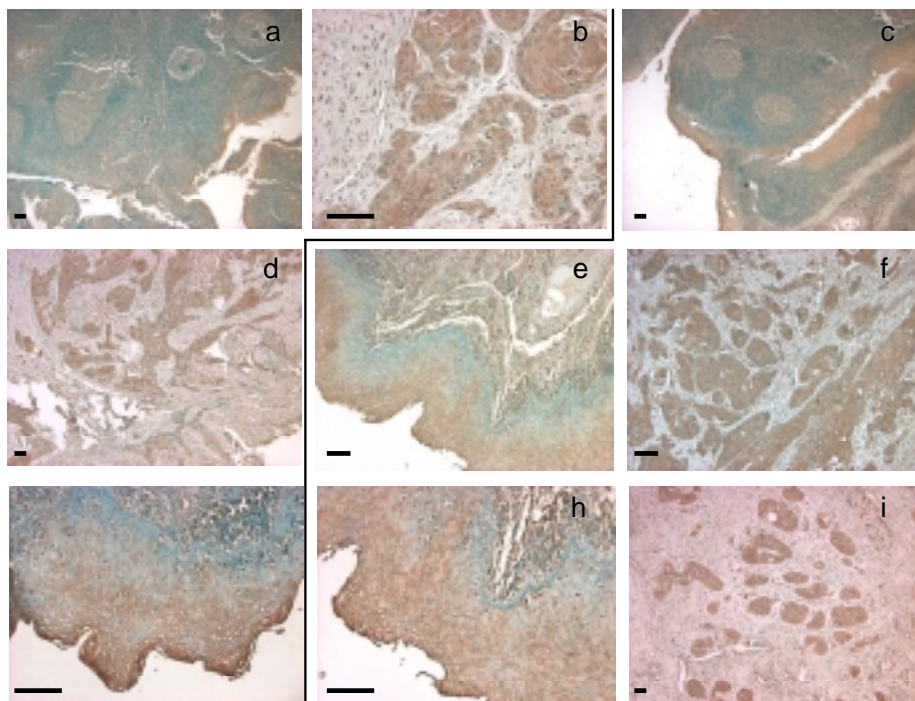
Des exemples de marquage par le sérum pré immun (a, d, g), le sérum purifié 2027 (b, c, e, h) et le FB 2027 (f, i) sont présentés.

Figure VII.15



Immunohistochimie pour 99b78; tumeur 25004T  
Des exemples de marquage par le sérum pré immun (a,b, d, e, g), le sérum purifié 2026 (c, f, h, i) sont présentés.

Figure VII.16



Immunohistochimie pour 99b78; tumeur 25005T  
Des exemples de marquage par le sérum pré immun (a, b, d, g), le sérum purifié 2026 (c, e, f, h, i) sont présentés.

Figure VII.17

Dans le cas des deux sérums 2026 et 2027, les résultats de l'analyse immunohistochimique sont délicats à interpréter car les sérums pré immuns produisent un signal assez fort. Il pourrait être utile de purifier ces sérums de manière analogue aux sérums purifiés, de manière à se éliminer de la détection de bandes n'ayant aucun rapport avec 99b78, liées à l'immunité originelle des lapins injectés avec les peptides immunogènes. Le blocage des lames avec du sérum de chèvre plutôt qu'avec l'albumine sérique bovine ainsi qu'une optimisation de la dilution d'anticorps utilisée et des conditions de démasquage des épitopes, pourrait contribuer à diminuer le bruit de fond.

## 13. Etude fonctionnelle préliminaire

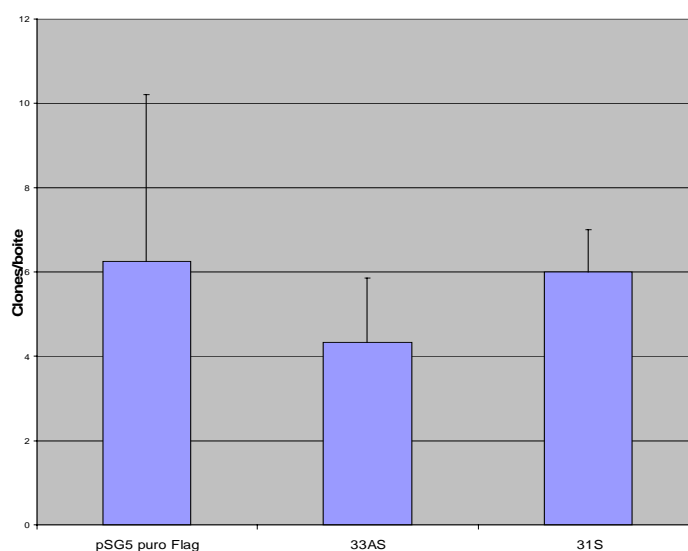
---

### 1. Analyse par FACS

Un effet modeste de diminution de la prolifération a pu être mesuré, de manière peu reproductible, sur un nombre limité d'expérimentations, pour le clone 33AS exprimant l'ARN de 99b78 en orientation antisens. Ce résultat ne constitue qu'une indication, attendu sa faible reproductibilité et les effets presque aussi forts obtenus avec le plasmide contrôle pBSK.

### 2. Test de clonogénicité et clones stables

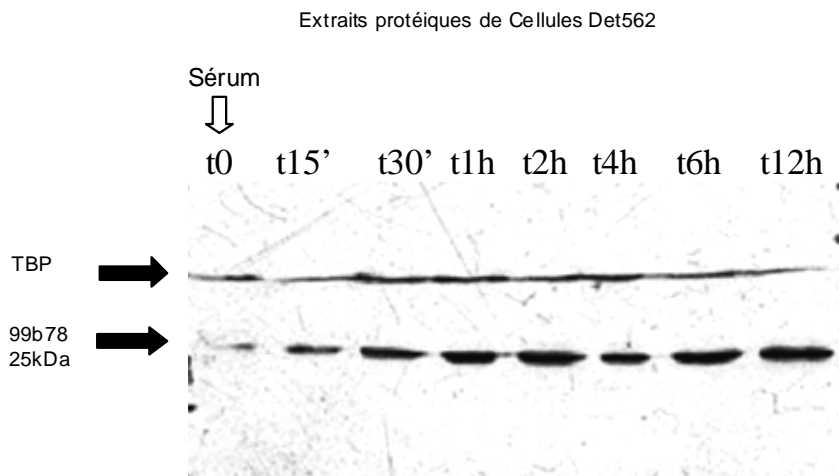
J'ai effectué un test de clonogénicité dans les cellules Det562 (FigVII.18). Aucune différence significative n'a pu être observée entre les clones 31S et 33AS exprimant 99b78 en orientation sens et antisens, et, le contrôle pSG5 puro flag. Toutefois, peu de clones ont été obtenus (méthode JetPEI), avec une assez forte variabilité dans le contrôle. Cette expérience est préliminaire et n'exclue pas un rôle potentiel d'oncogène de 99b78.



Tests de clonogénicité pour 99b78  
Clones sens (31S), Antisens (33AS) et contrôle (pSG5 puro Flag)  
dans la lignée Det562

Figure VII.18

### 3. Etude de la régulation par le sérum



Induction par le sérum dans les cellules Det562

Figure VII.19

De façon très préliminaire, j'ai pu montrer en Western blot une augmentation du niveau de la protéine 99b78, après induction par le sérum dans les cellules Det562 (FigVII.19). Cette augmentation semble débuter dès 30 minutes après induction, et, reste forte jusqu'à 12h après induction. Cette induction indique peut être un rôle de 99b78 dans la prolifération cellulaire. Cependant, cette information fonctionnelle préliminaire demeure vague car de nombreux gènes sont induits par le sérum.

### 4. Etude de l'induction de kinase en aval de la signalisation membranaire

La transfection du clone sens codant pour 99b78 semble être capable d'activer les kinases ERK1/2, comme révélé par une analyse en Western blot avec l'anticorps spécifique des formes phosphorylées actives de ces kinases. Ce résultat a été obtenu en conditions de déplétion du sérum, de manière à ne pas subir d'interférences avec l'induction endogène des kinases par le sérum. L'induction étant faible, ce résultat doit être répété mais il est cohérent avec une fonction de tyrosine kinase membranaire putative de 99b78.



## C. Matériel biologique

Clone 99b78 Résumé	Séquence	Localisation
fragment DD dans pGEM-T EASY (Clontech)	AATCCTACTACCCCAAAGACAAGATGATCATTAGTCATA TAGTATCATACATATATAATATGTACACATATATTTAATGTA AGTGTAGTGTCAAAATCATGTATAGTCTTTCAAAAACAC GTTGGCACTTAATATCTTATATGAACAACCTTCTAGAGCAAC AAGCTTAGTTACAGCAATTGTAATGTAAGAGTACTCATTTA CATAGCTACAGCATCCTTTGTTTACCTCATCTCTTATGGTTG AT	Taille 318bp// clone RP4-705P8 on chromosome 1p33-34.2, Accession Number AL356455, nt (11426-11114) / IMAGE:5178133 , 1387-1704 (317bp, 1 T-A mismatch en 760) / Clone TIGR THC882896, nt (1373 - 1685)// IMAGE:3916394 (650-967) mismatch T-A (948)
Séquence codante (CDS) clonée	ATGGGAATAACATGCTTAAAGGTGGATAIGAGATGGCCAA AGTCCCATGAGGAATGGCCTCCAGGTATGATGATGAAATC CAGTGCAGGAGGGAAGGTCTCTTCTCATAGTATCCAATTG CAAACTGGTCAAAATCAACCATTTTCTCAGGGTTATTCTA AAAGCTGTTGGACTTGTACTGCTTCTTTGGCTTCTCTGGCTC TTGTTCCAGCTTTTGCAGTTACAAAATGAGAAAGACCTCGC TGTGCTGTCC	579 letters/ Clone TIGR THC882896, nt (173-751) // IMAGE:5178133, 769-192 (+/-, 1 mismatch en 545)
Orientation Sens 31S	Sens	
Orientation Antisens 3AS	Antisens	
Protéine correspondante	MGITCLKVDMRWPKSHEEWPPGMMMKSSAGGKVSSHSI QLQNWSKFNHFLRVILKAVGLVLLWLLWLLFQLLQLQNEK DLAVLSSIQWVLDGLPDDVKGSWSWLVLFPVGRMPAPAPA LLGAPECASFPSSSVLNLWHGIGICAFRTSFLPILQDVCPST GQDPAGVGNTGTNPRHQTSALG MANWSCVL	193aa
Peptides injectés		
PG183	SHEEWPPGMMKSSAGGKVSSC	Lapins 2026 & 2034, aa 15 à 34
PG184	CPSTGQDPAGVGNTGTNPRHQ	Lapins 2027 & 2035, aa 158 à 178

Matériel biologique obtenu pour l'étude de 99b78

Figure VII.20

# **Partie VIII :** **Perspectives**

## A. Perspectives les analyses de masse

---

### 1. Perspectives sur l'étude globale en Differential Display

Nous avons réalisé une analyse du profil d'expression des gènes dans les tumeurs VADS avec les techniques du Differential Display (DD), afin principalement d'isoler de nouveaux acteurs inconnus de la carcinogenèse, et, des puces à ADN Affymetrix, afin d'assurer une couverture des gènes connus avec un biais expérimental différent. Une collection d'environ 1'200 gènes uniques (1'400 avant élimination récente de la redondance) a été obtenue en DD, correspondant, pour l'essentiel, à des différences d'expression entre tissu normal et tumoral. Une expression différentielle entre tissu tumoral et normal a été confirmée pour 70 gènes avec des sondes complexes focalisant le marquage sur la population d'ADNc étudiée grâce aux amorces arbitraires de DD. Le taux de clones inconnus identifiés dans cette étude est élevé (Lemaire et al., 2003). Nous avons donc mis en évidence des nouveaux acteurs potentiels de la carcinogenèse. Nous avons confirmé l'expression différentielle d'un nombre significatif de gènes par des techniques adaptées de Reverse Northern blot, Northern blot, Northern slot blot, Virtual Northern blot et PCR quantitative en temps réel (Lemaire et al., 2003). Parmi les gènes connus, l'identification d'acteurs avérés de la carcinogenèse soutient la validité de notre étude. Le caractère pronostique de la collection DD est en cours d'évaluation sur une série de 50 tumeurs issues de patients souffrant de tumeurs de l'hypopharynx.

Il est important de resituer l'utilisation d'un tel criblage en DD dans le contexte industriel sous-jacent. Les résultats obtenus en DD ont pu permettre d'évaluer l'intérêt des clones isolés en DATAS par *ExonHit Therapeutics*. Même si le DD peut isoler des événements d'épissage alternatif, en relation avec les limites d'extension des fragments en PCR et l'éventail des tailles de fragments résolus sur gels de DD, le DD permet principalement d'isoler des différences quantitatives d'expression. La comparaison des résultats obtenus en DD et en DATAS peut permettre de mieux sélectionner les séquences modifiées strictement qualitativement (épissage alternatif), ou, au contraire d'isoler des gènes différentiellement exprimés tant au niveau quantitatif que qualitatif entre les tumeurs et les tissus normaux. D'un point de vue de la propriété intellectuelle et du dépôt éventuel de

brevets, la technique du DD, favorisant l'isolement de séquences encore peu ou pas caractérisées, présente un intérêt indubitable.

Une validation plus vaste des gènes à titre individuel a été menée grâce à la technique du Northern blot, à l'adoption de la technique du Virtual Northern blot, et à l'émergence de la PCR quantitative. Parmi les clones présentant les profils les plus intéressants -robustes et/ou spécifiques du pronostic-, j'ai tenté d'approfondir la caractérisation du profil d'expression de 4 clones favoris (F1.4, B14.4, 99b78 et P5.1). Pour les 3 clones présentant les profils les plus robustes j'ai appliqué la technique d'hybridation *In Situ*, et, tenté d'obtenir des anticorps spécifiques à fin d'analyse en immunohistochimie. J'ai également mené une analyse fonctionnelle préliminaire de l'effet de la transfection de la séquence codante de ces clones en orientation sens et antisens, entre autres sur la formation de clones sélectionnés par la puromycine, et, sur la répartition des cellules dans le cycle cellulaire par la technique FACS.

La mise en évidence dans notre publication d'une amplification de deux régions du génome, contenant un cluster de gènes impliqués dans la différenciation terminale des kératinocytes (EDC en 1q21) et dans l'immunologie (NK complex en 12p13), pourrait être confirmée dans des échantillons hypopharyngés par PCR quantitative sur ADN génomique.

En parallèle à l'étude des tumeurs VADS, les sondes focalisées et la sélection des clones correspondant à ce marquage particulier ont été mis à profit pour l'étude de tumeurs de prostate, pour lesquelles les contraintes en quantité d'ARN sont comparables à celles rencontrées dans notre étude VADS. Après hybridation des sondes à partir d'ARN tumoral et de tissu bénin prostatique, quelques gènes ont été validés comme associés à la carcinogenèse du cancer de la prostate par Chunhua Zhao. Un de ces gènes, inconnu, est en cours d'analyse fonctionnelle par le Dr Alberto Zambrano. De même, la plateforme de DD, initiée dans notre étude, a été mise à profit pour une autre étude de pharmacogénomique, menée par le Dr Senghua Hao, avec des échantillons de tumeurs présentant ou non des réponses significatives à la chimiothérapie combinée 5-FU/cisplatine. Les clones isolés par cette étude ont été rassemblés sur un macro array spécifique (S3) dont la valeur prédictive de la chimiorésistance doit encore être évaluée.

A la différence de l'étude par puces à ADN Affymetrix, des fragments de séquences ont été clonés et peuvent être mis à profit, tant pour la constitution de puces à ADN et de macro-arrays, que pour une utilisation pour des tests fonctionnels de masse. Nous avons déjà

initié une analyse de la valeur pronostique de la collection de marqueurs issus du DD sur macro array constitués de membranes de nylon. Les résultats préliminaires semblent indiquer que, compte tenu de la sensibilité des sondes disponibles à l'heure actuelle, la force du Differential Display -sa capacité à isoler des séquences différentiellement exprimées de faible abondance- constitue sa plus grande limite en terme de validation de masse. L'utilisation de sondes mettant à profit la technologie des oligodendrimères ou les dernières évolutions de sondes linéaires développées par Affymetrix pour ses puces pourrait éventuellement permettre de lever cette limite. Le clonage des séquences complètes ou la constitution d'une puce avec les clones IMAGE correspondant à nos clones pourrait peut être permettre d'augmenter la sensibilité du macro array. En attendant ces améliorations techniques, les clones faiblement exprimés doivent être validés à titre individuel.

Les clones isolés en DD les plus informatifs en terme d'évolution tumorale devraient être réunis sur une membrane, ou, plus vraisemblablement une plateforme de PCR quantitative en 96 puits, aux autres clones possédant une valeur pronostique issus de l'analyse sur puces Affymetrix « VADS » et validés en étude prospective. Une autre approche pourrait consister en une validation individuelle, ou, en masse par PCR quantitative en 96 puits, des séquences codant pour des protéines inconnues, prédites comme membranaires (domaine transmembranaire) ou sécrétées (peptide signal) par bio-informatique parmi la collection DD, à des fins d'application diagnostiques ou thérapeutiques.

Une analyse complémentaire des membranes de nylon haute densité sur un nombre restreint de lignées cellulaires (Fadu, Det562, RPMI2650, HaCaT) et de cellules primaires (HEK-a) pourrait être menée. De manière à mieux caractériser les évènements associés à l'initiation tumorale, une analyse sur des lésions pré malignes VADS, comme les échantillons indiens étudiés par le Dr Rahlan, pourrait être menée. Cette analyse pourrait également permettre d'isoler des altérations spécifiques de l'association alcool/tabac par opposition à la consommation de Béthel.

Après ré-amplification des séquences DD et clonage dans un vecteur d'expression eucaryote, des micro-arrays fonctionnels pourraient être constitués. Des micro-arrays « fonctionnels » ont été décrits (Ziauddin & Sabatini, 2001). Ils permettent, après culture de cellules sur la puce contenant des clones enrobés dans un milieu de transfection, l'expression locale d'un gène donné. L'analyse fonctionnelle de centaines de gènes devient possible en parallèle. Le point critique de l'analyse de ce type de puce fonctionnelle est la détermination du phénotype observé dans les cellules transfectées (induction de l'apoptose, morphologie des

cellules, taille des colonies, expression d'un marqueur, etc... limitée par l'imagination de l'expérimentateur et les contraintes techniques). Ce type d'array pourrait être utilisé, une fois constitué aussi bien avec des lignées cellulaires dérivées de tumeurs VADS que d'autres origines comme de la prostate.

L'ensemble des clones pourrait être ré-amplifié et cloné dans un vecteur d'expression de type rétrovirus, pour analyser l'effet antisens de ces clones et de fragments possédant un effet dominant négatif, et identification de Gene Suppressor Elements {GSE ; (Gudkov & Roninson, 1997; Roninson & Gudkov, 2003)}, éléments régulateurs négatifs d'expression génique. La technique MaRX (Hannon et al., 1999), relativement similaire, pourrait également être utilisée.

## **2. Perspectives sur l'étude globale par puces à ADN Affymetrix**

Le mérite de l'analyse Affymetrix est d'avoir pu isoler une signature permettant de distinguer les tumeurs pourvues ou non d'un potentiel métastatique (Cromer et al., 2003). La robustesse de cette signature doit encore être confirmée par une étude prospective sur un nombre important de patients. Les meilleurs marqueurs pronostiques devront être rassemblés sur une membrane pronostique ou étudiés par PCR quantitative en plaques 96 puits. Une signature « métastatique » apparaît à partir des tumeurs primaires et non d'échantillons métastatiques : elle pourrait permettre la détection des patients à risque dès leur traitement initial par chirurgie. Comme pour l'étude DD, les meilleurs marqueurs pronostiques doivent encore être évalués à titre individuel ou en combinaison en nombre limité.

Les données issues d'études sur le même support étant relativement comparables, l'ensemble des résultats obtenus sur les tumeurs hypopharyngées analysées pourraient être mis en ligne sur un site dédié afin de participer à la caractérisation des tumeurs d'origine variée.

Comme évoqué dans la partie I, les techniques de puces à ADN ne constituent pas une panacée. Encore récemment, des auteurs (Tan et al., 2003) ont constaté de vastes divergences lors de l'analyse multiple d'échantillons identiques avec les plateformes Affymetrix, Amersham (oligonucléotides) et Agilent (ADNc). En fait, ces auteurs concluent que les différences identifiées sont dépendantes de la plateforme utilisée. Une partie des discordances observées pourrait correspondre, outre des biais intrinsèques différents, à la détection de

variants d'épissage alternatif différents et à un traitement des données brutes différents (algorithmes). Malgré leurs limites, ces technologies sont particulièrement adaptées à la définition d'un groupe de gènes sur lesquels les efforts d'investigation peuvent être concentrés.

### **3. Perspectives technologiques et applications diagnostiques de nos études**

Tout d'abord, d'un point de vue technologique, il apparaît clairement que si l'étude DD devait être initiée aujourd'hui, l'accès de plus en plus aisée à la technologie Affymetrix nous pousserait à choisir cette plateforme. Cette technologie permet, entre autres, d'obtenir rapidement des résultats par rapport au DD, qui nécessite de nombreux clonages et des expériences de validation secondaires laborieuses. Il n'en demeure pas moins que le DD possède encore, pour un temps de plus en plus restreint, la plus grande capacité à isoler des événements demeurés inconnus. Notre collection de marqueurs potentiels de la carcinogenèse serait moins exhaustive sans la contribution du DD, en particulier pour les gènes faiblement exprimés. Il semble, d'après les résultats de Tan et al (Tan et al., 2003), qu'aucune plateforme commerciale de puces à ADN ne peut assurer une analyse exhaustive du fait de ses biais propres. Une stratégie visant à une exhaustivité maximale pourrait consister en une analyse simultanée sur plusieurs plateformes commerciales. Par rapport à notre étude, la plupart des études en DD sont réalisées sur un champ plus limité, avec quelques amorces arbitraires, de manière à isoler rapidement quelques gènes inconnus dont la fonction est ensuite étudiée. Remarquons que si les étapes du protocole de DD sont réalisées assez rapidement, les étapes de validation secondaire deviennent extrêmement longues et laborieuses quand le DD est entrepris à grande échelle. La plupart des études sont réalisées dans des lignées cellulaires plutôt que sur des biopsies tumorales, de façon à faciliter les étapes de validation secondaire. De plus, l'étude de la variation de l'expression de gènes en réponse à un stimulus précis dans une lignée cellulaire, permet de retourner plus facilement à la fonction des gènes isolés.

Après combinaison des marqueurs pronostiques issus de DD et de l'étude par puces Affymetrix, une étude prospective doit être réalisée sur un nombre étendu de patients. Un transfert technologique au niveau clinique doit être envisagé. Pour cela, une membrane prédictive de l'évolution tumorale pourrait être constituée. La stratégie méthodologique pourrait consister à réaliser un macro-array sur membrane de nylon, de taille restreinte et

utilisable en laboratoire hospitalier, à partir des séquences les plus spécifiques de chaque transcrit (hors des domaines conservés de familles de gènes). Le marquage des sondes pourrait être réalisé de manière analogue aux membranes Atlas, avec un mélange d'amorces correspondant à la collection de gènes analysée. Une autre option pourrait être la fabrication d'une puce Affymetrix à façon (« customisée »). Cependant ce type de puce reste très onéreux. Dernière option, la PCR quantitative à grande échelle (96 puits) pourrait être mise à profit. Une difficulté particulière apparaît lors du développement de tests diagnostiques : les tissus normaux ne sont généralement pas disponibles (pas de ratio T/N, ni de normalisation par rapport au tissu normal), et, il faut donc analyser le niveau absolu d'expression du marqueur pronostique dans la biopsie tumorale -ce qui suppose éventuellement le choix d'un marqueur invariant pour la normalisation- par rapport à un seuil d'expression « pathologique ».

La caractérisation de marqueurs pronostiques circulants, détectables dans des échantillons sanguins (ou urinaires pour des cancers comme ceux de la vessie), est un but fortement recherché à l'heure actuelle. En cas de très forte surexpression dans les tumeurs, comme une différence de type absence/présence, la détection de protéines sécrétées, ou même de cellules cancéreuses circulantes, (PCR quantitative sur ARN, technique très sensible) pourrait être envisagée malgré la dilution dans le flux sanguin. Une focalisation sur les protéines sécrétées, marqueurs humoraux potentiels de la carcinogenèse, pourrait être menée après tri bioinformatique des protéines potentiellement sécrétées.

Si une signature du potentiel métastatique a peut être été déterminée dans l'étude Affymetrix -l'étude prospective restant à mener- il serait également important de caractériser une signature du potentiel de récurrence locorégionale (SPT), particulièrement fréquentes dans les cancers VADS soumis à la cancérisation par champs. La valeur prédictive de cet outil diagnostique pourrait être évaluée sur d'autres types de cancers, en particulier des carcinomes. Enfin, l'étude fonctionnelle de quelques gènes prédictifs de l'évolution tumorale pourrait devenir une priorité, si un ou plusieurs gènes s'avéraient être des marqueurs robustes du pronostique.

L'ensemble de notre stratégie expérimentale a été centré sur l'analyse du transcriptome, en lien avec la disponibilité de techniques d'analyses sensibles et globales. Le développement plus récent des techniques d'analyse protéomique pourrait justifier une analyse complémentaire au niveau du protéome. Une stratégie risquée –peu de marqueurs étant sans doute isolés en lien avec des phénomènes de dilution- mais proche du



développement d'outils diagnostiques utilisables au niveau clinique, pourrait être une analyse protéomique d'échantillons sanguins issus de patients atteints de tumeurs VADS bien caractérisées. La mise au point d'un test de type ELISA (Enzyme-Linked Immunosorbent Assay) pourrait être entreprise dans la foulée.

## B. Perspectives sur les applications thérapeutiques potentielles : gènes cibles favoris

---

### 1. Perspectives sur le gène HOP, gène suppresseur de tumeurs potentiel

HOP ayant été bien caractérisé fonctionnellement dans le développement embryonnaire cardiaque (boucle de régulation négative de Nkx2.5 via interaction avec le SRF), et, ayant été mis en relation avec une fonction de gène suppresseur de tumeurs dans le choriocarcinome, l'étude de ce clone ne va vraisemblablement pas être poursuivie beaucoup plus avant. En effet, l'analyse de gènes inconnus ou non publiés est favorisée au laboratoire, de manière à maximiser l'impact de nos publications.

En tant que communication courte, la fonction de HOP ne sera reliée à la différenciation que par les données existantes dans le développement cardiaque et le profil d'expression en Hybridation *In Situ*. Si cette communication courte devait être étoffée et soumise en tant que publication classique, une expérimentation de différenciation des cellules HaCaT par le CaCl<sub>2</sub>, avec analyse des variations éventuelles de l'expression des kératines 1, 10 et éventuellement 14, resterait à compléter.

La protéine HOP est un gène suppresseur de tumeurs dont la taille exceptionnellement réduite pourrait refléter une économie d'énergie sélectionnée au cours de l'évolution. Le gène étant de petite taille, il constitue une cible également réduite pour les mutations. Les hélices 3/4 sont conservées bien que ne possédant pas de fonction dans la carcinogenèse et dans le développement cardiaque. Nous avons démontré une expression de HOP dans divers tissus du cerveau (multiple tissue array ; non présenté). Il est possible que ces hélices conservées soient impliquées dans d'autres procédés, comme dans une fonction au niveau du cerveau. Une fonction éventuelle dans le cerveau ou dans les tumeurs du système sympathique périphérique tels les neuroblastomes (démonstration éventuelle d'une expression différentielle) pourrait être étudiée.

Si l'étude devait être poursuivie, une tentative de clonage de la séquence codante de l'isoforme B de HOP/mOB1 devra être entreprise.

La caractérisation fonctionnelle pourrait mettre à profit l'existence de clones stables, en combinaison avec la technique ARN interférence (RNAi). Si des tests de clonogénicité ont

été entrepris, ils pourraient être répétés avec le vecteur pCMV. Des tests de migration (Boyden chamber) et de croissance en absence d'adhésion au support (soft agar growth assay) pourraient être réalisés après transfection de HOP en orientation sens dans les lignées tumorales, ou, d'un RNAi pour HOP dans la lignée HaCaT. En cas de différences significatives de comportement (test de clonogénicité, croissance en soft agar, migration en chambre de Boyden, temps de doublement), l'analyse de la tumorigénicité pourrait être poursuivie par injection dans des souris nude.

Des souris inactivées pour le gène HOP ont été produites et ont été étudiées pour leur phénotype au niveau du développement cardiaque embryonnaire. Aucune formation accrue de tumeurs n'a été rapportée. Ceci peut être lié à deux causes : les souris n'ont pas été analysé pour un tel phénotype et n'ont pas été observées intensivement à un âge suffisamment avancé, ou, la formation de tumeurs nécessite une initiation par exposition à des substances cancérigènes. L'étude de l'effet de l'exposition à des applications de 7,12-diméthylbenz-[alpha]-anthracène (DMBA) / 12-O-tétradécanoylphorbol-13-acétate (TPA), sur la peau ou ingérées dans la nourriture pourrait être étudié chez ces souris, en relation avec la formation de carcinomes au niveau de la peau et des muqueuses VADS. Le modèle « Pouch model » d'introduction des substances carcinogènes (introduction par incision cutanée d'un réservoir diffusant lentement ces substances dans la joue de souris) pourrait aussi être utilisé dans ce contexte. Des modèles de souris présentant spontanément des tumeurs, comme les souris p53 -/-, pourraient être croisées avec les souris HOP -/- de manière à évaluer une accentuation potentielle du phénotype chez les doubles mutants.

Quelques gènes cibles de HOP (indirects via interaction avec le SRF) ont été décrits par analyse sur micro-array au niveau cardiaque (myocarde). Des gènes cibles supplémentaires pourraient être isolés par comparaison de l'expression génique dans les muqueuses VADS de souris déficientes pour HOP à l'âge adulte. L'expression des gènes dans les tumeurs formées dans des souris, à différents stades de progression (y compris l'épithélium hypopharyngé normal) pourrait être comparée entre souris sauvages et mutantes.

Nous avons initialement fait l'hypothèse d'une interaction entre HOP (alors décrit comme SMAP31, très homologue de l'homéodomaine de Pax6) et le facteur de transcription à potentiel oncogénique Pax6 pour expliquer le rôle de gène suppresseur de tumeurs de HOP. Une interaction physique pourrait être analysée par des études de co-immunoprécipitation, et, les effets de la surexpression de HOP sur l'expression d'un gène rapporteur régulé par Pax6 pourraient être mesurés.

D'après les données sur le développement cardiaque, la protéine SRF interagit physiquement avec HOP (Chen et al., 2002; Shin et al., 2002). Les facteurs TCF (Ternary Complex Factors) sont des régulateurs de la fonction du SRF. Une interaction fonctionnelle entre HOP et les TCF pour la régulation des gènes induits par le SRF pourrait être investiguée, mettant à profit l'expertise du laboratoire concernant cette famille de protéines.

Une étude par criblage en double hybride pourrait permettre d'identifier exhaustivement les protéines interagissant avec HOP. Le même but pourrait être atteint par fixation de la protéine HOP purifiée sur un support d'analyse en spectrométrie de masse de type SELDI-TOF (Surface enhanced laser desorption/ionization time-of-flight mass spectrometry), et, mise en contact avec les protéines isolées d'épithélium normal VADS (issu d'opérations d'ablation de la lèvre par exemple) ou de kératinocytes primaires en culture.

D'un point de vue thérapeutique, à long terme, HOP étant exprimé à un niveau relativement élevé dans les cellules épithéliales normales (détection en NB et en ISH techniques relativement peu sensibles), la surexpression de HOP pourrait être bien tolérée par les cellules normales, tandis qu'elle pourrait préférentiellement inhiber la prolifération des cellules cancéreuses, comme indiqué par Asanoma dans le cadre du choriocarcinome (Asanoma et al., 2003): une inhibition totale de la formation de tumeurs par injection sous cutanée de clones stables exprimant HOP dans les souris nude a en effet été démontrée, les rares tumeurs se formant ayant éliminé le transgène. Il demeure à démontrer la capacité de HOP à induire la différenciation des cellules cancéreuses (le temps de doublement est allongé pour les clones stables exprimant HOP dans des lignées de choriocarcinomes), voir leur apoptose. Dans cette perspective, la courte taille de HOP contribuant peut être à une production efficace de la protéine, il serait envisageable de mettre en œuvre une thérapie génique des tumeurs VADS, tumeurs de mauvais pronostic, facilement accessible, et, ne nécessitant pas nécessairement de ciblage des cellules transfectées (traitement locorégional rendant l'utilisation de vecteurs de type liposome).

Si une analyse par PCR quantitative au niveau de l'ADN génomique indiquait une perte d'expression sans perte allélique (LOH) de HOP, une étude de la régulation du promoteur de HOP pourrait être entreprise dans l'espoir d'identifier des régulateurs positifs de son expression. Il serait alors intéressant de tester l'effet thérapeutique éventuel d'une ré-induction de l'expression de HOP.

## 2. Perspectives sur les oncogènes potentiels

Les oncogènes putatifs isolés dans notre étude doivent être analysés plus profondément avant d'être développés comme cibles thérapeutiques potentielles. En particulier, leur profil d'expression, en particulier leur éventuelle valeur pronostique, doit être validée sur un nombre étendu de patients par la technique de PCR quantitative en temps réel. Outre la valeur pronostique, la fréquence d'altération de ces gènes est critique dans l'optique d'applications thérapeutiques au niveau clinique.

La caractérisation de l'expression devra être poursuivie en optimisant les conditions d'analyse par immunohistochimie. Diverses dilutions des sérums, le blocage des interactions Antigène/Anticorps par le sérum de chèvre (heat inactivated goat serum) plutôt que la BSA, les conditions de démasquage des épitopes pourraient être testées. Certains sérums pré immuns produisant un bruit de fond important, en lien avec une immunité intrinsèquement active et inexplicée des lapins sans lien avec la problématique, ils pourraient être purifiés de manière identique aux sérums purifiés. Si le gène s'avérait indicatifs du pronostique, l'amélioration des anticorps pourrait permettre une tentative de détection de ces protéines en tant que marqueur circulant, et, permettre la constitution de tests de type ELISA. Dans ce cadre, la production d'anticorps monoclonaux pourrait devenir critique.

Divers modèles de surexpression doivent être mis en œuvre afin d'assurer l'analyse fonctionnelle de ces gènes.

L'analyse du cycle cellulaire par FACS pourrait être répétée avec les variants nouvellement caractérisés des gènes d'intérêt, après re-clonage. Un clonage dans les vecteurs pCMV et pCDNA3 pourrait permettre une surexpression plus marquée et des effets plus nets. L'étude du comportement de clones stables surexprimant les gènes d'intérêt, leur antisens ou un ARNi (vecteur pSUPER) pourrait être analysée par une batterie de tests fonctionnels (tests de clonogénicité, croissance en soft agar, chambre de Boyden, mesure de l'apoptose par la technique TUNNEL, etc...), de manière à confirmer leur potentiel oncogénique et mieux cerner leur fonction.

La tumorigénicité des clones stables pourrait être analysée dans la lignée HaCaT, non tumorigène mais capable de former des tumeurs (injections sous-cutanées) dans les souris

nude après transformation par H-Ras activé. A l'inverse, l'effet de l'antisens/ du ARNi de ces gènes pourrait être analysé dans des lignées tumorigéniques, comme Fadu et Det562.

Sur un plus long terme, la fonction oncogénique de ces protéines devra être analysée dans des modèles murins formant spontanément ou non de tumeurs.

Une série de constructions générées par le Dr Ganguli permettant l'expression d'un transgène sous le contrôle du promoteur de la kératine 14 K14 dans les couches basales de la peau (Ganguli et al., 2000) ainsi que, de manière plus générale, dans les cellules basales des épithéliums à un moindre niveau d'expression, pourraient être utilisées chez la souris. Une tumorigenèse éventuellement induite par surexpression dans les cellules de la couche basale de la peau pourrait être analysée. L'apparition de tumeurs pourrait être facilitée par traitement au DMBA/TPA ou par utilisation de souris présentant spontanément des tumeurs comme les souris p53 -/-.

Le promoteur de l'involucrine permet une expression dans les cellules suprabasales des épithéliums squameux stratifiés (Carroll et al., 1993). Ces cellules étant déjà en cours de différenciation, il est probable que la surexpression de nos oncogènes putatifs (FLJ10261 et 99b78) sous le contrôle de ce promoteur ne permette pas une formation accrue de tumeurs, sauf si leur effet est assez puissant pour inverser le processus de différenciation. Outre le promoteur de K14, le promoteur ED-L2 du virus Epstein-Barr est capable de diriger l'expression d'un transgène dans les épithéliums squameux stratifiés de la cavité buccale et de l'oesophage. Ce modèle a été validé par expression de la cycline D1 dans des souris déficientes pour p53 pour la formation de carcinomes invasifs, tandis que les souris sauvages développent des hyperplasies, bénignes (Opitz et al., 2002). Ce modèle pourrait nous permettre d'analyser la fonction de nos gènes surexprimés dans les tumeurs.

Pour les 2 oncogènes putatifs présentés dans cette thèse, possédant tous deux des domaines transmembranaires, les ligands ou les substrats potentiels devront être déterminés. Une série d'une dizaine de gènes rapporteurs spécifiques de voies classiques de signalisation pourrait être mise à profit afin de déterminer les voies de signalisation situées en aval de ces récepteurs potentiels.

Si la nature d'oncogène et la fonction de nos oncogènes putatifs pouvaient être ainsi validés, l'identification de composés chimiques capables d'inhiber leurs capacités tumorigéniques, par criblage de bibliothèques de drogues, pourrait être entreprise, de façon à envisager des applications thérapeutiques. Les composés pourraient être testés, dans les modèles murins de surexpression de nos cibles, pour leur capacité à inhiber la formation de

tumeurs, puis dans des modèles murins sans surexpression formant des tumeurs (souris p53<sup>-/-</sup>, souris sauvages traitées au DMBA/TPA, « pouch model »).

*ExonHit Therapeutics S.A.* pourrait mettre en œuvre son œuvre son outil prédictif *Safe-Hit*, de manière à sélectionner les composés actifs présentant la toxicité la plus réduite, avant la mise en place d'essais cliniques chez l'homme.

### **3. Perspectives sur le gène FLJ10261**

La caractérisation de l'expression de FLJ10261 dans les tumeurs VADS, et, en particulier, de sa nature pronostique éventuelle, est actuellement poursuivie au Centre Paul Strauss par analyse PCR quantitative en temps réel sur un nombre plus étendu de patients. Des amorces correspondant aux variants 7 et 8 transmembranaires ont été dessinées et sont en cours d'analyse de manière à confirmer une expression différentielle, pour l'une et/ou l'autre de ces formes. Les premiers résultats confirment une très forte surexpression de FLJ10261 (ratio T/N d'environ 10) dans une proportion importante de tumeurs (environ 80% des patients). Le locus 11q13 étant fréquemment amplifié dans les tumeurs VADS, comme démontré par l'étude de l'amplification de la cycline D1 (Muller et al., 1994; Muller et al., 1997), l'amplification du gène FLJ10261 va être analysée sur l'ADN génomique de patients surexprimant ce gène. La valeur pronostique de FLJ10261, éventuellement en combinaison avec la cycline D1, va être évaluée. L'obtention d'anticorps spécifiques de la version 8 transmembranaire pourrait être entreprise, une fois son expression différentielle tumorale confirmée.

La caractérisation des anticorps devra être poursuivie. Les sérums pré immuns produisant un fort bruit de fond (présence de bandes avant immunisation des lapins), ils vont être purifiés de manière strictement identique aux sérums purifiés, de façon à mieux pouvoir gagner en spécificité en immunohistochimie. La répétition de ces expériences, avec un sérum pré immun plus faible, pourrait permettre de tirer des conclusions plus claires sur la localisation des cellules exprimant FLJ10261 dans les tumeurs. La découverte récente d'une version longue 8 transmembranaire de FLJ10261 nous amène à reconsidérer l'exclusion éventuelle de sérums détectant une bande « trop » grande sur des extraits protéiques endogènes. Il est possible que certains sérums aient été exclus bien qu'ils détectent le produit de grande taille 8 transmembranaire.

La forme 8 –transmembranaire de la protéine FLJ10261 pourrait posséder des propriétés régulatrices (régulation indirecte *a priori*) du cycle cellulaire ou de l'apoptose que la version 7 transmembranaire ne posséderait pas. Elle devra être clonée et être analysée au niveau fonctionnel (FACS et autres tests fonctionnels).

La protéine FLJ10261 étant potentiellement localisé dans des « speckles » nucléaires des cellules HaCaT, il pourrait être utile de tenter de co-localiser FLJ10261 avec un facteur d'épissage précis par immunocytochimie avec le sérum 2069 et un série d'anticorps spécifiques de différents facteurs d'épissage. Le premier candidat pourrait être le facteur d'épissage SC35, par analogie avec la protéine des jonctions serrées ZO-2. Si une co-localisation pouvait être observée avec ce facteur d'épissage, il serait intéressant d'évaluer un rôle éventuel de FLJ10261 dans les jonctions serrées et dans la perméabilité cellulaire.

La construction de vecteurs codant pour des protéines de fusion avec la GFP à l'extrémité 3' de la séquence codante (localisation du peptide 2069) pourrait permettre de suivre, dans les clones stables correspondants ou dans des cellules transfectées transitoirement, le clivage et la relocalisation éventuelle de FLJ10261 en réponse à des stimuli comme la densité cellulaire (par analogie à ZO-2), les UV ou d'autres mutagènes, l'induction par le sérum, des inducteurs de p53 comme l'anisomycine, des facteurs de croissance comme EGF ou encore des drogues de chimiothérapie (sélection d'une présence nucléaire ou membranaire).

Une fois le profil d'expression, l'éventuelle valeur pronostique confirmée et des effets fonctionnels démontrés, il sera particulièrement intéressant, mais aussi difficile, d'identifier les ligands ou les substrats de ce récepteur/transporteur transmembranaire. Pour cela, la technique double hybride dans la levure pourrait être mise à profit dans l'hypothèse où le ligand/substrat de FLJ10261 serait une protéine.

L'immobilisation de FLJ10261 sur un support et un traitement par la méthode de protéomique SELDI-TOF récemment installée à l'IGBMC, pourrait permettre d'identifier le ligand/substrat.

#### **4. Perspectives sur le clone 99b78**

Du fait de l'évolution des bases de données, l'expression différentielle des diverses isoformes potentielles de 99b78 devra être confirmée sur un plus grand nombre de patients par Realtime quantitative PCR. En cas d'expression différentielle d'isoformes non clonées, le



clonage de la séquence codante devra être entrepris. Pour s'assurer d'isoler toutes les isoformes possibles, une expérience de 5' RACE pourrait être menée.

Après purification des sérums pré immuns, l'expérience devra être répétée avec potentiellement un bruit de fond plus faible.

Une localisation nucléaire pouvant être observée pour ce clone, tant avec une version étiquetée qu'avec la protéine endogène (en admettant la spécificité du sérum 2026), l'étude de la régulation de la localisation de ce clone en réponse à divers stimuli pourrait être effectuée dans des clones stables exprimant une protéine de fusion avec la GFP (fusion en 5' de la séquence codante de 99b78). L'induction d'un clivage par des stress cellulaires pourrait être étudié. Une tentative d'identification de la protéase impliquée dans ce clivage pourrait être menée par analyse *in vitro* dans un premier temps, avec les protéines TACE et SP1. La surexpression par transfection de ces protéines pourrait éventuellement induire une re-localisation nucléaire de 99b78 dans les cellules en culture. L'effet d'une modulation du niveau des stérols, responsables du déclenchement du système ERAD pourrait être évalué au niveau du réticulum endoplasmique.

Le rôle potentiel de kinase ou dans la transduction du signal en général de 99b78 pourrait être étudié, en accord avec les résultats préliminaires obtenus sur la phosphorylation des MAP kinases ERK1/2. Une série d'une dizaine de gènes rapporteurs correspondant à des voies de signalisation classiques sont disponibles dans le laboratoire, et, pourraient permettre d'évaluer le rôle de 99b78 dans ces voies de signalisation. La fonction éventuelle d'alcool dehydrogenase pourrait être évaluée par suivi de l'activité de la protéine purifiée *in vitro*, ou, *in vivo* dans les clones stables surexprimant 99b78, par suivi des variations des taux d'alcool par HPLC ou des flux par un hydrogène tritié fixé sur les alcools.

Après démonstration du potentiel oncogénique de 99b78, élucidation de son mode d'action, voir de ses partenaires, la recherche de substances chimiques ou d'anticorps inhibant sa fonction devra être entreprise, dans la perspective d'applications thérapeutiques.

## **5. Perspectives générales sur les traitements futurs des cancers**

Deux de mes quatre clones favoris ont fait l'objet d'une publication (une complète avec élucidation d'aspects fonctionnels et une purement bioinformatique *In Silico*) au cours

de mon projet de thèse. Ceci illustre la très grande compétitivité existant dans la caractérisation des acteurs moléculaires de la carcinogenèse, et, des pathologies humaines en général. Plusieurs de mes anciens gènes favoris, initialement inconnus, se sont vus inclure à des brevets dans d'autres cancers (carcinome hépatocellulaire notamment). Ceci illustre le dynamisme de la recherche contre le cancer. En même temps, ce fait démontre un éventuel risque de sclérose de la recherche du fait du dépôt de brevets, parfois généraux, obscurs, sans suites, et sans perspective réelle d'applications thérapeutiques (blanket patent).

De manière positive, ce fait illustre également l'avancée des connaissances sur le cancer dont le transfert à des applications cliniques commence à porter ses fruits, comme mis en évidence par le développement de drogues spécifiques de cellules cancéreuses comme le Gleevec et l'herceptin. Nous espérons donc que la signature de l'expression dans les tumeurs VADS que nous avons définie par une approche combinant deux techniques d'analyse à large échelle, aura des applications cliniques à deux niveaux : au niveau du pronostique, permettant une meilleure adaptation du protocole thérapeutique ; au niveau thérapeutique lui même, en offrant de nouvelles cibles potentielles au drug design. Dans cette perspective, la forte fréquence de certaines altérations identifiées dans notre étude est particulièrement prometteuse. Les événements de perte d'expression de potentiels gènes suppresseurs de tumeurs étant très fréquents, une application de la thérapie génique pourrait être, à terme, envisagée pour ces tumeurs de mauvais pronostique, dont la localisation superficielle permet l'utilisation de vecteurs de type liposome. De nombreux clones isolés en DD ne sont pas associés à des gènes de fonction connue et constituent donc une réserve importante de nouvelles cibles potentielles.

Par analogie au SIDA, il est probable que le traitement des tumeurs, possédant fréquemment des mécanismes de résistance aux traitements, sera grandement amélioré par la combinaison de plusieurs drogues possédant des cibles –type cellulaire comme les cellules endothéliales intratumorales, cellules cancéreuses, cellules immunitaires, mais aussi des cibles protéiques spécifiques impliquées dans divers processus cellulaire comme l'apoptose, la migration, l'adhésion, etc...-, des mécanismes d'action et des mécanismes de résistance différents. De même, la spécificité d'action des drogues en cours de développement pour les cellules cancéreuses, ou, les cellules endothéliales impliquées dans la néoangiogenèse tumorale, devrait entraîner une moindre toxicité que la majorité des agents actuels. Il est donc

particulièrement important de fournir le plus de cibles possible de manière à augmenter l'arsenal de composés chimiques anti-tumoraux.

Le cancer étant, en partie, une maladie du vieillissement, son traitement aurait sans doute pour conséquence un accroissement de la durée de vie, l'émergence de nouvelles limites, et, une importance croissante de certaines maladies comme, par exemple, la maladie d'Alzheimer. Une amélioration significative du traitement des cancers VADS, et des cancers en général, devrait mener, comme dans le cas du SIDA, à un changement de perception de la maladie, à un moindre impact psychologique du diagnostic de cancer, et, à une meilleure acceptation sociale des malades atteints de cancer.

# Bibliographie

- Abiko, Y., Nishimura, M., Kusano, K., Yamazaki, M., Arakawa, T., Takuma, T. & Kaku, T.** (2003). Upregulated expression of human beta defensin-1 and -3 mRNA during differentiation of keratinocyte immortalized cell lines, HaCaT and PHK16-0b. *J Dermatol Sci*, **31**, 225-8.
- Adams, M.D., Dubnick, M., Kerlavage, A.R., Moreno, R., Kelley, J.M., Utterback, T.R., Nagle, J.W., Fields, C. & Venter, J.C.** (1992). Sequence identification of 2,375 human brain genes. *Nature*, **355**, 632-4.
- Ah-See, K.W., Cooke, T.G., Pickford, I.R., Soutar, D. & Balmain, A.** (1994). An allelotype of squamous carcinoma of the head and neck using microsatellite markers. *Cancer Res*, **54**, 1617-21.
- Al Moustafa, A.E., Alaoui-Jamali, M.A., Batist, G., Hernandez-Perez, M., Serruya, C., Alpert, L., Black, M.J., Sladek, R. & Foulkes, W.D.** (2002). Identification of genes associated with head and neck carcinogenesis by cDNA microarray comparison between matched primary normal epithelial and squamous carcinoma cells. *Oncogene*, **21**, 2634-40.
- Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K., Watson, J.D.** (1998). *Molecular Biology of the Cell*. Garland publishing, New York, 528.
- Alevizos, I., Mahadevappa, M., Zhang, X., Ohyama, H., Kohno, Y., Posner, M., Gallagher, G.T., Varvares, M., Cohen, D., Kim, D., Kent, R., Donoff, R.B., Todd, R., Yung, C.M., Warrington, J.A. & Wong, D.T.** (2001). Oral cancer in vivo gene expression profiling assisted by laser capture microdissection and microarray analysis. *Oncogene*, **20**, 6196-204.
- Ames, B.N., Shigenaga, M.K. & Gold, L.S.** (1993a). DNA lesions, inducible DNA repair, and cell division: three key factors in mutagenesis and carcinogenesis. *Environ Health Perspect*, **101 Suppl 5**, 35-44.
- Ames, B.N., Shigenaga, M.K. & Hagen, T.M.** (1993b). Oxidants, antioxidants, and the degenerative diseases of aging. *Proc Natl Acad Sci U S A*, **90**, 7915-22.

- Ang, K.K., Berkey, B.A., Tu, X., Zhang, H.Z., Katz, R., Hammond, E.H., Fu, K.K. & Milas, L.** (2002). Impact of epidermal growth factor receptor expression on survival and pattern of relapse in patients with advanced head and neck carcinoma. *Cancer Res*, **62**, 7350-6.
- Asanoma, K., Matsuda, T., Kondo, H., Kato, K., Kishino, T., Niikawa, N., Wake, N. & Kato, H.** (2003). NECC1, a candidate choriocarcinoma suppressor gene that encodes a homeodomain consensus motif. *Genomics*, **81**, 15-25.
- Bachem, C.W., van der Hoeven, R.S., de Bruijn, S.M., Vreugdenhil, D., Zabeau, M. & Visser, R.G.** (1996). Visualization of differential gene expression using a novel method of RNA fingerprinting based on AFLP: analysis of gene expression during potato tuber development. *Plant J*, **9**, 745-53.
- Basu, S., Binder, R.J., Ramalingam, T. & Srivastava, P.K.** (2001). CD91 is a common receptor for heat shock proteins gp96, hsp90, hsp70, and calreticulin. *Immunity*, **14**, 303-13.
- Belbin TJ, S.B., Barber I, Socci N, Wenig B, Smith R, Prystowsky MB, Childs G.** (2002). Molecular classification of head and neck squamous cell carcinoma using cDNA microarrays. *Cancer Res*, **62**, 1184-90.
- Belli, F., Testori, A., Rivoltini, L., Maio, M., Andreola, G., Sertoli, M.R., Gallino, G., Piris, A., Cattelan, A., Lazzari, I., Carrabba, M., Scita, G., Santantonio, C., Pilla, L., Tragni, G., Lombardo, C., Arienti, F., Marchiano, A., Queirolo, P., Bertolini, F., Cova, A., Lamaj, E., Ascani, L., Camerini, R., Corsi, M., Cascinelli, N., Lewis, J.J., Srivastava, P. & Parmiani, G.** (2002). Vaccination of metastatic melanoma patients with autologous tumor-derived heat shock protein gp96-peptide complexes: clinical and immunologic findings. *J Clin Oncol*, **20**, 4169-80.
- Bertelsen, A.H., Velculescu, V.E.** (1998). High-Throughput gene expression analysis using SAGE. *Drug Discovery Today*, **3**, 152-159.
- Bertioli, D.J., Schlichter, U.H., Adams, M.J., Burrows, P.R., Steinbiss, H.H. & Antoniw, J.F.** (1995). An analysis of differential display shows a strong bias towards high copy number mRNAs. *Nucleic Acids Res*, **23**, 4520-3.
- Bird, A.P.** (1986). CpG-rich islands and the function of DNA methylation. *Nature*, **321**, 209-13.
- Bishop, J.O., Morton, J.G., Rosbash, M. & Richardson, M.** (1974). Three abundance classes in HeLa cell messenger RNA. *Nature*, **250**, 199-204.
- Bland, C.E., Kimberly, P. & Rand, M.D.** (2003). Notch-induced proteolysis and nuclear localization of the Delta ligand. *J Biol Chem*, **278**, 13607-10.
- Bockmuhl, U. & Petersen, I.** (2002). DNA ploidy and chromosomal alterations in head and neck squamous cell carcinoma. *Virchows Arch*, **441**, 541-50.
- Bonaldo, M.F., Lennon, G. & Soares, M.B.** (1996). Normalization and subtraction: two approaches to facilitate gene discovery. *Genome Res*, **6**, 791-806.
- Bouffler, S.D., Kemp, C.J., Balmain, A. & Cox, R.** (1995). Spontaneous and ionizing radiation-induced chromosomal abnormalities in p53-deficient mice. *Cancer Res*, **55**, 3883-9.
- Bowler, L.D., Hubank, M. & Spratt, B.G.** (1999). Representational difference analysis of cDNA for the detection of differential gene expression in bacteria: development using a model of iron-regulated gene expression in *Neisseria meningitidis*. *Microbiology*, **145** ( Pt 12), 3529-37.
- Boyle, J.O., Hakim, J., Koch, W., van der Riet, P., Hruban, R.H., Roa, R.A., Correo, R., Eby, Y.J., Ruppert, J.M. & Sidransky, D.** (1993). The incidence of p53 mutations increases with progression of head and neck cancer. *Cancer Res*, **53**, 4477-80.

- Breitkreutz, D., Boukamp, P., Ryle, C.M., Stark, H.J., Roop, D.R. & Fusenig, N.E.** (1991). Epidermal morphogenesis and keratin expression in c-Ha-ras-transfected tumorigenic clones of the human HaCaT cell line. *Cancer Res*, **51**, 4402-9.
- Breitkreutz, D., Stark, H.J., Plein, P., Baur, M. & Fusenig, N.E.** (1993). Differential modulation of epidermal keratinization in immortalized (HaCaT) and tumorigenic human skin keratinocytes (HaCaT-ras) by retinoic acid and extracellular Ca<sup>2+</sup>. *Differentiation*, **54**, 201-17.
- Brennan, J.A., Boyle, J.O., Koch, W.M., Goodman, S.N., Hruban, R.H., Eby, Y.J., Couch, M.J., Forastiere, A.A. & Sidransky, D.** (1995). Association between cigarette smoking and mutation of the p53 gene in squamous-cell carcinoma of the head and neck. *N Engl J Med*, **332**, 712-7.
- Brown, M.G., Fulmek, S., Matsumoto, K., Cho, R., Lyons, P.A., Levy, E.R., Scalzo, A.A. & Yokoyama, W.M.** (1997). A 2-Mb YAC contig and physical map of the natural killer gene complex on mouse chromosome 6. *Genomics*, **42**, 16-25.
- Bryan, T.M. & Cech, T.R.** (1999). Telomerase and the maintenance of chromosome ends. *Curr Opin Cell Biol*, **11**, 318-24.
- Buess, M., Moroni, C. & Hirsch, H.H.** (1997). Direct identification of differentially expressed genes by cycle sequencing and cycle labelling using the differential display PCR primers. *Nucleic Acids Res*, **25**, 2233-5.
- Cahill, D.P., Lengauer, C., Yu, J., Riggins, G.J., Willson, J.K., Markowitz, S.D., Kinzler, K.W. & Vogelstein, B.** (1998). Mutations of mitotic checkpoint genes in human cancers. *Nature*, **392**, 300-3.
- Cairns, J.** (1975). Mutation selection and the natural history of cancer. *Nature*, **255**, 197-200.
- Califano, J., van der Riet, P., Westra, W., Nawroz, H., Clayman, G., Piantadosi, S., Corio, R., Lee, D., Greenberg, B., Koch, W. & Sidransky, D.** (1996). Genetic progression model for head and neck cancer: implications for field cancerization. *Cancer Res*, **56**, 2488-92.
- Callard, D., Lescure, B. & Mazzolini, L.** (1994). A method for the elimination of false positives generated by the mRNA differential display technique. *Biotechniques*, **16**, 1096-7, 1100-3.
- Capone, A., Visco, V., Belleudi, F., Marchese, C., Cardinali, G., Bellocchi, M., Picardo, M., Frati, L. & Torrisi, M.R.** (2000). Up-modulation of the expression of functional keratinocyte growth factor receptors induced by high cell density in the human keratinocyte HaCaT cell line. *Cell Growth Differ*, **11**, 607-14.
- Carpenter, G.** (2003). Nuclear localization and possible functions of receptor tyrosine kinases. *Curr Opin Cell Biol*, **15**, 143-8.
- Carroll, J.M., Albers, K.M., Garlick, J.A., Harrington, R. & Taichman, L.B.** (1993). Tissue- and stratum-specific expression of the human involucrin promoter in transgenic mice. *Proc Natl Acad Sci U S A*, **90**, 10270-4.
- Caspersson, T., Zech, L. & Johansson, C.** (1970). Differential binding of alkylating fluorochromes in human chromosomes. *Exp Cell Res*, **60**, 315-9.
- Cavenee, W.K., Dryja, T.P., Phillips, R.A., Benedict, W.F., Godbout, R., Gallie, B.L., Murphree, A.L., Strong, L.C. & White, R.L.** (1983). Expression of recessive alleles by chromosomal mechanisms in retinoblastoma. *Nature*, **305**, 779-84.
- Chavakis, T., Kanse, S.M., Lupu, F., Hammes, H.P., Muller-Esterl, W., Pixley, R.A., Colman, R.W. & Preissner, K.T.** (2000). Different mechanisms define the antiadhesive function of high molecular weight kininogen in integrin- and urokinase receptor-dependent interactions. *Blood*, **96**, 514-22.

- Chen, F., Kook, H., Milewski, R., Gitler, A.D., Lu, M.M., Li, J., Nazarian, R., Schnepf, R., Jen, K., Biben, C., Runke, G., Mackay, J.P., Novotny, J., Schwartz, R.J., Harvey, R.P., Mullins, M.C. & Epstein, J.A.** (2002). Hop is an unusual homeobox gene that modulates cardiac development. *Cell*, **110**, 713-23.
- Chen, Y., Medvedev, A., Ruzanov, P., Marvin, K.W. & Jetten, A.M.** (1997). cDNA cloning, genomic structure, and chromosome mapping of the human epithelial membrane protein CL-20 gene (EMP1), a member of the PMP22 family. *Genomics*, **41**, 40-8.
- Chen, Y.J., Ko, J.Y., Chen, P.J., Shu, C.H., Hsu, M.T., Tsai, S.F. & Lin, C.H.** (1999). Chromosomal aberrations in nasopharyngeal carcinoma analyzed by comparative genomic hybridization. *Genes Chromosomes Cancer*, **25**, 169-75.
- Cho, K.R. & Vogelstein, B.** (1992a). Genetic alterations in the adenoma--carcinoma sequence. *Cancer*, **70**, 1727-31.
- Cho, K.R. & Vogelstein, B.** (1992b). Suppressor gene alterations in the colorectal adenoma-carcinoma sequence. *J Cell Biochem Suppl*, **16G**, 137-41.
- Chu, J., Jeffries, S., Norton, J.E., Capobianco, A.J. & Bresnick, E.H.** (2002). Repression of activator protein-1-mediated transcriptional activation by the Notch-1 intracellular domain. *J Biol Chem*, **277**, 7587-97.
- Cigudosa, J.C., Parsa, N.Z., Louie, D.C., Filippa, D.A., Jhanwar, S.C., Johansson, B., Mitelman, F. & Chaganti, R.S.** (1999). Cytogenetic analysis of 363 consecutively ascertained diffuse large B-cell lymphomas. *Genes Chromosomes Cancer*, **25**, 123-33.
- Craven, R.A. & Banks, R.E.** (2001). Laser capture microdissection and proteomics: possibilities and limitation. *Proteomics*, **1**, 1200-4.
- Crawford, L.V., Pim, D.C., Gurney, E.G., Goodfellow, P. & Taylor-Papadimitriou, J.** (1981). Detection of a common feature in several human tumor cell lines--a 53,000-dalton protein. *Proc Natl Acad Sci U S A*, **78**, 41-5.
- Cromer, A., Carles, A., Millon, R., Ganguli, G., Chalmel, F., Lemaire, F., Young, J., Dembele, D., Thibault, C., Muller, D., Poch, O., Abecassis, J. & Wasylyk, B.** (2003). Identification of genes associated with tumorigenesis and metastatic potential of hypopharyngeal cancer by microarray analysis. *Oncogene*, **E pub ahead of print**.
- Datson, N.A., van der Perk-de Jong, J., van den Berg, M.P., de Kloet, E.R. & Vreugdenhil, E.** (1999). MicroSAGE: a modified procedure for serial analysis of gene expression in limited amounts of tissue. *Nucleic Acids Res*, **27**, 1300-7.
- Debouck, C.** (1995). Differential display or differential dismay? *Current Opinion in Biotechnology*, **6**, 597-599.
- DeLeo, A.B., Jay, G., Appella, E., Dubois, G.C., Law, L.W. & Old, L.J.** (1979). Detection of a transformation-related antigen in chemically induced sarcomas and other transformed cells of the mouse. *Proc Natl Acad Sci U S A*, **76**, 2420-4.
- Derynck, R., Akhurst, R.J. & Balmain, A.** (2001). TGF-beta signaling in tumor suppression and cancer progression. *Nat Genet*, **29**, 117-29.
- Diatchenko, L., Lau, Y.F., Campbell, A.P., Chenchik, A., Moqadam, F., Huang, B., Lukyanov, S., Lukyanov, K., Gurskaya, N., Sverdlov, E.D. & Siebert, P.D.** (1996). Suppression subtractive hybridization: a method for generating differentially regulated or tissue-specific cDNA probes and libraries. *Proc Natl Acad Sci U S A*, **93**, 6025-30.

- Dickson, C., Fantl, V., Gillett, C., Brookes, S., Bartek, J., Smith, R., Fisher, C., Barnes, D. & Peters, G.** (1995). Amplification of chromosome band 11q13 and a role for cyclin D1 in human breast cancer. *Cancer Lett*, **90**, 43-50.
- Doll, R. & Peto, R.** (1981). The causes of cancer: quantitative estimates of avoidable risks of cancer in the United States today. *J Natl Cancer Inst*, **66**, 1191-308.
- Drets, M.E. & Shaw, M.W.** (1971). Specific banding patterns of human chromosomes. *Proc Natl Acad Sci U S A*, **68**, 2073-7.
- Elenbaas, B. & Weinberg, R.A.** (2001). Heterotypic signaling between epithelial tumor cells and fibroblasts in carcinoma formation. *Exp Cell Res*, **264**, 169-84.
- El-Naggar, A.K., Kim, H.W., Clayman, G.L., Coombes, M.M., Le, B., Lai, S., Zhan, F., Luna, M.A., Hong, W.K. & Lee, J.J.** (2002). Differential expression profiling of head and neck squamous carcinoma: significance in their phenotypic and biological classification. *Oncogene*, **21**, 8206-19.
- Emmert-Buck, M.R., Bonner, R.F., Smith, P.D., Chuaqui, R.F., Zhuang, Z., Goldstein, S.R., Weiss, R.A. & Liotta, L.A.** (1996). Laser capture microdissection. *Science*, **274**, 998-1001.
- Endege, W.O., Steinmann, K.E., Boardman, L.A., Thibodeau, S.N. & Schlegel, R.** (1999). Representative cDNA libraries and their utility in gene expression profiling. *Biotechniques*, **26**, 542-8, 550.
- Feinberg, A.P.** (2001). Methylation meets genomics. *Nat Genet*, **27**, 9-10.
- Field, J.K., Kiaris, H., Risk, J.M., Tsiriyotis, C., Adamson, R., Zoumpourlis, V., Rowley, H., Taylor, K., Whittaker, J., Howard, P. & et al.** (1995). Allelotype of squamous cell carcinoma of the head and neck: fractional allele loss correlates with survival. *Br J Cancer*, **72**, 1180-8.
- Franceschi, S., Bidoli, E., Herrero, R. & Munoz, N.** (2000). Comparison of cancers of the oral cavity and pharynx worldwide: etiological clues. *Oral Oncol*, **36**, 106-15.
- Franz, O., Bruchhaus, I.I. & Roeder, T.** (1999a). Verification of differential gene transcription using virtual northern blotting. *Nucleic Acids Res*, **27**, e3.
- Franz, T., Vingron, M., Boehm, T. & Dear, T.N.** (1999b). Capn7: a highly divergent vertebrate calpain with a novel C-terminal domain. *Mamm Genome*, **10**, 318-21.
- Frost, M.R. & Guggenheim, J.A.** (1999). Mammalian polyadenylation sites: implications for differential display. *Nucleic Acids Res*, **27**, 1386-91.
- Fusenig, N.E. & Boukamp, P.** (1998). Multiple stages and genetic alterations in immortalization, malignant transformation, and tumor progression of human skin keratinocytes. *Mol Carcinog*, **23**, 144-58.
- Ganguli, G., Abecassis, J. & Waslyk, B.** (2000). MDM2 induces hyperplasia and premalignant lesions when expressed in the basal layer of the epidermis. *Embo J*, **19**, 5135-47.
- Gao, Y. & Pimplikar, S.W.** (2001). The gamma -secretase-cleaved C-terminal fragment of amyloid precursor protein mediates signaling to the nucleus. *Proc Natl Acad Sci U S A*, **98**, 14979-84.
- Gillison, M.L., Koch, W.M., Capone, R.B., Spafford, M., Westra, W.H., Wu, L., Zahurak, M.L., Daniel, R.W., Viglione, M., Symer, D.E., Shah, K.V. & Sidransky, D.** (2000). Evidence for a causal association between human papillomavirus and a subset of head and neck cancers. *J Natl Cancer Inst*, **92**, 709-20.
- Gleich, L.L. & Salamone, F.N.** (2002). Molecular genetics of head and neck cancer. *Cancer Control*, **9**, 369-78.



- Golub, T.R., Slonim, D.K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J.P., Coller, H., Loh, M.L., Downing, J.R., Caligiuri, M.A., Bloomfield, C.D. & Lander, E.S.** (1999). Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science*, **286**, 531-7.
- Green, C.D., Simons, J.F., Taillon, B.E. & Lewin, D.A.** (2001). Open systems: panoramic views of gene expression. *J Immunol Methods*, **250**, 67-79.
- Greenblatt, M.S., Bennett, W.P., Hollstein, M. & Harris, C.C.** (1994). Mutations in the p53 tumor suppressor gene: clues to cancer etiology and molecular pathogenesis. *Cancer Res*, **54**, 4855-78.
- Gudkov, A.V. & Roninson, I.B.** (1997). Isolation of genetic suppressor elements (GSEs) from random fragment cDNA libraries in retroviral vectors. *Methods Mol Biol*, **69**, 221-40.
- Ha, P.K. & Califano, J.A.** (2003). The molecular biology of mucosal field cancerization of the head and neck. *Crit Rev Oral Biol Med*, **14**, 363-9.
- Hahn, W.C., Counter, C.M., Lundberg, A.S., Beijersbergen, R.L., Brooks, M.W. & Weinberg, R.A.** (1999). Creation of human tumour cells with defined genetic elements. *Nature*, **400**, 464-8.
- Han, E.K., Ng, S.C., Arber, N., Begemann, M. & Weinstein, I.B.** (1999). Roles of cyclin D1 and related genes in growth inhibition, senescence and apoptosis. *Apoptosis*, **4**, 213-219.
- Hanahan, D. & Weinberg, R.A.** (2000). The hallmarks of cancer. *Cell*, **100**, 57-70.
- Hannon, G.J., Sun, P., Carnero, A., Xie, L.Y., Maestro, R., Conklin, D.S. & Beach, D.** (1999). MaRX: an approach to genetics in mammalian cells. *Science*, **283**, 1129-30.
- Heid, C.A., Stevens, J., Livak, K.J. & Williams, P.M.** (1996). Real time quantitative PCR. *Genome Res*, **6**, 986-94.
- Hennings, H., Michael, D., Cheng, C., Steinert, P., Holbrook, K. & Yuspa, S.H.** (1980). Calcium regulation of growth and differentiation of mouse epidermal cells in culture. *Cell*, **19**, 245-54.
- Hittelman, W.N., Kim, H.J., Lee, J.S., Shin, D.M., Lippman, S.M., Kim, J., Ro, J.Y. & Hong, W.K.** (1996). Detection of chromosome instability of tissue fields at risk: in situ hybridization. *J Cell Biochem Suppl*, **25**, 57-62.
- Huang, X., Gollin, S.M., Raja, S. & Godfrey, T.E.** (2002). High-resolution mapping of the 11q13 amplicon and identification of a gene, TAOS1, that is amplified and overexpressed in oral cancer cells. *Proc Natl Acad Sci U S A*, **99**, 11369-74.
- Hubank, M. & Schatz, D.G.** (1994). Identifying differences in mRNA expression by representational difference analysis of cDNA. *Nucleic Acids Res*, **22**, 5640-8.
- Islas, S., Vega, J., Ponce, L. & Gonzalez-Mariscal, L.** (2002). Nuclear localization of the tight junction protein ZO-2 in epithelial cells. *Exp Cell Res*, **274**, 138-48.
- Jackson, A.L. & Loeb, L.A.** (1998). The mutation rate and cancer. *Genetics*, **148**, 1483-90.
- Jallepalli, P.V. & Lengauer, C.** (2001). Chromosome segregation and cancer: cutting through the mystery. *Nat Rev Cancer*, **1**, 109-17.
- Jares, P., Fernandez, P.L., Campo, E., Nadal, A., Bosch, F., Aiza, G., Nayach, I., Traserra, J. & Cardesa, A.** (1994). PRAD-1/cyclin D1 gene amplification correlates with messenger RNA overexpression and tumor progression in human laryngeal carcinomas. *Cancer Res*, **54**, 4813-7.
- Jeffries, S., Robbins, D.J. & Capobianco, A.J.** (2002). Characterization of a high-molecular-weight Notch complex in the nucleus of Notch(ic)-transformed RKE cells and in a human T-cell leukemia cell line. *Mol Cell Biol*, **22**, 3927-41.

- Jin, Y., Mertens, F., Jin, C., Akervall, J., Wennerberg, J., Gorunova, L., Mandahl, N., Heim, S. & Mitelman, F.** (1995). Nonrandom chromosome abnormalities in short-term cultured primary squamous cell carcinomas of the head and neck. *Cancer Res*, **55**, 3204-10.
- Kaplan, A.S., Caldarelli, D.D., Chacho, M.S., Bruce, D.R., Hutchinson, J., Conway, S. & Coon, J.S.** (1986). Retrospective DNA analysis of head and neck squamous cell carcinoma. *Arch Otolaryngol Head Neck Surg*, **112**, 1159-62.
- Kashiwazaki, H., Tonoki, H., Tada, M., Chiba, I., Shindoh, M., Totsuka, Y., Iggo, R. & Moriuchi, T.** (1997). High frequency of p53 mutations in human oral epithelial dysplasia and primary squamous cell carcinoma detected by yeast functional assay. *Oncogene*, **15**, 2667-74.
- Katoh, M.** (2003). FLJ10261 gene, located within the CCND1-EMS1 locus on human chromosome 11q13, encodes the eight-transmembrane protein homologous to C12orf3, C11orf25 and FLJ34272 gene products. *Int J Oncol*, **22**, 1375-81.
- Kim, J.A.** (2003). Targeted therapies for the treatment of cancer. *Am J Surg*, **186**, 264-8.
- Kim, N.W., Piatyszek, M.A., Prowse, K.R., Harley, C.B., West, M.D., Ho, P.L., Coviello, G.M., Wright, W.E., Weinrich, S.L. & Shay, J.W.** (1994). Specific association of human telomerase activity with immortal cells and cancer. *Science*, **266**, 2011-5.
- Kinzler, K.W. & Vogelstein, B.** (1997). Cancer-susceptibility genes. Gatekeepers and caretakers. *Nature*, **386**, 761, 763.
- Knight, J.** (2001a). Geneticists' work in disarray as DNA-chip producer pulls the plug. *Nature*, **414**, 135-6.
- Knight, J.** (2001b). When the chips are down. *Nature*, **410**, 860-1.
- Knudson, A.G., Jr.** (1978). Retinoblastoma: a prototypic hereditary neoplasm. *Semin Oncol*, **5**, 57-60.
- Knudson, A.G., Jr., Hethcote, H.W. & Brown, B.W.** (1975). Mutation and childhood cancer: a probabilistic model for the incidence of retinoblastoma. *Proc Natl Acad Sci U S A*, **72**, 5116-20.
- Land, H., Parada, L.F. & Weinberg, R.A.** (1983). Cellular oncogenes and multistep carcinogenesis. *Science*, **222**, 771-8.
- Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczy, J., LeVine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J.P., Miranda, C., Morris, W., Naylor, J., Raymond, C., Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J.C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R.H., Wilson, R.K., Hillier, L.W., McPherson, J.D., Marra, M.A., Mardis, E.R., Fulton, L.A., Chinwalla, A.T., Pepin, K.H., Gish, W.R., Chissole, S.L., Wendl, M.C., Delehaunty, K.D., Miner, T.L., Delehaunty, A., Kramer, J.B., Cook, L.L., Fulton, R.S., Johnson, D.L., Minx, P.J., Clifton, S.W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J.F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., et al.** (2001). Initial sequencing and analysis of the human genome. *Nature*, **409**, 860-921.
- Larramendy, M.L., Tarkkanen, M., Valle, J., Kivioja, A.H., Ervasti, H., Karaharju, E., Salmivalli, T., Elomaa, I. & Knuutila, S.** (1997). Gains, losses, and amplifications of DNA sequences evaluated by comparative genomic hybridization in chondrosarcomas. *Am J Pathol*, **150**, 685-91.

- Lavery, D.J., Lopez-Molina, L., Fleury-Olela, F. & Schibler, U.** (1997). Selective amplification via biotin- and restriction-mediated enrichment (SABRE), a novel selective amplification procedure for detection of differentially expressed mRNAs. *Proc Natl Acad Sci U S A*, **94**, 6831-6.
- Lee, P.D., Sladek, R., Greenwood, C.M. & Hudson, T.J.** (2002). Control genes and variability: absence of ubiquitous reference transcripts in diverse mammalian expression studies. *Genome Res*, **12**, 292-7.
- Leethanakul, C., Patel, V., Gillespie, J., Pallente, M., Ensley, J.F., Koontongkaew, S., Liotta, L.A., Emmert-Buck, M. & Gutkind, J.S.** (2000). Distinct pattern of expression of differentiation and growth-related genes in squamous cell carcinomas of the head and neck revealed by the use of laser capture microdissection and cDNA arrays. *Oncogene*, **19**, 3220-4.
- Lemaire, F., Millon, R., Young, J., Cromer, A., Wasyluk, C., Schultz, I., Muller, D., Marchal, P., Zhao, C., Melle, D., Bracco, L., Abecassis, J. & Wasyluk, B.** (2003). Differential expression profiling of head and neck squamous cell carcinoma (HNSCC). *Br J Cancer*, **89**, 1940-9.
- Lengauer, C., Kinzler, K.W. & Vogelstein, B.** (1997a). DNA methylation and genetic instability in colorectal cancer cells. *Proc Natl Acad Sci U S A*, **94**, 2545-50.
- Lengauer, C., Kinzler, K.W. & Vogelstein, B.** (1997b). Genetic instability in colorectal cancers. *Nature*, **386**, 623-7.
- Lengauer, C., Kinzler, K.W. & Vogelstein, B.** (1998). Genetic instabilities in human cancers. *Nature*, **396**, 643-9.
- Liang, P. & Pardee, A.B.** (1992). Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. *Science*, **257**, 967-71.
- Liang, P.** (1998). Factors ensuring successful use of differential display. *Methods*, **16**, 361-4.
- Liang, P.** (2002). A decade of differential display. *Biotechniques*, **33**, 338-44, 346.
- Liang, P., Zhu, W., Zhang, X., Guo, Z., O'Connell, R.P., Averboukh, L., Wang, F. & Pardee, A.B.** (1994). Differential display using one-base anchored oligo-dT primers. *Nucleic Acids Res*, **22**, 5763-4.
- Lin, S.Y., Makino, K., Xia, W., Matin, A., Wen, Y., Kwong, K.Y., Bourguignon, L. & Hung, M.C.** (2001). Nuclear localization of EGF receptor and its potential new role as a transcription factor. *Nat Cell Biol*, **3**, 802-8.
- Lingen, M., Sturgis, E.M. & Kies, M.S.** (2001). Squamous cell carcinoma of the head and neck in nonsmokers: clinical and biologic characteristics and implications for management. *Curr Opin Oncol*, **13**, 176-82.
- Lisitsyn, N. & Wigler, M.** (1993). Cloning the differences between two complex genomes. *Science*, **259**, 946-51.
- Liss, C., Fekete, M.J., Hasina, R., Lam, C.D. & Lingen, M.W.** (2001). Paracrine angiogenic loop between head-and-neck squamous-cell carcinomas and macrophages. *Int J Cancer*, **93**, 781-5.
- Loeb, L.A.** (1991). Mutator phenotype may be required for multistage carcinogenesis. *Cancer Res*, **51**, 3075-9.
- Loeb, L.A.** (2001). A mutator phenotype in cancer. *Cancer Res*, **61**, 3230-9.
- Madsen, P., Anant, S., Rasmussen, H.H., Gromov, P., Vorum, H., Dumanski, J.P., Tommerup, N., Collins, J.E., Wright, C.L., Dunham, I., MacGinnitie, A.J., Davidson, N.O. & Celis, J.E.** (1999). Psoriasis upregulated phorbolin-1 shares structural but not functional similarity to the mRNA-editing protein apobec-1. *J Invest Dermatol*, **113**, 162-9.

- Maki, R.G., Old, L.J. & Srivastava, P.K.** (1990). Human homologue of murine tumor rejection antigen gp96: 5'-regulatory and coding regions and relationship to stress-induced proteins. *Proc Natl Acad Sci U S A*, **87**, 5658-62.
- Malkin, D., Li, F.P., Strong, L.C., Fraumeni, J.F., Jr., Nelson, C.E., Kim, D.H., Kassel, J., Gryka, M.A., Bischoff, F.Z., Tainsky, M.A. & et al.** (1990). Germ line p53 mutations in a familial syndrome of breast cancer, sarcomas, and other neoplasms. *Science*, **250**, 1233-8.
- Malumbres, M. & Barbacid, M.** (2001). To cycle or not to cycle: a critical decision in cancer. *Nat Rev Cancer*, **1**, 222-31.
- Manolov, G. & Manolova, Y.** (1972). Marker band in one chromosome 14 from Burkitt lymphomas. *Nature*, **237**, 33-4.
- Marenholz, I., Volz, A., Ziegler, A., Davies, A., Ragoussis, I., Korge, B.P. & Mischke, D.** (1996). Genetic analysis of the epidermal differentiation complex (EDC) on human chromosome 1q21: chromosomal orientation, new markers, and a 6-Mb YAC contig. *Genomics*, **37**, 295-302.
- Marshall, A. & Hodgson, J.** (1998). DNA chips: an array of possibilities. *Nat Biotechnol*, **16**, 27-31.
- Marshall, E.** (2001). DNA arrays. Affymetrix settles suit, fixes mouse chips. *Science*, **291**, 2535.
- Marshall, L.J., Perks, B., Ferkol, T. & Shute, J.K.** (2001). IL-8 released constitutively by primary bronchial epithelial cells in culture forms an inactive complex with secretory component. *J Immunol*, **167**, 2816-23.
- Martin, K.J., Kwan, C.P. & Sager, R.** (1997). A direct-sequencing-based strategy for identifying and cloning cDNAs from differential display gels. *Methods Mol Biol*, **85**, 77-85.
- Martin, K.J., Kwan, C.P., O'Hare, M.J., Pardee, A.B. & Sager, R.** (1998). Identification and verification of differential display cDNAs using gene-specific primers and hybridization arrays. *Biotechniques*, **24**, 1018-26.
- Maser, R.S. & DePinho, R.A.** (2002). Keeping telomerase in its place. *Nat Med*, **8**, 934-6.
- Matz, M., Shagin, D., Bogdanova, E., Britanova, O., Lukyanov, S., Diatchenko, L. & Chenchik, A.** (1999). Amplification of cDNA ends based on template-switching effect and step-out PCR. *Nucleic Acids Res*, **27**, 1558-60.
- Mendez, E., Cheng, C., Farwell, D.G., Ricks, S., Agoff, S.N., Futran, N.D., Weymuller, E.A., Jr., Maronian, N.C., Zhao, L.P. & Chen, C.** (2002). Transcriptional expression profiles of oral squamous cell carcinomas. *Cancer*, **95**, 1482-94.
- Menegoz, F., Leseq, H.J., Rame, J.P., Reyt, E., Bauvin, E., Arveux, P., Buemi, A., Raverdy, N. & Schaffer, P.** (2002). [Lip, oral cavity and pharynx cancers in France: incidence, mortality and trends (period 1975-1995)]. *Bull Cancer*, **89**, 419-29.
- Menghi-Sartorio, S., Mandahl, N., Mertens, F., Picci, P. & Knuutila, S.** (2001). DNA copy number amplifications in sarcomas with homogeneously staining regions and double minutes. *Cytometry*, **46**, 79-84.
- Menon, G.K., Grayson, S. & Elias, P.M.** (1985). Ionic calcium reservoirs in mammalian epidermis: ultrastructural localization by ion-capture cytochemistry. *J Invest Dermatol*, **84**, 508-12.
- Meyer, G.A. & Radsak, K.D.** (2000). Identification of a novel signal sequence that targets transmembrane proteins to the nuclear envelope inner membrane. *J Biol Chem*, **275**, 3857-66.
- Michel, L.S., Liberal, V., Chatterjee, A., Kirchwegger, R., Pasche, B., Gerald, W., Dobles, M., Sorger, P.K., Murty, V.V. & Benezra, R.** (2001). MAD2 haplo-insufficiency causes premature anaphase and chromosome instability in mammalian cells. *Nature*, **409**, 355-9.

- Mikulowska-Mennis, A., Taylor, T.B., Vishnu, P., Michie, S.A., Raja, R., Horner, N. & Kunitake, S.T.** (2002). High-quality RNA from cells isolated by laser capture microdissection. *Biotechniques*, **33**, 176-9.
- Miller, J.A. & Miller, E.C.** (1975). Metabolic activation and reactivity of chemical carcinogens. *Mutat Res*, **33**, 25-6.
- Mork, J., Lie, A.K., Glattre, E., Hallmans, G., Jellum, E., Koskela, P., Moller, B., Pukkala, E., Schiller, J.T., Youngman, L., Lehtinen, M. & Dillner, J.** (2001). Human papillomavirus infection as a risk factor for squamous-cell carcinoma of the head and neck. *N Engl J Med*, **344**, 1125-31.
- Motokura, T., Bloom, T., Kim, H.G., Juppner, H., Ruderman, J.V., Kronenberg, H.M. & Arnold, A.** (1991). A novel cyclin encoded by a bcl1-linked candidate oncogene. *Nature*, **350**, 512-5.
- Muller, D., Millon, R., Lidereau, R., Engelmann, A., Bronner, G., Flesch, H., Eber, M., Methlin, G. & Abecassis, J.** (1994). Frequent amplification of 11q13 DNA markers is associated with lymph node involvement in human head and neck squamous cell carcinomas. *Eur J Cancer B Oral Oncol*, **30B**, 113-20.
- Muller, D., Millon, R., Velten, M., Bronner, G., Jung, G., Engelmann, A., Flesch, H., Eber, M., Methlin, G. & Abecassis, J.** (1997). Amplification of 11q13 DNA markers in head and neck squamous cell carcinomas: correlation with clinical outcome. *Eur J Cancer*, **33**, 2203-10.
- Myers, J.M., Martins, G.G., Ostrowski, J. & Stachowiak, M.K.** (2003). Nuclear trafficking of FGFR1: a role for the transmembrane domain. *J Cell Biochem*, **88**, 1273-91.
- Navarro, J.M. & Jorcano, J.L.** (1999). The use of arbitrarily primed polymerase chain reaction in cancer research. *Electrophoresis*, **20**, 283-90.
- Nawroz, H., van der Riet, P., Hruban, R.H., Koch, W., Ruppert, J.M. & Sidransky, D.** (1994). Allelotype of head and neck squamous cell carcinoma. *Cancer Res*, **54**, 1152-5.
- Nowell, P.C. & Hungerford, D.A.** (1960). Chromosome studies on normal and leukemic human leukocytes. *J Natl Cancer Inst*, **25**, 85-109.
- Nowell, P.C.** (1976). The clonal evolution of tumor cell populations. *Science*, **194**, 23-8.
- O-charoenrat, P., Rhys-Evans, P.H., Modjtahedi, H. & Eccles, S.A.** (2002). The role of c-erbB receptors and ligands in head and neck squamous cell carcinoma. *Oral Oncol*, **38**, 627-40.
- Opitz, O.G., Harada, H., Suliman, Y., Rhoades, B., Sharpless, N.E., Kent, R., Kopelovich, L., Nakagawa, H. & Rustgi, A.K.** (2002). A mouse model of human oral-esophageal cancer. *J Clin Invest*, **110**, 761-9.
- Ornitz, D.M. & Itoh, N.** (2001). Fibroblast growth factors. *Genome Biol*, **2**, REVIEWS3005.
- Parkin, D.M., Pisani, P. & Ferlay, J.** (1999). Estimates of the worldwide incidence of 25 major cancers in 1990. *Int J Cancer*, **80**, 827-41.
- Patel, A.M., Incognito, L.S., Schechter, G.L., Wasilenko, W.J. & Somers, K.D.** (1996). Amplification and expression of EMS-1 (cortactin) in head and neck squamous cell carcinoma cell lines. *Oncogene*, **12**, 31-5.
- Patel, K.J., Yu, V.P., Lee, H., Corcoran, A., Thistlethwaite, F.C., Evans, M.J., Colledge, W.H., Friedman, L.S., Ponder, B.A. & Venkitaraman, A.R.** (1998). Involvement of Brca2 in DNA repair. *Mol Cell*, **1**, 347-57.
- Poirier, G.M. & Erlander, M.G.** (1998). Postdifferential display: parallel processing of candidates using small amounts of RNA. *Methods*, **16**, 444-52.
- Poirier, G.M., Pyati, J., Wan, J.S. & Erlander, M.G.** (1997). Screening differentially expressed cDNA clones obtained by differential display using amplified RNA. *Nucleic Acids Res*, **25**, 913-4.

- Prashar, Y. & Weissman, S.M.** (1996). Analysis of differential gene expression by display of 3' end restriction fragments of cDNAs. *Proc Natl Acad Sci U S A*, **93**, 659-63.
- Proctor, A.J., Coombs, L.M., Cairns, J.P. & Knowles, M.A.** (1991). Amplification at chromosome 11q13 in transitional cell tumours of the bladder. *Oncogene*, **6**, 789-95.
- Quon, H., Liu, F.F. & Cummings, B.J.** (2001). Potential molecular prognostic markers in head and neck squamous cell carcinomas. *Head Neck*, **23**, 147-59.
- Ramaswamy, S., Ross, K.N., Lander, E.S. & Golub, T.R.** (2003). A molecular signature of metastasis in primary solid tumors. *Nat Genet*, **33**, 49-54.
- Reed, A.L., Califano, J., Cairns, P., Westra, W.H., Jones, R.M., Koch, W., Ahrendt, S., Eby, Y., Sewell, D., Nawroz, H., Bartek, J. & Sidransky, D.** (1996). High frequency of p16 (CDKN2/MTS-1/INK4A) inactivation in head and neck squamous cell carcinoma. *Cancer Res*, **56**, 3630-3.
- Reilly, J.F. & Maher, P.A.** (2001). Importin beta-mediated nuclear import of fibroblast growth factor receptor: role in cell proliferation. *J Cell Biol*, **152**, 1307-12.
- Reiss, M.** (1997). Transforming growth factor-beta and cancer: a love-hate relationship? *Oncol Res*, **9**, 447-57.
- Renan, M.J.** (1993). How many mutations are required for tumorigenesis? Implications from human cancer data. *Mol Carcinog*, **7**, 139-46.
- Rondinelli, R.H., Epner, D.E. & Tricoli, J.V.** (1997). Increased glyceraldehyde-3-phosphate dehydrogenase gene expression in late pathological stage human prostate cancer. *Prostate Cancer Prostatic Dis*, **1**, 66-72.
- Roninson, I.B. & Gudkov, A.V.** (2003). Genetic suppressor elements in the characterization and identification of tumor suppressor genes. *Methods Mol Biol*, **222**, 413-36.
- Rowley, J.D.** (1973). Letter: A new consistent chromosomal abnormality in chronic myelogenous leukaemia identified by quinacrine fluorescence and Giemsa staining. *Nature*, **243**, 290-3.
- Rowley, J.D.** (1975). Nonrandom chromosomal abnormalities in hematologic disorders of man. *Proc Natl Acad Sci U S A*, **72**, 152-6.
- Sakr, W., Hussan, M., Zarbo, R.J., Ensley, J. & Crissman, J.D.** (1989). DNA quantitation and histologic characteristics of squamous cell carcinoma of the upper aerodigestive tract. *Arch Pathol Lab Med*, **113**, 1009-14.
- Salomon-Nguyen, F., Della-Valle, V., Mauchauffe, M., Busson-Le Coniat, M., Ghysdael, J., Berger, R. & Bernard, O.A.** (2000). The t(1;12)(q21;p13) translocation of human acute myeloblastic leukemia results in a TEL-ARNT fusion. *Proc Natl Acad Sci U S A*, **97**, 6757-62.
- Sankaranarayanan, R., Masuyer, E., Swaminathan, R., Ferlay, J. & Whelan, S.** (1998). Head and neck cancer: a global perspective on epidemiology and prognosis. *Anticancer Res*, **18**, 4779-86.
- Sargent, T.D.** (1987). Isolation of differentially expressed genes. *Methods Enzymol*, **152**, 423-32.
- Sato, Y., Kobayashi, H., Suto, Y., Olney, H.J., Davis, E.M., Super, H.G., Espinosa, R., 3rd, Le Beau, M.M. & Rowley, J.D.** (2001). Chromosomal instability in chromosome band 12p13: multiple breaks leading to complex rearrangements including cytogenetically undetectable sub-clones. *Leukemia*, **15**, 1193-202.
- Schoop, V.M., Mirancea, N. & Fusenig, N.E.** (1999). Epidermal organization and differentiation of HaCaT keratinocytes in organotypic coculture with human dermal fibroblasts. *J Invest Dermatol*, **112**, 343-53.

- Schuuring, E., Verhoeven, E., Litvinov, S. & Michalides, R.J.** (1993). The product of the EMS1 gene, amplified and overexpressed in human carcinomas, is homologous to a v-src substrate and is located in cell-substratum contact sites. *Mol Cell Biol*, **13**, 2891-98.
- Seo, E.Y., Piao, Y.J., Kim, J.S., Suhr, K.B., Park, J.K. & Lee, J.H.** (2002). Identification of calcium-induced genes in HaCaT keratinocytes by polymerase chain reaction-based subtractive hybridization. *Arch Dermatol Res*, **294**, 411-8.
- Serrano, M., Lee, H., Chin, L., Cordon-Cardo, C., Beach, D. & DePinho, R.A.** (1996). Role of the INK4a locus in tumor suppression and cell mortality. *Cell*, **85**, 27-37.
- Shawver, L.K., Slamon, D. & Ullrich, A.** (2002). Smart drugs: tyrosine kinase inhibitors in cancer therapy. *Cancer Cell*, **1**, 117-23.
- Sheikh, M.S. & Huang, Y.** (2003). The FADD is going nuclear. *Cell Cycle*, **2**, 346-7.
- Shin, C.H., Liu, Z.P., Passier, R., Zhang, C.L., Wang, D.Z., Harris, T.M., Yamagishi, H., Richardson, J.A., Childs, G. & Olson, E.N.** (2002). Modulation of cardiac growth and development by HOP, an unusual homeodomain protein. *Cell*, **110**, 725-35.
- Simone, N.L., Bonner, R.F., Gillespie, J.W., Emmert-Buck, M.R. & Liotta, L.A.** (1998). Laser-capture microdissection: opening the microscopic frontier to molecular analysis. *Trends Genet*, **14**, 272-6.
- Singh, B., Gogineni, S.K., Sacks, P.G., Shaha, A.R., Shah, J.P., Stoffel, A. & Rao, P.H.** (2001). Molecular cytogenetic characterization of head and neck squamous cell carcinoma and refinement of 3q amplification. *Cancer Res*, **61**, 4506-13.
- Spinella, D.G., Bernardino, A.K., Redding, A.C., Koutz, P., Wei, Y., Pratt, E.K., Myers, K.K., Chappell, G., Gerken, S. & McConnell, S.J.** (1999). Tandem arrayed ligation of expressed sequence tags (TALEST): a new method for generating global gene expression profiles. *Nucleic Acids Res*, **27**, e22.
- Stehelin, D., Guntaka, R.V., Varmus, H.E. & Bishop, J.M.** (1976a). Purification of DNA complementary to nucleotide sequences required for neoplastic transformation of fibroblasts by avian sarcoma viruses. *J Mol Biol*, **101**, 349-65.
- Stehelin, D., Varmus, H.E., Bishop, J.M. & Vogt, P.K.** (1976b). DNA related to the transforming gene(s) of avian sarcoma viruses is present in normal avian DNA. *Nature*, **260**, 170-3.
- Sudbo, J., Bryne, M., Johannessen, A.C., Kildal, W., Danielsen, H.E. & Reith, A.** (2001). Comparison of histological grading and large-scale genomic status (DNA ploidy) as prognostic tools in oral dysplasia. *J Pathol*, **194**, 303-10.
- Sutcliffe, J.G., Foye, P.E., Erlander, M.G., Hilbush, B.S., Bodzin, L.J., Durham, J.T. & Hasel, K.W.** (2000). TOGA: an automated parsing technology for analyzing expression of nearly all genes. *Proc Natl Acad Sci U S A*, **97**, 1976-81.
- Tan, P.K., Downey, T.J., Spitznagel, E.L., Jr., Xu, P., Fu, D., Dimitrov, D.S., Lempicki, R.A., Raaka, B.M. & Cam, M.C.** (2003). Evaluation of gene expression measurements from commercial microarray platforms. *Nucleic Acids Res*, **31**, 5676-84.
- Tarone, G., Hirsch, E., Brancaccio, M., De Acetis, M., Barberis, L., Balzac, F., Retta, S.F., Botta, C., Altruda, F., Silengo, L. & Retta, F.** (2000). Integrin function and regulation in development. *Int J Dev Biol*, **44**, 725-31.
- Temam, S., Flahault, A., Perie, S., Monceaux, G., Coulet, F., Callard, P., Bernaudin, J.F., St Guily, J.L. & Fouret, P.** (2000). p53 gene status as a predictor of tumor response to induction chemotherapy of patients with locoregionally advanced squamous cell carcinomas of the head and neck. *J Clin Oncol*, **18**, 385-94.

- Tesfaigzi, J. & Carlson, D.M.** (1999). Expression, regulation, and function of the SPR family of proteins. A review. *Cell Biochem Biophys*, **30**, 243-65.
- Trenkle, T., Welsh, J. & McClelland, M.** (1999). Differential display probes for cDNA arrays. *Biotechniques*, **27**, 554-60, 562, 564.
- van Belzen, N., Dinjens, W.N., Diesveld, M.P., Groen, N.A., van der Made, A.C., Nozawa, Y., Vlietstra, R., Trapman, J. & Bosman, F.T.** (1997). A novel gene which is up-regulated during colon epithelial cell differentiation and down-regulated in colorectal neoplasms. *Lab Invest*, **77**, 85-92.
- Van Gelder, R.N., von Zastrow, M.E., Yool, A., Dement, W.C., Barchas, J.D. & Eberwine, J.H.** (1990). Amplified RNA synthesized from limited quantities of heterogeneous cDNA. *Proc Natl Acad Sci U S A*, **87**, 1663-7.
- van Steeg, H. & Kraemer, K.H.** (1999). Xeroderma pigmentosum and the role of UV-induced DNA damage in skin cancer. *Mol Med Today*, **5**, 86-94.
- van 't Veer, L.J., Dai, H., van de Vijver, M.J., He, Y.D., Hart, A.A., Mao, M., Peterse, H.L., van der Kooy, K., Marton, M.J., Witteveen, A.T., Schreiber, G.J., Kerkhoven, R.M., Roberts, C., Linsley, P.S., Bernards, R. & Friend, S.H.** (2002). Gene expression profiling predicts clinical outcome of breast cancer. *Nature*, **415**, 530-6.
- Velculescu, V.E., Zhang, L., Vogelstein, B. & Kinzler, K.W.** (1995). Serial analysis of gene expression. *Science*, **270**, 484-7.
- Venter, J.C.** (2003). A part of the human genome sequence. *Science*, **299**, 1183-4.
- Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A., Gocayne, J.D., Amanatides, P., Ballew, R.M., Huson, D.H., Wortman, J.R., Zhang, Q., Kodira, C.D., Zheng, X.H., Chen, L., Skupski, M., Subramanian, G., Thomas, P.D., Zhang, J., Gabor Miklos, G.L., Nelson, C., Broder, S., Clark, A.G., Nadeau, J., McKusick, V.A., Zinder, N., Levine, A.J., Roberts, R.J., Simon, M., Slayman, C., Hunkapiller, M., Bolanos, R., Delcher, A., Dew, I., Fasulo, D., Flanigan, M., Florea, L., Halpern, A., Hannenhalli, S., Kravitz, S., Levy, S., Mobarry, C., Reinert, K., Remington, K., Abu-Threideh, J., Beasley, E., Biddick, K., Bonazzi, V., Brandon, R., Cargill, M., Chandramouliswaran, I., Charlab, R., Chaturvedi, K., Deng, Z., Di Francesco, V., Dunn, P., Eilbeck, K., Evangelista, C., Gabrielian, A.E., Gan, W., Ge, W., Gong, F., Gu, Z., Guan, P., Heiman, T.J., Higgins, M.E., Ji, R.R., Ke, Z., Ketchum, K.A., Lai, Z., Lei, Y., Li, Z., Li, J., Liang, Y., Lin, X., Lu, F., Merkulov, G.V., Milshina, N., Moore, H.M., Naik, A.K., Narayan, V.A., Neelam, B., Nusskern, D., Rusch, D.B., Salzberg, S., Shao, W., Shue, B., Sun, J., Wang, Z., Wang, A., Wang, X., Wang, J., Wei, M., Wides, R., Xiao, C., Yan, C., et al.** (2001). The sequence of the human genome. *Science*, **291**, 1304-51.
- Vila, M.R., Nicolas, A., Morote, J., de, I. & Meseguer, A.** (2000). Increased glyceraldehyde-3-phosphate dehydrogenase expression in renal cell carcinoma identified by RNA-based, arbitrarily primed polymerase chain reaction. *Cancer*, **89**, 152-64.
- Vogelstein, B., Fearon, E.R., Kern, S.E., Hamilton, S.R., Preisinger, A.C., Nakamura, Y. & White, R.** (1989). Allelotype of colorectal carcinomas. *Science*, **244**, 207-11.
- Vokes, E.E., Weichselbaum, R.R., Lippman, S.M. & Hong, W.K.** (1993). Head and neck cancer. *N Engl J Med*, **328**, 184-94.
- Wan, J.S., Sharp, S.J., Poirier, G.M., Wagaman, P.C., Chambers, J., Pyati, J., Hom, Y.L., Galindo, J.E., Huvar, A., Peterson, P.A., Jackson, M.R. & Erlander, M.G.** (1996). Cloning differentially expressed mRNAs. *Nat Biotechnol*, **14**, 1685-91.
- Wang, X. & Feuerstein, G.Z.** (1995). Direct sequencing of DNA isolated from mRNA differential display. *Biotechniques*, **18**, 448-53.



- Weiskirchen, R., Moser, M., Weiskirchen, S., Erdel, M., Dahmen, S., Buettner, R. & Gressner, A.M.** (2001). LIM-domain protein cysteine- and glycine-rich protein 2 (CRP2) is a novel marker of hepatic stellate cells and binding partner of the protein inhibitor of activated STAT1. *Biochem J*, **359**, 485-96.
- Welsh, J., Chada, K., Dalal, S.S., Cheng, R., Ralph, D. & McClelland, M.** (1992). Arbitrarily primed PCR fingerprinting of RNA. *Nucleic Acids Res*, **20**, 4965-70.
- Wennerberg, J., Baldetorp, B. & Wahlberg, P.** (1998). Distribution of non-diploid flow-cytometric DNA indices and their relation to the nodal metastasis in squamous cell carcinomas of the head and neck. *Invasion Metastasis*, **18**, 184-91.
- Williams, M.E., Gaffey, M.J., Weiss, L.M., Wilczynski, S.P., Schuurin, E. & Levine, P.A.** (1993). Chromosome 11Q13 amplification in head and neck squamous cell carcinoma. *Arch Otolaryngol Head Neck Surg*, **119**, 1238-43.
- Xie, L., Xu, L., He, Z., Zhou, W., Wang, L., Zhang, L., Lan, K., Ren, C., Liu, W. & Yao, K.** (2000). Identification of differentially expressed genes in nasopharyngeal carcinoma by means of the Atlas human cancer cDNA expression array. *J Cancer Res Clin Oncol*, **126**, 400-6.
- Xu, X., Weaver, Z., Linke, S.P., Li, C., Gotay, J., Wang, X.W., Harris, C.C., Ried, T. & Deng, C.X.** (1999). Centrosome amplification and a defective G2-M cell cycle checkpoint induce genetic instability in BRCA1 exon 11 isoform-deficient cells. *Mol Cell*, **3**, 389-95.
- Yedavelli, S.P., Guo, L., Daou, M.E., Srivastava, P.K., Mittelman, A. & Tiwari, R.K.** (1999). Preventive and therapeutic effect of tumor derived heat shock protein, gp96, in an experimental prostate cancer model. *Int J Mol Med*, **4**, 243-8.
- Yoshida, T., Sakamoto, H. & Terada, M.** (1993). Amplified genes in cancer in upper digestive tract. *Semin Cancer Biol*, **4**, 33-40.
- Yousef, G.M. & Diamandis, E.P.** (2002). Expanded human tissue kallikrein family--a novel panel of cancer biomarkers. *Tumour Biol*, **23**, 185-92.
- Yousef, G.M., Chang, A. & Diamandis, E.P.** (2000). Identification and characterization of KLK-L4, a new kallikrein-like gene that appears to be down-regulated in breast cancer tissues. *J Biol Chem*, **275**, 11891-8.
- Yuspa, S.H., Kilkenny, A.E., Steinert, P.M. & Roop, D.R.** (1989). Expression of murine epidermal differentiation markers is tightly regulated by restricted extracellular calcium concentrations in vitro. *J Cell Biol*, **109**, 1207-17.
- Zaphiropoulos, P.G., Unden, A.B., Rahnema, F., Hollingsworth, R.E. & Toftgard, R.** (1999). PTCH2, a novel human patched gene, undergoing alternative splicing and up-regulated in basal cell carcinomas. *Cancer Res*, **59**, 787-92.
- Zhang, L., Zhou, W., Velculescu, V.E., Kern, S.E., Hruban, R.H., Hamilton, S.R., Vogelstein, B. & Kinzler, K.W.** (1997). Gene expression profiles in normal and cancer cells. *Science*, **276**, 1268-72.
- Zhang, Y.J., Jiang, W., Chen, C.J., Lee, C.S., Kahn, S.M., Santella, R.M. & Weinstein, I.B.** (1993). Amplification and overexpression of cyclin D1 in human hepatocellular carcinoma. *Biochem Biophys Res Commun*, **196**, 1010-6.
- Zhao, S., Ooi, S.L. & Pardee, A.B.** (1995). New primer strategy improves precision of differential display. *Biotechniques*, **18**, 842-6, 848, 850.
- Zhu, Y.Y., Machleder, E.M., Chenchik, A., Li, R. & Siebert, P.D.** (2001). Reverse transcriptase template switching: a SMART approach for full-length cDNA library construction. *Biotechniques*, **30**, 892-7.

- Zhumabayeva, B., Diatchenko, L., Chenchik, A. & Siebert, P.D.** (2001). Use of SMART-generated cDNA for gene expression studies in multiple human tumors. *Biotechniques*, **30**, 158-63.
- Ziauddin, J. & Sabatini, D.M.** (2001). Microarrays of cells expressing defined cDNAs. *Nature*, **411**, 107-10.
- Zimonjic, D., Brooks, M.W., Popescu, N., Weinberg, R.A. & Hahn, W.C.** (2001). Derivation of human tumor cells in vitro without widespread genomic instability. *Cancer Res*, **61**, 8838-44.
- Zimonjic, D.B., Keck, C.L., Thorgeirsson, S.S. & Popescu, N.C.** (1999). Novel recurrent genetic imbalances in human hepatocellular carcinoma cell lines identified by comparative genomic hybridization. *Hepatology*, **29**, 1208-14.

## Abstract

Upper aerodigestive tract cancers are the fifth most common cancer worldwide and account for 780'000 new cases per year. Despite research on alternative therapeutical protocols the 5 year survival rate has not significantly improved over the past three decades. Individual prognosis markers are yet insufficient and histopathological features of the tumour (TNM, differentiation index) constitute the best indicator of prognosis. In order to generate a comprehensive collection of differentially expressed genes in these tumours, we have analyzed the transcriptome of early as well as further advanced hypopharyngeal tumours bearing or not metastatic propensity. We have analyzed the tumour and normal tissue from the same patients. Performing a large scale (56 primers) Differential Display (DD) analysis has allowed us to generate a collection of 1200 differentially expressed genes in tumours. The differential expression of a subset of 70 genes has been demonstrated with complex probes (Lemaire et al., 2003). The prognostic value of our DD collection is currently evaluated on 50 individual patients. The DD technique has been used to maximize isolation of genes independently of their abundance, and thus the isolation of unknown or key regulatory genes expressed at low copy number. An Affymetrix DNA chip analysis has been performed (Cromer et al., 2003) in order to provide a supplementary screening for prognostic markers on essentially known genes. ExonHit Therapeutics S.A. has performed a qualitative profiling of alternatively spliced transcripts of the same tumours using its patented DATAS technique. The characterization of the expression of 4 target genes (1 under expressed in tumours and 3 over expressed in tumours) has been furthered using a variety of techniques (Virtual Northern blot, Northern blot, PCR, immunocytochemistry, in situ hybridization, immunohistochemistry). The coding sequences of these genes have been cloned into mammalian expression vectors in sens and antisens orientation. The FACS based cell cycle analysis technique, clonogenicity assays, and western blotting have been used to evaluate potential effect of ectopic expression of these genes. The functional study of these genes as potential therapeutic targets is in progress.

## Résumé

Les cancers des voies aérodigestives supérieures (VADS) sont le 5<sup>ème</sup> cancer le plus commun dans le monde et représentent 780'000 cas nouveaux par an. Malgré des recherches de nouveaux protocoles thérapeutiques la survie à 5 ans des patients n'a pas été significativement améliorée lors des trois dernières décennies. L'utilisation de marqueurs pronostiques individuels est insuffisante à ce jour et les grade histopronostique de la tumeur (TNM, index de différenciation) est à ce jour le seul indicateur pronostique avancé. De manière à générer une collection aussi exhaustive que possible de gènes différentiellement exprimés dans les tumeurs VADS, nous avons analysé le transcriptome de tumeurs hypopharyngées de stade précoce et de stade plus avancé, présentant ou non une propension à la métastase. Nous avons comparé l'expression des gènes pour les tissus tumoraux ainsi que les tissus normaux issus du même patient. La technique du Differential Display (DD), entreprise à grande échelle (56 amorces arbitraires) nous a permis d'isoler une collection de 1'200 gènes potentiellement différentiellement exprimés dans les tumeurs. Un sous-groupe de 70 gènes a été mise en évidence comme différentiellement exprimé avec une sonde complexe (Lemaire et al., 2003). L'étude de la valeur pronostique de l'ensemble de notre collection DD est en cours sur une cinquantaine de patients. La technique du DD a été utilisée de manière à isoler des gènes indépendamment de leur niveau absolu d'expression, afin d'isoler des gènes inconnus ou des gènes-clés, régulateurs exprimés à faible niveau. La technique des puces à ADN Affymetrix (Cromer et al., 2003) a été utilisée de manière à effectuer un criblage supplémentaire essentiellement orienté sur le criblage des gènes connus, pouvant posséder une valeur pronostique. ExonHit Therapeutics SA a mis en oeuvre la technique brevetée DATAS de façon à isoler les transcrits de l'épissage alternatif exprimés dans les tumeurs de différents types tumoraux. La caractérisation de l'expression de 4 gènes favoris (1 sous exprimé et 3 surexprimés dans les tumeurs) a été entreprise par diverses techniques (Virtual Northern blot, Northern blot, PCR, immunocytochimie, hybridation *in situ* et immunohistochimie). Les séquences codantes de ces gènes ont été clonées dans des vecteurs d'expression pour une caractérisation fonctionnelle par transfection en orientation sens et antisens avec les techniques de FACS, des tests de clonogénicité, et de Western blot. L'étude de ces gènes est poursuivie dans le but d'identifier de nouvelles cibles thérapeutiques.