

Université de Montréal

Isolement d'îlots CpG différentiellement méthylés

par

Miguel Landers

Département de médecine  
Faculté des études supérieures

Thèse présentée à la Faculté des études supérieures  
en vue de l'obtention du grade de  
Philosophiæ Doctor (Ph.D.)  
en biologie moléculaire

Avril, 1999

© Miguel Landers, 1999



PH  
506  
U54  
2000  
V.003

Université de Montréal  
Faculté des études supérieures

Cette thèse intitulée

Isolement d'îlots CpG différentiellement méthylés

présentée par:

Miguel Landers

a été évaluée par un jury composé des personnes suivantes:

**Dr. Abdella Belmaaza**

Président-rapporteur

**Dr. Edward Bradley**

Directeur de recherche

**Dr. Daniel Sinnett**

Membre du jury

**Dr. Maria Zannis-Hadjopoulos**

Examinatrice externe

**Dr. Pierre Chartrand**

Répresentant du doyen de la FES

Thèse acceptée le: 28 avril 2000

## SOMMAIRE

Des niveaux normaux de méthylation sont importants dans le développement et la survie des mammifères. Cette importance est soulignée par le fait que l'inactivation par recombinaison de l'ADN méthyltransferase ou de la protéine MeCP2 aboutissent à une léthalité embryonnaire. La méthylation du génome des mammifères a lieu dans la cytosine du dinucléotide CpG. La distribution de ce dinucléotide n'est pas aléatoire, sa fréquence étant supprimée à près de 25% de celle attendue dans la plupart du génome, mais elle atteint près de 100% de ce pourcentage aux séquences riches en GC connues comme "îlots CpG" (ICGs). Ces régions se trouvent généralement sur ou près des régions promotrices des gènes. Tandis que la plupart de CpGs dans le reste du génome sont méthylés, ceux étant sur les ICGs eux-mêmes ne le sont pas et lorsque les ICGs des promoteurs sont méthylés, la transcription du gène en question est généralement inhibée. La méthylation des ICGs est impliquée dans deux phénomènes épigénétiques normaux, l'inactivation du chromosome X et l'empreinte génomique parentale. Les promoteurs des gènes situés sur le chromosome X inactif chez les femelles sont typiquement méthylés tandis que leurs allèles sur le chromosome X actif sont libres de méthylation. L'empreinte génomique parentale est un phénomène montré par un nombre limité de gènes où leur expression monoallélique est déterminée par son origine parentale. La plupart des gènes ayant une empreinte génomique parentale sont regroupés dans des domaines chromosomiques de quelques megabases. Les ICGs des gènes de ces domaines sont très souvent méthylés à degrés différents sur les deux allèles créant ainsi des régions méthylées différenciellement (RMDs). La méthylation est un candidat intéressant pour être l'empreinte, la marque qui distingue un allèle de l'autre car elle est susceptible d'être effacée dans la lignée germinale pour permettre le rétablissement de patrons d'empreinte dans la génération suivante et dans quelques cas précis il a été effectivement démontré que la méthylation est l'empreinte génomique. Pour isoler des séquences méthylées différenciellement (SMDs) des endroits au génome humain pouvant être impliqués dans l'empreinte génomique

parentale, nous avons conçu une méthode, FP-CGI, qui se sert de la technique F-PERT et qui dépend de la formation d'hétéroduplexes d'ADN provenant d'ICGs hémiméthylés. Les hétéroduplexes sont digérés deux fois, moyennant une étape de dénaturation, par une enzyme de restriction sensible à la méthylation ayant au moins un CpG dans le site de reconnaissance. La deuxième digestion permet finalement de relâcher de l'ADN simple brin facile à amplifier. Nous avons effectué cinq expériences différentes de FP-CGI sur de l'ADN de granulocytes humains. Pour chacune de ces expériences plusieurs bandes ont été repérées sur gel et quarante-trois ont été isolées, clonées et séquencées. Les clones provenant des séquences uniques ou inconnues ont été testés par Southern Blot pour déterminer leur patron de méthylation différentielle. Cinq des quatorze clones qui ont donné des résultats de Southern Blot convaincants détectent des SMDs et quatre d'entre eux localisent sur ou près de domaines ayant certaines caractéristiques d'empreinte génomique parentale ou de maladies avec un fort biais parental.

## TABLE DES MATIÈRES

Sommaire.....	III
Table des matières.....	V
Liste des figures.....	IX
Liste des tableaux.....	XIV
Liste des sigles et abréviations.....	XV
Remerciements.....	XVII

<b>1 INTRODUCTION.....</b>	<b>1</b>
1.1 Avant-propos.....	2
1.2 Mécanismes épigénétiques.....	3
1.3 La méthylation.....	5
1.3.1 La machinerie de la méthylation: méthylases, deméthylases et mécanismes de régulation.....	7
1.3.2 Pourquoi l'ADN génomique des cellules des mammifères est-il méthylé?.....	11
1.3.3 La structure de la chromatine et les enzymes qui lient les groupements méthyl.....	12
1.3.4 Réplication et recombinaison.....	16
1.3.5 Îlots CpG .....	17
1.4 L'empreinte génomique parentale.....	19
1.4.1 Comment l'empreinte est-elle établie dans les cellules germinales?.....	23
1.4.2 L'empreinte génomique parentale et la méthylation: est-ce que la méthylation est l'empreinte?.....	28
1.4.3 L'empreinte génomique parentale et les maladies.....	31
1.4.4 Pourquoi existe-t-il l'empreinte génomique parentale?.....	32
1.5 L'inactivation du chromosome X.....	33

1.5.1 L'inactivation du chromosome X et la méthylation.....	33
1.6 La méthylation et la progression tumorale.....	34
1.7 À la recherche de séquences méthylées différemment.....	38
1.7.1 Étude de la méthylation.....	38
1.7.2 Comment analyse-t-on des séquences inconnues?.....	40
1.7.3 L'isolement d'îlots CpG différemment méthylés par FP-CGI.....	41
<b>2 MATÉRIEL ET MÉTHODES.....</b>	<b>45</b>
2.1 FP-CGI.....	46
2.1.1 F-PERT sur de l'ADN de granulocytes humains.....	48
2.1.2 Ligations des adaptateurs <i>MseI</i> .....	49
2.1.3 PCRs avec des amorces arbitraires.....	50
2.1.4 Clonages.....	50
2.1.4.1 Préparation du vecteur T.....	50
2.1.4.2 Clonages.....	51
2.1.5 Séquençages.....	51
2.1.6 Recherches des banques de données.....	52
2.2 Analyses d'ADN.....	52
2.2.1 Transferts Southern.....	52
2.3 FP-CGI sur des séquences connues.....	52
2.3.1 PCRs des promoteurs des gènes MAOA et PGK-1 et transferts Southern.....	52
2.4 Localisation chromosomique.....	53
2.4.1 Panel de Coriell et panel d'hybrides de radiation G3 ("radiation hybrids").....	53
2.4.2 Cartes génétique et physique.....	55
2.5 Obtention de fragments génomiques humains de très haut poids moléculaire.....	56
2.5.1 Identification de clones génomiques, BACs.....	56
2.5.2 Identification de clones génomiques, YACs.....	58
2.6 Utilisation des BACs et des YACs.....	58

2.6.1	"Fluorescence in situ hybridization".....	58
2.6.2	Technique d'amplification d'exons à partir d'ADN génomique ("Exon trapping").....	59
2.6.2.1	Transfection de cellules Cos-1.....	60
2.7	Études de méthylation et d'expression avec des croisements interspécifiques.....	63
2.7.1	Technique SnuPE ("Single nucleotide primer extension").....	63
2.7.1.1	Études de Ms-SnuPE ("Methylation-sensitive single nucleotide primer extension") sur la séquence génomique de L59cand#2.....	64
2.7.1.2	Études d'expression de TMCL59 par SnuPE.....	65
2.7.2	Études de RT-PCR sur la séquence L59cand#1.....	65
<b>3</b>	<b>RÉSULTATS</b> .....	<b>67</b>
3.1	Clones obtenus par FP-CGI.....	68
3.2	Validation du FP-CGI sur des séquences de RMDs connues.....	68
3.3	Analyses d'ADN.....	71
3.3.1	Transferts Southern de granulocytes humains.....	71
3.3.1.1	Transferts Southern de triples digestions.....	80
3.3.2	Transferts Southern de différents tissus.....	87
3.4	Localisation chromosomique des clones méthylés différemment obtenus par FP-CGI.....	90
3.5	Recherche et utilisation de fragments génomiques de haut poids moléculaire: BACs et YACs contenant des clones FP-CGI.....	91
3.5.1	Technique d'amplification d'exon ou la recherche de gènes exprimés différemment .....	104
3.6	Analyses des séquences.....	104
3.6.1	Clone L59.....	104
3.6.2	Clone L88.....	108
3.6.3	D'autres clones.....	114



3.7 Études de méthylation et d'expression différentielles du clone L59.....	114
3.7.1 SNUPEs.....	114
3.7.2 RT-PCRs.....	115
<b>4 DISCUSSION.....</b>	<b>119</b>
4.1 FP-CGI.....	120
4.2 Localisations chromosomiques.....	122
4.3 Méthodes pour identifier des gènes avec empreinte génomique parentale.....	126
4.4 Importance des RMDs.....	128
4.5 FP-CGI et les études de méthylation dans le cancer.....	130
<b>5 CONCLUSIONS.....</b>	<b>131</b>
5.1 Perspectives.....	132
5.2 Régions méthylées différemment.....	132
<b>6 RÉFÉRENCES.....</b>	<b>133</b>

## LISTE DES FIGURES

Figure 1:	Structure de la 5-méthylcytosine.....	6
Figure 2:	Mécanismes de l'ADN Méthyltransférase.....	8
Figure 3:	Mécanisme pour la déméthylation de la 5-méthyl- cytosine.....	10
Figure 4:	Formation du complexe répresseur méthyl CpG – MeCP2 – Histone déacetylase.....	13
Figure 5:	Îlots CpG.....	18
Figure 6:	Régions chromosomiques murines portant des gènes potentiellement associés avec une empreinte génomique parentale.....	21
Figure 7:	Contrôle de l'empreinte parentale selon le modèle de Reik et Walter.....	24
Figure 8:	Modèle d'expression et compétition de Barlow .....	26
Figure 9:	Modèle de la compétition des "enhancers" de Tilghman.....	27
Figure 10:	Modèle d'inactivation de gènes suppresseurs de tumeurs en deux étapes: Rôle de la méthylation .....	37
Figure 11:	Technique de F-PERT.....	44

Figure 12:	Protocole FP-CGI.....	47
Figure 13:	Intégration des cartes physiques et génétique d'une région du chromosome 6.....	57
Figure 14:	Réprésentation schématique de la technique d'amplification d'exon.....	61
Figure 15:	Vecteur pSPL3.....	62
Figure 16:	Méthode SNUPE pour détecter de la méthylation ou de l'expression allèle-spécifique.....	66
Figure 17:	Produits de PCR obtenus par FP-CGI visualisés sur gel d'agarose 1,2%.....	69
Figure 18:	Transferts Southern des PCRs MAOA (a) et PGK-1 (b) faits sur de l'ADN de granulocytes femelle et mâle.....	72
Figure 19:	Hybridation Southern avec le clone L163.....	73
Figure 20:	Hybridation Southern avec le clone L51.....	75
Figure 21:	Carte de restriction du clone FP-CGI L88.....	76
Figure 22:	Hybridation Southern avec le clone L88.....	77
Figure 23:	Hybridation Southern avec le clone L88.....	78
Figure 24:	Hybridation Southern avec le clone L59.....	79

Figure 25: Carte de restriction du clone FP-CGI L211.....81

Figure 26: Hybridation Southern avec le clone L211.....82

Figure 27: Carte de restriction du clone FP-CGI L3.....83

Figure 28: Hybridation Southern avec le clone L3.....84

Figure 29: Hybridation Southern avec le clone L62.....85

Figure 30: Hybridation Southern avec le clone L61.....86

Figure 31: Hybridation Southern avec le clone L88.....88

Figure 32: Hybridation Southern d'ADN de différents tissus avec le clone L88 .....89

Figure 33: Criblage par PCR du panel 2 de Coriell menant à la localisation chromosomique du clone L88.....92

Figure 34: Criblage par PCR du panel 2 de Coriell menant à la localisation chromosomique du clone L3 .....93

Figure 35: Criblage par PCR du panel 2 de Coriell menant à la localisation chromosomique du clone L211 .....94

Figure 36: Criblage par PCR du panel 2 de Coriell menant à la localisation chromosomique du clone L59.....95

Figure 37: Criblage par PCR du panel G3 d'hybrides de radiation menant à la localisation chromosomique du clone L88.....96

Figure 38:	Intégration des cartes physique et génétique de la région du chromosome 22 entourant le clone L88.....	97
Figure 39:	Criblage par PCR du panel G3 d'hybrides de radiation menant à la localisation chromosomique du clone L3.....	98
Figure 40:	Criblage par PCR du panel G3 d'hybrides de radiation menant à la localisation chromosomique du clone L59.....	99
Figure 41:	Criblage d'une librairie de BACs avec un STS dérivé du clone L59.....	102
Figure 42:	Résultats de FISH effectué avec le clone L59 ou BAC 421M16....	103
Figure 43:	Homologies de séquences obtenues en utilisant le clone L59 dans une recherche BLAST.....	105
Figure 44:	Séquence nucléotidique du clone L59.....	107
Figure 45:	Homologies de séquences murines obtenues en utilisant THCL59 dans une recherche BLAST.....	109
Figure 46:	Homologies de séquences non-murines et non-humaines obtenues en utilisant THCL59 dans une recherche BLAST.....	110
Figure 47:	Des homologies de séquence entre le clone L59 et des ESTs de différentes espèces.....	112
Figure 48:	Positions des clones L53 et L88 dans un contig de clones génomiques situé sur le chr 22q11.....	113

Figure 49:	Études de méthylation différentielle du clone L59cand#2 par MS-SNuPE.....	116
Figure 50:	Études d'expression différentielle du clone L59cand#2 par RT-PCR-SNuPE.....	117
Figure 51:	Essais d'expression différentielle par RT-PCR pour le clone L59cand#1.....	118
Figure 52:	Régions chromosomiques humaines contenant des clones isolés par FP-CGI.....	123

**LISTE DES TABLEAUX**

Tableau I:	Quelques protéines qui lient préférentiellement l'ADN méthylé chez la souris ou chez l'homme.....	15
Tableau II:	Liste de gènes avec l'empreinte génomique parentale.....	22
Tableau III:	Caractéristiques de l'empreinte génomique parentale et ressemblances avec la méthylation de l'ADN génomique.....	29
Tableau IV:	Gènes suppresseurs de tumeurs inactivés par la méthylation du promoteur .....	36
Tableau V:	Liste des clones obtenus par FP-CGI associés à des séquences non-répétitives.....	70
Tableau VI:	Localisation chromosomique des clones L3, L59, L88 et L211 à l'aide des hybrides de radiation .....	100

**LISTE DES SIGLES ET ABRÉVIATIONS**

<	Moins de
>	Plus de
%	Pourcentage
μ	micro
μl	microlitre
<sup>32</sup> P	isotope radioactif du phosphore
'	minute
''	seconde
°C	degré Celsius
1p36	bande 36 du bras court du chromosome 1
6q21	bande 21 du bras long du chromosome 6
A	adénine
ADN	acide désoxyribonucléique
AMD	“allelic message display”
ARN	acide ribonucléique
ARNm	ARN messenger
BAC	chromosome artificiel de bactérie
C	cytosine
CEPH	Centre d'études des polymorphismes humains
Ci	Curie
coll.	Collaborateurs
chr	chromosome
d22s427	marqueur s427 sur le chromosome 22
dCTP	désoxycytidine-5'-triphosphate
dNTP	désoxynucléotide-5'-triphosphate
EDTA	éthylènediamine tetraacétate



EST	“expressed sequence tag”
F1	première génération
Fig.	figure
FP-CGI	“F-PERT on CpG islands”
F-PERT	“Formamide phenol-enhanced reassociation technique”
g	gramme
G	guanine
h	heure
http	“hypertext transfer protocol”
kb	kilobase
M	molaire
mb	megabase
mg	milligramme
ml	millilitre
mM	millimolaire
NaCl	chlorure de sodium
NaOH	hydroxyde de sodium
NCBI	National Center for Biotechnology Information
ng	nanogramme
pb	paire de bases
PCR	réaction en chaîne de la polymérase (“polymerase chain reaction”)
RLGS	“restriction landmark genomic scanning”
RMD	région méthylée différenciellement
RT-PCR	“reverse transcriptase polymerase chain reaction”
SMD	séquence méthylée différenciellement
T	thymine
TE	tris-EDTA
Tris	Tris (hydroxyméthyl) aminométhane
U	unité d’activité enzymatique
YAC	chromosome artificiel de levure

J'aimerais bien remercier tous ceux qui ont rendu possible la réalisation de cette thèse.

Je voudrais tout d'abord remercier mon directeur de recherche, Ted Bradley, pour sa patience, son appui et ses encouragements depuis que j'ai débuté ma maîtrise dans son laboratoire et qui ont continué tout au long de mon Doctorat. Je voudrais remercier Johane Morin qui m'a transmis à la veille de mon départ du labo une partie de son savoir-faire. Ses conseils et la facilité avec laquelle elle fait des manip ont donné un nouveau souffle au projet dans lequel je travaillais.

Je remercie Hélène et Martin pour leur aide et collaboration, Abdelmajid Belouchi, pour son aide inestimable, Maja Krajinovic et Daniel Sinnett pour leur aide avec les hybrides de radiation. Je remercie également des membres de l'ancien et du nouveau labo, spécialement Patrick, Raëd et Nathalie qui m'ont amicalement transmis jour après jour une partie de leurs connaissances et de leur expérience. Je tiens à remercier Philippe Robert et Aude LeGoffic qui ont investi une partie du temps qu'ils ont passé à West Hartford, CT, pour apporter des corrections à cette thèse.

Je tiens aussi à remercier l'Université de Montréal de m'avoir accordé une bourse d'exemption des droits de scolarité différentiels pour des étudiants étrangers sans laquelle je n'aurais jamais été capable de faire mes études de Doctorat.

Je voudrais remercier mes parents qui ont fait beaucoup d'efforts pour offrir de l'éducation à ma soeur, à mes frères et à moi, une éducation qui pourrait nous être utile dans l'avenir.

*“The truth is out there”*

**CHAPITRE 1**

**INTRODUCTION**

## 1.1 AVANT-PROPOS

Le séquençage du génome humain aura fini d'ici quelques années. Ce séquençage était vu il n'y a pas longtemps comme la source de presque toutes les réponses à des questions sur les gènes et sur leurs mécanismes et fonctions. Cependant, il est de plus en plus accepté que l'information héritable du génome humain ne se trouve pas que dans la séquence de bases d'ADN. Elle se trouve aussi chez des modifications que l'on appelle épigénétiques qui ajoutent d'autres niveaux d'information à celui fourni par la séquence sans que le code génétique soit changée.

Depuis quelques années nous assistons à une explosion de travaux et découvertes sur les rôles de la méthylation de l'ADN dans un grand nombre de processus épigénétiques. Cependant, les études sur la méthylation ne sont pas nouvelles, il faut remonter jusqu'en 1948, quand la première base modifiée, la 5-méthyl cytosine (Weisbach 1993) a été découverte. 1955 fut l'année de la découverte de la N6-méthyladénine, dans de l'ADN bactérien et 1964 celle de l'ADN méthyltransférase, chez *E.coli*, capable de méthyler de l'ADN non modifié. Après d'autres percées importantes concernant des aspects biochimiques plutôt que fonctionnels, nous arrivons à la fin des années quatre-vingt, où l'on commence à remarquer l'importance de la méthylation dans trois processus très importants pour la cellule: l'inactivation des gènes suppresseurs dans la progression tumorale, l'inactivation du chromosome X chez les femmes et l'empreinte parentale ou "imprinting". Plus spécifiquement, dans le cas de l'empreinte parentale, tous les gènes affectés par ce phénomène jusqu'ici montrent une méthylation différentielle allélique à la hauteur de leurs îlots CpG.

Selon Thomas (1995) il existerait dans le génome humain une centaine de gènes sous empreinte dont le quart ont été déjà trouvés. Il y a à présent des méthodes destinées à chercher de nouveaux gènes sous empreinte génomique parentale. Ces méthodes sont cependant très compliquées à réaliser et nulle d'entre elles ne cherche directement des gènes chez l'humain. Il est par conséquent important de souligner que la mise au point d'une méthode accessible pour chercher de nouveaux gènes sous empreinte chez l'humain constituerait un apport considérable à ce domaine.

L'étape initiale de ce travail a donc été entreprise dans le but de chercher des séquences méthylées différemment dans le génome humain. À l'aide de ces séquences nous serions en mesure de trouver de nouveaux gènes sous empreinte compte tenu du lien étroit qu'il existe entre ces gènes et la méthylation différentielle allélique de leurs îlots CpG.

## 1.2 MÉCANISMES ÉPIGÉNÉTIQUES

Selon ce qui est admis actuellement, les phénomènes épigénétiques sont ceux qui entraînent des changements d'expression dus à des modifications héréditaires mais réversibles au niveau de la structure de la chromatine, de l'ADN ou de protéines sans changer la séquence de l'ADN (Lewin, 1998; Henikoff et Matzke, 1997; Riggs et coll. 1996). La modification la plus connue de l'ADN est la méthylation de la position 5 de la cytosine du dinucléotide CpG. Il existe d'autres phénomènes qui jouent aussi des rôles importants dans l'inactivation génique. Il est préférable de noter que bien que ces phénomènes soient expliqués séparément ou bien qu'il soient souvent étudiés chez des organismes aussi différents que la Drosophile, la souris ou la levure, leurs mécanismes sous-jacents pourraient être étroitement liés.

L'acétylation des histones est un sujet récent mais qui attire de plus en plus d'intérêt. L'acétylation des groupements  $\epsilon$ -amino des lysines des histones H3 et H4, catalysée par un système d'enzymes comprenant les histones déacétylases et les histones acétyltransférases, est impliquée dans la formation de structures de chromatine active et inactive. L'acétylation est associée à l'activation de l'expression (Pazin et Kadonaga 1997). Il a été postulé que la neutralisation des charges qui a lieu après l'acétylation causerait une diminution de l'affinité des interactions ADN-histones, ce qui augmenterait l'accessibilité de certains facteurs de transcription à la chromatine (Kadonaga 1998). À l'opposé, des histones déacétylases telles que Rpd3, HDAC1 ou HDAC2, avec des represseurs transcriptionnels tels que Mad:Max ou Ume6 formeraient des complexes avec des protéines comme mSin3A et mSin3B. Ces complexes causeraient une inhibition de la transcription en amenant les histones déacétylases vers les régions promotrices.

La *paramutation* est un phénomène héritable que l'on retrouve plutôt chez les plantes et qui a son origine dans le changement d'expression d'un allèle qui est induit par un autre allèle (Hollick et coll. 1997). Les allèles qui sont sensibles à la paramutation sont appelés "paramutables" et ceux qui induisent la paramutation sont appelés "paramutagènes".

Dans le phénomène connu comme *PEV* (pour "position effect variegation" ou variation selon la position), l'hétérochromatine s'étend sur l'ADN avoisinant (euchromatine), et ce à des niveaux variables chez des cellules différentes (d'où vient le terme "variegation"), chez des mouches qui héritent des translocations chromosomiques. Ceci entraîne la perte d'expression des gènes de l'euchromatine, le degré de la perte dépendant de l'ampleur de l'étendue de l'hétérochromatine. De façon générale, *PEV* est un phénomène récessif, limité au gène situé sur le chromosome réarrangé (inactivation en "cis"). Cependant, les allèles variables ("variegating") de trois gènes, *brown+* (*bw+*), *Punch+* et *karmoisin+* sont dominants, entraînant une inactivation en "trans" de l'allèle non-réarrangé sur le chromosome homologue. L'hypothèse actuelle, émise par Henikoff et Matzke (1997), pour expliquer ce type d'inactivation en "trans", suggère que l'insertion de l'hétérochromatine dans l'une des copies de *bw* (l'allèle *bw<sup>D</sup>*) entraîne son déplacement et celui de l'autre copie (*bw+*) vers une position anormale au noyau où l'expression de la copie normale serait réprimée.

Récemment il a été démontré que le nombre de copies de séquences répétées, telles que des transgènes pourrait influencer le niveau d'expression de ces séquences et la structure de la chromatine qui les contient (Garrick et coll. 1998). En particulier, la présence d'un grand nombre de copies homologues en forme de concatémères pourrait avoir un effet répressif sur la transcription en menant à la formation d'une structure de chromatine inactive.

Un autre exemple d'implication d'informations autres que celle de la séquence est donné par la chromatine des centromères (Wiens et Sorger 1998). Chez les levures, les kinétochores sont assemblés sur des séquences centromériques qui ne sont

présentes qu'une fois sur les chromosomes. Néanmoins, il semblerait que l'information de ces séquences, chez la *Drosophile* et chez l'humain ne soit pas suffisante pour l'assemblage, les kinétochores pouvant s'assembler sur des séquences qui peuvent varier même dans une population de cellules génétiquement identiques.

Il existe chez la *Drosophile* un autre mécanisme épigénétique constitué par la repression de la chromatine qui dépend des protéines du groupe Polycomb (Pirrotta 1998). Ces protéines ne déterminent pas le profil initial d'expression des gènes sur lesquels elles agissent mais elles sont nécessaires pour le maintenir. Ceci implique que ces protéines reconnaissent des caractéristiques de ces gènes réprimés et qu'elles sont capables de les perpétuer. Un autre groupe de protéines, identifiées aussi chez la *Drosophile*, joue un rôle contraire, c'est à dire qu'elles contribuent à maintenir une expression forte de certains gènes.

Pour finir cette petite revue des mécanismes épigénétiques nous allons parler des prions. La définition la plus courante d'un prion est celle d'une particule infectieuse protéinique qui n'a pas d'acide nucléiques (Prusiner et coll. 1998). La forme modifiée ou infectieuse du prion (PrP) appelée PrP<sup>Sc</sup> est capable, une fois introduite dans l'animal hôte, d'induire la conversion de la forme normale (PrP<sup>C</sup>) en forme pathogène ou infectieuse. C'est donc un mécanisme épigénétique presque poussé à l'extrême où une forme déterminée peut causer une transformation juste en induisant un changement de conformation protéique.

### 1.3 LA MÉTHYLATION

La méthylation couvre un spectre d'au moins une demi-douzaine de modifications covalentes de l'ADN génomique de cellules procaryotes ou eucaryotes. Les modifications les plus communes chez les procaryotes sont la N<sup>6</sup>-méthyladénine (m<sup>6</sup>A) et la 5-méthylcytosine (m<sup>5</sup>C) (Rein et coll. 1998). Chez les eucaryotes la m<sup>5</sup>C (Fig. 1) constitue la base modifiée la plus abondante que l'on retrouve presque toujours dans



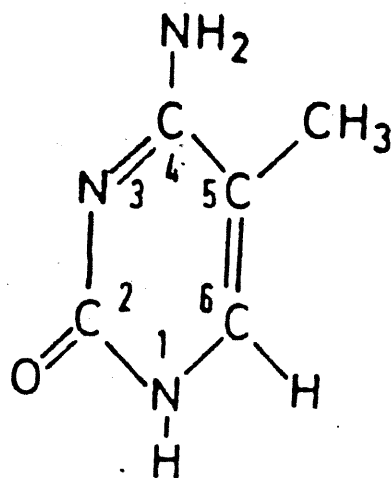


Fig. 1: Structure de la 5-méthylcytosine.

les dinucléotides CpG avec quelques évidences de sa présence dans les trinucleotides CpNpG (Clark et coll. 1995). Chez les mammifères, la m<sup>5</sup>C représente près de 1% des bases du génome. Elle se trouve distribuée d'une manière à peu près uniforme à travers le génome à l'exception des régions appelées "îlots CpG" où les cytosines ne sont pas toujours méthylées. Malgré leur caractère aléatoire les méthyl-CpGs ne représentent que 25% de la proportion attendue. Quelques hypothèses ont été proposées pour expliquer cette suppression des CpG dans le génome des mammifères mais celle qui retient notre attention est celle du "mutational hotspot" qui postule que la méthylation de la cytosine suivie par une déamination la changeant en thymine pourrait convertir une paire CpG en une paire TpG ou CpA (Schorderet et Gartler 1992). Il existe chez les mammifères un système de réparation de glycosylases qui identifie et répare les misappariements G/T et G/U. Cependant, la glycosylase G/T semble être beaucoup moins efficace *in vitro* que la glycosylase G/U probablement dû au fait que la thymine est une composante normale de l'ADN tandis que l'uracile ne l'est pas (Zingg et Jones 1997). Cette diminution de l'efficacité de réparation des misappariements G/T se traduirait dans un nombre diminué de dinucléotides CpG et une fréquence augmentée des dinucléotides TpG ou CpA.

### *1.3.1 La machinerie de la méthylation: méthylases, deméthylases et mécanismes de régulation.*

L'enzyme chargée d'établir et de maintenir les niveaux de méthylation chez les mammifères est l'ADN (5-cytosine) méthyltransférase (ADN Mtase) (Bestor et Verdine 1994). La séquence du domaine carboxy terminal de l'ADN Mtase humaine est très similaire à celle des bactéries. Mais au contraire des enzymes bactériennes, l'ADN Mtase des mammifères aurait une préférence pour de l'ADN hémiméthylé comme substrat. Le mécanisme utilisé par l'ADN Mtase (Verdine 1994) est illustré dans la Fig. 2.

Quelques évidences suggèrent très fortement que l'ADN MTase n'est pas la seule enzyme impliquée dans l'établissement des profils de méthylation: D'abord, cette

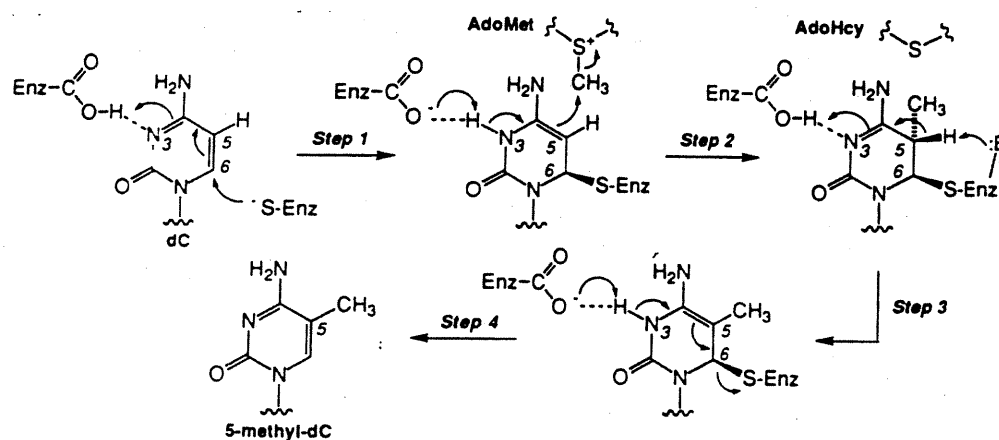


Fig. 2: Mécanisme de l'ADN méthyltransférase. Tiré de Bestor et Verdine 1994. La réaction est initiée par l'attaque d'une cystéine thiolate sur le carbone 6 de la cytosine, ce qui donne une énamine. L'énamine est attaquée à son tour par le groupement méthyl de la S-adénosyl-L-méthionine qui est finalement transféré au carbone 5.

enzyme a une préférence marquée pour l'ADN hemiméthylé suggérant qu'elle n'a pas d'activité déméthylase ou méthylase *de novo*, activités qui seraient importantes pour établir des profils de méthylation histospécifiques ou dans le développement embryonnaire. Ensuite, même si les souris qui ont les deux allèles déficients du gène de l'ADN Mtase (souris "knockout") ne sont pas viables, (ce qui accorde à la méthylation un rôle très important dans le développement) (Li et coll. 1992), l'inactivation de ces deux allèles apparemment n'a pas d'effet sur la viabilité des cellules ES qui sont homozygotes pour la mutation. Finalement, dans une expérience avec ces mêmes cellules ES infectées par un virus, l'infection déclenchait une méthylation *de novo* du provirus intégré (Lei et coll. 1996). Il semble y avoir aussi, au moins chez la souris, d'autres isoformes de l'ADN Mtase produits par épissage alternatif à partir de deux promoteurs du même gène (Pradhan et coll. 1997) et le clonage d'une nouvelle famille d'ADN MTases a aussi été récemment rapporté (Okano et coll. 1998). Aussi, un gène eucaryote cloné chez *Ascobolus* (un type de champignon) code pour une protéine semblable aux ADN Mtases mais qui ne contient pas le domaine amino terminal. L'inactivation de cette protéine n'a pas d'effet sur la viabilité ou le maintien des niveaux de méthylation mais elle empêche la méthylation *de novo* de quelques séquences répétées d'ADN (Malagnac et coll. 1997). Finalement, le groupe de Szyf (Battacharya et coll. 1999) a récemment cloné un ADNc codant pour ce qui semble être la première déméthylase à être caractérisée. Cette protéine contient un domaine de liaison au méthyl CpG et déméthyle un plasmide lorsque l'ADNc est transfecté transitoirement dans des cellules tumorales humaines *in vitro*. Cette déméthylase transforme les cytosines méthylées en cytosine et les produits de la réaction sont composés de cytosine et de méthanol (Fig 3).

Les mécanismes de régulation de l'ADN Mtase sont peu connus. Quelques auteurs, par exemple, ont émis l'hypothèse que cette enzyme fait partie de la voie de signalisation de la protéine Ras (MacLeod et coll. 1995) dû au fait qu'on est capable de contrôler ses niveaux en changeant les niveaux de Ras chez des cellules transfectées et dû aussi au fait que le promoteur de l'ADN Mtase contient des éléments de réponse

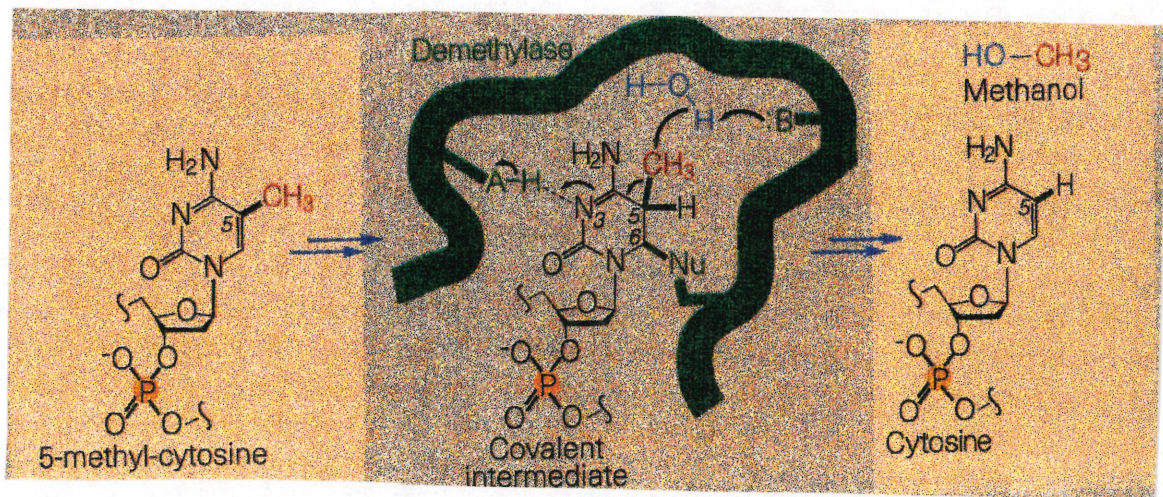


Fig. 3: Mécanisme pour la déméthylation de la 5-méthyl-cytosine. La déméthylase (en vert) forme un intermédiaire covalent par l'addition d'un nucléophile enzymatique ( $\text{Nu-H}$ ) qui attaque l'ion hydroxide qui est produit par l'activation "in situ" de l'eau. Les flèches doubles indiquent des réactions en deux étapes dont les intermédiaires ne sont pas montrés.

Tiré de Cedar et Verdine 1999.

pour AP-1, protéine qui fait partie de la voie de Ras. Un autre élément en faveur de cette hypothèse se trouve dans le fait que les niveaux d'expression de l'ADN Mtase sont trois fois plus élevés chez des cellules transformées avec *fos* (l'une des composantes d'AP-1) que chez des cellules normales (Bakin et Curran 1999). Il est aussi connu que l'ADN Mtase interagit avec des protéines telles que p21 (Chuang et coll. 1997), un régulateur du cycle cellulaire, et p23 (Zhang et Verdine 1996), une protéine qui se lie au récepteur de la progestérone mais dont la fonction exacte n'est pas encore connue.

### *1.3.2 Pourquoi l'ADN génomique des cellules des mammifères est-il méthylé?*

Deux hypothèses principales ont été proposées pour expliquer l'existence de la méthylation chez les mammifères. L'une d'elles propose que les changements de méthylation et de déméthylation qui ont lieu dans le génome pendant le développement ont un rôle direct dans le contrôle de l'activité génique (Riggs 1975). Elle se base sur des corrélations entre méthylation/déméthylation et répression/activation observées pour certains gènes et sur le fait que les profils de méthylation, sont, d'une façon générale, héréditaires. L'autre hypothèse, suggérée par Bestor et Tycko (1996), suggère que la méthylation fait partie d'un système de défense du génome qui servirait à inactiver des séquences parasites ou envahissantes telles que des transposons (Woodcock et coll. 1997), des virus ou même des séquences répétées (Kochanek et coll. 1993). Ce serait un système de défense de l'hôte similaire à celui qui existe chez les bactéries. La différence se trouvant dans le fait que chez les bactéries c'est le génome bactérien qui est méthylé à des séquences spécifiques, reconnues par certaines enzymes, pour le distinguer du génome étranger.

Les deux hypothèses sont appuyées par un certain nombre d'évidences et elles seraient plutôt non-exclusives. Cependant, de nos jours c'est la deuxième qui prend de plus en plus d'importance, dûe à un certain nombre de facteurs (Walsh et Bestor 1999). D'abord, à peu près 90% des résidus de m<sup>5</sup>C se trouvent au niveau des rétrotransposons et

des rétrovirus endogènes, des séquences qui représentent environ 35% du génome des mammifères (Yoder et coll. 1997). Ensuite, malgré la controverse, les gens estiment que malgré l'existence de bonnes corrélations entre la méthylation de certains gènes et la perte d'expression, personne n'a réussi à démontrer pour un gène déterminé qu'un certain profil de méthylation existant dans un tissu où le gène n'est pas exprimé est capable d'empêcher la transcription du même gène dans un tissu où le gène est transcrit.

Puis, Bestor et coll. (Walsh, C.P. et Bestor, T. H. 1999) croient que la méthylation a un rôle très important dans le génome mais pas celui de régulateur du développement des mammifères. Son rôle serait plutôt d'être impliqué dans des fonctions très spécialisées dans certains processus tels que l'empreinte génomique parentale, l'inactivation du chromosome X et la répression de séquences parasites.

Aussi attrayantes que soient ces hypothèses, il n'en reste pas moins que la méthylation est présente dans beaucoup de phénomènes qui sont importants pour les cellules de mammifères.

### *1.3.3 La structure de la chromatine et les enzymes qui lient les groupements méthyl.*

Il est connu depuis plusieurs années que la méthylation semble jouer un rôle dans la formation ou le maintien de la structure de la chromatine. Des expériences de microinjection et de transfection, par exemple, en utilisant des séquences méthylées *in vitro*, donnaient comme résultat la formation de structures de chromatine inactive (Keshet et coll. 1986). Récemment, il a été démontré qu'il existe un lien entre la méthylation et l'acétylation des histones dans la formation d'une structure de chromatine inactive ou active. Ce lien, c'est la protéine MeCP2, un inhibiteur de la transcription (Nan et coll. 1997) qui lie des méthyl-CpGs. Des travaux publiés récemment démontrent que la protéine MeCP2 réprime apparemment la transcription en liant des méthyl-CpG et en recrutant des histones déacetylases et d'autres facteurs de répression (Kass et coll, 1997; Nan et coll, 1998; Jones et coll., 1998) (Fig. 4).

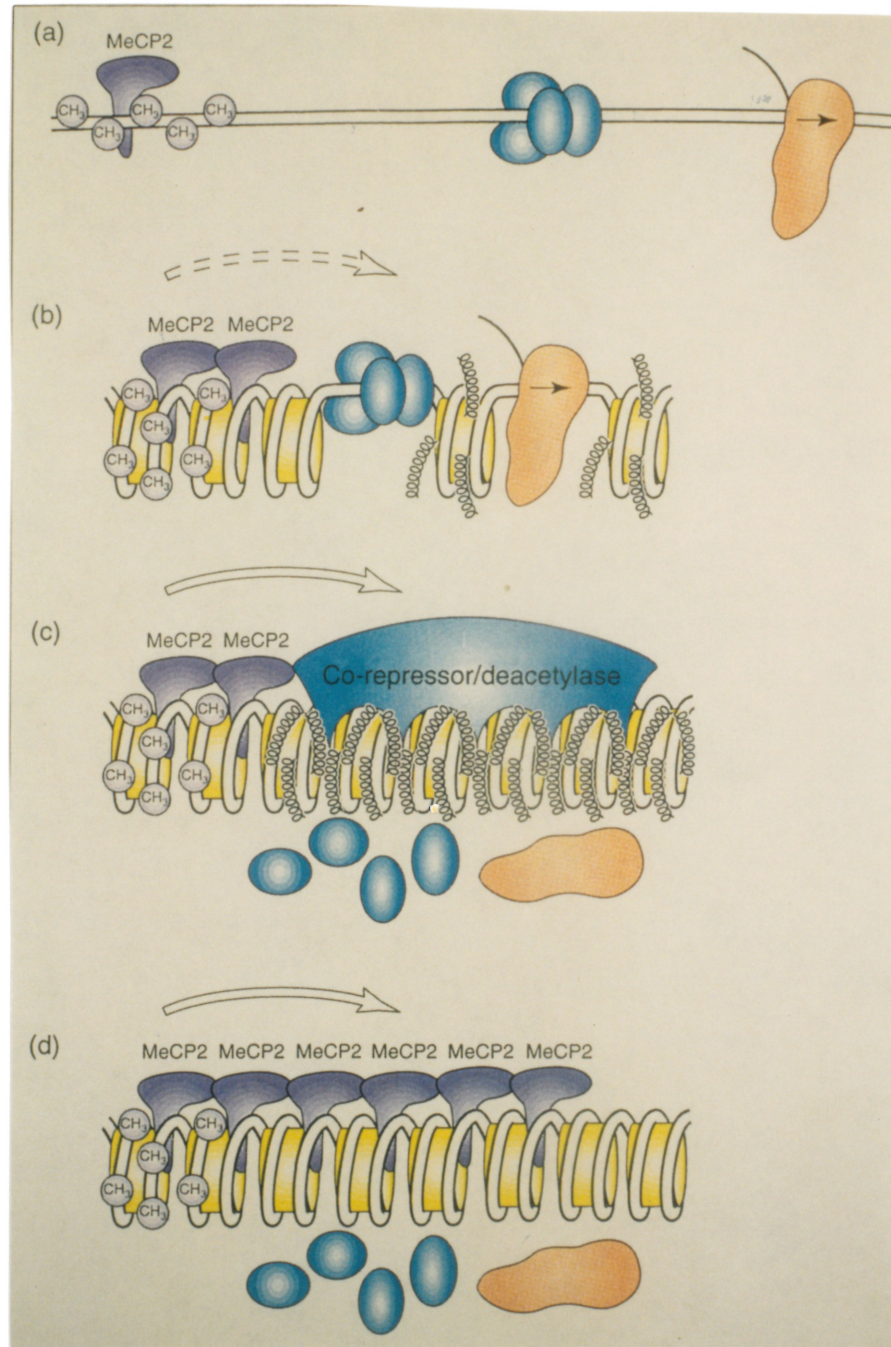


Fig. 4: Formation du complexe méthyl CpG - MeCP2 - histone déacétylase.



Depuis, d'autres travaux ont rapporté le même type d'interactions pour d'autres processus tels que l'empreinte parentale (Hu et coll. 1998), l'inactivation de transgènes (Pikaart et coll. 1998) et l'inactivation de gènes suppresseurs dans le cancer (Cameron et coll. 1999), sans oublier le fait connu de leur existence dans les mécanismes d'inactivation du chromosome X (Wakefield et coll. 1997). Dans le travail effectué sur les gènes suppresseurs, les auteurs expliquent quand même que dans les gènes étudiés: *p15*, *p16*, *hMLH1* et *Timp3*, la méthylation, et non la déacétylation des histones, est essentielle pour maintenir l'état de répression des promoteurs de ces îlots.

MeCP2 fait partie d'un groupe grandissant de protéines qui lient l'ADN méthylé (Tableau I), les plus connues étant MeCP1 et MeCP2 (Tate et Bird 1993). MeCP1 est un complexe de quelques chaînes polypeptidiques d'un poids moléculaire de 800 kd impliqué dans la répression de la transcription (Bird et coll. 1995). Le complexe de MeCP1 se lie spécifiquement à des séquences contenant de multiples CpG méthylés. Aussi, l'une des protéines de ce complexe (PCM1) contient un domaine qu'on trouve chez des ADN méthyltransférases et chez la protéine HRX de mammifères dont l'homologue chez la *Drosophile* est supposé réguler l'expression de gènes homéotiques (Cross et coll. 1997).

Le gène codant pour MeCP2 a été cloné et localisé sur le chromosome X et son domaine de liaison au méthyl-CpG a été caractérisé (Meehan et coll. 1992; Nan et coll. 1993). Contrairement à MeCP1, MeCP2 lie l'ADN méthylé d'une manière stœchiométrique, c'est-à-dire, que l'on n'a besoin que d'une paire CpG symétriquement méthylée pour que la liaison d'une protéine MeCP2 ait lieu. Il faut alors signaler qu'aucune de ces deux protéines ne lie de l'ADN hémiméthylé.

L'inactivation par recombinaison homologe ("knockout") du gène de MeCP2 chez la souris donne un phénotype très semblable à celui des souris "knockout" de l'ADN MTase et suggère que MeCP2 jouerait un rôle important dans le développement. Les souris homozygotes ne sont donc pas viables, avec une mort embryonnaire vers le jour

**Tableau I:** Quelques protéines qui lient préférentiellement l'ADN méthylé chez la souris ou chez l'homme<sup>(a)</sup>

Protéine	Séquence consensus de liaison	Références
MeCP1	(N <sub>x</sub> MGN <sub>x</sub> ) <sub>12</sub> (N <sub>x</sub> GMN <sub>x</sub> ) <sub>12</sub>	Tate et Bird 1993
MeCP2	N <sub>x</sub> MGN <sub>x</sub> N <sub>x</sub> GMN <sub>x</sub>	Tate et Bird 1993
MDBP1	RTMYRYYAMYRGMYRAY YAGRYRRT GR YCGR YTR	Tate et Bird 1993
MBD2, MBD3 MBD4	Pas rapporté	Hendrich et Bird 1998

(a) Tableau tiré et modifié de Tate et Bird 1993.

Abréviations: M= 5-méthylcytosine, R= A ou G, Y= C ou T, N= n'importe quelle base, x= n'importe quel nombre de bases.

10.5 (Tate et coll. 1996). Cependant, les cellules ES homozygotes pour la mutation, au même titre que les cellules ES mutantes pour l'ADN Mtase sont capables de survivre et de proliférer *in vitro*.

De nouvelles protéines s'ajoutent à la liste de protéines liant l'ADN méthylé. Chez le poulet, ARBP, une protéine nucléaire qui a une haute affinité pour les MAR/SARs a été récemment clonée (Weitzel et coll. 1997). ARBP contient un domaine de liaison au méthyl CpG semblable à celui du MeCP2 de rat. Et très récemment une nouvelle famille de protéines liant des méthyl-CpGs a été rapportée (Hendrich et Bird 1998) chez la souris et chez l'humain. Les trois protéines de cette famille, MBD2, MBD3 et MBD4 contiennent un domaine de liaison au méthyl-CpG et elles colocalisent avec l'ADN hautement méthylé (Tableau I).

#### 1.3.4 Réplication et recombinaison

La méthylation est probablement présente dans des fonctions aussi importantes pour la cellule que la réplication et la recombinaison. De récentes observations (Araujo et coll. 1998; Rein et coll. 1997) suggèrent que, au contraire de ce qui se passe chez *E. coli* où la méthylation a lieu quelque temps après la réplication, chez les mammifères la réplication et la méthylation auraient lieu simultanément. Comme il semble que l'ADN Mtase lierait des protéines impliquées dans la réplication ou le cycle cellulaire telles que PCNA et p21 (Chuang, L. S.-H. et coll. 1997), on pourrait croire que la méthylation jouerait un rôle dans la réplication en déterminant, par exemple, la position d'une origine de réplication (Rein et coll. 1997).

La méthylation serait aussi impliquée dans la recombinaison homologe. Selon des études faites chez *Ascobolus* (Maloisel et Rossignol 1998), la méthylation de séquences répétées servirait à diminuer leur fréquence de recombinaison homologe en réduisant l'efficacité des *crossing over*, ce qui serait un moyen de préserver la stabilité du génome. Malgré le fait que chez *Ascobolus* la méthylation n'est pas restreinte au dinucléotide CpG, cette même situation pourrait s'appliquer aux

mammifères étant donné que la plupart des méthyl-CpG se retrouve au niveau des séquences répétées.

### 1.3.5 Îlots CpG.

Les îlots CpG (Fig. 5) sont de petites régions (moins de 1-2 kb) riches en G+C (60-80%) constituant 1-2% du génome des vertébrés qui, contrairement au reste du génome, ne contiennent pas de méthyl-CpG (Antequera et Bird 1993; Cross et Bird 1995). Ces régions se trouvent généralement à la hauteur de l'extrémité 5' des gènes et ne montrent aucun type de suppression du dinucléotide CpG, il y a autant de CpGs que de GpCs. Les îlots CpG peuvent être identifiés en digérant l'ADN génomique avec des enzymes sensibles à la méthylation telles que *HpaII* ou *HhaI*. Le marquage et migration de ces digestions avait révélé des fragments très courts (HTFs pour "HpaII tiny fragment") en plus des fragment normaux de haut poids moléculaire qui contiennent des méthyl-CpG.

Selon quelques estimations (Antequera et Bird 1993), le nombre d'îlots CpG chez l'homme et chez la souris est de 45, 000 et 37, 000, respectivement. Comme il y a autant de gènes chez la souris que chez l'homme, cela indiquerait une carence en îlots CpG chez la souris comparativement à l'homme. Par contre, les îlots CpG de la souris sont moins riches en G+C que ceux de l'humain (Matsuo et coll. 1993). Il faut noter qu'on ne peut pas assumer l'existence d'un îlot dans un gène chez la souris ou une autre espèce si cet îlot existe chez l'homologue humain. Règle générale, tous les gènes "housekeeping" et 40% des gènes histospécifiques chez l'humain ont des îlots CpG. Cependant, les îlots ne sont pas distribués d'une manière uniforme sur le génome humain, ils se retrouvent sur ce qu'on appelle la chromatine active, euchromatine ou bandes R et T, certains chromosomes comme les chr 1, 9, 15, 16, 17, 19, 20 et 22 montrant une forte densité d'îlots (Craig et Bickmore 1994). Chez le poulet, la distribution est encore plus inégale, avec 75% des gènes se trouvant sur des

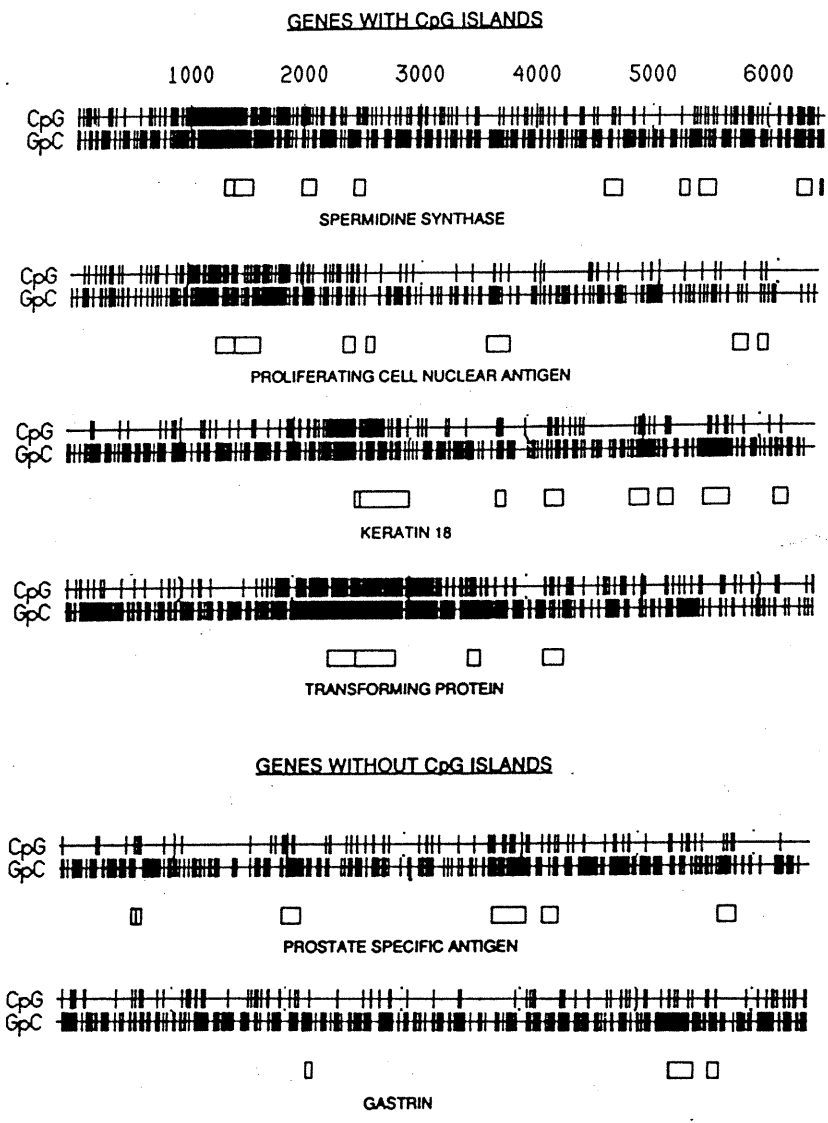


Fig. 5: Îlots CpG. Des cartes qui montrent les fréquences de CpGs et GpCs pour six gènes humains. Les exons sont indiqués par les carrés et chaque ligne verticale sur un exon signifie un début de traduction. Chaque CpG ou GpC est marqué par une ligne verticale.

Tiré de Antequera et Bird 1993b.

microchromosomes qui constituent à peu près 25% du génome (McQueen et coll. 1998).

Il y a des cas exceptionnels de méthylation des îlots CpG. Des cas ("imprinting", inactivation du chromosome X, progression tumorale) dont nous allons parler plus en détail après cette section. Mais, sinon, tout le reste des CpGs du génome sont méthylés. La question qu'on se pose, évidemment, c'est qu'est-ce qui empêche la machinerie de la méthylation de faire avec les îlots ce qu'elle fait avec les autres CpGs?. Certains auteurs (Cross et Bird 1995) y proposent quelques réponses. La première nous dit que l'ADN Mtase aurait des problèmes à méthyliser l'ADN qui est très riche en G+C. La deuxième suggère que les îlots CpG sont vraiment méthylés *de novo* mais qu'ils seraient déméthylés par une déméthylase spécifique des îlots. La troisième propose qu'il y a des facteurs tels que la protéine Sp1 (Macleod et coll. 1994), (Mummaneni et coll. 1998) qui se lient aux îlots et qui empêcheraient l'accès de l'ADN Mtase. Cependant, on ne connaît pas encore les raisons pour lesquelles les îlots CpG restent non-méthylés.

#### 1.4 L'EMPREINTE GÉNOMIQUE PARENTALE

Les génomes d'origine paternelle et maternelle ne sont pas équivalents et jouent un rôle complémentaire au cours du développement des mammifères. Des expériences de création d'embryons constitués de deux génomes mâles (androgénètes) ou femelles (gynogénètes ou parthénogénètes) montrent que les premiers participent de manière prédominante au développement des annexes extra-embryonnaires et les seconds à celui de l'embryon (Surani et coll. 1990). Cela suggérerait qu'il devrait y avoir un certain nombre de gènes qui ne seraient exprimés qu'à partir d'un des allèles selon l'origine parentale. Les défauts dans le développement refléteraient donc l'altération de la fonction d'un ou plusieurs gènes. Cattanach et Beechey (1990) ont fait des expériences avec des lignées de souris dont les zygotes avaient un contenu chromosomique diploïde normal mais chez lesquels tout ou une partie de chromosomes d'une paire d'homologues provenait d'un seul des deux parents. Ils ont

observé que des zygotes androgénotes ou parthénogénotes pour certaines régions aboutissaient à un développement embryonnaire anormal voire léthal. Ces travaux ont donc permis de décélérer une dizaine de régions ayant une différence fonctionnelle selon qu'elles étaient héritées du père ou de la mère (Fig. 6). Les génomes parentaux ont donc un marquage différentiel, une "empreinte" génomique, qui les distingue l'un de l'autre. Les gènes dont l'un des allèles est exprimé selon l'origine parentale doivent ainsi posséder une empreinte génomique parentale ("imprinting"). Par convention, l'allèle qui a l'empreinte génomique parentale (allèle sous empreinte) est celui qui n'est pas exprimé.

À ce jour, une trentaine de gènes sous empreinte parentale (Tableau II) ont été découverts, le premier étant *IGF2* ("insulin-like growth factor 2") en 1991. La liste croissante de gènes sous empreinte parentale et leur étude permettent de mettre en évidence quelques caractéristiques communes (Tableau III) dont certaines peuvent être impliquées dans les fonctions et mécanismes de l'empreinte génomique parentale. Une caractéristique importante est la présence de gènes sous empreinte parentale à l'intérieur de domaines chromosomiques de quelques megabases de longueur (Paulsen et coll. 1998) chez l'humain. Les domaines les plus connus étant celui sur le chr 15q11 impliqué dans les syndromes d'Angelman et de Prader Willi et celui sur 11p15.1 qui contient parmi d'autres les gènes *H19*, *IGF2*, *INS-2* et *p57*. Cette particularité pourrait faciliter l'action d'un mécanisme de contrôle régional partagé par les membres d'un même domaine. Par exemple, les domaines sous empreinte parentale se répliquent de manière asynchrone par rapport au reste du génome et par rapport à la même région du chromosome homologue (Kitsberg et coll. 1993; Knoll et coll. 1994; Grealley et coll. 1998). À cela s'ajoute le fait qu'il y a une association entre les homologues des domaines sous empreinte parentale du chromosome 15 paternel et maternel (LaSalle et Lalande 1996) et le fait qu'il y ait apparemment des différences dans la fréquence de recombinaison pour les méioses mâle et femelle des domaines sous empreinte (Pàldi et coll. 1995). La conservation de l'empreinte parmi différentes espèces de mammifères (Casparly et coll. 1998; Lighten et coll. 1997; Nezer et coll. 1999), sauf quelques

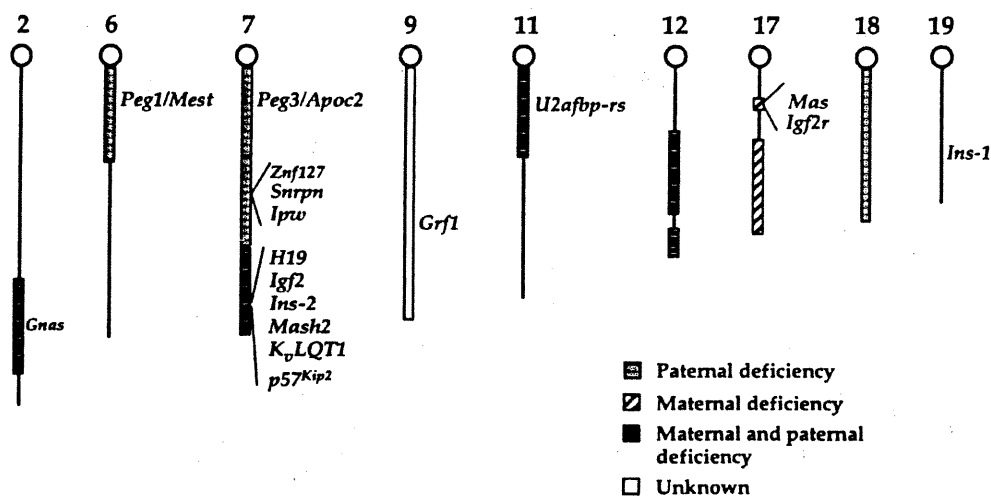


Fig. 6: Régions chromosomiques murines portant des gènes potentiellement associés avec une empreinte génomique parentale. Les rectangles montrent des régions associées avec un phénotype suite à des réarrangements de type maternel ou paternel, ou les deux.

Tiré de Bartolomei et Tilghman 1997.



**Tableau II:** Liste de gènes avec l'empreinte génomique parentale<sup>(a)</sup>.

Gène	Allèle exprimé	Chr (souris)	Empreinte chez l'homologue humain <sup>(b)</sup>	Références
<i>γ2-COP</i>	PR	6	Paternelle	Blagitko et coll. 1999
<i>Gnas</i>	Paternel	2	Oui	Hayward et coll. 1998
<i>Grf1</i>	Paternel	9	PR	Plass et coll. 1996
<i>H19</i>	Maternel	7	Oui	Bartolomei et coll. 1993
<i>Htr2</i>	Maternel	14	Oui	Kato et coll. 1998
<i>Igf2</i>	Paternel	7	Oui	DeChiara et coll. 1991
<i>Igf2r</i>	Maternel	17	qq cas	Barlow et coll. 1991
<i>Igf2r-as</i>	Paternel	17	PR	Wutz et coll. 1997
<i>IMPT1</i>	Maternel	7	Oui	Dao et coll. 1998
<i>Ins1</i>	Paternel	19	PR	Giddings et coll. 1994
<i>Ins2</i>	Paternel	7	PR	Giddings et coll. 1994
<i>IPL</i>	Maternel	7	Oui	Qian et coll. 1997
<i>Ipw</i>	Paternel	7	Oui	Wevrick et Francke 1994
<i>KvlQT1</i>	Maternel	7	Oui	Gould et Pfeifer 1998
<i>Magel2</i>	Paternel	7	Oui	Boccaccio et coll. 1999
<i>Mas</i>	Paternel	17	qq cas	Villar et Pedersen 1994
<i>Mash2</i>	Maternel	7	PR	Guillemot et coll. 1995
<i>Meg1</i>	Maternel	11	PR	Miyoshi et coll. 1998
<i>Necdin</i>	Paternel	7	Oui	MacDonald et coll. 1997
<i>p57</i>	Maternel	7	Oui	Hatada et Mukai 1995
<i>PAR-1</i>	PR	PR	Paternelle	Saito et coll. 1996
<i>PAR-5</i>	PR	PR	Paternelle	Saito et coll. 1996
<i>Peg1</i>	Paternel	6	Oui	Kanek-Ishino et coll. 1995
<i>Peg3</i>	Paternel	7	PR	Relaix et coll. 1998
<i>Peg5</i>	Paternel	2	PR	Kagitani et coll. 1997
<i>Snrpn</i>	Paternel	7	Oui	Leff et coll. 1992
<i>Ube3a</i>	Maternel	7	Oui	Rougeulle et coll. 1997
<i>Ube3a-as</i>	Paternel	7	Oui	Rougeulle et coll. 1998
<i>U2afbp-rs</i>	Paternel	11	Non	Hatada et coll. 1993
<i>Wt1</i>	PR	2	Maternelle	Jinno et coll. 1994
<i>Zfp127</i>	Paternel	7	Oui	Jong et coll. 1999
<i>Zfp127-as</i>	Paternel	7	Oui	Jong et coll. 1999
<i>Zim1</i>	Maternel	7	PR	Kim et coll. 1999

(a) Tableau tiré et modifié de Bartolomei et Tilghman 1997.

Abréviations: Chr= chromosome, PR= pas rapporté, as=antisens

exceptions (Riesewijk et coll. 1996), est aussi assez frappante. Il est également intéressant d'observer que les gènes sous empreinte contiennent des répétitions directes (Neumann et coll. 1995), que leur structure au contraire des autres gènes en général est composée de petits introns et de longs exons (Hurst et coll. 1996) et qu'ils codent dans certains cas pour des ARNs sens ou antisens non-traduits (Pfeifer et Tilghman 1994; Wutz et coll. 1997; Dittrich et coll. 1996; Moore et coll. 1997; Rougeulle et coll. 1998). Mais probablement l'une des caractéristiques les plus importantes des gènes sous empreinte est la présence d'une méthylation différentielle allélique à la hauteur de leurs îlots CpG (Jaenisch 1997).

#### *1.4.1 Comment l'empreinte est-elle établie dans les cellules germinales?*

Pour que l'empreinte génomique parentale (quelle qu'elle soit) soit fonctionnelle et qu'elle soit capable de marquer, de faire la distinction entre l'allèle maternel et l'allèle paternel ou viceversa, elle doit répondre à certaines exigences pour le gène en question (Bartolomei et Tilghman 1997): Premièrement, elle doit être établie dans les gamètes, la seule place où les allèles se trouvent dans des compartiments séparés et conséquemment peuvent être modifiés différentiellement. Deuxièmement, l'empreinte doit être stable et maintenue dans les tissus où l'empreinte est requise. Troisièmement, il faut que l'empreinte puisse être effacée et rétablie dans les cellules germinales pour la prochaine génération selon le sexe du porteur de ces cellules.

Quelques modèles ont été proposés pour essayer d'expliquer les mécanismes de l'empreinte génomique et la façon dont elle est établie, préservée et héritée. Nous allons en réviser les plus importants à la lumière des connaissances courantes.

Reik et Walter (1998) proposent la présence de "centres d'empreinte" (CIs) et d'éléments de maintien de l'empreinte" (EMIs) qui contiennent (Fig. 7) des régions riches en CpG qui sont méthylées différentiellement et se trouvant très souvent près de groupes de répétitions directes. Le CI serait un début de transcription d'ARNs, traduits ou non-traduits, qui pourraient contrôler les niveaux de méthylation de l'IME

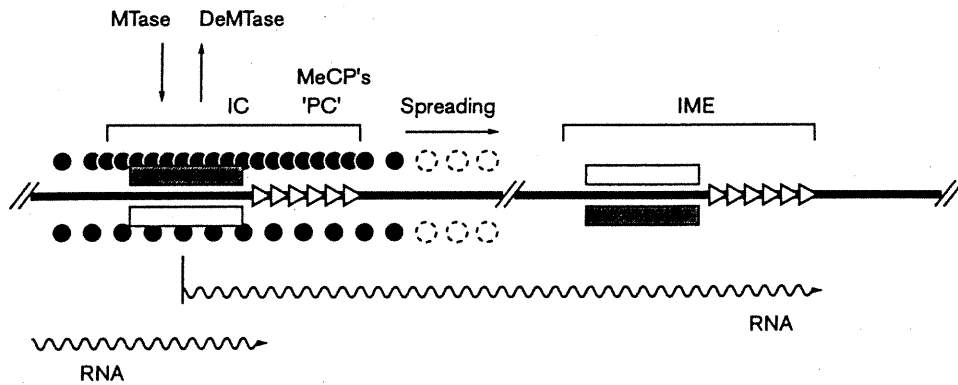


Fig. 7: Contrôle de l'empreinte parentale selon le modèle de Reik et Walter . Le centre d'empreinte (IC) est composé d'une RMD (boîte noire, méthylée; boîte blanche, non méthylée) et un groupe de séquences répétées ( $\Delta$ ). L'élément de maintien de l'empreinte (IME) a aussi des RMDs et des séquences répétées mais il n'a pas besoin d'avoir une RMD centrale.

Tiré de Reik et Walter 1998.

le plus proche. Ces éléments agiraient en concert avec les systèmes de méthylation (*de novo* et de maintien) et de déméthylation ainsi qu'avec des facteurs de transactivation pour s'adapter aux changements de méthylation et d'expression qui ont lieu durant le développement.

De son côté Barlow (1997) propose un modèle d'expression-compétition (Fig 8) où l'expression d'un gène sous empreinte n'est pas le résultat d'un allèle actif et d'un autre inactif, mais plutôt la conséquence d'une compétition-expression entre un gène empreinteur et sa cible, son voisin le gène sous empreinte. Dans ce cas-ci, le seul gène qui aurait l'empreinte serait le gène empreinteur. Ce modèle se base sur certaines évidences: D'abord, les deux allèles d'un gène sous empreinte parentale se retrouvent fréquemment associés à la chromatine active. Ensuite, très souvent les gènes sous empreinte très près l'un de l'autre montrent des profils contraires d'expression parentale-spécifique. Les gènes empreinteurs n'auraient pas besoin d'avoir de cadre lecture ouverte (comme c'est le cas pour *H19* ou *KVLQTI*). Ce modèle permet également d'expliquer certains résultats obtenus chez des souris déficientes pour l'ADN Mtase où certains gènes sous empreinte tels que *Igf2* et *Igf2r* ont été inactivés tandis que d'autres tels que *H19* ont montré une expression biallélique. Finalement si l'on essaie de corréliser seulement la méthylation avec l'expression des gènes empreinteurs, la corrélation est très bonne, c'est à dire, l'allèle non-méthylé est exprimé tandis que l'allèle méthylé n'est pas exprimé. Cela accentue l'importance des RMDs de ces gènes (Thorvaldsen et coll. 1998) à la lumière des résultats du groupe de Razin (Birger et coll. 1999) et des ressemblances assez frappantes avec le mécanisme d'inactivation du chromosome X telles que l'implication d'ARNs non traduits (Pfeifer et Tilghman 1994), la méthylation différentielle des promoteurs, et les différences dans le temps de réplication par rapport à d'autres régions (Xiong et coll. 1998).

Le modèle d'expression-compétition est une modification du premier modèle postulé pour expliquer le phénomène de l'empreinte génomique parentale, celui du "enhancer competition" (Fig. 9) du groupe de Tilghman (Bartolomei et coll. 1993). Ce modèle

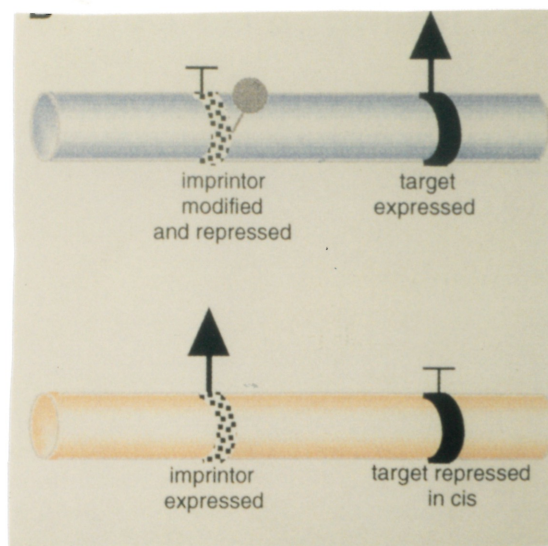


Fig. 8: Modèle d'expression et compétition de Barlow. Modification allèle-spécifique d'un gène "imprintor" (blanc et points noirs) qui se trouve en "cis" de sa cible, un gène sous empreinte (noir). Flèche, gène exprimé. Barre en forme de T, gène réprimé.

Tiré de Barlow 1997.

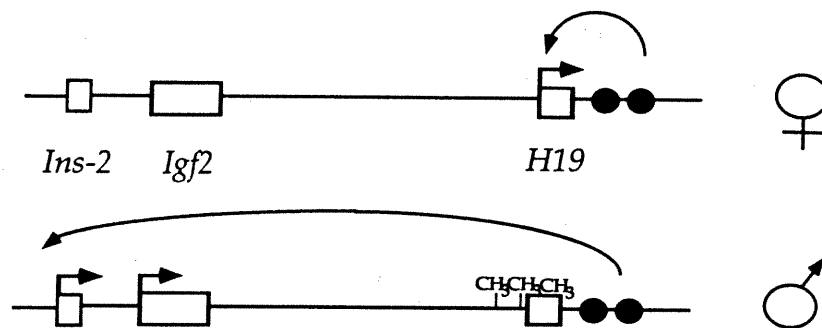


Fig. 9: Modèle de la compétition des "enhancers" de Tilghman. Les gènes sont indiqués par des "boîtes" et les allèles transcrits par des flèches horizontales. Les flèches partant des "enhancers" (cercles noirs) et la présence de la méthylation sur l'allèle paternel de *H19* (symbole CH<sub>3</sub>) indiquent leur interaction avec soit le gène *H19* dans l'allèle maternel, soit les gènes *Igf2* et *Ins-2* dans l'allèle paternel.

Tiré de Bartolomei et Tilghman 1997.

est basé sur les études faites sur les gènes *Igf2* et *H19*. Ces deux gènes sont exprimés paternellement et maternellement, respectivement, mais dans les deux cas, c'est l'allèle paternel qui est méthylé. Les auteurs proposent que les deux gènes utilisent les mêmes *enhancers*, mais le promoteur de *H19* étant un promoteur plus "fort" que celui de *Igf2*, *H19* est exprimé. Dans le cas où le promoteur de *H19* serait méthylé (allèle paternel), la transcription de *H19* serait inhibée et les *enhancers* seraient "libres" pour activer la transcription à partir du promoteur de *Igf2*.

#### *1.4.2 L'empreinte génomique parentale et la méthylation: est-ce que la méthylation est l'empreinte?*

Le meilleur candidat pour être l'empreinte génomique parentale est la méthylation. Elle répond aux critères d'établissement, de maintien et d'effacement de l'empreinte (Tableau III). Elle pourrait être établie dans les cellules germinales par des ADN Mtases sexe-spécifiques et/ou par certains facteurs activateurs ou repressors toujours inconnus. Son maintien serait assuré par la machinerie de méthylation à travers les différentes divisions des cellules somatiques. Et finalement elle pourrait être effacée dans la génération suivante par un système de déméthylation (Tilghman 1993).

En fait, la plupart des gènes avec l'empreinte génomique parentale identifiés jusqu'ici montrent une méthylation différentielle allélique à la hauteur de leurs îlots CpG (Tremblay et coll. 1995; Wutz et coll. 1997; Sutcliffe et coll. 1994). L'expérience la plus convaincante qui démontre l'implication de la méthylation dans l'empreinte génomique parentale a été fournie par le groupe de Jaenisch (Li et coll. 1993). Ils ont étudié chez des embryons de souris déficients pour l'ADN Mtase (Li et coll. 1992) l'expression de quelques gènes avec empreinte génomique parentale: *H19*, *Igf2* et *Igf2r* qui ont tous un profil caractéristique de méthylation différentielle. *H19*, méthylé dans l'allèle inactif, était exprimé d'une façon biallélique chez ces embryons tandis que *Igf2* et *Igf2r*, méthylés dans leur allèle actif, étaient reprimés.

Tableau III: Caractéristiques de l'empreinte génomique parentale et ressemblances avec la méthylation de l'ADN génomique<sup>(a)</sup>.

Méthylation	Empreinte parentale
Déméthylation chez les cellules germinales et les embryons précoces.	Signal de l'empreinte effacé chez les cellules germinales, retenu chez les embryons.
Méthylation <i>de novo</i> chez les cellules germinales et les embryons précoces.	Signal de l'empreinte établi chez les cellules germinales ou les embryons dicaryotes.
Profils de méthylation copiés par la mnDMNT au même allèle d'une façon stable après la réplication.	Signal restreint à un allèle parental d'une façon stable après la réplication.
Liée à des changements en <i>cis</i> de l'expression génique et de la structure de la chromatine.	Différences alléliques parentales de l'expression génique et du comportement des chromatides.

(a) Tiré de Neumann et Barlow 1996



Une question importante concernant l'établissement de l'empreinte et la méthylation est la suivante: si la méthylation constitue l'empreinte, que lui arrive-t-il lors de la déméthylation généralisée du génome des mammifères qui a lieu pendant le développement, plus spécifiquement avant l'implantation de l'embryon dans l'utérus? Certaines exceptions pourraient apporter des explications pour la déméthylation des îlots que l'on vient de décrire. Par exemple, des CpGs de l'îlot du gène antisens de *Igf2r* (Wutz et coll. 1997), du promoteur de *H19* (Tremblay et coll. 1995) et de celui du gène *Xist* (Zuccotti et Monk 1995) résistent à cette déméthylation globale de l'embryon. Il est intéressant de noter que ces trois gènes sont justement ceux qui, selon le modèle de Barlow, appartiendraient à la classe de gènes empreinteurs. Ces gènes codant pour des ARNs non-traduits, à différence de leurs cibles les gènes sous empreinte qui n'auraient pas besoin d'être méthylés, devraient garder des profils de méthylation différentielle pendant le développement. D'autres expériences concernant les gènes empreinteurs effectuées sur des cellules ES donnent des résultats très intéressants. Des cellules ES homozygotes pour l'inactivation de l'ADN Mtase (Li et coll. 1992) ont été transfectées avec un ADNc de l'ADN MTase. Cette transfection a élevé les niveaux de méthylation de l'ADN génomique et même ceux des gènes *Xist* et *Igf2* à des niveaux normaux (Tucker et coll. 1996). Cependant, elle n'a pas rétabli la méthylation des gènes *H19* et *Igf2r* ni les niveaux d'expression de *H19*, *Igf2* et *Igf2r*. Ces cellules ES ont été utilisées pour produire des chimères adultes et le rétablissement des niveaux normaux d'expression et de méthylation de ces trois gènes n'a été achevé qu'après un passage à travers la lignée germinale (lorsqu'il y a eu une nouvelle progéniture).

De récents travaux (Birger et coll. 1999) donnent un aperçu important du rôle de la méthylation dans l'empreinte génomique. En utilisant des délétions d'une RMD du gène *Igf2r* de souris (Wutz et coll. 1997), un gène qui est exprimé maternellement, ils ont identifié deux séquences: ADS ("allele-discrimination signal") et DNS ("de novo methylation signal") qui se trouvent dans une région de 113 pb qui est nécessaire pour permettre une méthylation *de novo* allèle-spécifique. DNS est une séquence de 8 pb (TCACCTCG) qui serait suffisante pour diriger une méthylation *de novo* de la RMD

dans le pronoyau maternel tandis que ADS est une séquence de 6pb (GTAGCC) qui empêcherait l'allèle paternel de subir une méthylation *de novo*. DNS et ADS se trouvent aux extrémités 3' et 5', respectivement, de la région de 113 pb. Aussi, selon ces mêmes travaux, cette "boîte" permettrait d'induire une méthylation monoallélique d'une séquence RMD (*Snrpn*) et non pas d'une séquence sans empreinte génomique parentale. Pour conclure, il semble y avoir des protéines spécifiques qui lient les séquences DNS et ADS. Ces protéines, encore inconnues, pourraient participer aux processus de méthylation *de novo* et de protection de l'allèle paternel d'*Igf2r*.

#### 1.4.3 L'empreinte génomique parentale et les maladies

L'empreinte parentale semble aussi être impliquée dans certaines maladies dont le cancer. Dans le cas du cancer il y a le phénomène dénommé LOI ("loss of imprinting") ou perte de l'empreinte (Issa et Baylin 1996) observé chez les gènes *Igf2* et *H19* dans certains types de tumeurs (Eggenschwiler et coll. 1997). Nous pourrions résumer la perte de l'empreinte dans ce cas-là comme une surexpression d'*Igf2* à cause de divers facteurs dont la méthylation biallélique de *H19* (Hashimoto et coll. 1995; Feinberg 1993).

L'un des principaux liens entre l'empreinte parentale et d'autres maladies telles que le syndrome d'Angelman ou le syndrome de Prader Willi n'est probablement pas inhérent à l'empreinte parentale mais à ses conséquences. En effet, un gène avec l'empreinte génomique est en désavantage par rapport aux autres gènes des mammifères qui se présentent toujours en deux copies. Pour un gène sous empreinte il n'y a pas de solutions de réchange si la copie active est perdue, l'autre copie est déjà inactivée. Cependant, si d'un point de vue évolutif la cellule a décidé de vivre avec le risque d'avoir une seule copie active de ces gènes c'est parce que les avantages l'emportent probablement sur les désavantages. C'est cette question qui nous amène au prochain titre de cette section sur l'empreinte génomique parentale.

#### 1.4.4 Pourquoi existe-t-il l'empreinte génomique parentale?

L'empreinte génomique parentale est observée chez les mammifères mais pas chez d'autres vertébrés et ces dernières années quelques théories ont été proposées pour expliquer l'évolution de ce phénomène. Nous allons vous en donner un aperçu des plus importantes.

Varmuza et Mann (1994) ont suggéré que l'empreinte génomique parentale est un moyen de protéger les femelles de mammifères de développer des trophoblastes invasifs. Le trophoblaste étant le tissu qui médie l'invasion de l'épithélium utérin maternel par l'embryon, ils suggèrent qu'une façon d'empêcher de développer une maladie invasive par le trophoblaste est d'inactiver chez les oocytes les gènes responsables de son développement.

Thomas (1995) a proposé que l'empreinte génomique soit un mécanisme de surveillance pour la perte de chromosomes afin de maintenir la diploïdie au cours des divisions cellulaires. On estime qu'il y a environ une centaine de gènes sous empreinte chez la souris et que ces gènes sont distribués aléatoirement dans le génome. Par conséquent, n'importe quel événement de perte chromosomique aboutirait à la perte complète d'expression d'au moins un de ces gènes. Une telle perte serait détectée par la machinerie de la cellule. Ce mécanisme de surveillance serait important dans au moins deux cas: la progression tumorale et les déséquilibres chromosomiques fœtaux.

Moore et Haig (1991) proposent que l'empreinte parentale soit la conséquence d'un conflit évolutif entre les gènes qui proviennent de la mère et ceux du père. À l'opposé d'autres phyla tels que les *oviparous* où les nutriments pour l'embryon sont déjà alloués et se trouvent dans l'œuf, les embryons des mammifères reçoivent leur allocation quotidienne directement de la mère. Les auteurs proposent que cette allocation soit dépendante du comportement reproductif de l'espèce. Les gènes soumis à l'empreinte paternelle auraient un effet négatif sur la croissance du fœtus,

contrairement aux gènes soumis à l'empreinte maternelle. Le parent mâle pourrait ainsi plus facilement perpétuer son génome dans la population s'il peut promouvoir de grands placentas et une grande descendance. Le parent femelle aurait une action plus évolutionniste en limitant la taille du placenta et de la descendance. Ainsi, la charge nutritionniste de chaque grossesse est réduite et des grossesses successives peuvent être supportées.

Le modèle de Moore et Haig implique que les gènes avec une empreinte maternelle soient des gènes qui favorisent la croissance tandis que ceux qui ont une empreinte génomique paternelle soient des gènes inhibiteurs de la croissance. C'est effectivement le cas pour des gènes tels que *Igf2* (exprimé paternellement) et de *H19* et *Igf2r* (exprimés maternellement) qui sont des inhibiteurs d'*Igf2*. Cependant, plus on identifie des gènes avec empreinte génomique, plus on voit que peu sont impliqués dans le développement et les cas d'*Igf2* et de *Igf2r* et *H19* semblent l'exception plutôt que la règle.

## 1.5 L'INACTIVATION DU CHROMOSOME X

Les femelles des mammifères inactivent l'un de leurs deux chromosomes X. C'est un mécanisme de compensation de dosage pour s'assurer que le niveau d'expression de leurs gènes sera équivalent à celui du seul chromosome X chez les mâles. Ce n'est pas le seul mécanisme possible. Chez la *Drosophila*, par exemple, la compensation est atteinte en doublant les niveaux de transcription du chromosome X chez les mâles. Et chez le nématode *C. elegans* l'expression des animaux XX est réduite de moitié.

### 1.5.1 L'inactivation du chromosome X et la méthylation.

Le processus d'inactivation du chromosome X (Heard et coll. 1997) implique quelques étapes dont celles d'initiation, de propagation et de maintien de l'inactivation. Ce n'est pas notre but de faire une revue exhaustive des mécanismes d'inactivation du chromosome X. Nous sommes plutôt intéressés à quelques détails de ces mécanismes et à voir ce qu'ils ont en commun avec d'autres processus et événements dont vous avez pris connaissance tout au long de l'Introduction.

L'inactivation du chromosome X, sauf quelques exceptions (Miller et Willard 1998), concerne le chromosome au complet. La méthylation joue probablement un rôle important dans ce processus. Cela est appuyée par une série d'évidences (Singer-Sam et Riggs 1993): Les îlots CpG des gènes sur le chr X inactif sont méthylés (Lock et coll. 1986; Toniolo et coll. 1988) ce qui constitue un autre cas de méthylation différentielle allélique. Les gènes sur le chromosome X inactif peuvent être reactivés avec des inhibiteurs de l'ADN Mtase tels que la 5-azacytidine. La méthylation de l'îlot CpG du gène *Xist* semble être importante pour l'inactivation du chr X.

*Xist* est un gène avec de petits introns et de longs exons codant pour un ARN non-traduit localisé dans le noyau (Brown et coll. 1992). Il est postulé que *Xist* fait partie du Xic ("X-inactivation center") qui est l'endroit où l'inactivation du chromosome X est déclenchée. Le promoteur de *Xist* sur le chr X actif, contrairement aux autres promoteurs sur le même chromosome, est méthylé, empêchant la transcription. À l'opposé, sur le chromosome X inactif, *Xist* est déméthylé et le gène est transcrit. L'ARN du *Xist* est retenu au noyau et des études de FISH démontrent qu'il couvre littéralement le chromosome X inactif (Sheardown et coll. 1997). Des études d'inactivation du gène de *Xist* (souris "knockout") montrent que *Xist* est nécessaire pour l'inactivation du chromosome X (Penny et coll. 1996).

## 1.6 LA MÉTHYLATION ET LA PROGRESSION TUMORALE

Les premières évidences sur le rôle de la méthylation dans le cancer proviennent d'études sur des lignées cellulaires faites aux années soixante-dix et au début des années quatre-vingt (Balmain 1995) où il a été démontré que la méthylation des sites CpG inhibait l'expression. Dans les années quatre-vingt l'idée courante était que la progression tumorale était favorisée par une hypométhylation du génome. Cette hypométhylation mènerait à l'activation de la transcription de certains oncogènes. Ce n'est qu'en 1995 qu'on a finalement eu des évidences très solides sur l'implication de la méthylation dans la progression tumorale *in vivo*. Laird et coll. (1995) ont croisé des souris déficientes pour l'ADN Mtase avec des souris *Min* (Su et coll. 1992) qui développent de multiples tumeurs intestinales à ~120 jours d'âge. La progéniture de

ces croisements a développé beaucoup moins de tumeurs intestinales et avait une espérance de vie plus longue que leurs parents. Ces résultats appuyaient la thèse qu'une méthylation *de novo* des îlots CpG (ICGs) de certains gènes suppresseurs favoriserait la progression tumorale et que l'hypométhylation déjà observée dans certains cancers proviendrait d'une déméthylation généralisée qui aurait lieu dans le reste du génome et non pas dans les ICGs. Peu de temps après, des travaux sur le gène suppresseur *p16*, inactivé par la méthylation de son îlot CpG (Merlo et coll. 1995) étaient publiés. Depuis, nous assistons à une explosion d'articles décrivant un grand nombre de gènes suppresseurs inactivés par une méthylation *de novo* de leur promoteur (Tableau IV).

Jones et Laird (1999) proposent que l'inactivation des gènes suppresseurs par méthylation constitue une alternative à l'hypothèse de Knudson. Ce dernier a proposé que deux événements mutationnels sont requis pour une inactivation complète d'un gène suppresseur de tumeurs. Les mutations les plus connues étant des mutations ponctuelles et/ou des pertes chromosomiques. Ainsi, la méthylation des séquences promotrices (Fig. 10) des gènes suppresseurs constituerait l'un de ces deux événements menant à l'inactivation complète de ces gènes .

En réalité, la progression tumorale serait juste un cas parmi d'autres où la méthylation serait impliquée dû au potentiel mutagénique des méthyl-CpG. C'est donc une situation qui pourrait concerner d'autres maladies. Par exemple le dinucléotide CpG constitue l'un des sites privilégiés de mutations du gène *p53* (Denissenko et coll. 1997) et d'autres gènes (Mancini et coll. 1997).

La méthylation favoriserait d'une manière indirecte certaines maladies en réprimant l'expression de gènes liés au contrôle de l'instabilité génomique. L'inactivation par méthylation du promoteur du gène *hMLH1* impliqué dans le système de réparation des misappariements d'ADN (Veigl et coll. 1998) et le taux élevé de mutations et de délétions chez des cellules ES déficientes pour l'ADN Mtase (Chen et coll. 1998) illustrent bien ce point. Il est quand même intéressant de noter la possibilité que grand

Tableau IV: Gènes suppresseurs de tumeurs inactivés par la méthylation du promoteur.

Gène	Type de tumeur	Référence
<i>APC</i>	colon	Hiltunen et coll. 1997
<i>BRCA1</i>	sein	Dobrovic et Simpfend 1997
<i>E-cadherin</i>	sein, prostate	Graff et coll. 1995
<i>ER</i>	sein	Ottaviano et coll. 1994
<i>HIC-1</i>	poumon	Wales et coll. 1995
<i>hMLH1</i>	colon	Kane et coll. 1997
<i>p16</i>	mélanomes et d'autres	Merlo et coll. 1995
<i>p15</i>	leucémies	Herman et coll. 1996
<i>RAR<math>\beta</math></i>	colon	Côté et Momparler 1997
<i>Rb</i>	rétinoblastome	Stirzaker et coll. 1997
<i>VHL</i>	rein	Herman et coll. 1994

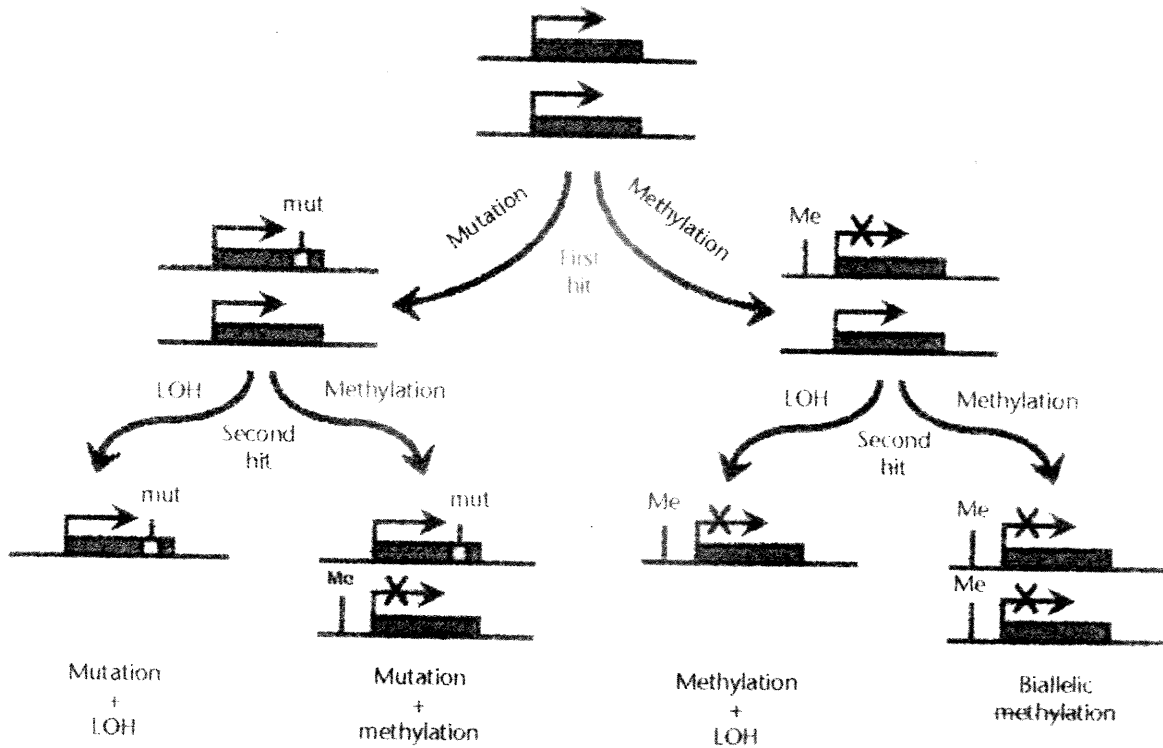


Fig. 10: Modèle d'inactivation de gènes suppresseurs de tumeurs en deux étapes: Rôle de la méthylation. Les deux allèles actifs d'un gène suppresseur sont indiqués par les deux rectangles. La première étape de l'inactivation du gène est montrée comme une mutation localisée à gauche ou par une répression de la transcription par la méthylation du promoteur à droite. Le second événement sera une délétion détectable par perte d'hétérozygotie, une répression de la transcription par méthylation ou une mutation ponctuelle.

Tiré de Jones et Laird 1999.



nombre d'événements qu'on vient de décrire soient des événements secondaires, des effets ou des conséquences d'autres événements déclenchants ou précoces trouvant peut-être leur origine dans des altérations au niveau du système de méthylation/déméthylation de la cellule (Szyf, 1996; Ramchandani et coll. 1997).

## 1.7 À LA RECHERCHE DE SÉQUENCES MÉTHYLÉES DIFFÉRENTIELLEMENT

Avant de commencer cette partie nous voudrions souligner quelques points que nous considérons très importants par rapport aux trois premières parties de l'introduction.

D'abord, par rapport à l'empreinte parentale. Il est connu que dans beaucoup de cas, l'empreinte est variable. Les rapports d'expression allèle exprimé/allèle non-exprimé n'étant pas de l'ordre de 100%/0% mais plutôt de 70%/30%, 80%/20% ou même de 50%/50% (biallélique normal). L'histospecificité pourrait être aussi très restreinte. *UBE3A*, par exemple, a une empreinte seulement au niveau du cerveau (Rougeulle et coll. 1997). L'empreinte peut varier aussi selon le stade de développement.

### 1.7.1 Étude de la méthylation

Grand nombre d'expériences décrites concernant l'importance de la méthylation dans la vie des mammifères n'auraient pas été possibles sans la disponibilité de techniques conçues pour étudier les méthyl-CpG. Les plus connues se basent sur l'impossibilité de certaines enzymes de digérer lorsque le(s) CpG(s) de leur site de reconnaissance se trouve méthylé (Nelson et McClelland 1991) et sur la capacité de certains agents chimiques de faire la distinction entre la cytosine et la 5-méthyl cytosine.

Généralement, dans la méthode basée sur l'utilisation d'enzymes de restriction, l'ADN est d'abord digéré avec une enzyme qui n'est pas sensible à la méthylation du CpG ou dont le site de reconnaissance ne contient pas de CpGs. Une partie de l'échantillon est ensuite digérée avec des enzymes sensibles à la méthylation ou préférablement avec

une paire d'enzymes qui ont le même site de reconnaissance mais l'une d'elles étant sensible à la méthylation et l'autre insensible (e.g. *HpaII* et *MspI*). Les deux digestions sont ensuite migrées en parallèle. Après des transferts Southern et des hybridations avec des sondes spécifiques, les profils de méthylation sont déterminés en comparant les longueurs des fragments provenant des doubles digestions avec celles provenant de la première digestion (une seule enzyme). Cette technique à quelques désavantages dont l'utilisation de grandes quantités d'ADN, ce qui empêche son utilisation dans des essais où les quantités d'ADN sont limitantes. Elle est aussi longue à faire et, parfois, les profils de méthylation sont difficiles à interpréter lorsqu'on soupçonne que l'enzyme utilisée n'a digéré que d'une manière partielle les échantillons à étudier. Des variations de cette technique (Gonzalzo et Jones 1997) utilisent la PCR (réaction en chaîne de la polymérase) avec des amorces spécifiques pour amplifier les fragments provenant des digestions avec des enzymes sensibles à la méthylation. L'amplification ne fonctionne que lorsque le(s) méthyl-CpG(s) se trouvant dans le site de reconnaissance de l'enzyme utilisée sera méthylé. Malgré le fait qu'elle peut être appliquée à de petites quantités d'ADN, pour cette technique il est cependant nécessaire de connaître la séquence contenant les CpGs pour être capable de faire le design des amorces. Et elle porte encore deux désavantages de la technique dont elle est dérivée: elle ne permet d'étudier que les CpGs contenus dans le site de reconnaissance de l'enzyme utilisée et il faut tenir en compte que les profils de méthylation obtenus pourraient provenir d'une digestion partielle.

Au cours des dernières années on a vu l'émergence de nouvelles méthodes (Clark et coll. 1994, Kubota et coll. 1997, Rein et coll. 1998) basées sur la capacité du bisulphite à convertir la cytosine, et non pas la 5-méthyl cytosine, en uracile dans de l'ADN simple brin. Les 5-méthyl cytosines sont amplifiées comme des cytosines avec des amorces spécifiques aux brins modifiés. Ensuite, le séquençage des produits de PCR nous donne une idée claire des profils de méthylation de TOUS les CpGs se trouvant sur la région modifiée et amplifiée. Elle constitue donc une méthode très puissante pour déterminer le degré de méthylation de régions très riches en CG telles que les îlots CpG. Un autre avantage de cette méthode de PCR comparée à celles qui

utilisent l'hybridation Southern provient du fait que l'on peut travailler avec de très petites quantités d'ADN, ce qui est critique pour certaines expériences. Cette méthode est cependant ardue et les résultats sont difficiles à interpréter lorsque la réaction avec le bisulphite n'a pas été complète et un certain nombre de cytosines ne sont pas converties en uracile.

### *1.7.2 Comment analyse-t-on des séquences inconnues?*

De nos jours, l'analyse informatisée est une partie importante du travail quotidien du chercheur dont le but est d'identifier de nouvelles séquences. Lorsqu'il a des clones séquencés, la première question qu'il se pose est: est-ce que c'est quelque chose de connue?. Pour répondre à cette question l'on dispose aujourd'hui d'outils informatiques qui généralement sont disponibles pour n'importe quel chercheur disposant d'un appareil informatique de base. Cette retombée du Projet du Génome permet, malgré des moyens limités, de concurrencer jusqu'à un certain point avec des labos qui possèdent des ressources astronomiques à leur disposition.

L'un des logiciels les plus utilisés est BLAST (Altschul et coll. 1997) disponible sur le web du NCBI ("National center for biotechnology information", <http://www.ncbi.nlm.nih.gov/>), qui permet de comparer nos séquences avec des millions de séquences (ADN, protéines) se trouvant dans des banques telles que GENBANK ou EMBL. L'augmentation du nombre de données concernant les séquences exprimées (ESTs) (Marra et coll. 1998) ainsi que les séquences provenant d'autres espèces telles que le rat, la souris ou la levure permet de déterminer si notre séquence est exprimée, si elle fait partie de certaines familles de séquences répétées ou de déterminer sa présence et conservation parmi les espèces. L'abondance de ESTs permet aussi de construire de séquences exprimées plus longues, des THCs (ou "tentative human consensus" dans le cas des séquences humaines) à partir des chevauchements des ESTs. Il existe aussi d'autres types de logiciels qui permettent de chercher des cadres de lecture ouverte (e.g. GMAIL) (Claverie 1997) ou qui comparent notre séquence (si elle est exprimée) avec des domaines connus de protéines pour

déterminer s'il y a des domaines conservés (e.g. BLOCKS, PROSITE) (Bork et Koonin, 1998; Rastan et Beeley, 1997).

Un autre type de ressource informatique à la disposition des chercheurs est la cartographie par hybrides de radiation (McCarthy et coll. 1997). Les hybrides de radiation sont des lignées de hamster fusionnées avec des lignées humaines. La lignée humaine a été traitée avec un type de radiation qui cause des cassures dans l'ADN donnant des fragments de quelques megabases de longueur. Chaque lignée de hamster retient un certain pourcentage de ces fragments. Donc, une séquence humaine quelconque sera représenté dans un certain nombre de ces fragments mais probablement pas dans toutes les lignées. Toutes ces lignées hybrides, soit de hamster, de souris ou humaines ont été testées pour un certain nombre de marqueurs (10,000-20,000) dont on connaît bien la localisation chromosomique. En faisant un PCR spécifique pour un clone donné sur ces lignées hybrides, il est possible de localiser un clone en faisant parvenir au site web les résultats du PCR. On obtient une réponse sous la forme d'un "LOD Score" (logarithme decimal du rapport de vraisemblance) indiquant l'identité du marqueur qui est le plus étroitement lié ("linkage") avec notre clone.

### *1.7.3 L'isolement d'îlots CpG différenciellement méthylés par FP-CGI.*

Le but de notre projet il y a quelques années était de chercher une façon rapide et simple d'identifier de nouveaux gènes avec l'empreinte parentale. Une telle méthode constituerait un outil indispensable et un apport considérable au domaine. Nous avons décidé de faire face au problème en tirant profit d'une caractéristique des gènes avec empreinte parentale: la méthylation différentielle de leurs îlots CpG. Nous avons donc développé une méthode, FP-CGI (pour "F-PERT on CpG Islands") basée sur deux caractéristiques (FP-CGI va être expliquée en détail dans la section de matériel et méthodes):

La première était l'utilisation de deux groupes d'enzymes: Le premier groupe étant composé d'enzymes sensibles à la méthylation dont le site de reconnaissance est riche

en C+G e.g. *BssHII* (GCGCGC), *EagI* (CGGCCG), *SacII* (CCGCGG) . Elles vont donc couper préférentiellement à l'intérieur des îlots CpG (ce sont des "rare-cutters"). Le deuxième groupe est constitué d'enzymes qui coupent préférentiellement à l'extérieur des îlots à cause de leur site de reconnaissance, riche en A+T e.g. *Tsp509I* (AATT), *MseI* (TTAA), *NlaIII* (CATG).

La deuxième était d'utiliser une technique publiée il y a une quinzaine d'années, F-PERT (Casna et coll. 1986) pour "Formamide-phenol-enhanced reassociation technique", qui permet une réassociation rapide de séquences uniques qui ont été préalablement dénaturées (Fig.11). F-PERT est à son tour une modification d'une autre technique, PERT, publiée quelques années auparavant (Kohne et coll. 1977). L'avantage de F-PERT par rapport à PERT est qu'elle permet une augmentation substantielle des fragments réassociés provenant de séquences uniques sans avoir trop de dommages oxydatifs dus au phénol et ce probablement à cause de l'addition de la formamide.

Ces deux caractéristiques se combinent avec deux avantages des méthodes utilisées pour étudier la méthylation (section 1.7.1): l'utilisation d'enzymes sensibles à la méthylation et la PCR. Cela permettrait de chercher des séquences uniques méthylées différemment d'une façon allélique à la grandeur du génome humain (ou de n'importe quel génome de mammifères) sans utiliser de quantités importantes d'ADN.

Nous avons utilisé de l'ADN des granulocytes humains comme source d'ADN dans notre étude parce qu'il constitue un tissu homogène qui est facile à obtenir. Mais pour atteindre notre objectif principal, celui de trouver de séquences avec empreinte génomique parentale, il était nécessaire de caractériser ces séquences et déterminer si elles proviennent de régions démontrant un biais parental. Dans le cas de séquences inconnues, il fallait les localiser dans le génome, les cartographier afin de déterminer leur proximité des régions avec empreinte génomique. Il fallait aussi utiliser ces séquences pour isoler des fragments génomiques humains de haut poids moléculaire afin d'envisager des stratégies de clonage, et chercher des séquences codantes et/ou

faire des études d'expression. L'utilisation d'homologues murins de ces séquences était aussi indispensable à cause de la facilité de travailler avec les polymorphismes qui existent dans les croisements interspécifiques murins. Une autre stratégie envisagée était d'utiliser les séquences 5' codantes que nous avons identifiées pour chercher les promoteurs de ces gènes-là. Le but ultime, étant d'étudier le statut de méthylation différentielle allélo-spécifique des séquences clonées avec notre méthode pour ensuite évaluer l'expression monoallélique selon l'origine parentale.

## Principe du F-PERT

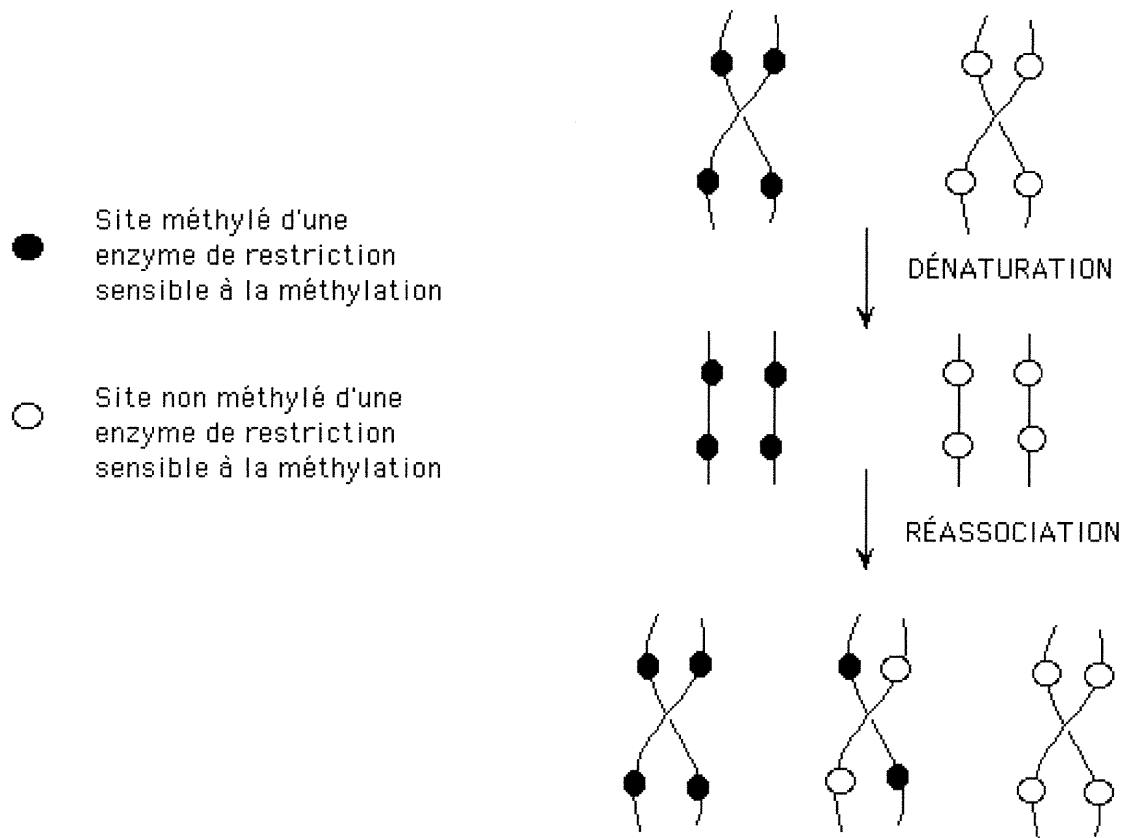


Fig. 11: Technique de F-PERT ("formamide-phenol enhanced reassociation technique"). La présence de phénol et de formamide permet une réassociation rapide de séquences uniques préalablement dénaturées. Dans le cas de séquences portant des profils différents de méthylation sur les deux allèles, la réassociation permet la formation d'une troisième population de molécules d'ADN hétéroduplexes en plus des deux populations originales d'homoduplexes méthylés et non-méthylés.

## CHAPITRE 2

### MATÉRIEL ET MÉTHODES



## 2.1 FP-CGI

Le protocole utilisé pour chercher des régions ou séquences méthylées différemment dans le génome humain est illustré à la Fig. 12. Le matériel de départ de chaque expérience consistait en 180 µg d'ADN extrait de granulocytes qui constituent une population relativement homogène de cellules faciles à obtenir. L'ADN a été digéré avec l'enzyme *MseI* qui reconnaît le site TTAA et qui a donc une tendance à couper à l'extérieur des ICGs, qui sont généralement riches en C+G à 55-80% (étape 1). Les fragments entre 0.6 et 7 kb ont été récupérés après une séparation électrophorétique, dénaturés et realignés en utilisant F-PERT (étape 2) dans le but de promouvoir une réassociation rapide de brins complémentaires uniques d'ADN dénaturés tout en diminuant la dégradation due aux dommages oxydatifs (Casna et coll. 1986). La réassociation devrait produire une population de molécules d'ADN hétéroduplexes générées à partir des ICGs qui portent des profils différents de méthylation sur les deux allèles. Dans le cas extrême, l'hétéroduplex serait hémiméthylé avec 100 % des CpGs méthylés sur un brin et 0 % sur l'autre brin.

L'ADN hybridé par F-PERT a été ligué à des adaptateurs *MseI* biotinylés (étape 3), mélangé avec des billes conjuguées à de la streptavidine et lavé. Cette étape enrichit notre population en molécules double brin issues de la réassociation de molécules simple brin parfaitement complémentaires. Les brins reassociés d'une façon imparfaite ne reformeraient pas le bout TA 5' requis pour la ligation aux adaptateurs. L'ADN a été ensuite digéré avec une enzyme sensible à la méthylation telle que *EagI* (étape 5), dont le site de restriction contient deux CpGs. L'ADN est ensuite dénaturé par un traitement alcalin (étape 6).

Après un lavage, l'ADN monocaténaire encore attaché aux billes a été hybridé avec une amorce de 12 nucléotides (12-mer) contenant le site de restriction pour *EagI*, ce qui permet de recréer localement un site de restriction *EagI*. Une nouvelle digestion avec *EagI* (étapes 7 et 8) permet de récupérer des fragments d'ADN simple brin enrichis en séquences qui étaient hémiméthylées après les étapes du F-PERT. Ces

**FP-CGI**  
**MÉTHODE POUR**  
**ISOLER DES RMDs**

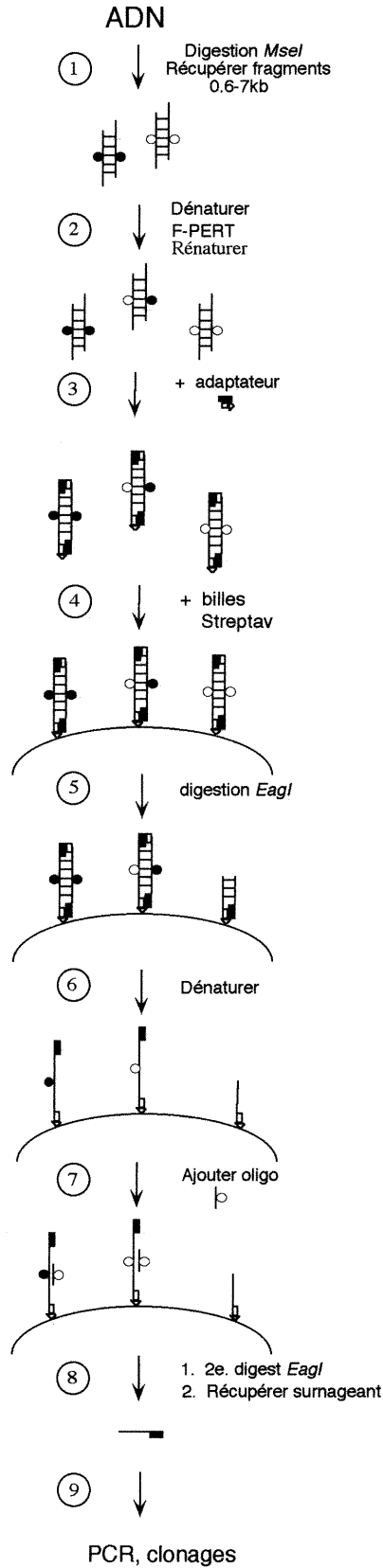
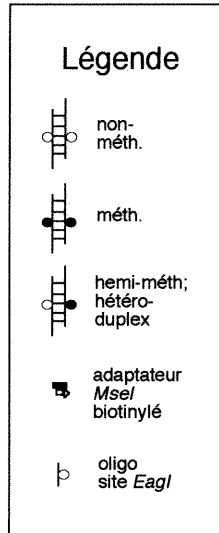


Fig. 12: Protocole FP-CGI.

séquences d'ADN simple brin porteraient une séquence ML-1, provenant de l'adaptateur *MseI* qui aurait été ligué à une extrémité, et devraient être riches en CG sur l'autre extrémité. Nous les avons donc amplifiées (étape 9) en utilisant l'amorce ML-2, complémentaire à ML-1, et un deuxième amorce arbitraire mais choisi quand même à cause de son haut pourcentage en C+G à l'extrémité 3'. Les produits résultants ont été séparés sur un gel d'agarose 1.2% et nous en avons essayé de cloner quelques fragments.

### 2.1.1 F-PERT sur de l'ADN de granulocytes humains

Le F-PERT d'ADN de granulocytes a été pratiqué selon le protocole de Casna et coll. (1986) et de Nelson et coll. (1993). Les granulocytes ont été préparés à partir de 50 ml de sang de donneurs femelles ou mâles en utilisant la trousse Polymorphprep<sup>TM</sup> de Nycomed Pharma (Oslo) selon les instructions du fabricant. L'examen microscopique de la préparation colorée avec la solution Turk a montré un degré de pureté de l'ordre de plus de 98%. L'ADN a été extrait par des procédures standard (Sambrook et coll. 1989), digéré pendant la nuit avec l'enzyme *MseI* (New England Biolabs, 1-2 U/ $\mu$ g ADN) et fractionné sur un gel 1.2 % d'agarose de bas point de fusion ("Low melting point agarose", Gibco BRL). La fraction d'ADN située entre 0.6 et 7 kb a été récupérée par une extraction phénol-chloroforme après avoir fait fondre l'agarose par une incubation à 65°C (Sambrook et coll. 1989). 5  $\mu$ g de cette fraction ont été résuspendus dans 100  $\mu$ l d'eau, dénaturés avec du NaOH 0.3 M pendant 15' à température de la pièce et neutralisés à l'aide d'un tampon (MOPS 0.4M). Le volume final a été ensuite ajusté à 400  $\mu$ l d'une solution dont la concentration finale était de thiocyanate de sodium 2M, Tris pH 8.0 10 mM, EDTA 0.1 mM et formamide 8%. Du phénol saturé avec de l'eau a été rajouté à cette solution jusqu'à ce que une émulsion soit formée, c'est à dire un volume d'entre 175-200  $\mu$ l. Les échantillons ont été ensuite agités constamment sur une plaque ADAMS Nutator, Modèle No. 1105 (Clay Adams, B-D) pour maintenir l'émulsion et ce pendant 24 heures à température de la pièce. L'ADN a été récupéré après deux extractions au chloroforme et une précipitation à l'éthanol, et resuspendu dans 80  $\mu$ l d'eau à une concentration finale de 0,25  $\mu$ g/ $\mu$ l.

### 2.1.2 Ligation des adaptateurs *MseI*

Les adaptateurs *MseI* ont été préparés en dénaturant à 95°C une solution de 1µg/µl des amorces ML-1 et ML-2 biotinylé (5'-TAG GGG CGA ATT CCG CTT AAC-3' et 5'-GTT AAG CGG AAT TCG CCC A-3') et en la laissant refroidir à T° de la pièce, sur une période de quelques heures. 8 µl (~2 µg) d'ADN F-PERT ont été ligaturés (0,5 U de T4 DNA ligase, BRL) dans un volume final de 15 µl pendant 12-16 heures à 14°C à 1 µg d'adaptateur *MseI* biotinylé, récupérés par extraction phénol-chloroforme, puis précipités à l'éthanol et resuspendus dans 40 µl de TE 7.5 ((Tris pH 7.5 10 mM, EDTA 1mM). Les oligonucléotides ont été synthétisés chez University Core DNA Services, University of Calgary, Alberta, Canada. 200 µg de billes magnétiques (Dynabeads M-280 streptavidin, Dynal, Inc. Norvège) ont été préparés selon les instructions du fabricant, c'est à dire en les sédimentant avec un aimant (Promega, Cat# Z5332) pour 1 min, en enlevant soigneusement tout le surnageant, en lavant avec 60 µl de solution de liaison ("binding solution", NaCl 3M, Tris pH 7.5 10 mM, EDTA 1mM) et en resuspendant finalement dans 40 µl de "binding solution". Cette étape permet de sélectionner des spécimens contenant l'adaptateur biotinylé: oligo seul, adaptateur seul ou adaptateur + ADN-FPERT. 40 µl d'ADN F-PERT (~3 µg) rehybridé et puis ligaturé à l'adaptateur biotinylé ont été rajoutés et incubés à 43°C pour 60 min en agitant occasionnellement. Les billes ont été sédimentées avec l'aimant, le surnageant enlevé, lavées avec la solution de liaison pour être finalement resuspendues dans 40 µl de TE pH 7.5. Cette suspension était désormais prête pour la première digestion avec *EagI*. Cette digestion va permettre de couper les spécimens contenant les deux brins non-méthylés et de sélectionner les homoduplexes et les hétéroduplexes hémi-méthylés. Après la digestion avec *EagI*, les billes ont été sédimentées et le surnageant enlevé. Les billes ont été résuspendues dans 11 µl de NaOH 0.1 N, incubées à T° de la pièce pour 5 min et sédimentées. Les billes ont été ensuite lavées séquentiellement avec 20 µl de NaOH 0.1N, 50 µl de "binding solution" et 50 µl de TE 7.5 et finalement résuspendues dans 20 µl d'H<sub>2</sub>O. L'oligonucléotide 5'-NGC GGC CGC NC-3' contient l'un des brins du site de reconnaissance de *EagI* ce qui va lui permettre de former un double brin localisé avec les spécimens simple brin

ayant l'autre moitié de ce site. Cet oligo a été rajouté à une concentration de 34  $\mu\text{M}$  de même que le tampon approprié pour la digestion avec *EagI* (NEB 3). Le mélange a été chauffé à 50°C et refroidi à 36°C. L'enzyme *EagI* (10 U) a été rajoutée et 15 min plus tard le tube a été rechauffé à 45°C pour 1 min (pour éviter l'hybridation de l'amorce avec elle-même) et refroidi ensuite à 36°C. De l'enzyme fraîche a été ensuite rajoutée et la procédure a été répétée une fois. Finalement, le surnageant a été récupéré, extrait avec du phénol-chloroforme, précipité et resuspendu dans 100  $\mu\text{l}$  d'  $\text{H}_2\text{O}$ . 25  $\mu\text{l}$  de cet échantillon ont été utilisés pour les réactions de PCR avec des amorces arbitraires.

#### 2.1.3.1 PCR avec des amorces arbitraires

Les réactions ont été faites dans un volume total de 100  $\mu\text{l}$  suivant des conditions d'amplification arbitraire (Gonzalzo et coll. 1997), c'est à dire, sept cycles de faible stringence à 93°C pour 1 min, 37°C pour 1 min, 72°C pour 1 min avant d'augmenter la stringence du PCR à la température de placement des amorces. Dans chaque réaction, l'une des amorces est ML-2 (voir section 2.1.3), tandis que l'autre était sélectionnée à partir d'une librairie selon le contenu C+G élevé à l'extrémité 3' : BRA 357 (5'-AGA GAT GAC CGC GAC CGA-3'), BRA 378 (5'-ACC GAG GCC CAG CAG ACG GC-3'), ML-16 (5'-CAT GTA GAA TTC GAT CGC GC-3'), BRA 364 (5'-AGT CTA TCT AGG ATC CCC CC-3'), et BRA 391 (5'-ATG TAG AAT TCC GCG GCC G-3'). Les produits des PCRs ont été visualisés sur gel d'agarose 1.2% et ensuite clonés et séquencés.

#### 2.1.4 Clonages

##### 2.1.4.1 Préparation du vecteur T

30  $\mu\text{g}$  du plasmide pBluescript<sup>®</sup> SK+ ont été linéarisés avec l'enzyme EcoRV (Amersham Pharmacia Biotech) et isolés sur gel d'agarose 0,7%. 5  $\mu\text{g}$  du plasmide linearisé ont été incubés avec la Taq polymérase (1U/ $\mu\text{g}$  plasmide/20  $\mu\text{l}$  volume, Amersham Pharmacia Biotech) en présence du tampon d'amplification (50 mM KCl,

20 mM Tris-HCl pH 8.5, 3.5 mM MgCl<sub>2</sub>, 0.1% Tween 20) et de 2 mM dTTP dans un volume de 100 µl pour 2 heures à 70°C. Après des extractions phénol-chloroforme, chloroforme et une précipitation à l'éthanol, le plasmide était prêt à être utilisé pour des clonages de fragments de PCR. Le vecteur modifié tire profit de l'activité terminale-transférase matrice-indépendante de la Taq polymérase qui ajoute un nucléotide (de préférence dATP) à l'extrémité 3' du fragment de PCR. L'utilisation de ce vecteur T "maison" augmente donc l'efficacité de clonage de ces fragments.

#### 2.1.4.2 Clonages

2,0 µl des produits de PCR provenant des essais de FP-CGI (section 2.1.3) ont été ligaturés (0,5 U T4 DNA ligase, NEB) pour la nuit à 14°C à 1 µl (~100ng) de vecteur T "maison" (section 2.1.4.1) dans un volume total de 10 µl. Les produits des ligations ont été utilisés pour transformer des bactéries compétentes DH5α. Les colonies sélectionnées en présence d'ampicilline (100 µg/ml) et de X-gal (GIBCO BRL) ont été utilisées pour faire des préparations plasmidiques par la méthode de lyse alcaline (Sambrook et coll. 1989). 3 µl de ces préparations ont servi à faire des digestions avec les enzymes *BamHI* et *HindIII* (New England Biolabs), des digestions qui ont été migrées sur gel d'agarose 1,2% pour repérer les bandes d'intérêt.

#### 2.1.5 Séquençages

Les plasmides contenant les inserts sélectionnés ci-dessus ont été purifiés en utilisant des colonnes QIAGEN-tip 20 (QIAGEN Inc.) ou des colonnes GFX (Amersham Pharmacia Biotech). Ces préparations ont été utilisées pour séquencer les inserts avec la trousse *Autoread™ Sequencing Kit* (Amersham Pharmacia Biotech) et avec les amorces T7 (5'-AAT ACG ACT CAC TAT AG-3') ou Reverse (5'-AAC AGC TAT GAC CAT G-3') en suivant les instructions du fabricant. Les réactions de séquençage ont été migrées dans un séquenceur automatique A. L. F. (Amersham Pharmacia Biotech) selon les instructions du fabricant.

### 2.1.6 Recherches des banques de données

Les séquences des clones obtenus par FP-CGI ont été comparées avec les séquences des banques en utilisant le logiciel BLAST (Altschul et coll. 1990), (Altschul et coll. 1997) selon les instructions disponibles sur le site web: <http://www.ncbi.nlm.nih.gov> afin de déterminer quels clones provenaient de séquences uniques. Deux banques de données principales ont été consultées: nr (*non-redundant* GenBank, EMBL et DDBJ) et dbEST (banque de séquences exprimées ou ESTs, "expressed sequence tags"). Les clones qui montraient une identité de séquence avec aucune ou un petit nombre de séquences humaines furent considérés comme étant uniques ou non-répétitifs. Ces clones ont été utilisés pour hybrider des membranes Southern (voir ci-dessous) pour déterminer s'ils provenaient de séquences méthylées différemment.

## 2.2 ANALYSES D'ADN

### 2.2.1 Transferts Southern

Les clones obtenus par FP-CGI contenant des séquences apparemment non-répétées ont été utilisées pour hybrider des membranes Southern d'ADN de granulocyte humain d'origine femelle ou mâle, digéré avec *BamHI*, *EcoRI*, *HindIII* ou *PstI* seules ou en combinaison avec *EagI*, *HhaI*, *HpaII* ou *MspI*. Les sondes et les transferts Southern ont été préparés conformément à des protocoles standard (Sambrook et coll. 1989).

## 2.3 FP-CGI SUR DES SÉQUENCES CONNUES

### 2.3.1 PCRs des promoteurs des gènes MAOA et PGK-1 et transferts Southern

Pour amplifier 165 pb de l'îlot CpG du promoteur du gène MAOA (Hendriks et coll. 1992) à partir de l'ADN provenant d'un essai de FP-CGI réalisé avec l'enzyme *HpaII* avec de l'ADN mâle ou femelle, nous avons utilisé les amorces ML-5 (5'-GAT AGA

AGG GTC CTT CCC AC-3') et ML-6 (5'-CTA CGG TCC ACA CTG ACC TG-3') avec les conditions suivantes: 1 cycle à 93°C pour 3 min; 35 cycles de 93°C pour 1 min, 58°C pour 1 min, 72°C pour 1 min; et 1 cycle à 72°C pour 10 min. Les produits de PCR ont été migrés sur gel d'agarose 2,0%.

Pour amplifier 175 pb de l'îlot CpG du promoteur du gène PGK-1 (Pfeifer et coll. 1990) à partir de l'ADN provenant d'un essai de FP-CGI réalisé avec l'enzyme EagI, nous avons utilisé les amorces ML-7 (5'-GTG TGG GGC GGT AGT GTG GG-3') et ML-8 (5'-GTC CAG CTT GTC CAG CGT CAG-3') avec les conditions suivantes: 1 cycle à 93°C pour 3 min; 35 cycles de 93°C pour 1 min, 62°C pour 1 min, 72°C pour 1 min; et 1 cycle à 72°C pour 10 min. Les produits de PCR ont été migrés sur un gel d'agarose 2,0%.

Les produits des PCRs MAOA et PGK-1 fractionnés sur gel d'agarose 2.0% furent transférés sur une membrane de Nylon en suivant des protocoles standard (Sambrook et coll. 1989). Les membranes ont été hybridées avec les amorces ML-5 ou ML-7 marquées à l'extrémité 5' avec du  $\gamma$ -<sup>32</sup>P ATP à l'aide de la T4 Polynucléotide Kinase (Amersham Pharmacia Biotech) selon des protocoles standard (Sambrook et coll. 1989).

## 2.4 LOCALISATION CHROMOSOMIQUE

### 2.4.1 Panel de Coriell et panel d'hybrides de radiation G3 ("Radiation Hybrids")

De l'ADN du "somatic cell hybrid panel 2" (SCHP2, Dubois et Naylor 1993) a été obtenu de la compagnie Coriell Cell Repositories (Candem, NJ). Le panel SCHP2 consiste en de l'ADN provenant de 27 lignées cellulaires: 24 lignées cellulaires hybrides hamster/humain et 3 contrôles pour l'ADN de souris, humain et de hamster. Les lignées hybrides sont supposées de ne contenir qu'un chromosome humain mais quelques-unes d'entre elles ont été montrées d'avoir des régions d'autres chromosomes humains en plus de celui qu'elles ont retenu (Dubois et Naylor 1993). Des aliquots de



20 ng ont été amplifiés par PCR en utilisant des amorces (des "STS") que nous avons faits synthétiser en nous basant sur la séquence des fragments à être localisés (PCR clone-spécifique).

De l'ADN du "radiation hybrid Stanford G3 panel" a été obtenu de la compagnie Research Genetics (Huntsville, AL). Ce panel est composé de quatre-vingt trois lignées cellulaires hybrides humain/hamster et deux contrôles pour l'ADN humain et celui de hamster. En utilisant ce panel, il est possible de localiser un marqueur quelconque par rapport à des marqueurs connus avec une résolution de près de 500 kb. Nous avons criblé ce panel par PCR à l'aide de STSs spécifiques à chacun des clones (voir ci-dessus et les Figs 31 à 33). Les résultats des quatre-vingt trois PCRs ont été envoyés au serveur du SHGC en forme de 0 pour les négatifs, 1 pour les positifs et R pour ceux pour lesquels nous avons des doutes. En plus, nous avons fourni au serveur le chromosome sur lequel le clone est probablement localisé, ce qui augmente la probabilité d'avoir une réponse précise pour la localisation. Les résultats ont été analysés par un logiciel se trouvant au site Web du Stanford Human Genome Center (SHGC, <http://shgc-www.stanford.edu>).

Pour les réactions de PCR servant à localiser les STSs spécifiques aux clones L88, L211, L59 et L3, le mélange de 30 µl consistait de: MgCl<sub>2</sub> 1.5 mM, Tris-HCl 20 mM pH 8,5, KCl 50 mM, Tween 20 0,1 %, 2 µCi de α-<sup>32</sup>P dCTP (ICN Radiochemicals, Irvine, CA), 1U de Taq Pol (Amersham Pharmacia Biotech), dNTP 200 µM et 200 nM de chaque amorce.

Pour localiser le clone L88, nous avons utilisé les amorces ML-13 (5'-GGG CAT TTT CAT AAG GTG TG-3') et ML-14 (5'-TGT TCC CTG TTA GCC CTT TAG-3') pour amplifier un fragment de 408 pb avec les conditions suivantes: 1 cycle à 93°C pour 3 min; 30 cycles à 93°C pour 1 min, à 55°C pour 1 min et à 72°C pour 1 min et 1 cycle à 72°C pour 10 min.

Pour localiser le clone L211, nous avons utilisé les amorces BRA 423 (5'-AAC TAT TGG GTC CAG CAC CAG-3') et BRA 424 (5'-CTG GCT AGA TTC AGC AAC ACC-3') pour amplifier un fragment de 156 pb avec les conditions suivantes: 1 cycle à 93°C pour 3 min; 35 cycles à 93°C pour 1 min, à 55°C pour 1 min et à 72°C pour 1 min et 1 cycle à 72°C pour 10 min. Lors du criblage des hybrides de radiation G3, l'amorce BRA 424 a été remplacé par l'amorce BRA 456 (5'-GGA CTG GCC TAC TTC TAA ACA G-3') pour amplifier une bande d'une longueur de 134 pb avec les conditions suivantes: 1 cycle à 93°C pour 3 min; 33 cycles à 93°C pour 45 sec, à 55°C pour 45 sec et à 72°C pour 45 sec et 1 cycle à 72°C pour 10 min.

Dans le cas du clone L59, les amorces BRA 428 (5'-AGC TTA GAG AAC AGG GAA CTG-3') et BRA 446 (5'-CTC AGA TAC TGA CCA AGC AAC-3') ont été utilisés pour amplifier un fragment de 308 pb en suivant les conditions: 1 cycle à 93°C pour 3 min; 35 cycles à 93°C pour 1 min, à 53°C pour 1 min et à 72°C pour 1 min et 1 cycle à 72°C pour 10 min.

Le clone L3 a été localisé en amplifiant un fragment de 157 pb avec les amorces BRA 457 (5'-ATT AGA GTC TAA TAA GTC ATT C-3') et BRA 458 (5'-ATT GAC GAT AGA AGT AGA ATA C-3'). Les conditions suivies étaient: 1 cycle à 93°C pour 3 min; 33 cycles à 93°C pour 45 sec, à 46°C pour 45 sec et à 72°C pour 45 sec; et 1 cycle à 72°C pour 10 min.

3 µl de chaque produit d'amplification a été migré sur un gel de polyacrylamide 5% non-dénaturant pendant ~ trois heures à 770 volts. Les gels séchés ont été exposés en présence de films KODAK XAR-5 pour la nuit et/ou pour quelques jours, selon le cas.

#### 2.4.2 *Cartes génétique et physique*

Les résultats du criblage des hybrides de radiation G3 sont présentés par le SHGC sous la forme de marqueurs chromosomiques liés. Ensuite, nous avons utilisé des cartes

physiques de chromosomes humains pour localiser ces marqueurs et d'autres marqueurs se trouvant près de nos clones. Nous avons surtout tiré profit du logiciel *Entrez* (<http://www3.ncbi.nlm.nih.gov/Entrez/>) qui permet une comparaison graphique rapide entre différentes cartes de marqueurs telles que celles de Généthon, du Stanford Human Genome Center, du MIT, du NCBI ou du CHLC (Fig. 13). Nous avons ainsi généralement localisé chacun de nos clones dans le génome humain.

## 2.5 OBTENTION DE FRAGMENTS GÉNOMIQUES HUMAINS DE TRÈS HAUT POIDS MOLÉCULAIRE

### 2.5.1 Identification de clones génomiques, BACs

De l'ADN du "Human BAC DNA Pools-Release III" a été obtenu de la compagnie Research Genetics (Huntsville, AL). 10 µl de chacun des 72 échantillons de *superpools* d'ADN ont été amplifiés en utilisant les mêmes amorces employés pour le panel de hybrides de radiation G3 sauf pour le clone L59 (voir ci-dessous). Chaque "superpool" consiste en de l'ADN provenant de 8 plaques de 384 puits de colonies de bactéries, ce qui donne un grand total de 221184 colonies. Lorsqu'on identifie les "superpools" contenant le(s) clone(s) d'intérêt, une seconde série de PCRs est effectuée pour identifier la plaque ("plate pool") contenant le BAC correspondant.

Après avoir identifié le "plate pool" nous avons entrepris un troisième PCR sur ce qu'on appelle les "row pool" (16 PCRs) et les "column pool" (24 PCRs) qui donnent les coordonnées finales du BAC. Un BAC possible est alors identifié par le # de la plaque (résultat du "plate pool") et par les résultats du "row" et du "column pool" à la suite de 120 PCRs .

Les réactions de PCR furent réalisées comme à la section 2.4.1, à l'exception du clone L59 où les amorces BRA447 (5'-TTG TCA ACG AGC TGT AGC AG-3') et BRA 448 (5'-TTG AGG CTG CCT TGA GGG AG-3') ont été utilisés pour amplifier un

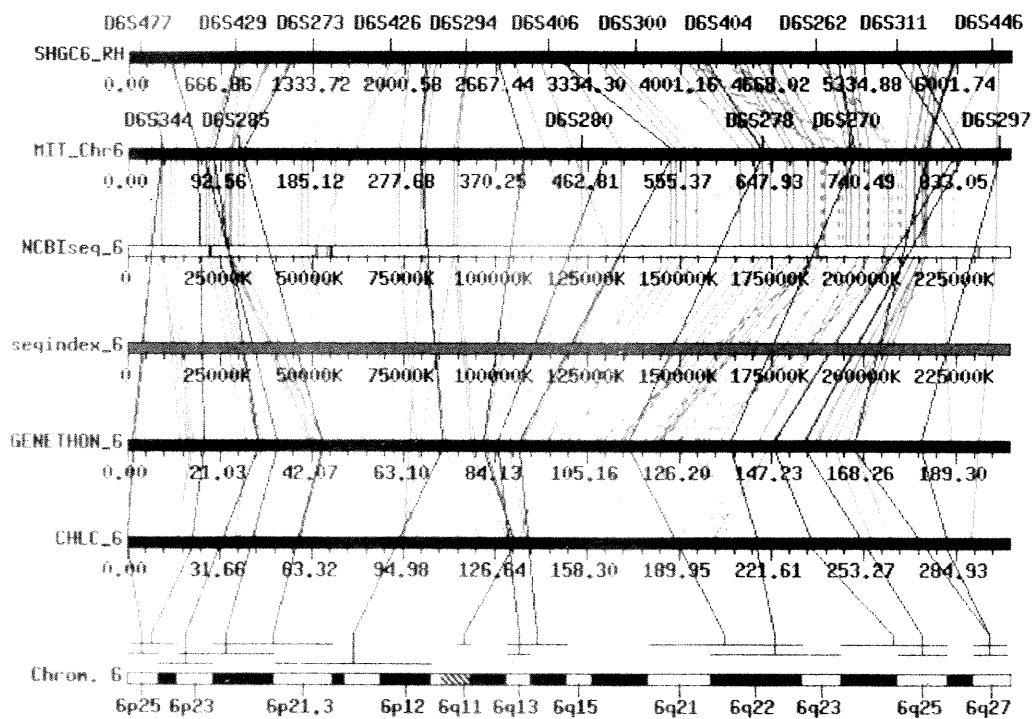


Fig. 13: Intégration des cartes physique et génétique d'une région du chromosome 6 .

fragment d'une longueur de 327 pb en suivant les conditions: 1 cycle à 93°C pour 3 min; 33 cycles à 93°C pour 45 sec, à 51°C pour 45 sec et à 72°C pour 45 sec et 1 cycle à 72°C pour 10 min.

~3 µl de chaque produit de PCR a été migré sur un gel de polyacrylamide 5 % non-dénaturant pendant trois heures à 770 voltes. Les gels ont été exposés en présence de films KODAK XAR-5 pour la nuit à -80°C.

### 2.5.2 *Identification de clones génomiques, YACs*

Des YACs localisés dans la région 22q11 ont été obtenus directement du CEPH (Paris, France; yacmanager@cephb.fr) et ceux portant le STS spécifique à L88 furent identifiés par PCR utilisant l'ADN du "CEPH 'A' Human YAC DNA pools" qui a été obtenu de la compagnie Research Genetics (Huntsville, AL). Chacun des 24 échantillons de "superpools" d'ADN a été amplifié en utilisant les amorces ML-13 et ML-14. Chaque "superpool" consiste en de l'ADN provenant de 8 plaques de 96 puits de colonies de levures. Lorsqu'on identifie les "superpools" contenant le(s) clone(s) d'intérêt sont identifiés, un deuxième PCR de 8 échantillons a été effectué pour identifier la plaque ("plate pool") contenant le YAC associé au STS. Après avoir identifié le "plate pool", un troisième PCR a été entrepris sur ce qu'on appelle les "row pool" et les "column pool". Les résultats des PCRs sur les "row pool" et les "column pool" donnent les axes qui permettent d'identifier le clone YAC correspondant. Après 52 PCRs, l'adresse d'un YAC est donnée par le # de la plaque ("plate pool") et par les "row" et le "plate pool".

## 2.6 UTILISATION DES BACs ET DES YACs

### 2.6.1 *"Fluorescence in situ hybridization"*.

Les expériences de localisation chromosomique par "Fluorescence in situ hybridization" (FISH) du BAC 421M16 contenant le clone L59 ont été effectuées au

labo de la Dr. Nicole Lemieux en utilisant ses protocoles décrits précédemment (Lemieux et coll. 1992)

### 2.6.2 Technique d'amplification d'exons à partir d'ADN génomique ("exon trapping").

L'amplification d'exons est une méthode de RT-PCR, illustrée à la figure 14, dont le but est de cloner des exons directement à partir d'ADN génomique (Church et coll. 1994). Elle se base sur la sélection *in vivo* de sites d'épissage fonctionnels entourant des séquences exoniques dans l'ADN génomique. L'ADN génomique de haut poids moléculaire provenant de cosmides, BACs ou YACs, est sous-cloné dans le plasmide d'épissage pSPL3 (Fig. 15). Ce vecteur contient un site de clonage multiple entouré de deux sites de reconnaissance pour l'enzyme *BstXI*. Lorsque l'ADN cloné contient un exon dans la bonne orientation, il pourrait se produire de l'épissage entre les séquences du vecteur et des séquences de l'insère. Le sous-clonage est suivi par la transfection de cellules de mammifères COS-1 par les récombinantes. Après une expression transitoire, l'ARN est recolté et reverse-transcrit en utilisant des amorces spécifiques au vecteur. Après un premier PCR, une digestion avec *BstXI* aide à se débarrasser des produits d'épissage dérivés seulement du plasmide. Cette digestion est suivie par une deuxième réaction de PCR ("nested") avec des amorces internes spécifiques à la séquence du vecteur. Cette réaction de PCR est analysée sur gel pour déterminer les longueurs des produits. La longueur des produits de PCR provenant de l'épissage des séquences exoniques du plasmide est de 177 pb. Tout produit ayant une taille supérieure à 177 pb pourrait contenir des exons.

Les insertions d'ADN génomique contenues dans les BACs 421M16 et 528D2 associées au clone L59 ont été sous-clonées dans le plasmide pSPL3. Bref, 1 µg d'ADN du BAC isolé par maxi-preps au CsCl (Sambrook et coll. 1989) et 1 µg d'ADN de pSPL3 ont été digérés séparément avec les combinaisons suivantes d'enzymes: *SmaI*, *SmaI* + *SacI*, *EagI* + *BamHI* ou *EagI* + *BglII*. Le plasmide pSPL3 linearisé a été dephosphorylé avec la CIP ("Calf intestinal alkaline phosphatase") selon les instructions du fabricant (Boehringer Mannheim). 250 ng de ce plasmide et 1 µg de

l'ADN génomique digéré ont été ligués (T4 DNA ligase, NEB) pendant quelques heures à température de la pièce. Nous avons utilisé les produits des ligations pour transformer des bactéries compétentes DH5 $\alpha$ . L'ADN plasmidique a été isolé par des minipreps de lyse alcaline.

#### 2.6.2.1 *Transfection de cellules Cos-1*

Seize heures avant les transfections, des cellules COS-1 (des cellules de rein de singe vert africain) ont étéensemencées à une confluence de 75-80% ( $\sim 5 \times 10^5$  cellules) dans des plaques de culture de six puits contenant du milieu DMEM avec 10% de sérum de veau fœtal. Pour chaque transfection, 5  $\mu$ l de Lipofectamine (Gibco BRL) ont été ajoutés à 100  $\mu$ l de milieu sans sérum dans un tube de polystyrène et le tout a été incubé pour 5 min à T° de la pièce. Dans un tube séparé, 1  $\mu$ g d'ADN plasmidique (sous-clones) a été mis pour chaque transfection dans 100  $\mu$ l de milieu DMEM sans sérum. Ces deux solutions ont été combinées et incubées 15 min à T° de la pièce pour permettre la formation de complexes ADN/lipide auxquels on a rajouté 0.8 ml de DMEM sans sérum à 37°C. Pour sa part, le milieu des cellules COS-1 a été enlevé pour ajouter 2 ml de DMEM sans sérum à 37°C suivi d'une incubation de 5 min à 37°C. Ensuite le milieu a été enlevé de la plaque et le complexe ADN/lipide y a été rajouté. Cela a été suivi d'une incubation de 5 à 6 heures puis enchaîné par l'ajout de 1 ml de milieu DMEM/20% SVF.

Les cellules COS-1 transfectées ont été récoltées par trypsinisation 24 heures plus tard, lorsque elles étaient 85% en confluence. L'ARN total des cellules transfectées a été isolé en utilisant du TRIZOL selon les instructions du fabricant (Gibco BRL, Cat No. 15596) et resuspendu dans 50  $\mu$ l d'eau ultrapure. L'ADNc a été synthétisé à partir de 3  $\mu$ l de l'ARN total à l'aide de l'amorce SA2(5'-ATC TCA GTG GTA TTT GTG AGC-3') qui est spécifique à la séquence du plasmide pSPL3, dans un mélange total de 50  $\mu$ l incluant DTT 10 mM, dNTP 0.5 mM, RNA guard (Amersham Pharmacia Biotech), tampon RT 1X (Tris pH 8,3 50 mM, MgCl<sub>2</sub> 8 mM), amorce SA2 400 nM et 100 U de l'enzyme *M-MLV reverse transcriptase*. Le mélange a été incubé pour une heure à

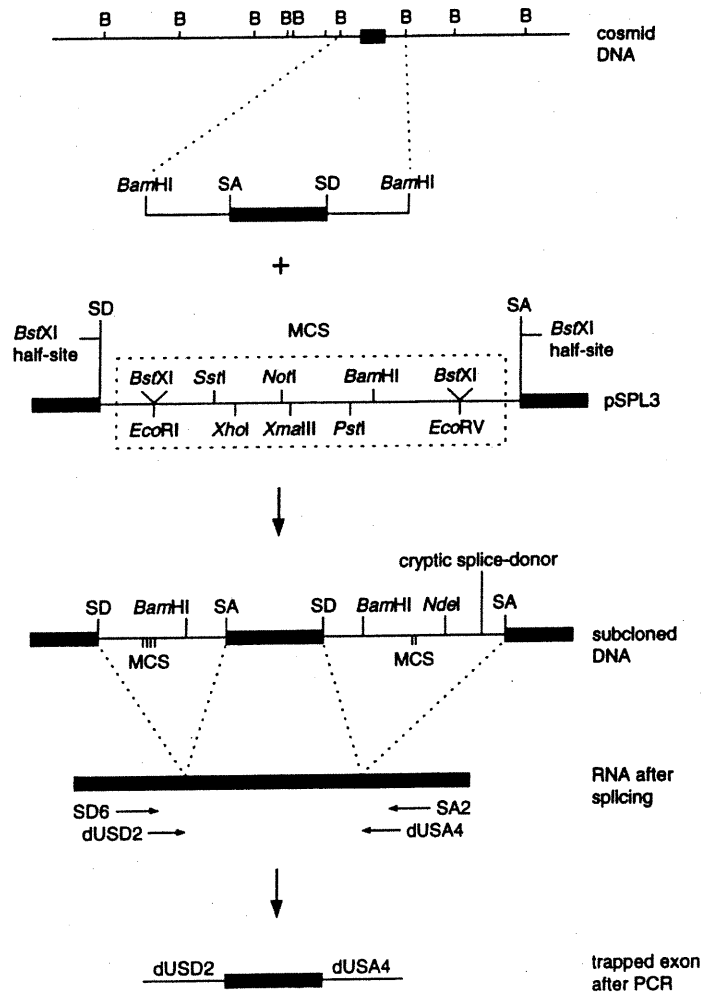


Fig. 14: Représentation schématique de la technique d'amplification d'exon.

Tiré de Dracopoli et coll. 1998



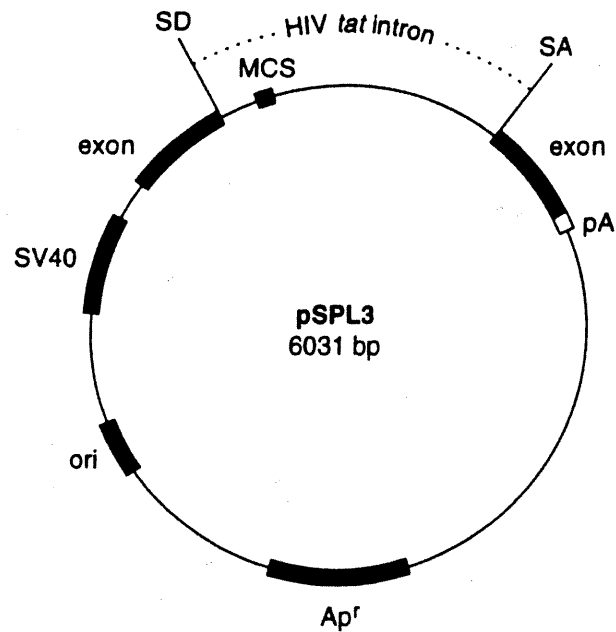


Fig. 15: Vecteur pSPL3. Le vecteur contient des séquences qui permettent sa réplication chez *E.coli* et chez les cellules COS-7 (ou COS-1). Un site de clonage multiple interrompt l'intron *tat* de HIV et fournit des sites de restriction pour le sous-clonage de l'ADN génomique.

Tiré de Dracopoli et coll. 1998

37°C. 5 µl de la réaction ont été employés pour faire des PCRs de 50 µl comprenant: dNTP 200 µM, MgCl<sub>2</sub> 1.5 mM, Tris-HCl 20 mM pH 8.5, KCl 50 mM, Tween 20 0.1 %, 2U de Taq polymérase et 500 nM des amorces SA2 et SD6 (5'-TCT GAG TCA CCT GGA CAA CC-3'). Le cycle de dénaturation initial était de 93°C pour 3 min. L'étape d'hybridations des amorces était de 6 cycles consistant de 1 min à 93°C, 1 min à 60°C et 1 min à 72°C. La dernière étape était d'un cycle de 10 min à 72°C. 25 U de l'enzyme BstXI ont été ajoutées à cette réaction de PCR et le mélange a été incubé pour la nuit à 55°C. 5U de BstXI ont été rajoutées à ce mélange pour un temps d'incubation additionnel de 2 heures à 55°C. 5 µl de cette réaction ont été ensuite ajoutés à une réaction de PCR de 50 µl consistant de: MgCl<sub>2</sub> 1.35 mM, Tris-HCl 20 mM pH 8.5, KCl 50 mM, Tween 20 0.1 %, dNTP 200 µM, 2U de Taq polymérase et 400 nM des amorces SA4 (5'-CAC CTG AGG AGT GAA TTG GTC G-3') et SD2 (5'-GTG AAC TGC ACT GTG ACA AGC-3') en suivant les conditions: 1 cycle à 93°C pour 3 min; 30 cycles à 93°C pour 1 min, à 60°C pour 1 min et à 72°C pour 3 min et 1 cycle à 72°C pour 10 min. Les produits de PCR ont été visualisés sur gel d'agarose 2.0%. Les fragments avec une taille nette de plus de 150 pb (en enlevant les séquences plasmidiques) ont été clonés et séquencés (comme décrit à la section 2.1.7).

## 2.7 ÉTUDES DE MÉTHYLATION ET D'EXPRESSION AVEC DES CROISEMENTS INTERSPÉCIFIQUES

### 2.7.1 Technique SNUPE ("Single nucleotide primer extension").

SNUPE est une technique d'analyse quantitative d'ARN par RT-PCR qui repose sur la spécificité d'incorporation par l'ADN polymérase dans une réaction d'extension de l'amorce (Fig.16) et dont le but est d'analyser des niveaux d'expression allèle-spécifique (Greenwood et Burke 1996). L'ARN reverse transcrit sert de matrice pour l'amplification par PCR d'une séquence où il n'y aura qu'une seule base différente entre les deux allèles. Les produits de PCR sont analysés pour la présence, absence ou les quantités relatives de chaque allèle en faisant apparier un oligo interne. Cet oligo se termine exactement en 5' du nucléotide polymorphique (SNP) et il sera prolongé par

l'addition, dans des tubes séparés, de bases différentes marquées. Cette incorporation à l'extrémité 3' de l'amorce ne marchera que lorsque la base correcte sera disponible dans la réaction. Les réactions sont ensuite analysées sur gel de polyacrylamide.

Ms-SNuPE est une variation de la technique SNuPE dont le but est d'analyser des niveaux de méthylation de sites CpG spécifiques (Fig. 16) (Gonzalvo et Jones 1997). Après une digestion avec une enzyme sensible à la méthylation, l'amplification par PCR ne fonctionne que lorsque le(s) méthyl-CpG(s) se trouvant dans le site de reconnaissance de l'enzyme utilisée sera méthylé. L'incorporation à l'extrémité 3' de l'amorce va donc permettre de distinguer l'allèle méthylé de l'allèle non-méthylé où, jusqu'à un certain degré, d'analyser leur méthylation relative.

#### 2.7.1.1 Études de Ms-SNuPE ("Methylation-sensitive single nucleotide primer extension") sur la séquence génomique de L59cand#2.

Nous avons extrait de l'ADN génomique de foie de souris C57Bl/6J, de souris *Mus musculus molossinus* et de souris F1 de croisements C57Bl/6J x *M. m. molossinus* (B6 x mol, femelle x mâle) par des procédures standard (Sambrook et coll. 1989). Les échantillons de foie des souris *M. m. molossinus* et des souris des croisements B6 x mol ont été gentiment fournis par la Dr. Danielle Malo tandis que les foies de souris B6 provenaient de souris B6 de notre colonie (Charles River). L'ADN a été coupé à 37°C pendant la nuit avec les enzymes HpaII et HhaI (New England Biolabs), seules ou en combinaison. Nous avons utilisé 100 ng de ces digestions dans des réactions de PCR de 50 µl consistant de: MgCl<sub>2</sub> 1.5 mM, Tris-HCl 20 mM pH 8.5, KCl 50 mM, Tween 20 0.1 %, dNTP 200 µM, 1U de Taq polymérase et 200 nM des amorces BRA 468 (5'-tgt gac gag ggt cag ccc t-3') et BRA 469 (5'-cct ggg cct tgc aaa acc t-3') en suivant les conditions: 1 cycle à 93°C pour 3 min; 28 cycles à 93°C pour 45", à 61°C pour 45" et à 72°C pour 45" et 1 cycle à 72°C pour 10 min. Les produits de PCR de ~800 pb ont été migrés sur un gel 1.2 % d'agarose de bas point de fusion (Gibco BRL) pour ensuite être récupérées par une extraction phénol-chloroforme après avoir préalablement fait fondre le gel à 65°C (Sambrook et coll. 1989). 25 ng de chaque

produit de PCR isolé sur gel ont été incubés avec 1U de Taq polymérase dans deux tubes séparés pour effectuer une réaction d'extension d'amorce de 20 µl contenant: MgCl<sub>2</sub> 1.5 mM, Tris-HCl 20 mM pH 8.5, KCl 50 mM, Tween 20 0.1 % et 1µM de l'amorce BRA 470 (5'-GCC ATA GAG AAN GGG GAT GA-3'). En plus, chacune des deux réactions contenait soit 1 µCi de α<sup>32</sup>P dTTP, soit 1 µCi de α<sup>32</sup>P dGTP. Les conditions de réaction étaient: 1 cycle à 95°C pour 1 min, à 50°C pour 2 min et à 72°C pour 1 min. Une solution stop (~10 µl) a été ajoutée à la réaction et 5 µl de celle-ci ont été migrés sur un gel dénaturant de polyacrylamide 15% (urée 7M).

#### 2.7.1.2 Études d'expression de *TMCL59* par *SNuPE*.

Nous avons utilisé 3 µl d'échantillons d'ADNc gentiment fournis par le Dr. Benjamin Tycko. Les échantillons provenaient de souris C57Bl/6J, *M. m. molossinus* et de souris F1 de croisements B6 x mol et mol x B6 (femelle x mâle). Ces ADNc ont été préparés en utilisant des oligos dT comme amorces. Les conditions et les amorces de la réaction de PCR (pour amplifier un fragment de 434 pb) et de celle d'extension d'amorce sont les mêmes que celles de la section 2.7.1.2.

#### 2.7.2 Études de RT-PCR sur la séquence *L59cand#1*.

De l'ADNc a été préparé à partir d'échantillons de foie de souris *M. m. molossinus* et B6 x mol selon le protocole décrit dans la section 2.6.2.1. Nous avons effectué des PCRs (fragment de 353 pb) avec 3 µl de ces échantillons ou avec 3 µl d'échantillons d'ADNc fournis par le Dr. Benjamin Tycko (section 2.7.1.2). Chaque réaction de 100µl était composée de MgCl<sub>2</sub> 1.5 mM, Tris-HCl 20 mM pH 8.5, KCl 50 mM, Tween 20 0.1 %, dNTP 200 µM, 1.5 U de Taq polymérase et 200 nM des amorces BRA 443 (5'-GAG AGT AGT GAG TAA TGT CAT TC-3') et BRA 445 (5'-ATG AGC TCG GCC TCT TCA GTC-3') avec ces conditions: 1 cycle à 93°C pour 3 min; 30 cycles à 93°C pour 1 min, à 55°C pour 1 min et à 72°C pour 1 min et 1 cycle à 72°C pour 10 min. Des digestions de quelques heures avec l'enzyme *XmnI* (NEB) ont été faites à partir de 20 µl de cette réaction et migrées sur gel d'agarose de 2.5%.

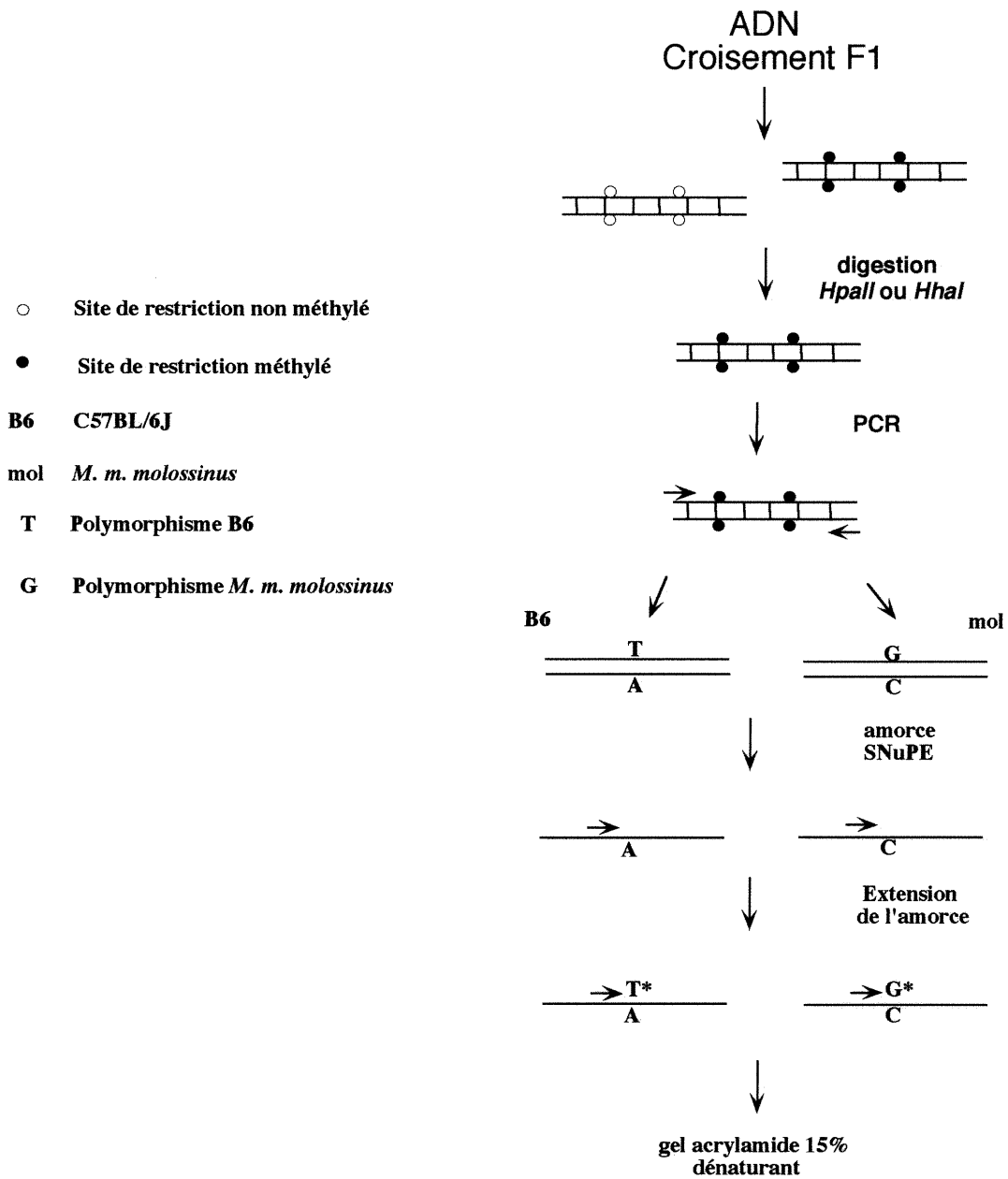


Fig. 16: Méthode SNUPE pour détecter de la méthylation ou de l'expression allèle-spécifique.

**CHAPITRE 3**

**RÉSULTATS**

### 3.1 CLONES OBTENUS PAR FP-CGI.

Nous avons effectué cinq expériences différentes de FP-CGI pour identifier des fragments d'ADN différentiellement méthylés chez des granulocytes humains. Pour chacune des expériences, plusieurs bandes ont été repérées sur gels d'agarose 1.2 % colorés au bromure d'éthidium (Figure 17). Quarante-trois fragments, d'une longueur moyenne de 300pb, ont été clonés puis séquencés. Selon les résultats des recherches BLAST, treize séquences (30%) étaient répétitives, une correspondait à de l'ADN ribosomal, et deux étaient des séquences de *E.coli*. Il reste donc vingt-sept séquences apparemment uniques vu qu'elles ne montraient aucune similitude avec des séquences uniques dans les banques de données utilisées. D'autres caractéristiques de ces 27 séquences uniques sont présentées au tableau V. On constate qu'un certain nombre d'entre elles répondent aux critères de base pour un îlot CpG, c'est à dire un haut contenu en C+G (~50-80%) et pas de suppression de CpG vs GpC ( $CpG/GpC \cong 0.6-1.0$ ). Six de ces fragments non-répétés ont plus de 80 % du nombre de CpGs attendu, calculé comme  $100/(proportion\ de\ C \times proportion\ de\ G)$  par 100 pb. Cinq autres avaient entre 50 % et 80 % de ce nombre.

### 3.2 VALIDATION DU FP-CGI SUR DES SÉQUENCES DE RMDs CONNUES.

Parallèlement aux essais de FP-CGI entrepris pour chercher de nouvelles séquences méthylées différentiellement, nous avons utilisé de l'ADN de granulocytes humains mâles et femelles, pour tester l'efficacité de la méthode à travers les différentes étapes. Des échantillons prélevés à partir des différentes étapes du protocole FP-CGI ont été utilisés pour suivre la présence et l'abondance de certaines séquences dont le profil de méthylation différentielle est déjà connu. Nous avons donc conçu des amorces pour amplifier des séquences promotrices des gènes MAOA et PGK-1 se trouvant sur le chromosome X qui montre une méthylation différentielle chez les femmes. Ces amorces amplifient une région contenant quelques sites *HpaII*, dans le cas du gène MAOA, et un site *EagI* dans le cas du gène PGK-1.

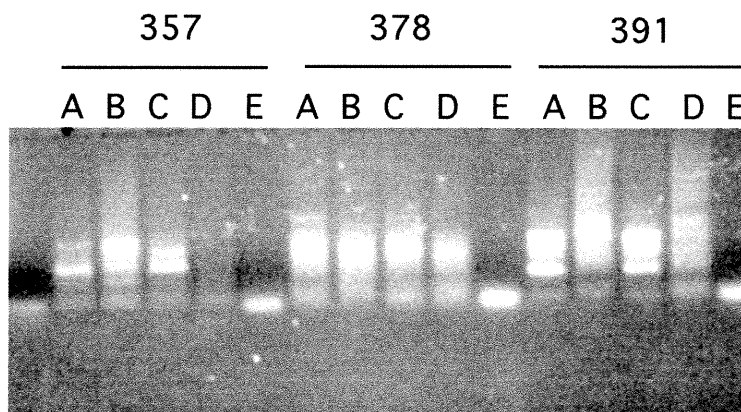


Fig.17: Produits de PCR obtenus par FP-CGI visualisés sur gel d'agarose 1,2%. A, B, C et D représentent quatre expériences de F-PERT différentes. 357=amorce BRA 357. 378=amorce BRA 378. 391= amorce BRA 391. E=H<sub>2</sub>O.



Clones	Grandeur (pb)	%C + G	CpG/GpC	Îlot?	Résultats d'hybridation Southern
L51	315	35	0,19(3/16)	Non	Non méthylé
L22	409	61	1,28(32/25)	Oui	Pas de signal
L211	288	53	0,36(4/11)	Oui	Méthylation différentielle
L31	310	35	0,30(3/10)	Non	À tester
L92	873	54	0,87(71/81)	Oui	Pas de signal
L11 (1 kb total)(T7)	426	65	0,97 (42/43)	Oui	Pas de signal
L11b (rev)*	423	46	0,73 (19/26)	Oui	Pas de signal
L20	153	44	0,25(1/4)	Non	À tester
L3	184	33	0,33(2/6)	Non	Méthylation différentielle
L5	273	53	1,27(14/11)	Oui	Pas de signal
L8	309	50	0,75(18/24)	Oui	Pas de signal
L12	111	49	0(0/5)	Non	Pas de signal
L511	193	52	0,87(7/8)	Oui	Méthylé
L52	85	46	0(0/4)	Non	À tester
L88	444	51	0,24(6/25)	Non	Méthylation différentielle
TR17	193	52	1 (7/7)	Oui?	Méthylé
TR19	79	42	0,25(1/4)	Non	À tester
TR51	167	61	0,93(14/15)	Oui	Pas de signal
TR53	163	47	0,71 (5/7)	Oui?	Méthylé
L59	1091	54.3	0,19(14/73)	Non	Méthylation différentielle
L62	315	62.5	0,42(11/26)	Oui?	Méthylation différentielle
L31	350	50.3	0,4 (10/25)	Oui?	Méthylé
L53	147	55.1	0,15 (2/13)	Non	Pas de signal
L61	136	65.4	0,625(10/16)	Oui	Méthylation partielle
L71(1.2kb tot)	525	43	0,26 (5/19)	Non	Méthylé?
L101	315	56	0,25(3/12)	Non	Méthylé
L151 (1.1kb tot)	314	39	0,86(6/7)	Non	Méthylé
L163(0.8kb tot)	724	54.3	0,11 (4/36)	Non	Méthylé

Tableau V: Liste des clones obtenus par FP-CGI associés à des séquences non -répétitives.

\* L11 et L11b font partie du même clone: L11. Îlot? = Est-ce que ce fragment peut être considéré un îlot CpG?

Dans la dernière étape du FP-CGI, l'intensité des produits de PCR était plus forte pour l'ADN femelle que pour l'ADN mâle. En réalité il était très difficile de distinguer la bande de l'ADN mâle par une coloration au bromure d'éthidium. Nous avons donc fait des transferts Southern pour mieux déterminer le rapport des deux bandes. Lorsque les membranes ont été hybridées (Fig.18) avec l'une des deux amorces utilisées pour amplifier les bandes, nous avons estimé que le rapport bande femelle/bande mâle était ~ 3.0.

### 3.3 ANALYSES D'ADN

#### 3.3.1 Transferts Southern de granulocytes humains

Vingt-deux des vingt-sept séquences non-répétées isolées par FP-CGI ont été analysées par hybridation de membranes contenant de l'ADN génomique (Tableau V) de granulocytes digérés avec différentes enzymes sensibles à la méthylation et avec *MspI* pour des fins de comparaison avec le profil de restriction de *HpaII*. Puisque près de la moitié des clones ont été obtenus à partir des donneurs femelles afin d'isoler des séquences se trouvant au chr X, les membranes ont été préparées en utilisant de l'ADN de granulocytes de donneurs femelles et mâles. Huit sondes n'ont pas donné des signaux interprétables. Parmi celles-ci, six lesquelles étaient celles qui avaient été classifiées comme des ICGs (Tableau V), et les deux autres lesquelles n'avaient qu'une longueur d'approximativement 100 bp. L'absence de signal pourrait refléter une combinaison de facteurs tels que la petite longueur de la sonde, le haut pourcentage de C + G et/ou d'autres caractéristiques inconnues de ces séquences.

Parmi les quatorze autres clones, sept desquels ont généré des profils d'hybridation suggérant une méthylation complète des séquences étudiées, c'est à dire digestion seulement avec *MspI* mais pas avec les enzymes sensibles à la méthylation comme *EagI*, *HhaI* ou *HpaII* (Fig. 19). Les sept autres se trouvent près d'ICGs semblant être au moins partiellement non-méthylés. De ce dernier groupe, le clone L51 a des bandes

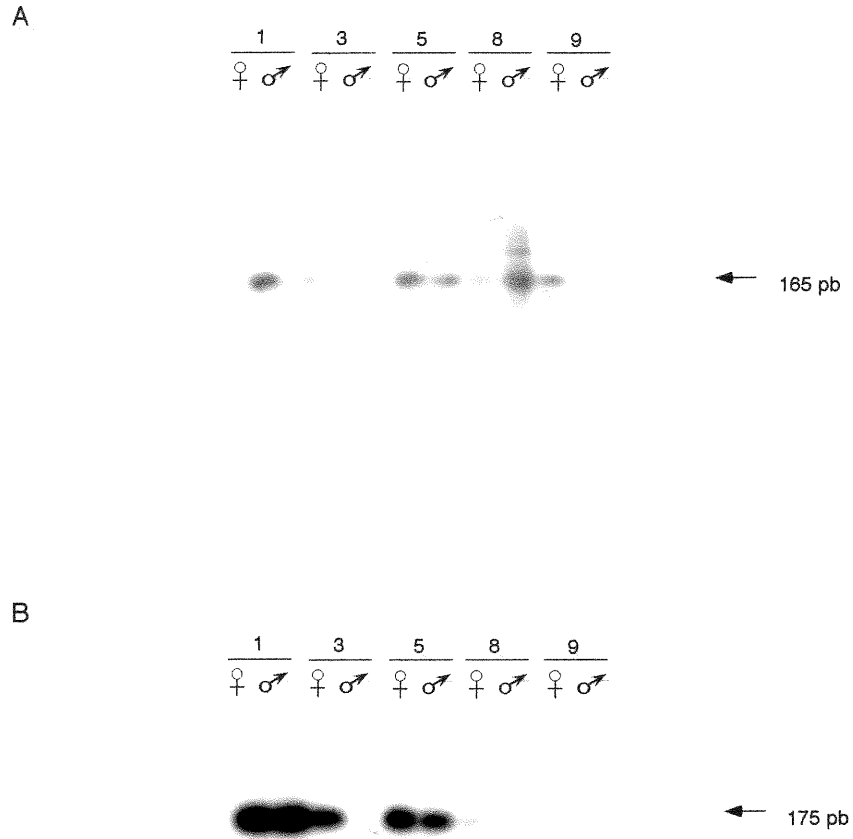


Fig. 18: Transferts Southern des PCRs MAOA (a) et PGK-1 (b) faits sur de l'ADN de granulocytes femelle et mâle. Les échantillons ont été isolés à partir des étapes 1, 3, 5, 8 et 9 du protocole FP-CGI (Fig. 14).

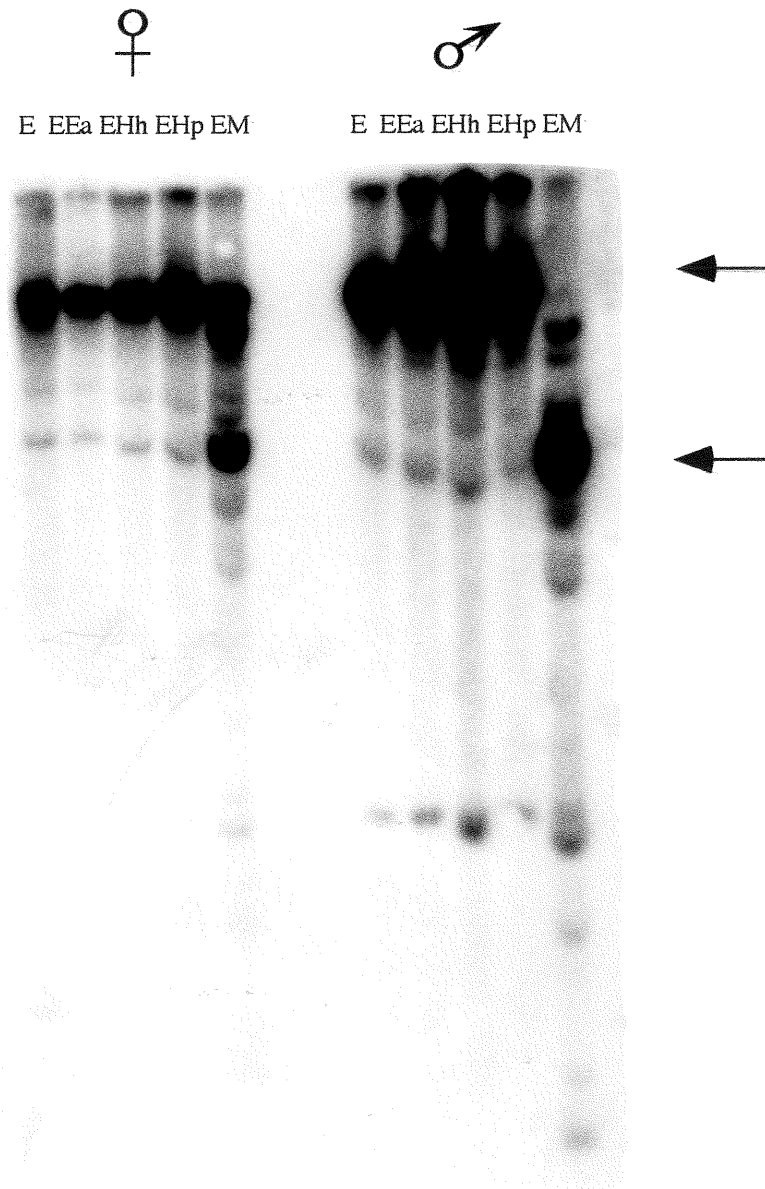


Fig. 19: Hybridation Southern avec le clone L163. L'ADN de donneurs mâles et femelles, isolé de granulocytes purifiés, a été digéré avec *EcoRI* (E), seule, ou suivie de l'une des enzymes suivantes: *EagI* (Ea), *HhaI* (Hh), *HpaII* (Hp) ou *MspI* (M). Les flèches à droite indiquent les bandes d'intérêt.

uniques dans chacune des pistes où il a été digéré avec une enzyme sensible à la méthylation (Fig 20) les bandes migrant plus rapidement que celle se trouvant sur la piste de la digestion *EcoRI* indiquant que l'ADN était au moins partiellement non-méthylé, et que le statut de méthylation était le même sur les deux allèles. Six clones: L3, L59, L61, L62, L88 et L211 ont généré des résultats de Southern suggérant un profil de méthylation d'ADN différentielle: rétention du fragment méthylé pleine longueur dans les échantillons qui avaient été digérés avec les enzymes sensibles à la méthylation plus l'apparition d'autres bandes de plus bas poids moléculaire suggérant que des copies de ce fragment portaient des sites de restriction non-méthylés. Les résultats étaient identiques tant pour l'ADN de femelles que pour celui de mâles.

Pour le clone L88 (Fig.21), la digestion avec *EcoRI* et *EagI* a donné deux bandes dont l'une migrait avec la bande *EcoRI* seule et l'autre, un peu plus intense, migrait plus vite (Fig. 22). Ce profil de restriction suggère la présence d'au moins un site *EagI* dans le fragment *EcoRI* qui était méthylé dans près de la moitié des copies. Le même profil est observé lorsqu'on utilise l'enzyme *HindIII* à la place d'*EcoRI* (Fig 23). Pour s'assurer que ce résultat n'était pas un artéfact de digestion partielle, l'expérience a été répétée avec des quantités croissantes de *EagI*, ayant comme résultat l'obtention de bandes d'intensité relative identiques (données non montrées). Des digestions avec soit *HpaII* (pistes 3 et 7, figure 22) ou *HhaI* ont généré beaucoup de bandes, encore avec l'une d'elles migrant avec la bande *EcoRI* indiquant que les sites *HpaII* (et *HhaI*, données non-montrées) étaient méthylés dans près de la moitié des copies. Les échantillons des pistes 4 et 8 ont été digérés avec *EcoRI* et *MspI* pour montrer le profil attendu des échantillons digérés avec *HpaII* si ces sites étaient complètement non-méthylés.

Le clone L59 (Fig. 47) montre une vingtaine d'identités de séquence avec des ESTs dans la base de données. Les hybridations Southern suggèrent une méthylation variable d'un fragment *PstI* (Fig. 24). *EagI* n'a pas généré des bandes plus petites probablement à cause de l'existence d'un site *PstI* entre le site *EagI* et la sonde. Tout

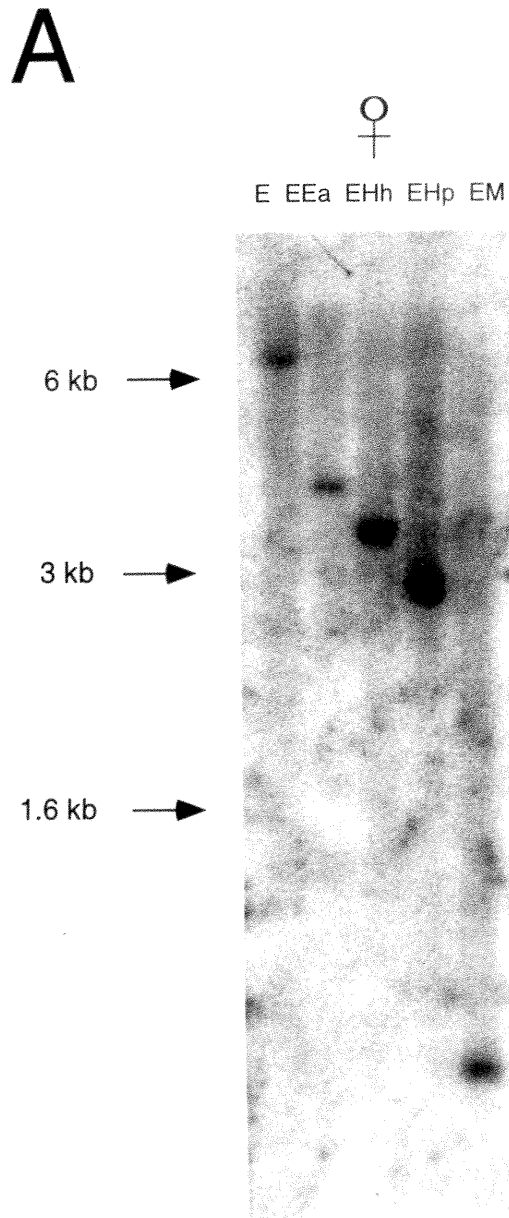


Fig. 20: Hybridation Southern avec le clone L51. L'ADN de donneurs mâles et femelles, isolé de granulocytes purifiés, a été digéré avec *EcoRI* (E), seule ou suivie de l'une des enzymes suivantes: *EagI* (Ea), *HhaI* (Hh), *HpaII* (Hp) ou *MspI* (M).

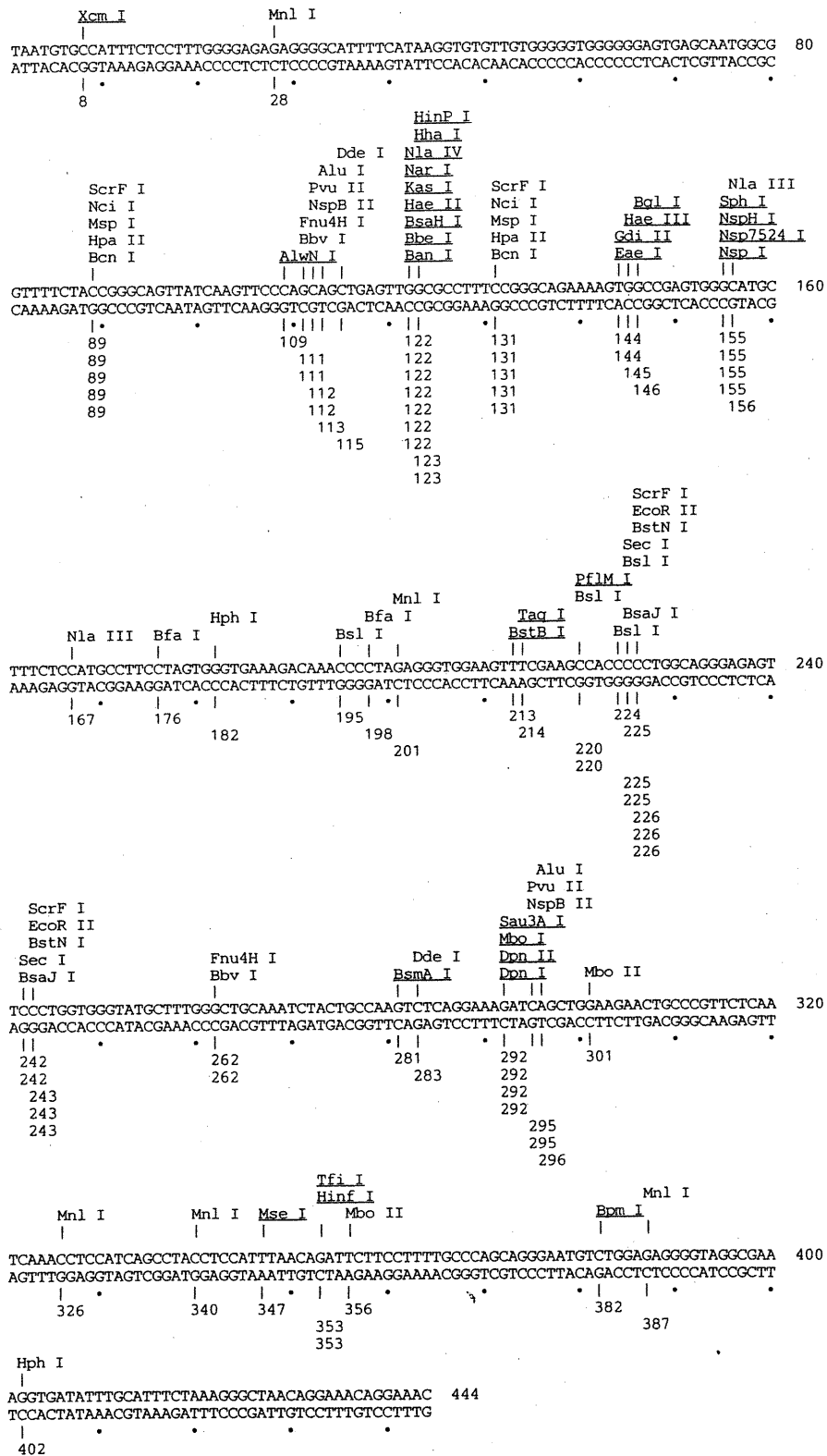


Fig. 21: Carte de restriction du clone FP-CGI L88.

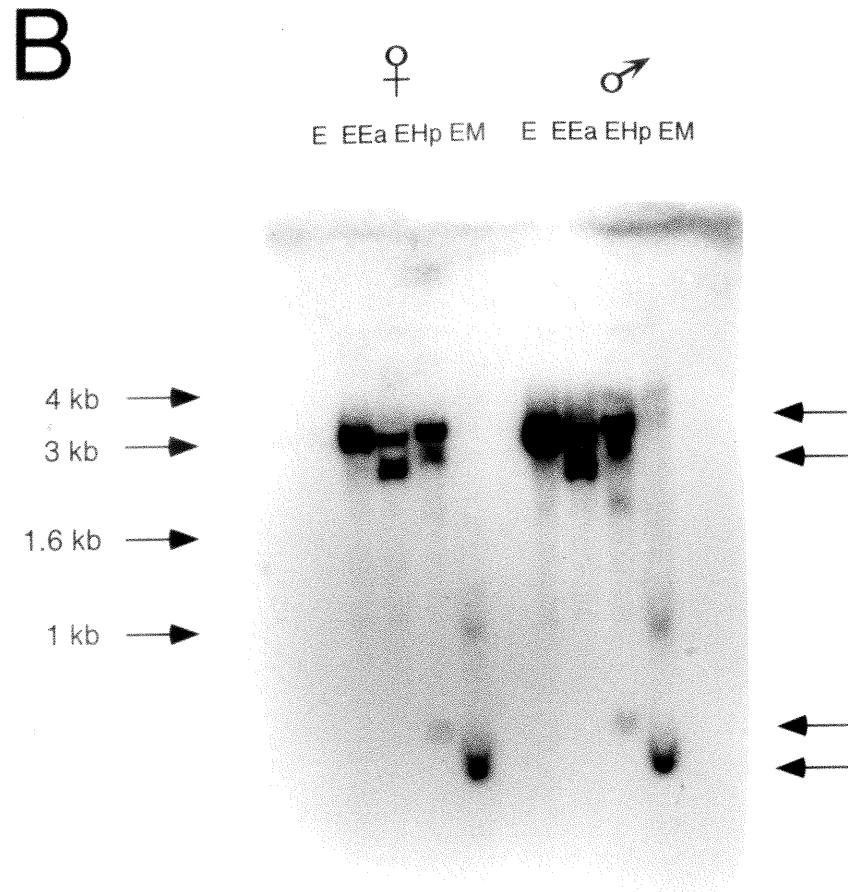


Fig. 22: Hybridation Southern avec le clone L88. L'ADN de donneurs mâles et femelles a été digéré avec *EcoRI* (E) seule, ou suivie de l'une des enzymes suivantes: *EagI* (Ea), *HpaII* (Hp) ou *MspI* (M). Les flèches à droite indiquent les bandes d'intérêt.



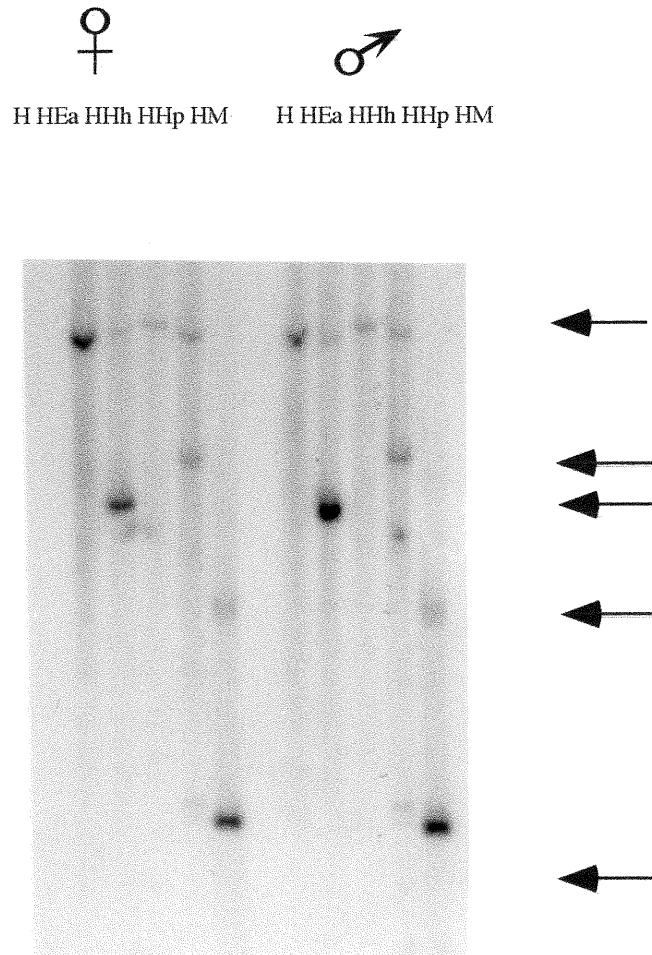


Fig. 23: Hybridation Southern avec le clone L88. L'ADN de donneurs mâles et femelles a été digéré avec *HindIII* (H) seule, ou suivie de l'une des enzymes suivantes: *EagI* (Ea), *HhaI* (Hh), *HpaII* (Hp) ou *MspI* (M). Les flèches à droite indiquent les bandes d'intérêt.



comme avec le clone L88, les intensités des bandes suggèrent que la proportion de la séquence détectée par L59 qui n'a pas été digérée par *HpaII* est de l'ordre de 50 %, ce qui est consistant avec des profils de méthylation allèle-spécifiques. Un profil de méthylation allèle-spécifique semblerait être présent pour la bande femelle digérée par *EagI* et absent chez le mâle mais il pourrait être le résultat d'une migration aberrante de la bande *PstI*.

Le clone L211 (Fig. 25) a donné comme résultat un profil semblable à celui de L59, avec une bande de pleine longueur retenue dans la piste *HpaII +EcoRI* en plus de deux bandes de plus bas poids moléculaire (Fig. 26). Les autres pistes avaient des bandes correspondant seulement au fragment *EcoRI* de pleine longueur.

Le clone L3 (Figs. 27 et 28), de même que L62 (Fig. 29), ont donné essentiellement le même résultat cette fois avec une bande de migration plus rapide dans la piste *EcoRI +HhaI*.

La comparaison des bandes *MspI* et *HpaII* du profil de méthylation du clone L61 (Fig. 30) est suggestif d'un cas de méthylation partielle où l'un des allèles serait méthylé sur un certain nombre de sites de restriction et non pas sur les autres. Vu que nous ne connaissons pas la séquence entourant L61 et donc son patron de restriction, il est plutôt difficile de déterminer s'il s'agit d'une méthylation partielle différentielle allélique.

### 3.3.1.1 Transferts Southern de triples digestions.

Les clones L88 et L59 isolés par FP-CGI ont été utilisés pour hybrider des membranes Southern d'ADN de granulocyte humain d'origine femelle digéré avec *EcoRI* (pour L88) ou *PstI* (pour L59, données non montrées) seules ou en combinaison avec *EagI*, *HhaI* et *HpaII*, *HpaII* et *EagI*, ou *MspI*. Ils ont généré un profil clair de méthylation différentielle où il y a un rapport d'environ 100% pour l'allèle méthylé et de 0% pour l'allèle non-méthylé, mais il restait à savoir si ce profil était celui d'une méthylation différentielle allélique plutôt qu'une méthylation variable sur chaque allèle. Pour

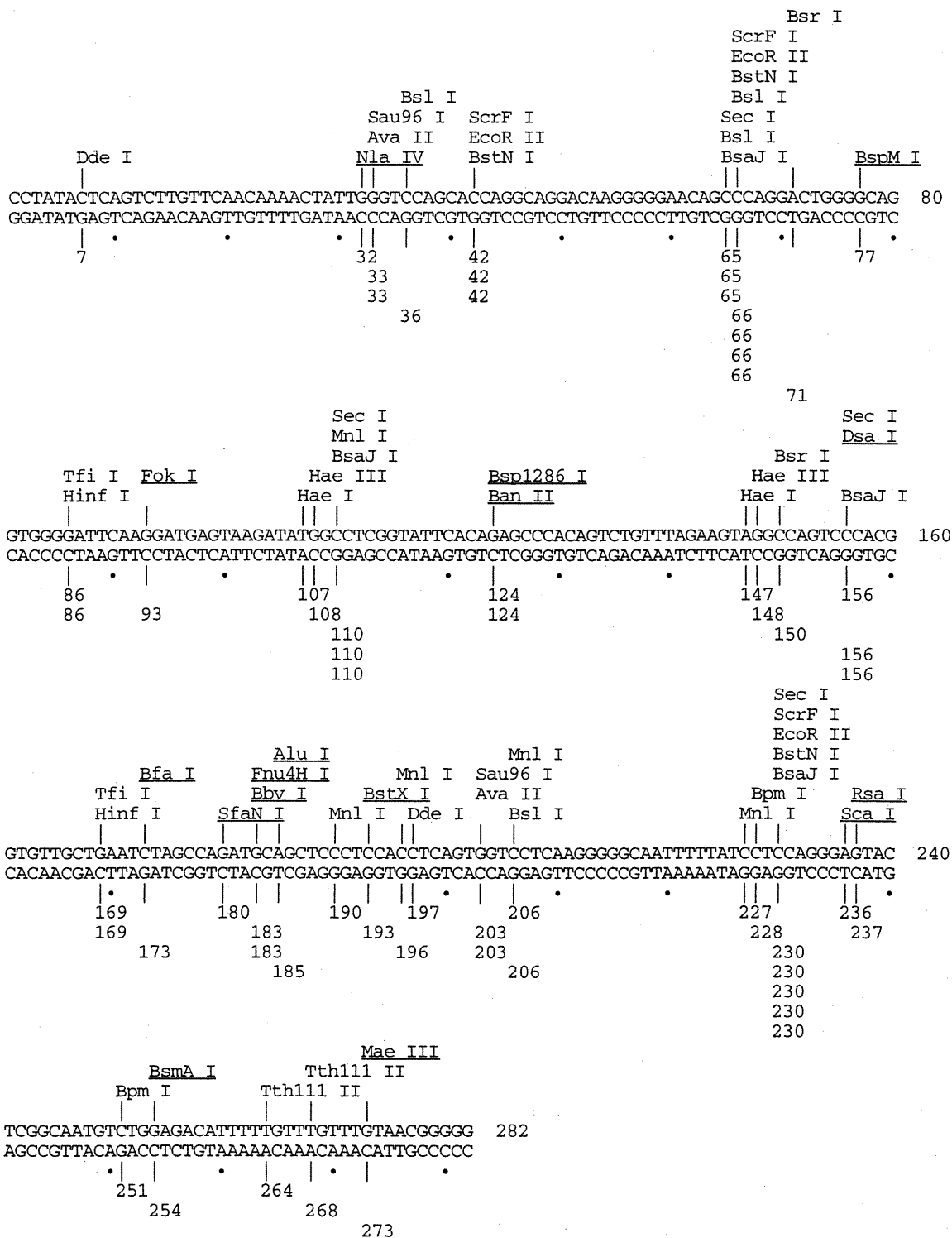


Figure 25: Carte de restriction du clone FP-CGI L211

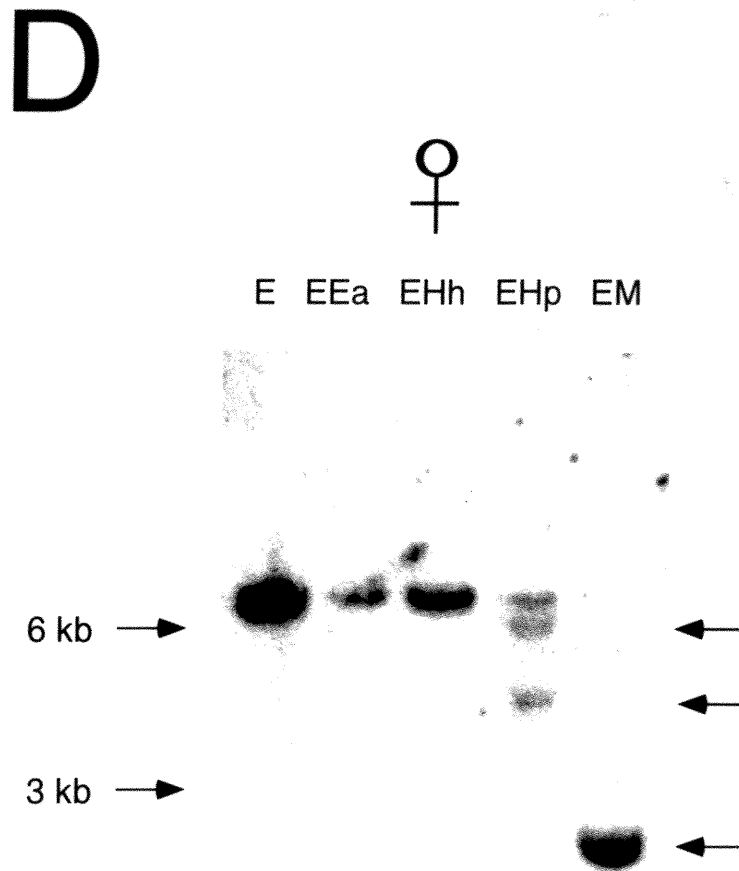


Fig. 26: Hybridation Southern avec le clone L211. De l'ADN de donneuses femelles a été digéré avec *EcoRI* (E) seule, ou suivie de l'une des enzymes suivantes: *EagI* (Ea), *HhaI* (Hh), *HpaII* (Hp) ou *MspI* (M). Les flèches à droite indiquent les bandes d'intérêt.

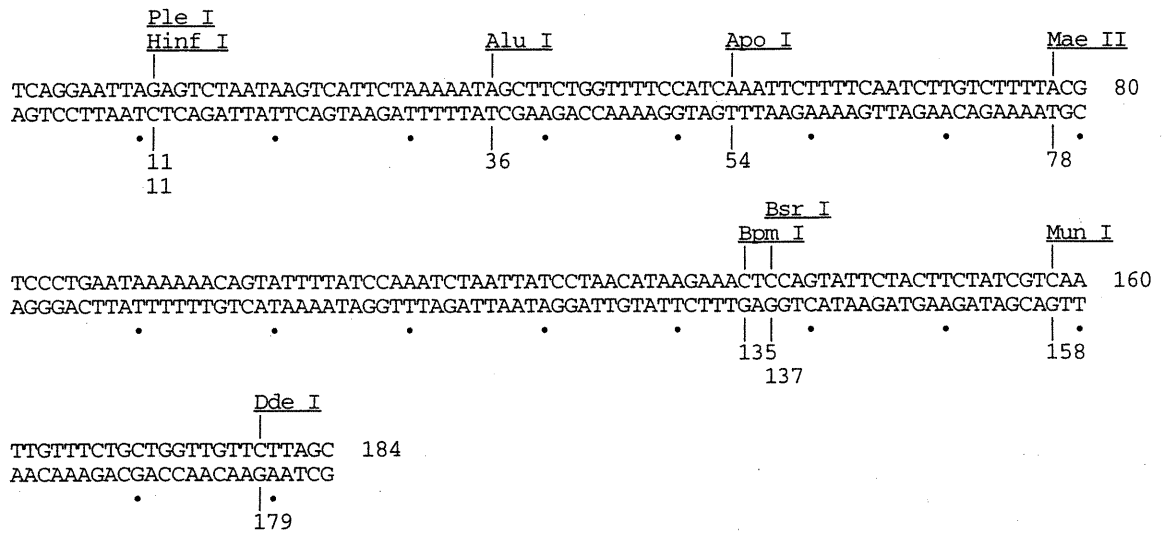


Figure 27: Carte de restriction du clone FP-CGI L3

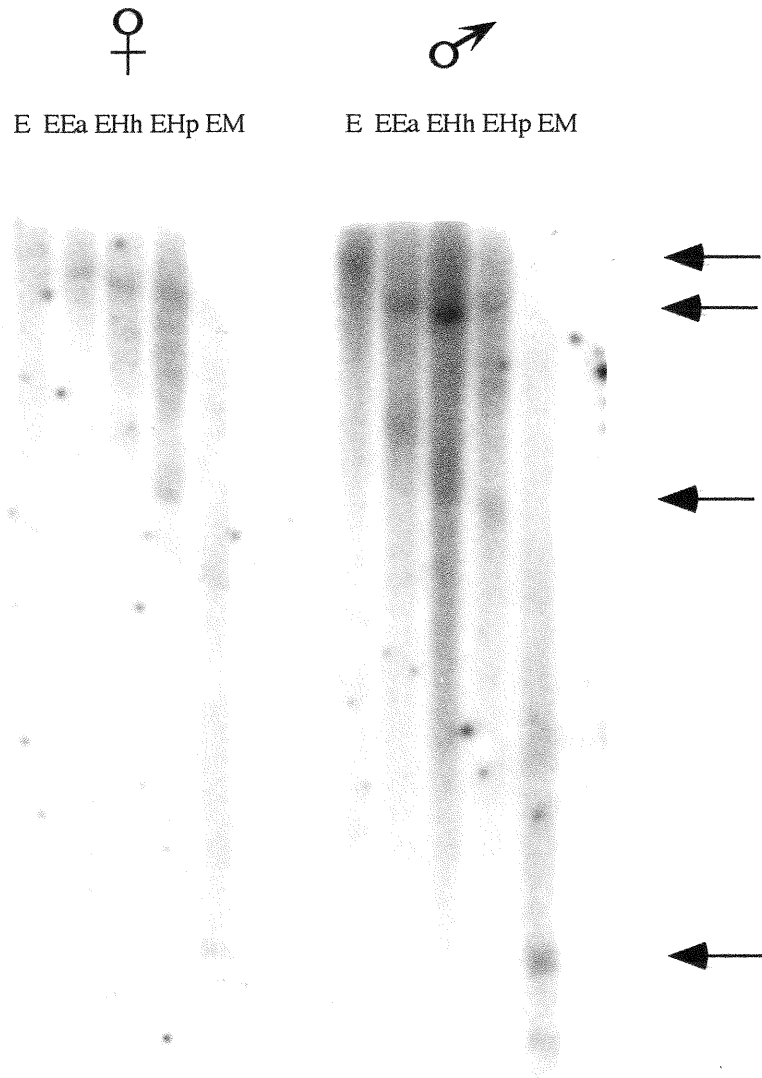


Fig. 28: Hybridation Southern avec le clone L3. L'ADN de donneurs mâles et femelles a été digéré avec *EcoRI* (E) seule, ou suivie de l'une des enzymes suivantes: *EagI* (Ea), *HhaI* (Hh), *HpaII* (Hp) ou *MspI* (M). Les flèches à droite indiquent les bandes d'intérêt.

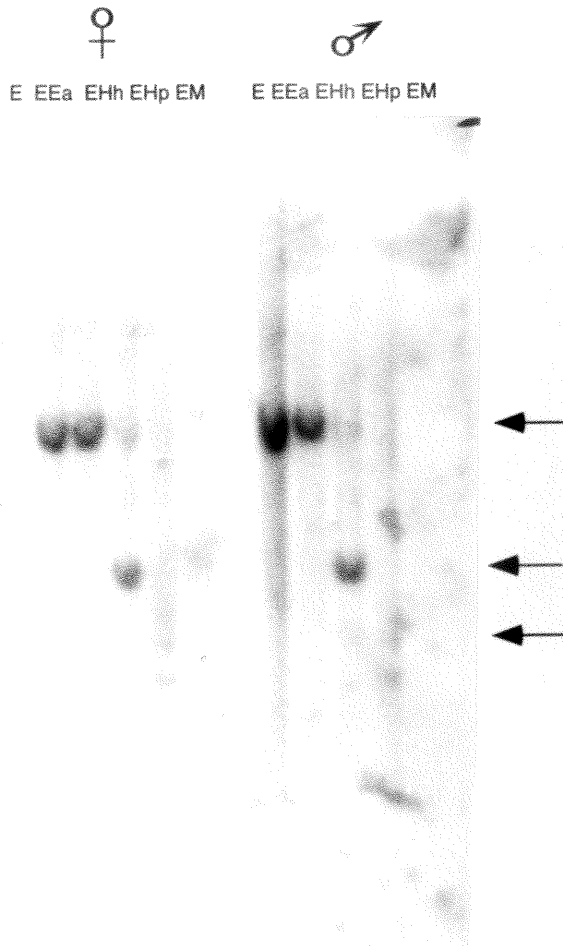


Fig. 29: Hybridation Southern avec le clone L62. De l'ADN de donneurs mâles et femelles a été digéré avec *EcoRI* (E) seule, ou suivie de l'une des enzymes suivantes: *EagI* (Ea), *HhaI* (Hh), *HpaII* (Hp) ou *MspI* (M). Les flèches à droite indiquent les bandes d'intérêt.



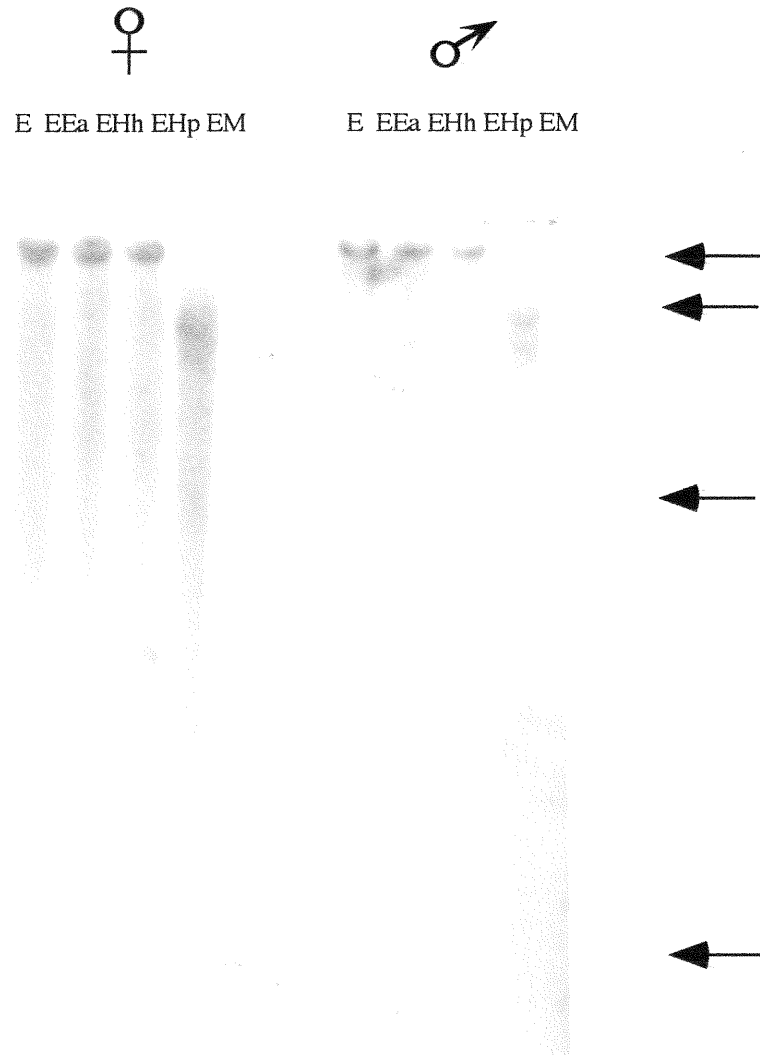


Fig. 30: Hybridation Southern avec le clone L61. L'ADN de donneurs mâles et femelles a été digéré avec *EcoRI* (E) seule, ou suivie de l'une des enzymes suivantes: *EagI* (Ea), *HhaI* (Hh), *HpaII* (Hp) ou *MspI* (M). Les flèches à droite indiquent les bandes d'intérêt.

répondre à cette question, nous avons utilisé ces clones comme sondes pour hybrider des membranes contenant de l'ADN de granulocytes digéré simultanément avec trois enzymes de restriction dont les sites de reconnaissance se trouvent dans le clone. La Fig. 31 montre donc l'hybridation avec le clone L88 qui a deux sites *HpaII* et deux sites *HhaI*. Toutes les pistes montrent rétention de la bande de haut poids moléculaire *EcoRI* observée antérieurement (Fig. 22) surtout celle contenant les digestions *EcoRI+HhaI+HpaII* ce qui suggère très fortement que la méthylation différentielle du clone L88 est allélique et non pas variable.

### 3.3.2 Transferts Southern de différents tissus.

Il y a des évidences que l'empreinte génomique parentale est souvent histospécifique. Étant donné qu'une telle empreinte est souvent associée à une méthylation différentielle allélique, nous avons examiné si le profil de méthylation différentielle de L88 était histospécifique. Pour répondre à cette question, nous avons hybridé avec L88 des membranes contenant de l'ADN de différents tissus digérés avec *EcoRI* et *EagI* de façon à garder un profil d'hybridation déjà connu (Figs. 22 et 31). Les échantillons d'ADN provenaient de granulocytes, de tissu ovarien, d'une tumeur ovarienne, de tissu musculaire lisse, de la lignée lymphoïde HL-60, de tissu pulmonaire et d'une tumeur pulmonaire de la même origine que le tissu pulmonaire. Ces hybridations (Fig. 32) ont donné des profils similaires à ceux attendus. Cependant, nous pouvons remarquer que dans la piste d'ADN de tissu ovarien, la méthylation différentielle est perdue telle qu'illustrée par la présence de la bande de plus haut poids moléculaire seulement. Cela est indicatif que la région entourant L88 est méthylé sur les deux allèles chez le tissu ovarien. Les profils différents donnés par les échantillons de tumeur ovarienne et de tissu pulmonaire sont probablement le produit d'une migration plus rapide des échantillons dûe à une plus grande concentration d'ADN pour ces échantillons.

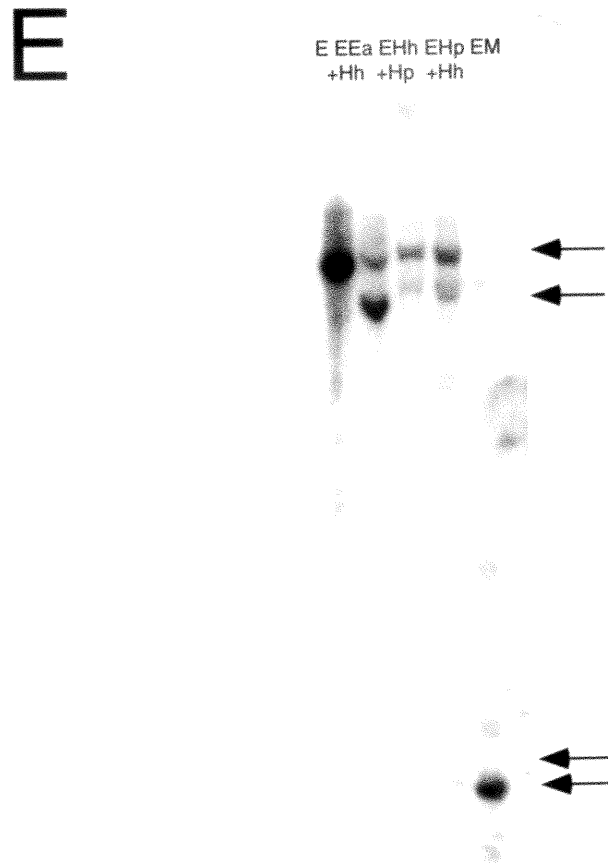


Fig. 31: Hybridation Southern avec le clone L88. L'ADN de donneuses femelles a été digéré avec *EcoRI* (E) seule, ou suivie des enzymes suivantes: *EagI* (Ea), *HhaI* (Hh) et *HpaII* (Hp) ou *MspI* (M). Les flèches à droite indiquent les bandes d'intérêt.

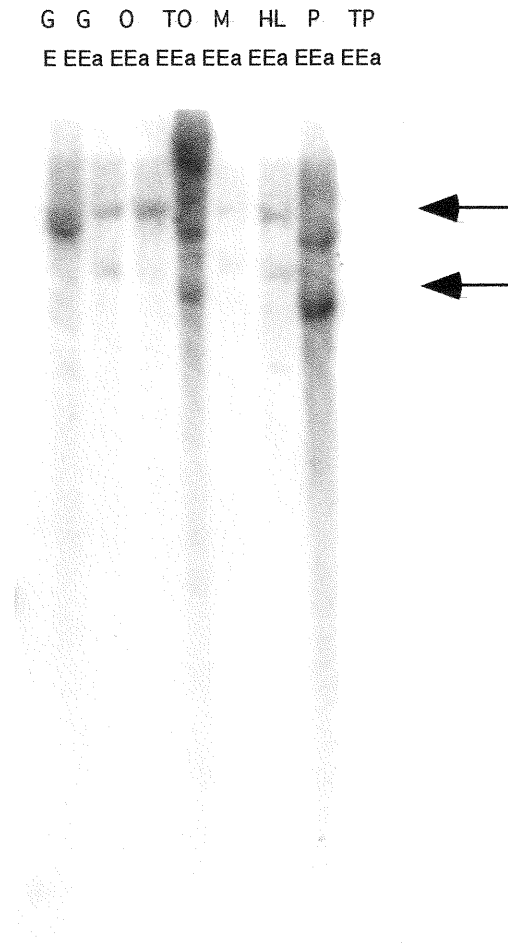


Fig. 32: Hybridation Southern d'ADN de différents tissus avec le clone L88. L'ADN a été digéré avec *EcoRI* (E) seule, ou en combinaison avec *EagI* (Ea). Les flèches à droite indiquent les bandes d'intérêt.

G=granulocytes, O=tissu ovarien, TO=tumeur ovarienne, M=tissu musculaire lisse, HL=lignée cellulaire HL-60, P=tissu pulmonaire, TP= tumeur pulmonaire.

### 3.4 LOCALISATION CHROMOSOMIQUE DES CLONES MÉTHYLÉS DIFFÉRENTIELLEMENT OBTENUS PAR FP-CGI.

Dans la section précédente, nous avons démontré la présence des profils de méthylation différentielle pour six des quatorze clones, sur les quarante-deux isolés par FP-CGI, pour lesquels nous avons des résultats analysables. Malheureusement et malgré le fait que trois de ces séquences (L59, L61 et L62) étaient des ESTs selon des recherches BLAST, nous n'avions aucune idée de leur position dans le génome humain. Comme les domaines des génomes humain et de souris qui ont une empreinte génomique parentale sont bien connus et caractérisés, la localisation de chaque clone candidat dans le génome humain ou sur celui de la souris est un objectif qui doit mettre en évidence la présence d'une quelconque empreinte parentale.

Nous avons donc essayé de localiser quatre clones, L3, L59, L88 et L211 sur les chromosomes humains à l'aide du panel SCHP2. Les 27 lignées cellulaires composant le panel furent criblées par PCR à l'aide de STSs spécifiques à chacun des clones (Figs. 33 à 36).

Suite à la localisation chromosomique de ces clones, nous avons voulu préciser leur position en utilisant le "panel de hybrides de radiation G3". Le STS du clone L88 a donné un signal positif sur des hybrides portant le chr 22, mais aussi sur ceux portant les chr 4 et 20 (Fig 33). Ce dernier hybride porte aussi un petit fragment du chr 22q11 qui est lié à IGCL2 (Dubois et Naylor 1993), ce qui suggérait que L88 était dans cette région sur le chr 22. Les PCRs sur les hybrides de radiation (Fig. 37) ont confirmé la localisation du STS L88 en obtenant du serveur SHGC une liaison (LOD score: 8.99) avec le marqueur D22S427 qui se trouve dans la région centromérique du chr 22 (Fig. 38). D'ailleurs, en criblant la librairie de YACs de Généthon (CEPH "B") à l'aide du STS L88, nous avons identifié les YACs 966A8 et 924B9 (données non montrées) qui sont localisés dans la région de Di George, près de D22S427 et de TUPLE1 (Lamour et coll. 1995).

Le clone L3 a été localisé sans équivoque au chr 18 (Fig. 34). La cartographie par hybrides de radiation (Fig. 39) a montré que L3 était lié au marqueur SHGC-37630 (LOD 3.0, Tableau VI), qui est à son tour lié aux marqueurs D18S1133 et D18S104E sur 18q11.

L211 a détecté des séquences sur deux hybrides du panel de Coriell, ceux ayant les chrs 1 et 12 (Fig. 35). Ce n'est pas inhabituel pour un hybride de porter de courtes séquences de chromosomes autres que ceux qui sont caryotypiquement visibles (e.g. voir pour L88, voir aussi (Brown et coll. 1994)). Donc, ce résultat ne veut pas nécessairement dire que L211 se trouve sur les deux chromosomes. Dans le cas de L211, la cartographie avec les hybrides de radiation a permis de démontrer un lien avec le marqueur CHLC.GAAT4D10 ou D1S1592 (LOD 3.6) qui localise au chr1p36 (Tableau VI). Nous n'avons pas trouvé de liaison avec des marqueurs sur le chr 12.

La sonde L59 a été localisée au chr 6 (Fig. 36). Le criblage des hybrides de radiation avec L59 a, cependant, donné des résultats ambigus (Fig.40, Tableau VI) dûs probablement à la faible intensité des bandes amplifiées à partir de quelques-uns des hybrides. Cela nous a empêché de localiser avec précision le clone L59. Cependant, nous avons été capables de localiser ce clone sans équivoque en utilisant une approche cytogénétique (voir section 3.5).

### **3.5 RECHERCHE ET UTILISATION DE FRAGMENTS GÉNOMIQUES DE HAUT POIDS MOLÉCULAIRE: BACs et YACs CONTENANT DES CLONES FP-CGI.**

Dans la section précédente nous avons procédé à la localisation des quatre clones méthylés différemment. Bien que l'un des clones testés (L59) corresponde à un EST, aucun de ces clones donnait de l'information sur un gène déjà identifié ou caractérisé. De plus, les gènes avec empreinte génomique parentale se trouvent presque toujours regroupés dans des domaines chromosomiques de quelques megabases. Donc, en travaillant avec de longs fragments génomiques, nous pourrions nous trouver dans une situation où nous pourrions travailler avec plus d'un gène ayant

Chromosome

BRL Hu So Ha Y X 1,X,6 5 4 8 3 21 19 4,20 14 17 16 9 6 7 2 12 22 13 10 11 18 15 Eau

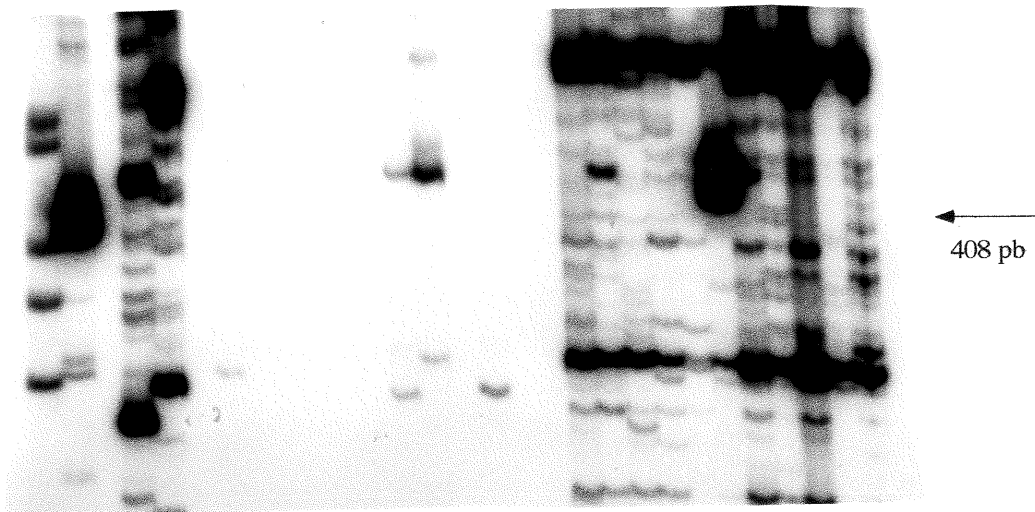


Fig. 33: Criblage par PCR du panel 2 de Coriell menant à la localisation chromosomique du clone L88.

Ha=Hamster chinois, Hu=humain, So=souris, BRL=marqueur de poids moléculaire.

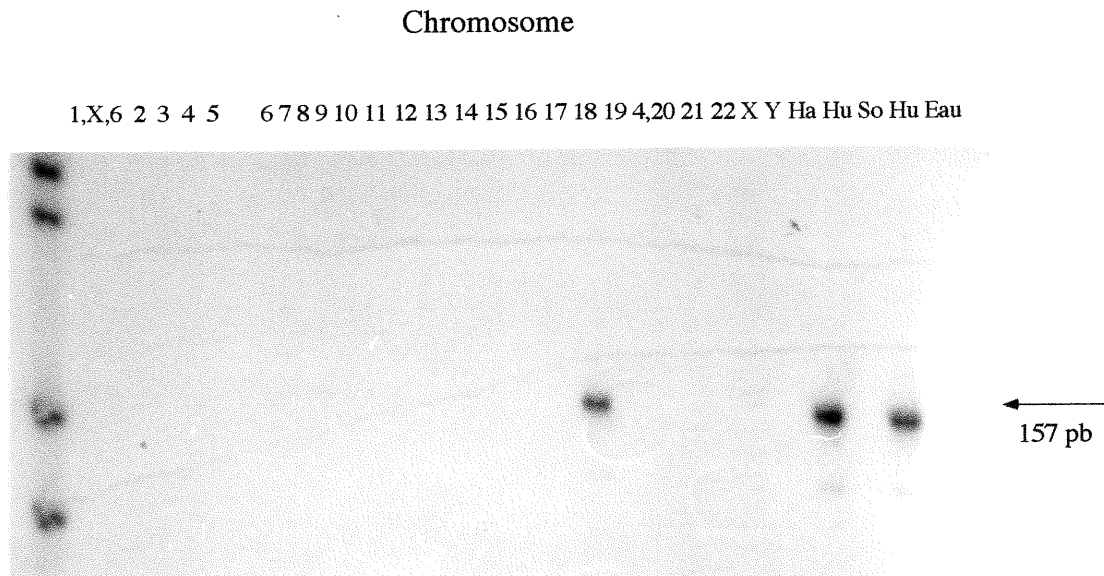


Fig. 34: Criblage par PCR du panel 2 de Coriell menant à la localisation chromosomique du clone L3.

Ha=Hamster chinois, Hu=humain, So=souris, BRL=marqueur de poids moléculaire.



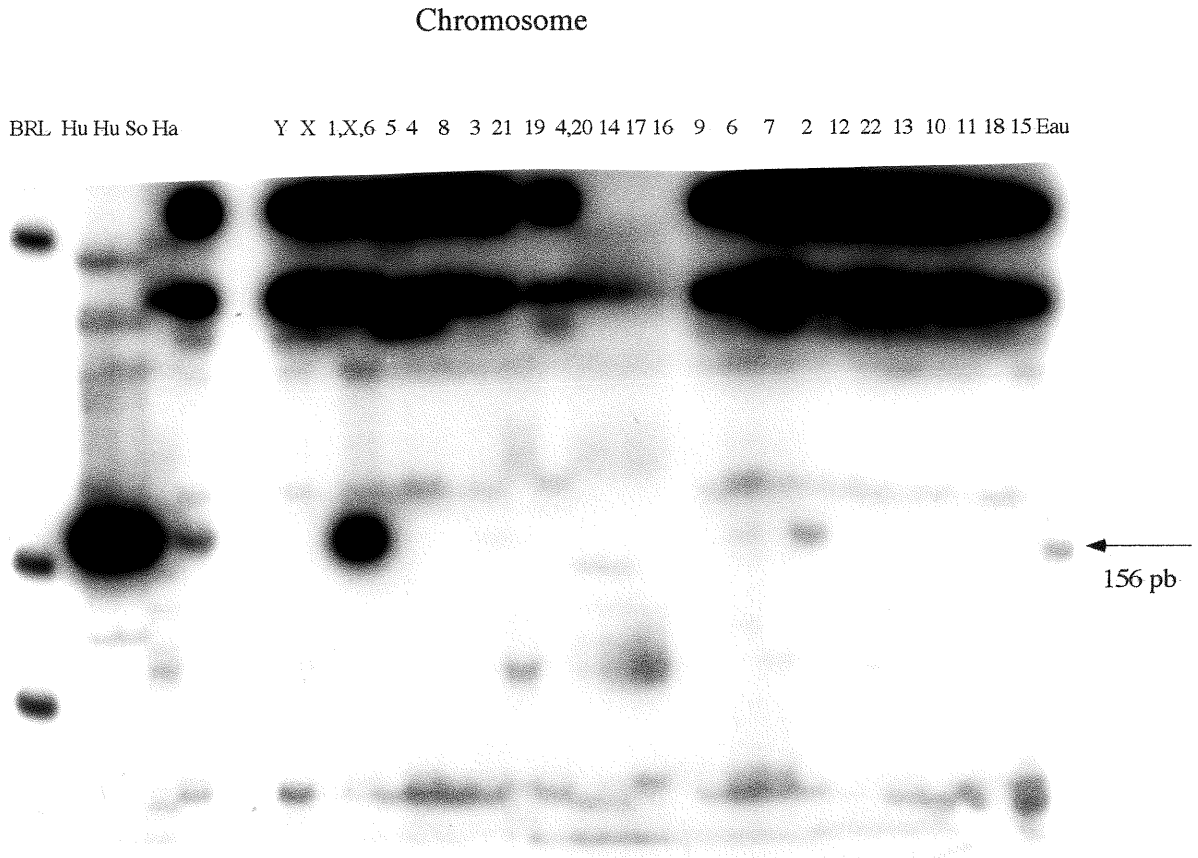


Fig. 35: Criblage par PCR du panel 2 de Coriell menant à la localisation chromosomique du clone L211.

Ha=Hamster chinois, Hu=humain, So=souris.

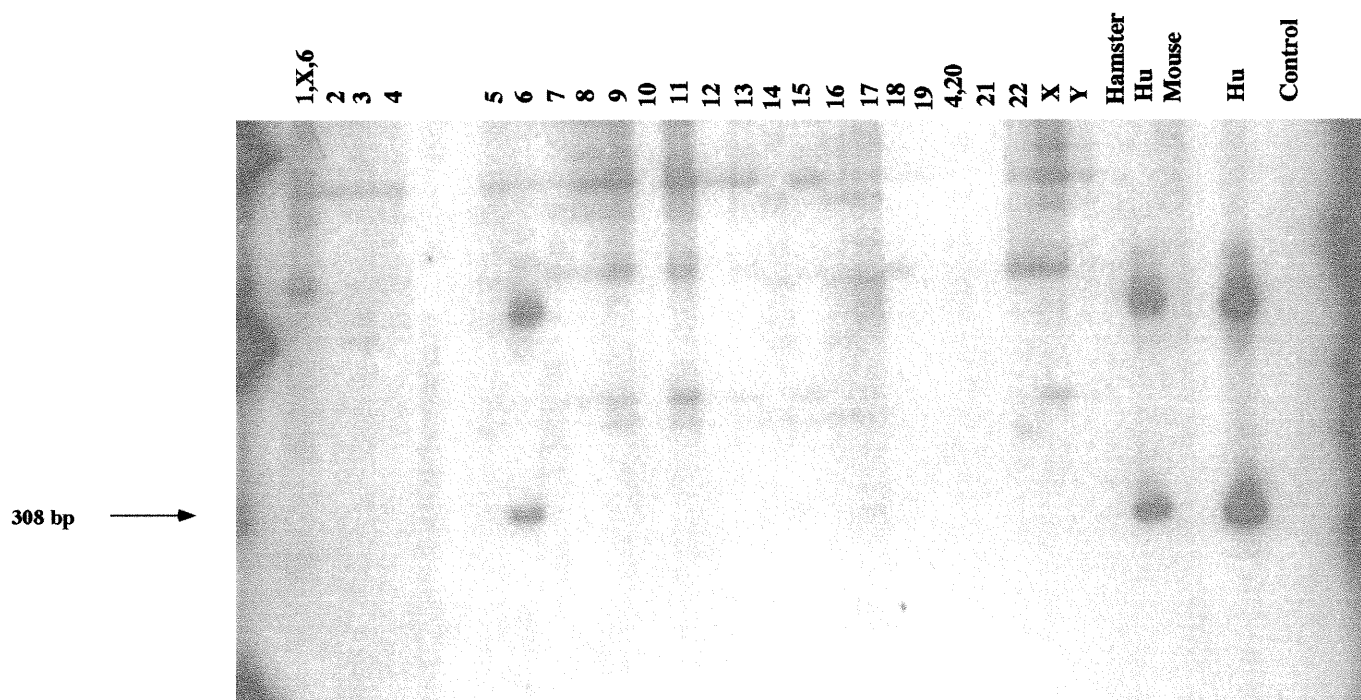


Fig. 36: Criblage par PCR du panel 2 de Coriell menant à la localisation chromosomique du clone L59.

Ha=Hamster chinois, Hu=humain, So=souris.

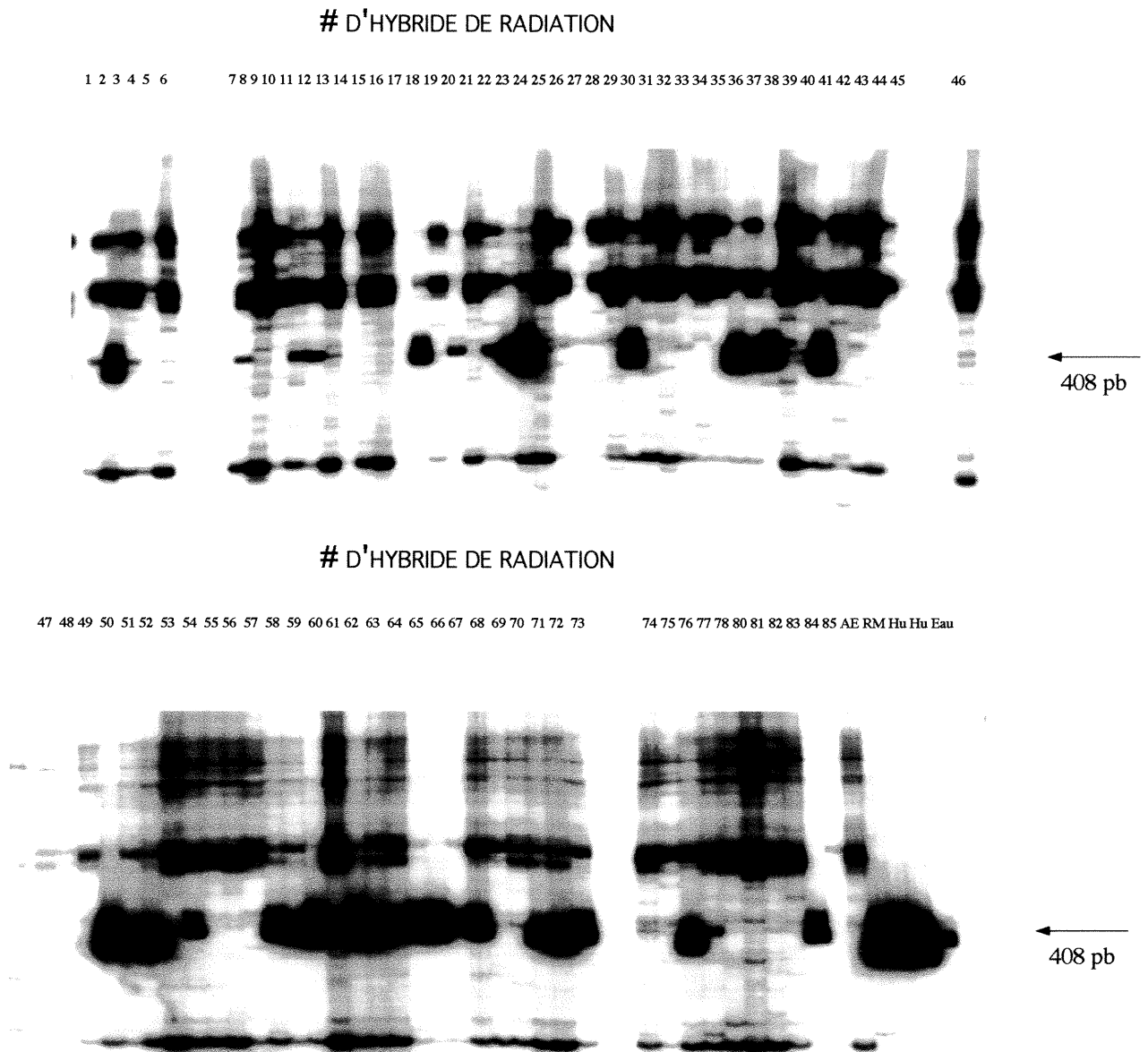


Fig. 37: Criblage par PCR du panel G3 d'hybrides de radiation menant à la localisation chromosomique du clone L88.

AE=lignée de Hamster, RM=lignée cellulaire humaine,  
Hu=granulocytes humains.

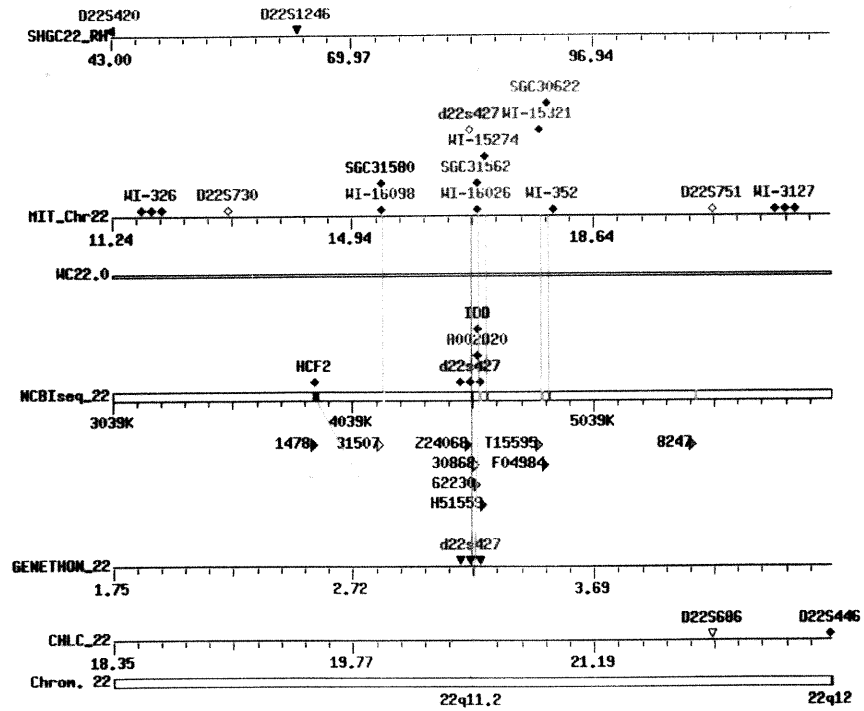
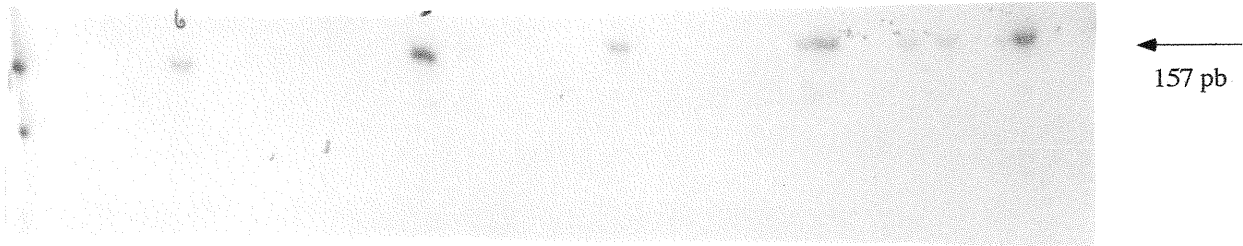


Fig. 38: Intégration des cartes physique et génétique de la région du chromosome 22 entourant le clone L88.

# D<sup>1</sup>HYBRIDE DE RADIATION

1 2 3 4 5 6

7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45



# D<sup>1</sup>HYBRIDE DE RADIATION

46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 80 81 82 83 84 85 AE RM Hu Eau Eau

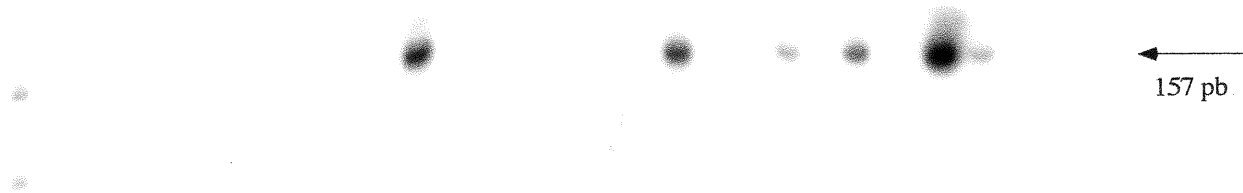


Fig. 39: Criblage par PCRs du panel G3 d'hybrides de radiation menant à la localisation chromosomique du clone L3.

AE=lignée de Hamster, RM=lignée cellulaire humaine, Hu=granulocytes humains.

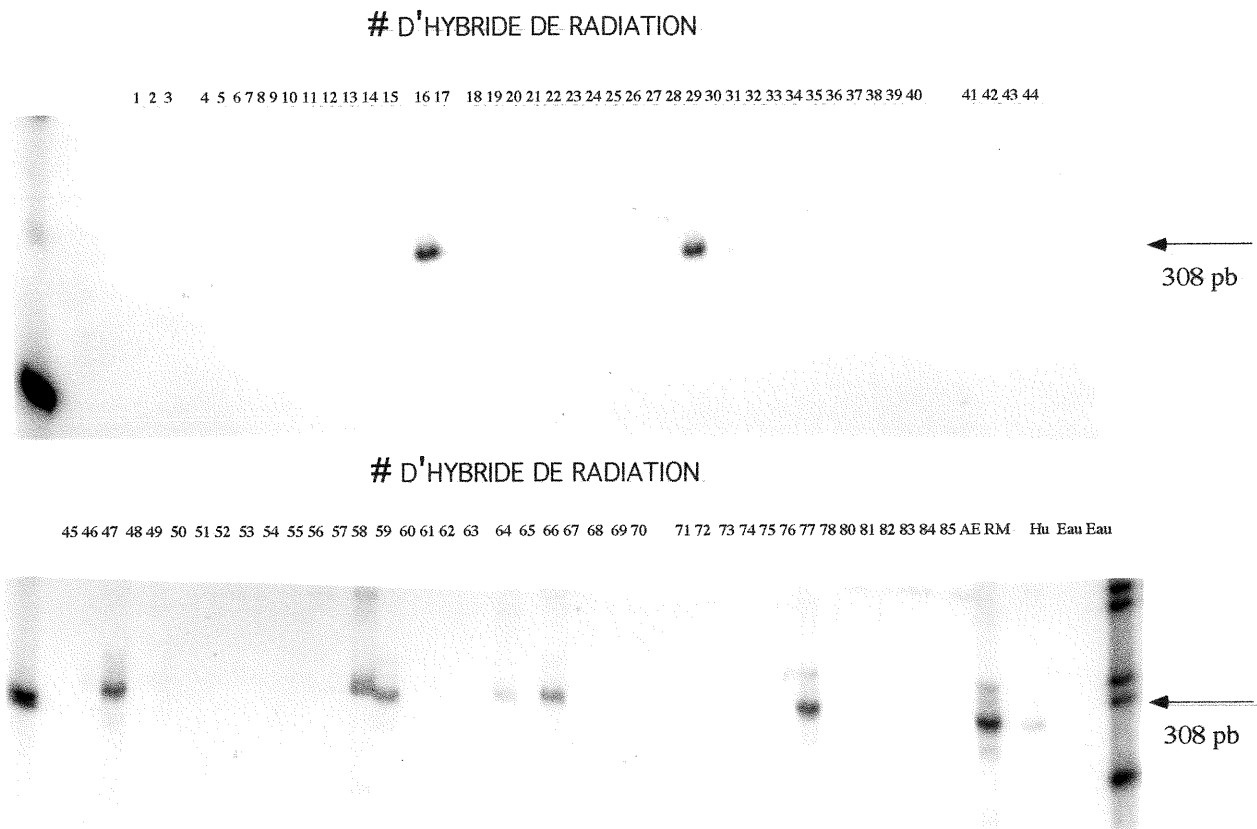


Fig. 40: Criblage par PCR du panel G3 d'hybrides de radiation menant à la localisation chromosomique du clone L59 .

AE=lignée de Hamster, RM= lignée cellulaire humaine, Hu=granulocytes humains.

**Résultats des hybrides de radiation**

<b>Clone</b>	<b>Marqueur lié</b>	<b>Marqueurs liés avoisinants</b>	<b>Localisation</b>
L3	SHGC-37630	D18S1133, D18S456	18q11
L59	D6S995	D6S284, D6S1646	6q16-6q24, 6q27
L88	D22S427	D22S1246, D22S751	22q11
L211	CHLC.GAAT4D10	D1S507, D1S483	1p36

Tableau VI: Localisation chromosomique des clones L3, L59, L88 et L211 à l'aide des hybrides de radiation.

une empreinte parentale. Nous avons donc entrepris de chercher de fragments génomiques de haut poids moléculaire.

Jusqu'à tout récemment le clone génomique de préférence était le YAC ("yeast artificial chromosome"). Cependant, il y a quelques désavantages lorsqu'on manipule ce vecteur. Souvent, ils subissent des réarrangements et ils ne sont pas très faciles à manipuler. Mais en 1992 (Shizuya et coll. 1992) un nouveau système de clonage bactérien (les BACs) a été décrit qui permet le clonage de grands fragments d'ADN (entre 80 et 300 kb) sans les inconvénients associés aux YACs. Nous avons donc décidé de travailler avec des BACs ou "bacterial artificial chromosomes" (exemple illustré à la figure 41).

Avec le STS L88, le BAC 273D14 a été identifié. Pour le clone L3, nous avons obtenu les BACs 518M24 et 290B7 et pour le clone L211 les BACs 349P21 et 321O20. Les BACs 528D2 et 421M16 sont associés avec L59.

Tel qu'il a été décrit dans la section antérieure, nous avons eu des résultats ambigus pour la localisation chromosomique du clone L59. Nous avons décidé de déterminer sa localisation par FISH en utilisant le BAC 421M16. L'essai de FISH, qui a été effectué au labo de la Dr. Nicole Lemieux à l'Université de Montréal a localisé le clone L59 sur la bande 6q21 (Fig. 42).

Nous avons ensuite passé à la prochaine étape de notre travail qui était orientée sur deux axes. Le premier étant la recherche, à l'aide des longues séquences génomiques (BACs ou YACs), de gènes ou de séquences exprimées se trouvant près de nos clones. La deuxième étant l'étude des caractéristiques de méthylation différentielle allélique ou d'expression différentielle allélique de nos clones et/ou des séquences environnantes. Nous avons décidé de concentrer nos efforts sur le clone L59.



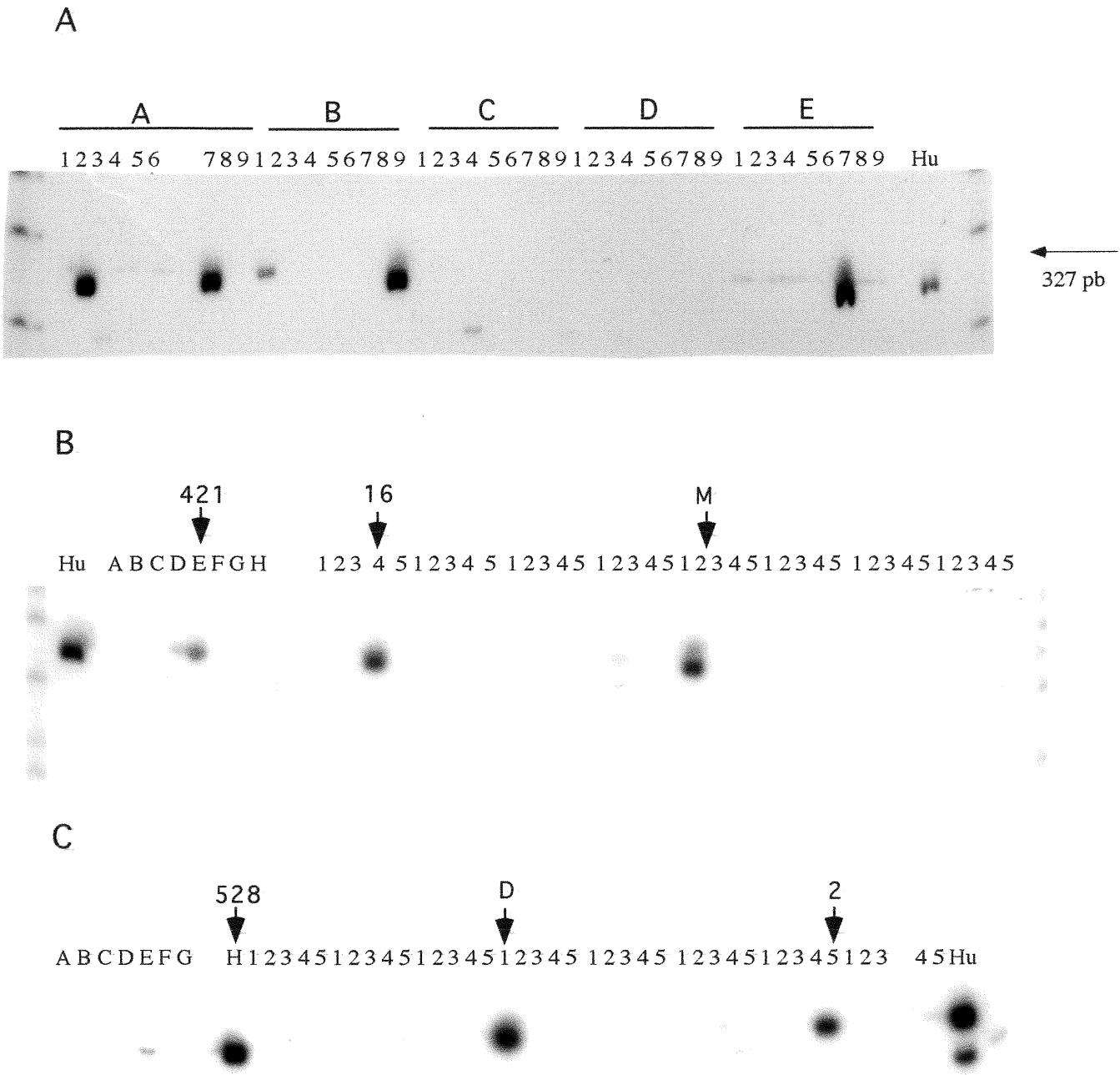


Fig. 41: Criblage d'une librairie de BACs avec un STS dérivé du clone L59. La partie A montre des PCRs sur des "superpools" de BACs. Les parties B et C montrent des PCRs sur les "plate pools" (A à H) et sur les "row" et "column" pools (groupes de 1 à 5) pour les superpools E7 (partie B) et B9 (partie C) de la partie A pour identifier les BACs 421M16 et 528D2.

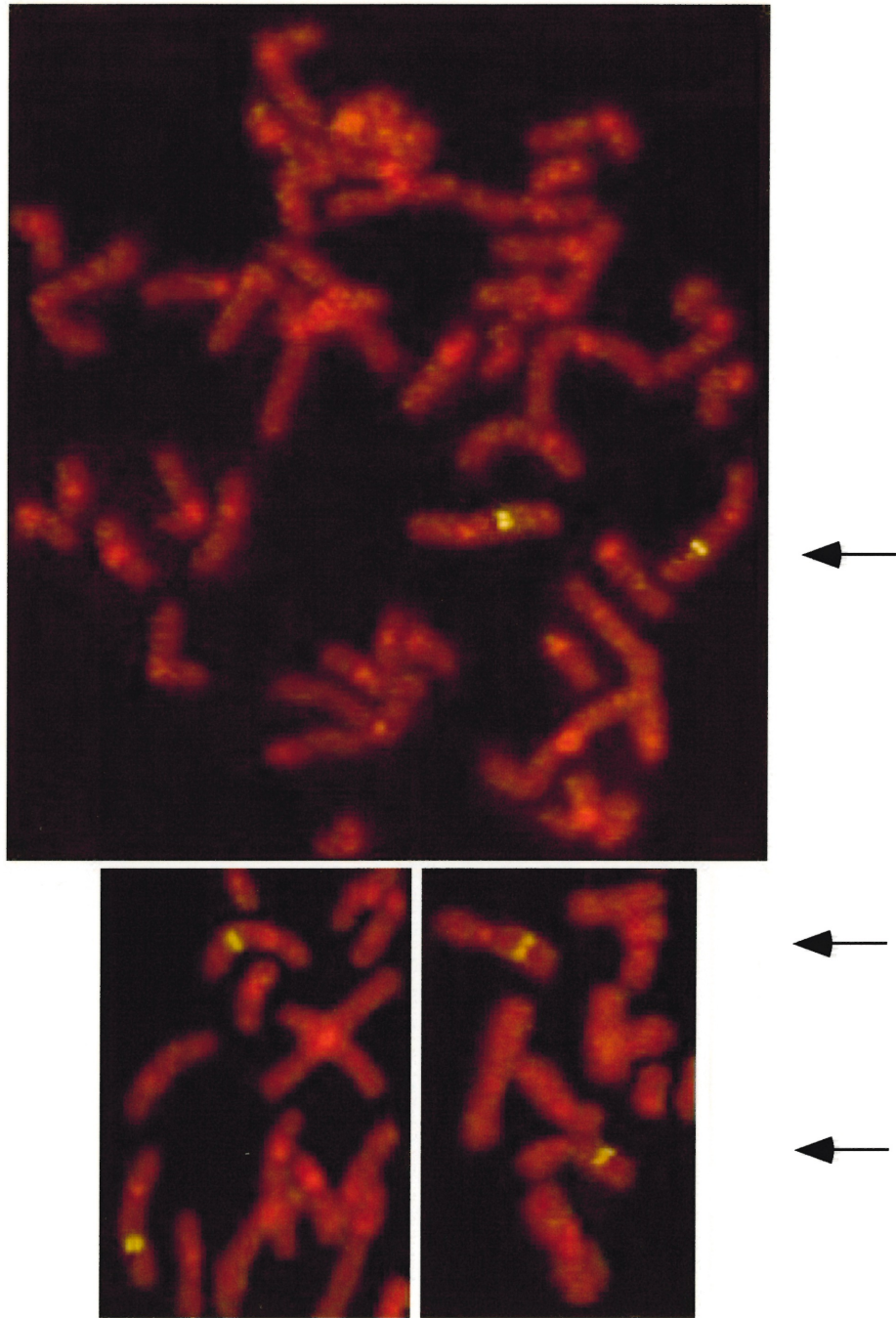


Fig. 42: Résultats de FISH effectué avec le clone L59 ou BAC 421M16. Un signal positif (couleur jaune) est observé chez les deux homologues du chromosome 6, bande q21.

### 3.5.1 *Technique de l'amplification d'exon ou la recherche de gènes exprimés différemment.*

Nous avons expliqué l'approche d'amplification d'exon sur les BACs 421M16 et 528D2 qui contiennent des séquences environnant le clone L59 dans les sections précédentes. Nous avons effectué un total de soixante transfections transitoires de cellules COS-1 qui ont mené à l'identification de onze produits de PCR, la digestion avec *Bst*XI et le deuxième PCR "nested" nous avons identifié onze bandes ayant au moins une centaine de bases de plus que la bande de 177 pb (données non montrées). Quatre de ces fragments ont été clonés et séquencés. Les comparaisons des séquences avec les bases de données (recherches BLAST) ont révélé que trois des quatre séquences provenaient du plasmide pSPL3. La quatrième séquence était une séquence unique selon les résultats BLAST. Vu le grand nombre de séquences exprimées qu'on trouve dans la banque dbEST, si cette quatrième séquence avait été une séquence exonique, elle aurait donné des homologues avec cette banque. Donc, elle n'a pas été jugée intéressante.

## 3.6 ANALYSES DES SÉQUENCES

### 3.6.1 *Clone L59*

Les séquences des clones obtenus par FP-CGI ont été utilisées pour faire des recherches BLAST. Trois des cinq clones méthylés différemment ont démontré certaines identités de séquence avec des ESTs. Le clone L59 (1091 pb) est identique à plusieurs ESTs humains. En comparant ces séquences (Figure 43) on peut constater que L59 fait probablement partie d'un gène contenant au moins des morceaux de quatre exons (illustrés en rouge).

Tel qu'on voit dans la figure 44 les trois introns de L59 contiennent des sites accepteur et donneur d'épissage consensus. En utilisant les ESTs chevauchant les extrémités 5' et 3' de L59, nous avons pu construire un THC ("tentative human consensus") de 1353 pb (Fig. 44) dont la partie 3' arrive à la queue poly A. Nous avons prouvé que la partie

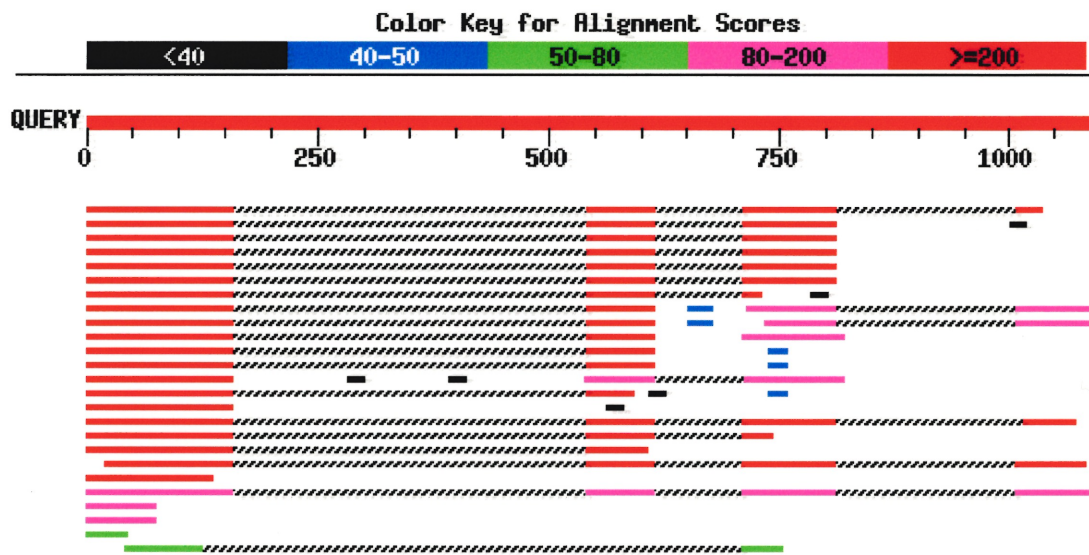


Fig 43: Homologies de séquences obtenues en utilisant le clone L59 ("query"=1091pb) dans une recherche BLAST (banque: dbEST) .

TTACCTCCTTAGATGGAGGCCGAGACCCCAAGTACTGAGGTGCCACCTGACCCAGAGCC  
TGGTGTACCCTGACACCCCATGCTAACACCAGGAGGCCGGTGTGGGGACTTGTGCA  
CTTTGTGGGGAACACCTCTATGTCCTGGAACGCCTCTGTGTCAACGGCCATGTCTCCAC  
CGGAGCTGCTTCCGCTGCCATACCTGTGAAGCCACACTGTGGCCAGGTGGCTACGAGCA  
GCACCCAGGAGATGGACATTTCTACTGCCTCCAGCACCTGCCCCAGACAGACCACAAAG  
CGGAAGGCAGCGATAGAGGCCCTGAGAGTCCGGAGCTCCCCACACCAAGTGAGAATAG  
CATGCCACCAGGCCTCTCAACTCCCACAGCCTCGCAGGAGGGGGCCGGTCTGTTCCA  
GATCCCAGCCAGCCCACCCGTCCGCAGATCCGCCTCTCCAGCCCAGGAGCGCCGTTGTC  
CTCCCTAACCTTACCCTGACCCGAAATGGAGCCGAGCCTCCACCCAAGCCTCCCCG  
CAGCTGCTCCGCTTGGCCCGCCACGCCCTGGAGAGCAGCTTTGTGGGCTGGGGCCTG  
CCAGTCCAGAGCCCTCAAGCTCTTGTGGCATGGGAAGGGGAAAAAGAGAGTCCCTTCTC  
CAGTGAAGAGGAAGAAGAAGATGTGCCTTTGGACTCAGATGTGGAACAGGCCCTGCAGA  
CCTTTGCCAAGACCTCAGGCACCATGAATAACTACCCAACATGGCGTCGGACTCTGCTGC  
GCCGTGCGAAGGAGGAGGAGATGAAGAGGTTCTGCAAGGCCAGACCATCCAACGGCG  
ACTAATGAGATTGAGGCTGCCTTGAGGGAGCTAGAGGCCGAGGGCGTGAAGTGGAGC  
TGGCCTTGAGGCGCCAGAGCAGTGAGTGACAAGCTAGGACACCCCTCTGCCTTCTAGCA  
CTGTGTTGAGCTCCACGTTCCAGCCCGTAGAGTCCCAGTGGCTATGCCAACAGCTAACA  
CATACCTGTGACACCCGGCTCTCCTGTCTCCCCTAAGCCTGTAACCTCCTGTGGTGCTGGC  
TGGAAGCTCCCTGATCACCCCTACCCATCTTCTTCTCAGGTTCCCCAGAACAGCAAA  
AGAAACTATGGGTAGGACAGCTGCTACAGCTCGTTGACAAGAAAAACCACCTGGTGGCT  
GAGGAGGCCGAGCTCATGATCACGTGAGTGACAAGCTAGGACACCCCTCTGCCTTCTAG  
CACTGTGTTGAGCTCCACGTTCCAGCCCGTAGAGTCCCAGTGGCTATGCCAACAGCTAA

CACATACCTGTGACACCCGGCTCTCCTGTCTCCCCTAAGCCTGTAACCTCCTGTGGTGCTG  
GCTGGAAGCTCCCTGATCACCTACCCATCTTCTTCTCAGGGTGCAGGAATTGAAT  
CTGGAGGAGAAACAGTGGCAGCTGGACCAGGAGCTACGAGGCTACATGAACCGGGAAG  
GTGAGTGACAAGCTAGGACACCCCTCTGCCTTCTAGCACTGTGTTGAGCTCCACGTTCCA  
GCCCGTAGAGTCCCAGTGGCTATGCCAACAGCTAACACATACCTGTGACACCCGGCTCT  
CCTGTCTCCCCTAAGCCTGTAACCTCCTGTGGTGCTGGCTGGAAGCTCCCTGATCACCTA  
CCCATCTTCTTCTTCTCAGAAAACCTAAAGACAGCTGCTGATCGGCAGGCTGAGGACCA  
GGTCCTGAGGAAGCTGGTGGATTTGGTCAACCAGAGAGATGCCCTCATCCGCTTCCAGG  
AGGAGCGCAGGCTCAGCGAGCTGGCCTTGGGGACAGGGGCCAGGGCTAGACGAGGG  
TGGGCCGTCTGCTTTGTTCCACAAAGAAAGCACCTCACCCAGCACAGTGCCACCCC  
TGTTTATCTGGGCTGCCTGGCAGAGAGCCTTGCTGTTTACAATTAATGTTTCTGCCAC  
AA

Figure 44: Séquence nucléotidique du clone L59.

Bases en noir=introns, bases en bleu=exons, bases en rouge= bases ajoutées au clone original à partir de ESTs trouvées dans les banques de séquences.

5' de notre THC fait effectivement partie du même messenger en séquençant un RT-PCR dont l'un des oligos est complémentaire à une région de L59 et l'autre à une région de l'un des ESTs qui sont à l'extrémité 5' de L59. Nous avons hybridé des membranes contenant de l'ARN de granulocytes humains avec le clone L59 pour évaluer la longueur du transcrit. Les hybridations n'ont pas donné de signaux clairs (données non montrées), nous pensons que la taille du transcrit de L59 est de l'ordre de 1.7 kb. Nous avons probablement presque la totalité du transcrit.

Les recherches BLAST ont donné, tel qu'on voit dans les figures 45 et 47, des homologies de l'ordre de 80-83% avec des ESTs de souris. Des ESTs que nous avons utilisés pour construire ce qu'on appelle un TMCL59 ("tentative mouse consensus").

De la même manière que nous l'avons fait avec le THCL59, nous avons démontré que la partie 5' du TMCL59 faisait effectivement partie du même messenger. Nous avons séquencé un RT-PCR dont l'un des oligos était complémentaire à une région du TMC correspondant à un EST et l'autre à une région d'un autre EST se trouvant en amont du premier EST. En comparant les séquences du THCL59 et du TMCL59 (qui a une longueur de 936 pb) on remarque la grande homologie qui existe entre ces deux séquences (Fig. 47).

En effectuant une recherche BLAST avec le THCL59, on observe (Figs. 46 et 47) des homologies de l'ordre de 80% avec des ESTs de rat (mauve) et de cochon (vert). Bref, L59 serait probablement un gène qui a un haut degré de conservation parmi les espèces.

### 3.6.2 Clone L88

En cherchant dans les banques de séquences, nous avons identifié un PAC, le clone 52f6 (accession number AC00500), séquencé au complet qui contient notre clone L88. Ce PAC se trouve sur le chr 22q11, confirmant davantage la localisation de L88. Le clone 52f6, d'une longueur de 192 kb, se trouve dans un contig formé par les PACs (BAC) b437g10, b433f11/f6, p52f6 et p\_M11 (Fig. 48). L88 est présent sur le PAC 52f6 entre les bases 169206-169651 précisément entre un site de restriction *EagI* et un

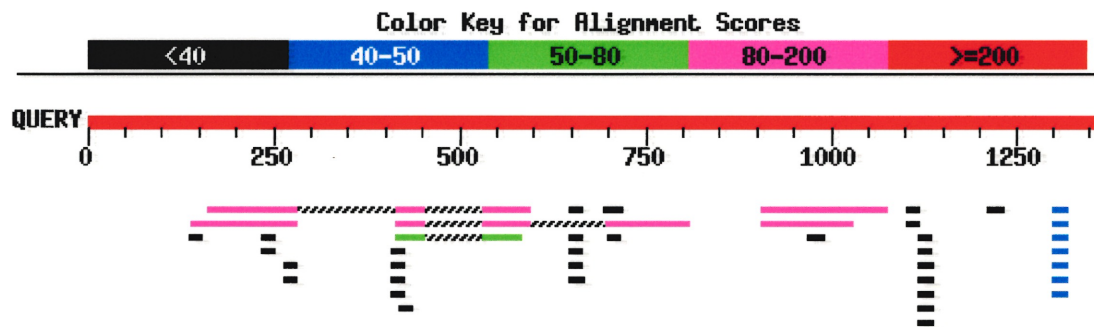


Fig. 45: Homologies de séquences murines obtenues en utilisant le THCL59 ("query"=1353 pb) dans une recherche BLAST (banques de séquences:"mouse ESTs").



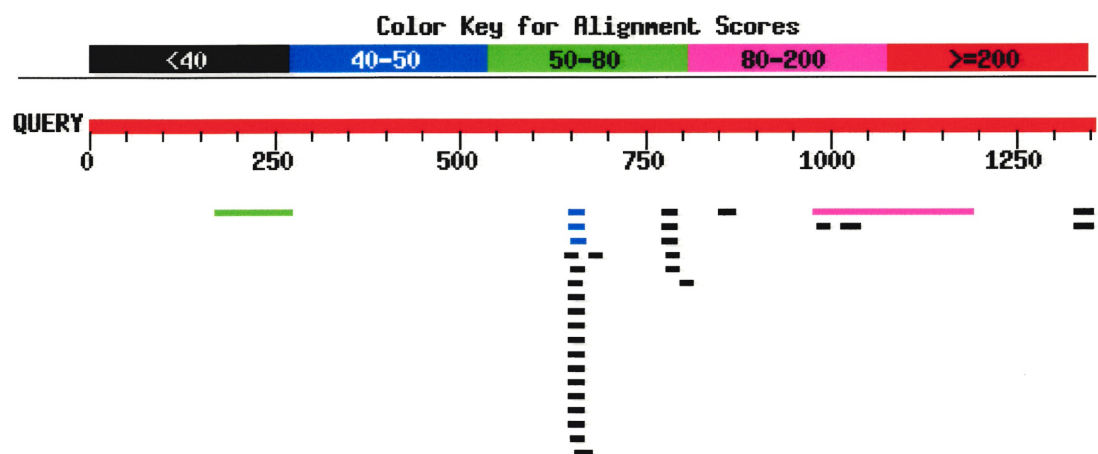


Fig 46: Homologies de séquences non-murines et non-humaines obtenues en utilisant le clone THCL59 ("query"=1353pb) dans une recherche BLAST (banque de séquences:"other ESTs"). En vert: EST de cochon. En mauve: EST de rat.



**B**

```

Humain:164 ggccatgtcttccaccggagctgcttccgctgccatacctgtgaagccacactgtggcca 223
          ||||| | ||||| ||||| ||||| ||||| | ||||| | | ||||| ||||| |||||
Souris:288 ggccatttctttcaccggagctgcttctgctgccatacctgtgaggccacattgt-gcga 346
          ||||| | ||||| ||||| ||||| ||||| ||||| | ||||| | ||||| ||||| |||||
Cochon: 3 ggccgtttcttccaccggagctgcttccgctgctgggtctgtaaggtcacactgtggcca 62

Humain:224 ggtggctacgagcagcaccaggagatg 251
          || | ||||| | ||||| ||||| ||||| |||||
Souris:347 ggtggctatgggcaacatccaggagatg 374
          || | ||||| | ||||| ||||| ||||| |||||
Cochon: 63 ggggcctacgggcagcaccagaagatg 90

Humain:251 ggacatttctactgcctccagc-acctgccccagacagaccacaaag 296
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Cochon:122 ggacatttctactgcctccagcaacctaccaagccaggccacaaag 168
    
```

**C**

```

Humain:959 aagaaaaaccacctggtggctgaggaggccgagctcatgatcacggtgcaggaattgaat 1018
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Rat: 381 aagaagaacagcctagtgactgaggaagccgaactcatgatcacggtgcaggagctggac 322

Humain:1019 ctggaggagaaacagtggcagctggaccaggagctacgaggctacatgaaccgggaagaa 1078
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Rat: 321 ctggaagagaagcagcggcaactggaccacgagtttcggggc---atcaaccgggaagaa 265

Humain:1079 aacctaaagacagctgctgatcggcaggctgaggaccaggctcctgaggaagctggtggat 1138
          | | ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Rat: 264 actctgaagaccaggccgaccgactgtctgaggaccgggtcctaaggaagctggtggat 205

Humain:1139 ttggtcaaccagagagatgccctcatccgcttccaggaggagcgcaggctcagcgagctg 1198
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Rat: 204 gtggtgaaccagcgggatgctctgatccaattccaggaggagcgaaggctcagagagatg 145

Humain:1199 gccttggggacaggggcccagggctagacgagg 1231
          || | ||||| | ||||| ||||| ||||| |||||
Rat: 144 cccgtatagatgggggcccagggctagaagagg 112
    
```

Fig. 47: Des homologies de séquence entre le clone L59 et des ESTs de différentes espèces.

A. L59 humain (ou THCL59, 1353 pb) et l'homologue murin de L59 (TMCL59, 936 pb). B. L59 humain, TMCL59 et un EST de cochon (234 pb). C. L59 humain et un EST de rat (381 pb).

**BAC, PAC and Cosmid Clones Spanning the Region DGCR through IGLC of Human Chromosome 22 containing BCRL2**

1-1-99

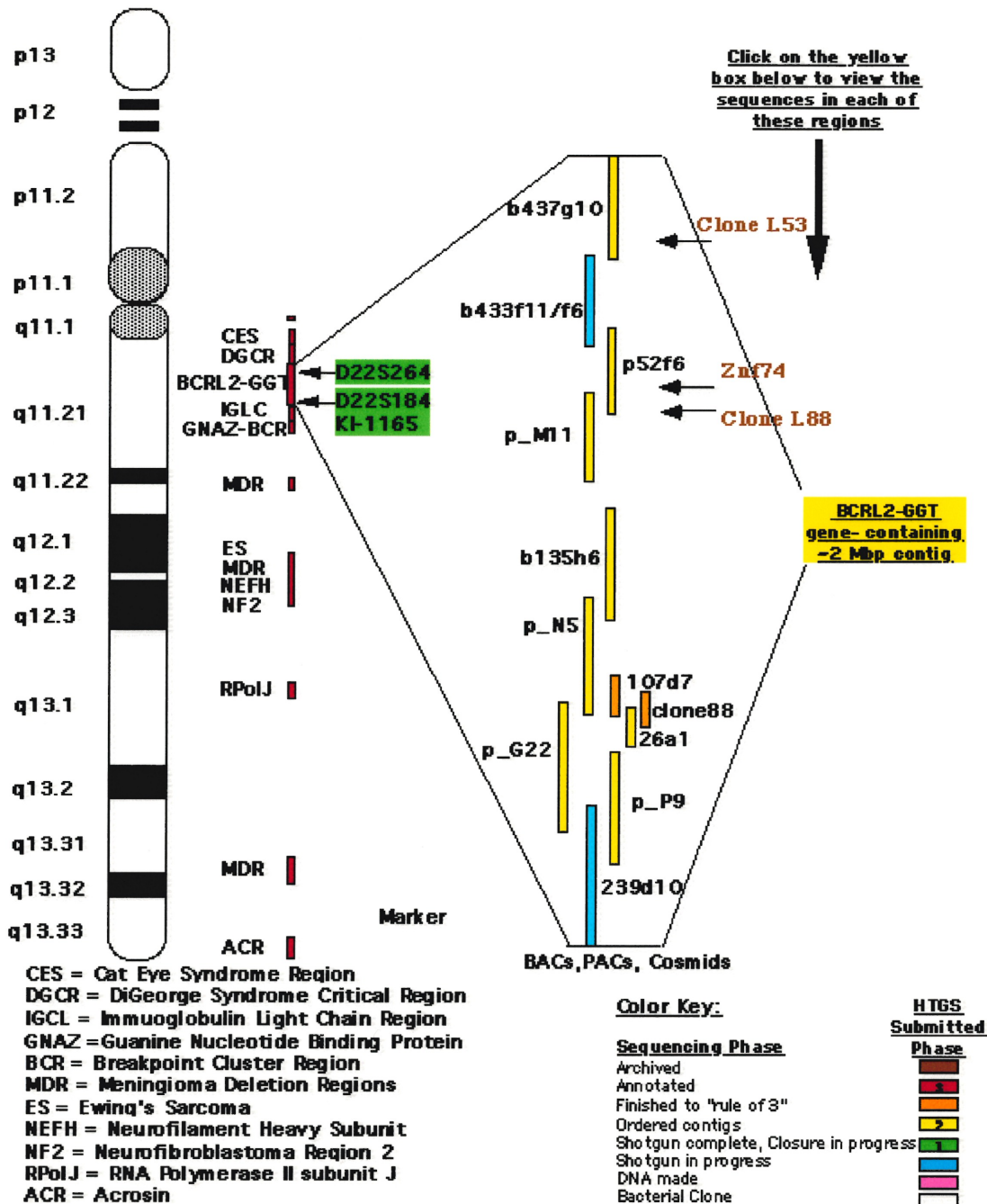


Fig. 48: Positions des clones L53 et L88 dans un contig de clones génomiques situé sur le chr 22q11.

site *MseI*. Il est à noter que le gène ZNF74, impliqué dans le syndrome de DiGeorge, se trouve sur le même PAC entre les bases 130,000 et 147,000 environ.

### 3.6.3 D'autres clones

En continuant nos recherches BLAST, nous avons obtenu des homologies avec d'autres clones.

Le clone L61 est identique au EST qui se trouve sur le PAC (*accession number* AC004596), d'une longueur de 92 kb, qui localise sur le Chr 17q21 à proximité de la région où se trouve le gène BRCA1.

Nous avons aussi eu des résultats BLAST pour les clones L53 et TR19, pour lesquels nous n'avons pas encore de données de méthylation différentielle (Tableau V). L53 se trouve dans le BAC b437g10 ("accession number" AC004032) de 200 kb, qui appartient au même contig que le PAC p52f6 contenant le clone L88 (Fig. 48). Finalement, le clone TR19 se trouve dans le PAC (*accession number* AC004891) de 112 kb qui est localisé au Chr 7q34-36.

## 3.7 ÉTUDES DE MÉTHYLATION ET D'EXPRESSION DIFFÉRENTIELLES DU CLONE L59.

En absence de polymorphismes du clone L59 humain (données non montrés), nous avons cherché et trouvé des polymorphismes au niveau du TMCL59, chez la souris.

### 3.7.1 SNUPEs

Nous avons utilisé de l'ADN et des cDNAs de croisements de souris B6 x mol ou mol x B6 pour déterminer si l'homologue murin de L59 est méthylé ou exprimé différemment et d'une manière allélique. Pour ce faire nous avons fait des essais de SNUPE ("single nucleotide primer extension").

Dans le cas de MS-SNUPE, nous avons testé le niveau de méthylation de certains sites de restriction HpaII se trouvant sur la séquence amplifiée (L59cand#2). Les

amplifications de PCR fonctionnent seulement si tous les sites HpaII sont méthylés. La Fig. 49 nous montre que les sites HpaII sont méthylés sur les deux allèles. Cependant, à juger par l'intensité des bandes, nous pourrions dire que l'allèle provenant de B6 est plus méthylé que celui venant de molossinus, peut-être dans un rapport de 3:1.

Pour les études d'expression allèle-spécifique de TMCL59 effectuées par RT-PCR SNUPE, nous avons utilisé le même polymorphisme que celui pour les MS-SNUPE en utilisant des échantillons provenant de croisements B6xmol et molxB6. Les résultats de la fig. 50 nous montrent que bien qu'il existe une variation dans l'expression de certains des échantillons, il est impossible de dire que cette variation est allélique.

### 3.7.2 RT-PCRs

Nous avons utilisé un polymorphisme de restriction pour l'enzyme *XmnI* entre les souris B6 et mol pour faire des études de RT-PCR allèle-spécifique d'un EST. Cet EST (que nous appelons L59cand#1) est l'homologue murin du THCL59 au niveau de son extrémité 3'.

Il est intéressant de remarquer que ce EST ne chevauche pas avec TMCL59. Des RT-PCRs essayés avec un oligo sur L59cand#1 et un autre sur TMCL59 pour déterminer s'ils font partie du même message n'ont jamais marché.

La figure 51 montre tout comme les résultats pour L59cand#2 (figure 51), dans les deux types de croisements, B6 mâle ou femelle ou l'inverse, l'absence d'une expression différentielle de type allélique.

B6	mol	B6xmol	B6xmol
G T	G T	G T	G T



Fig. 49: Études de méthylation différentielle du clone L59cand#2 par MS-SNuPE.  
 G=guanine, T=thymine, B6=C57 BL/6J, mol= *M. m. molossinus*.

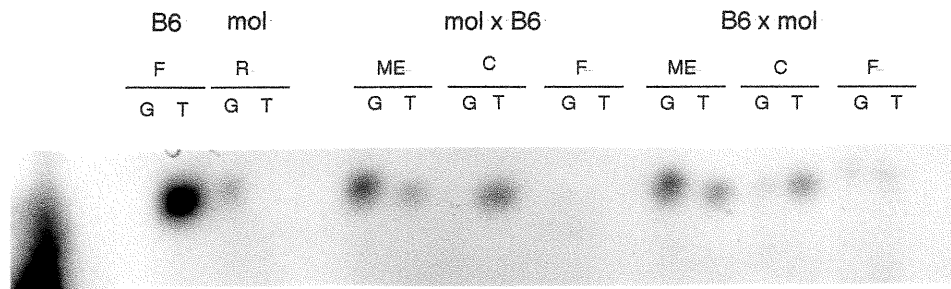


Fig. 50: Étude d'expression différentielle du clone L59cand#2 par RT-PCR-SNuPE.

B6= souris C57 B1/6J, mol=souris *M. m. molossinus*, F=foie, R=rein,

ME=membranes extraembryonnaires, C=cerveau, G=guanine, T=thymine.



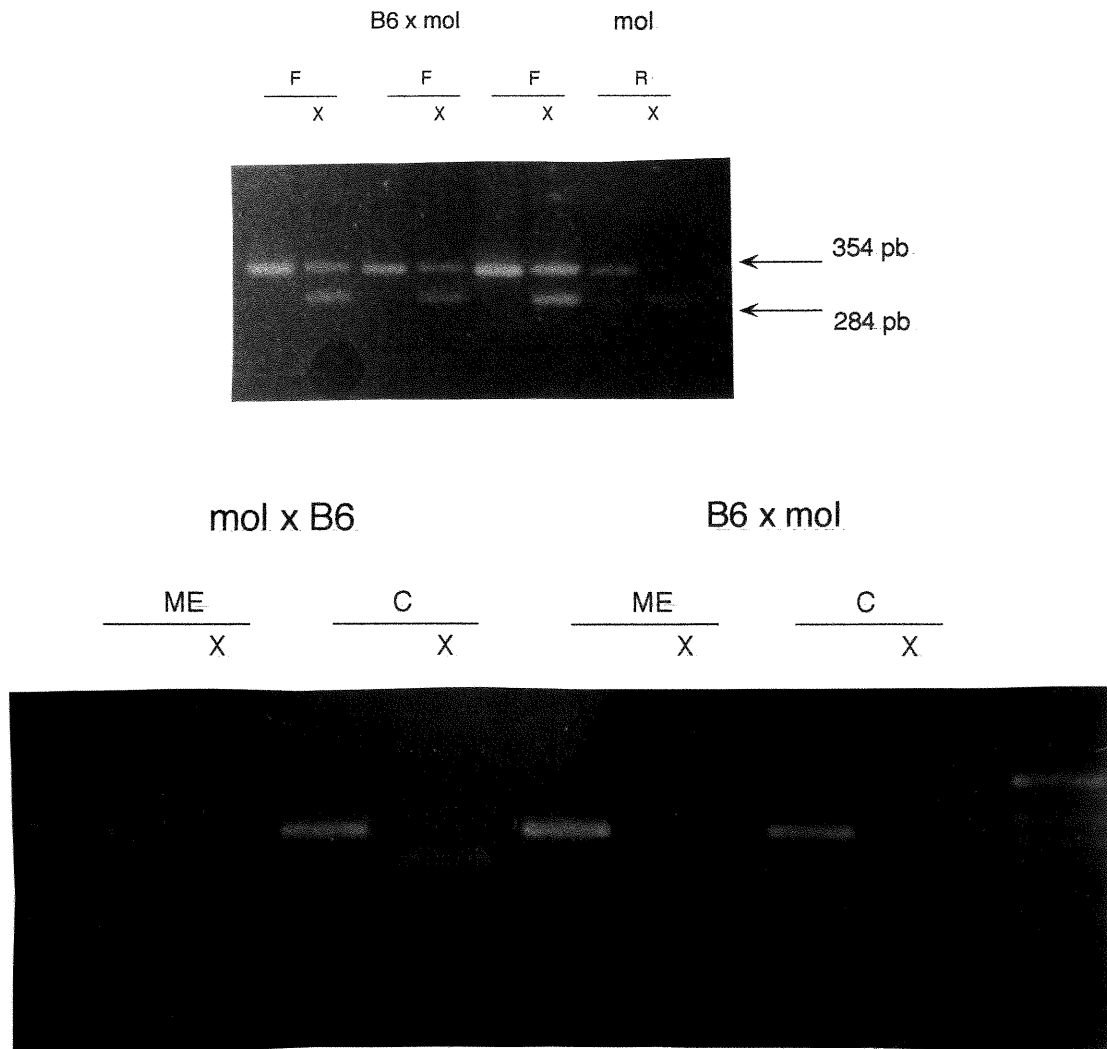


Fig. 51: Essais d'expression différentielle par RT-PCR pour le clone L59cand#1. B6=souris C57 Bl/6J, mol=souris *M.m. molossinus*, X=digestion avec l'enzyme *XmnI*, F=ADN de foie, R=ADNc de rein, ME=ADNc de membranes extraembryonnaires, C=cerveau.

**CHAPITRE 4**

**DISCUSSION**

#### 4.1 FP-CGI.

Nous avons mis au point une méthode, FP-CGI, dans le but de chercher des séquences méthylées différentiellement dans des génomes de mammifères. Nous avons effectué cinq essais de FP-CGI sur de l'ADN provenant de granulocytes humains qui nous ont permis d'isoler quarante-trois séquences dont vingt-huit semblent uniques. Quatorze de ces vingt-huit séquences non répétées ont été analysées par hybridation Southern. Sept de ces quatorze séquences ont montré des caractéristiques de méthylation complète sur les deux allèles. Sur les sept autres séquences, L51 était non-méthylée, L61 montrait un profil de méthylation partielle et les cinq restantes: L3, L59, L62, L88 et L211 ont donné des profils de méthylation différentielle: .

En développant la méthode nous estimions qu'il fallait s'attendre à un nombre assez élevé de clones provenant de séquences non-méthylées ou de séquences non-voulues. Ces séquences étant abondantes par rapport aux séquences méthylées différentiellement dans le génome humain. Mais la méthode semble être efficace pour les séquences recherchées. En d'autres mots, treize des quatorze séquences analysées par Southern sont soit méthylées, soit méthylées différentiellement. Donc, dans la dernière étape du FP-CGI il semble que l'on se soit complètement débarrassés des séquences non-méthylées. Cela est confirmé, malgré leur caractère préliminaire, par les tests de FP-CGI faits sur les promoteurs des gènes MAOA et PGK-1.

Concernant le grand nombre de séquences méthylées qui ont été isolées par FP-CGI, il est possible que ces séquences proviennent de billes présentes dans le surnageant de la dernière étape. Cependant, nous ne pouvons pas exclure la possibilité que quelques-unes des séquences identifiées comme méthylées puissent provenir d'une RMD. Un faux profil de méthylation complète par hybridation Southern pouvant être expliqué par l'existence d'un site de restriction du "six-cutter" (e.g. *EcoRI*) entre la sonde (STS) et le site de méthylation différentielle (e.g. *EagI*).

Le projet du séquençage du génome humain nous a fourni une autre preuve de l'efficacité du FP-CGI. En effet, quelques BACs ou PACs séquencés au complet contenant nos clones ont été identifiés dans les banques de séquences. Par exemple, dans le PAC AC005500 on a pu constater que L88 se trouve exactement entre un site *EagI* et un site *MseI*. Prenons aussi les cas des clones L53 et TR19 dont les extrémités correspondantes à l'adaptateur se trouvent sur un site *MseI* dans le PAC qui contient le clone. Donc, les oligos utilisés ont amplifié les séquences voulues. Ça pourrait aussi expliquer le grand nombre de séquences ramassés par FP-CGI qui ne rentrent pas dans les caractéristiques d'un îlot CpG, c'est à dire, probablement ces fragments amplifiés proviennent de la région entre les enzymes *EagI* et *MseI* qui n'est pas riche en C+G.

Un détail qui nous a étonné c'est le grand nombre de séquences répétées (quatorze sur quarante-trois) isolées par FP-CGI. Il pourrait s'agir d'une conséquence du très grand nombre de séquences répétées qui se trouvent sur les génomes des mammifères. Une autre explication serait que nous avons isolé ce type de séquences parce qu'elles constituent véritablement une cible de FP-CGI. Étant donné que des membres de la même famille de répétées ont des profils de méthylation différentielle (Rubin et coll. 1994; Woodcock et coll. 1997), ils ont probablement formé des hétéroduplexes dans l'étape 2 (Fig.14) remplissant les exigences pour la digestion de l'étape 8. Il est aussi connu que les profils de méthylation sont variables parmi les différents membres de la famille d'ADN ribosomal (Bird et Taggart 1980).

Quelques modifications à notre protocole pourraient être envisagées. On pourrait utiliser d'autres enzymes sensibles à la méthylation aux étapes 5 et 8, vu que le protocole tel que décrit permet seulement l'isolement de RMDs associées à des ICGs portant un site *EagI*. Lorsque les enzymes *HpaII* ou *HhaI* sont utilisées, par exemple, il ne sera pas nécessaire d'ajouter un amorce dans l'étape 8 car ces enzymes sont capable de digérer de l'ADN simple brin (Nishigaki et coll. 1985). Une autre modification que nous pourrions incorporer c'est une amplification linéaire par PCR en n'utilisant que l'amorce ML-2 pendant 10 cycles avant d'effectuer le PCR arbitraire de l'étape 9. Ceci génère quelques copies de chacun des fragments relâchés par la

digestion de l'étape 8, permettant l'utilisation de toute une gamme d'amorces arbitraires pour amplifier les produits de chaque expérience dans des tubes séparés. Finalement, nous pensons que la séparation des fragments après l'étape 9 serait plus efficace en marquant l'ADN pendant les réactions de PCR et en séparant les fragments sur un gel de séquençage. En choisissant les bandes les moins radioactives pour clonage, on pourrait éviter de cloner un grand nombre de séquences répétées. Une autre manière de diminuer le nombre de séquences répétées dans la population de séquences clonées par FP-CGI serait de hybrider des "colony lifts" avec de l'ADN répété.

## 4.2 LOCALISATIONS CHROMOSOMIQUES

Quatre des cinq clones méthylés différentiellement, L3, L59, L88 et L211 ont été localisés (Fig. 52) en utilisant des PCRs sur des lignées cellulaires somatiques hybrides et la technique de FISH (Fig. 42). Un clone, L61 (Fig. 52), pour lequel nous avons plutôt des données de méthylation partielle a été localisé en faisant des recherches dans les banques de séquences. Nous avons localisé de la même façon deux autres séquences qui n'avaient pas encore été testées par Southern, L53 et TR19.

Le clone L88 se trouve sur le chr 22q11 pour lequel des évidences d'un certain biais parental ont été déjà rapportés. Premièrement, il y a le syndrome de DiGeorge. Le syndrome de DiGeorge, connu aussi comme CATCH-22 ou comme "velocardiofacial syndrome" (VCFS) (Hong 1998) fait partie d'un ensemble de troubles comprenant des anomalies cardiaques, de l'hypoplasie thymique et de l'hypocalcémie qui sont associés à des délétions au niveau de 22q11. Six gènes ont été associés jusqu'ici à ces délétions qui couvrent une région d'environ 1.5 megabases dont *Tuple1* (Lamour et coll. 1995), *ES2* (Lindsay et coll. 1998), *Tbx1* (Chieffo et coll. 1997), *UFDIL* (Yamagishi et coll. 1999) et *ZNF74* (Grondin et coll. 1996). Ce dernier gène se trouve sur le PAC AA00500 à 32 kb du clone L88 (voir figure 42). Il a été décrit que les délétions de

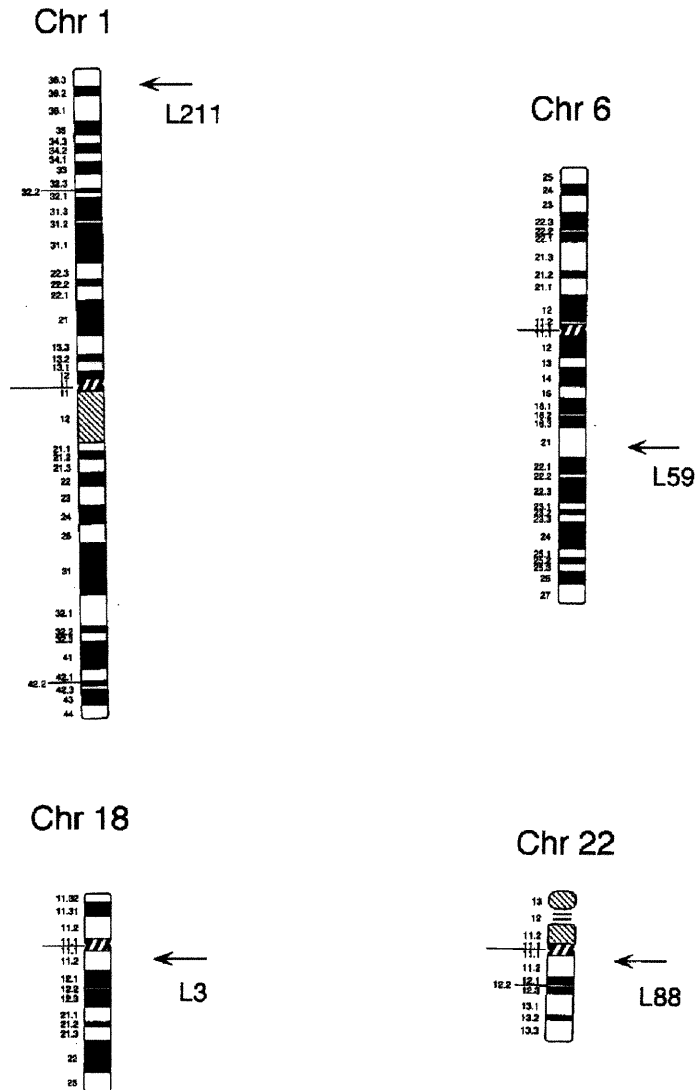


Fig. 52: Régions chromosomiques humaines contenant des clones isolés par FP -CGI.

DiGeorge affectent préférentiellement le chromosome maternel (Demczuk et coll. 1995; Ryan et coll. 1997). De plus, des séquences de cette région se répliquent d'une manière asynchrone (Kitsberg et coll. 1993); suggérant une empreinte parentale. Ensuite, il y a le gène BCR qui est télomérique à la région de délétions de DiGeorge et dont sa translocation avec le gène *Abl*, situé au chr 9 est associée à la leucémie myéloïde chronique (LMC). À ce sujet, un papier qui a suscité beaucoup de polémique a été publié en 1992 (Haas et coll. 1992). En résumé, les auteurs montrent que le chromosome 22 impliqué dans la translocation est d'origine maternelle tandis que le chromosome 9 vient du père suggérant qu'une empreinte parentale pourrait être impliquée dans ce type de réarrangements. Cependant, d'autres travaux publiés par la suite n'ont jamais confirmé ces résultats (Melo et coll. 1995; Melo et Goldman 1995). D'autres évidences du rôle de la méthylation dans la LMC vont dans un sens différent. Deux études (Zion et coll. 1994; Ohyashiki et coll. 1993) décrivent une méthylation progressive *de novo* au niveau du locus *bcr-abl* (c'est à dire après la translocation) et pendant le développement de la maladie.

L59 a été localisé par FISH sur le chr 6q21 qui ne possède pas d'empreinte parentale documentée. Par contre, le chromosome 6 porte deux domaines avec empreinte, 6q22-23 et 6q26-27 (Miller et coll. 1997). Il existe alors peut-être des possibilités que L59 se trouve près de 6q22. Cette région porte au moins un gène avec empreinte impliqué dans un type de diabète, le "transient neonatal diabetes mellitus" (TNDM), suite à des évidences d'isodisomie uniparentale paternelle du chr 6 chez quelques patients atteints de cette maladie (El, A. et coll. 1997; Temple et coll. 1996). Finalement, la région 6q21 est intéressante pour deux autres raisons. Le chromosome 6q21 est impliqué dans des délétions retrouvées dans un type de leucémie, la leucémie lymphoblastique aigüe (LLA) et pourrait contenir quelques gènes suppresseurs associés à cette maladie (Takeuchi et coll. 1998; Jackson et coll. 1998). Une patiente (Zneimer et coll. 1998) avec une duplication des bandes 6q21 et 6q23 (une trisomie partielle de 6q) montre certains troubles d'ordre neurologique.

Selon les premiers résultats des deux types d'essais SNUPE nous ne sommes pas arrivés à déterminer clairement si L59 est exprimé d'une façon monoallélique. Cependant, nous avons démontré la présence de méthylation différentielle chez la souris. De plus, cette méthylation différentielle serait, selon les résultats chez la souris, allèle-spécifique. Il reste donc à établir si L59 a une expression monoallélique allèle-spécifique chez l'humain dû à l'importance de ce clone. La comparaison de la séquence génomique de 1.1 kb du clone L59, avec la séquence dérivée des ESTs que la recherche BLAST a générés a montré que L59 a trois introns de 206, 95 et 383 pb, tous avec les séquences consensus d'épissage (Fig. 43). Ces petits introns sont l'une des caractéristiques retrouvées dans les gènes avec empreinte génomique (Hurst et coll. 1996). L59 ne semble pas avoir un cadre de lecture ouverte; il coderait donc pour un ARN non traduit, une caractéristique de certains gènes avec empreinte génomique (Moore et coll. 1997).

Il serait donc intéressant de continuer l'étude de L59. Non seulement à cause des raisons qu'on vient d'exprimer mais parce qu'il serait intéressant d'entreprendre sur L59 l'étude d'autres caractéristiques d'empreinte génomique telle que la réplication asynchrone et parce que l'empreinte génomique a été démontré dans certains cas d'avoir une histospécificité très restreinte (Rougeulle et coll. 1998), (Miller et coll. 1997).

Des gènes avec empreinte génomique parentale n'ont pas encore été rapportés sur le chr humain 18, qui héberge le clone L3. Cependant, *Impact*, un gène qui a une empreinte génomique chez la souris, se trouve dans la région pericentromérique du chr 18 murin près du gène *Lama* (Hagiwara et coll. 1997) qui est en synténie avec le chromosome humain 18q11, suggérant que L3 pourrait être dans une région sous empreinte parentale.

Le clone L211 est localisé sur le chromosome 1p36 où se trouve le gène p73 (un homologue du gène suppresseur p53), récemment rapporté d'avoir une empreinte génomique parentale (Kaghad et coll. 1997), ce qui a suscité une certaine controverse.



1p36 est une région qui est aussi soupçonnée d'avoir un gène suppresseur impliquée dans la leucémie myélocytique chronique (LMC) (Mori et coll. 1998). De plus, le syndrome connu comme "le syndrome de délétion du chromosome 1p36" a une prépondérance de délétions de l'allèle maternelle (Wu et coll. 1999). Il est donc probable que L211, tout comme les autres sondes, soit dans les environs d'une région avec un certain biais parental.

Nous n'avons pas encore de données claires de méthylation différentielle pour les autres clones. Cependant, L53 ne se trouve qu'à quelques 300 kb de L88. Il est très improbable de trouver deux clones sur vingt-huit à une distance de quelques centaines de kilobases sur le génome humain. Cela pourrait plutôt suggérer l'existence d'un grand domaine de méthylation différentielle. TR19 se trouve sur 7q34-q36, donc, à proximité du gène *PEG1/MEST* qui montre une empreinte génomique parentale paternelle et se trouve sur 7q31-34.

Nous n'avons pas encore de preuve directe que les RMDs détectées par nos sondes isolées par FP-CGI ont réellement une empreinte parentale. Néanmoins, les évidences fournies par les données de cartographie et par les analyses de séquences sont pointées dans cette direction. Nous pensons donc que la technique que nous avons décrite est une addition très utile aux méthodologies existentes pour l'analyse de ce phénomène.

#### **4.3 MÉTHODES POUR IDENTIFIER DES GÈNES AVEC EMPREINTE GÉNOMIQUE PARENTALE**

Jusqu'à présent, quatre méthodes systématiques de criblage ont été développées pour trouver des gènes avec empreinte génomique parentale, soustraction d'ADNc (Kaneko-Ishino et coll. 1995), "allelic message display" (Hagiwara et coll. 1997), "restriction landmark genomic scanning" (Hatada et coll. 1991) et un système d'hybrides cellulaires somatiques (Gabriel et coll. 1998).

La méthode de soustraction est basée sur la sélection d'ADNc après une hybridation entre ADNc d'embryons de souris normales d'une part, et d'embryons parthénogénotes ou androgénotes d'autre part. Le but étant de chercher des gènes qui sont exprimés

seulement à partir de l'allèle paternel, dans le cas des parthénogénètes ou seulement à partir de l'allèle maternel dans le cas des androgénètes. C'est une méthode assez puissante qui a permis de détecter quelques nouveaux gènes avec empreinte génomique (Lefebvre et coll. 1997; Relaix et coll. 1998; Kagitani et coll. 1997; Miyoshi et coll. 1998).

"Allelic message display" (AMD) est une méthode similaire au "differential display" (DD) qui fait appel au RT-PCR pour chercher des transcrits différentiellement exprimés en comparant des ADNc polymorphiques provenant des croisements interspécifiques. Néanmoins, ce qui distingue AMD du DD c'est ce qu'elle compare des ADNc de cellules ou de tissus identiques sous de différents fonds génétiques tandis que DD compare des ADNc provenant de tissus différents.

La technique de "restriction landmark genomic scanning" (RLGS) tire, elle aussi, profit des polymorphismes qu'on trouve dans les croisements interspécifiques. Mais contrairement aux techniques aux approches précédentes qui nécessitent des séquences exprimées, le RLGS utilise des enzymes "rare-cutter" (telles que *NotI*) sensibles à la méthylation pour chercher des polymorphismes au niveau de ces sites de restriction. Les séquences isolées se trouveront ainsi près d'un site "rare-cutter", donc très probablement tout près d'un îlot (Hatada et coll. 1993).

Le papier sur les hybrides cellulaires somatiques décrit que des hybrides cellulaires somatiques souris/humain portant parfois un seul chromosome humain soit paternel ou maternel conservent l'empreinte des gènes se trouvant sur le chromosome en question, les chromosomes testés étant les chr 11 et 15. Non seulement ces lignées conservent ces profils d'empreinte génomique mais elles ont conservé aussi les profils de méthylation différentielle connus pour ces gènes et ce pendant plus de dix ans et plus de cent passages en culture. Ce système pourrait donc être utile pour étudier l'empreinte parentale dans les années à venir.

En bref, de toutes les méthodes décrites ci-dessus seulement RLGS est capable de trouver des RMDs mais elle ne peut pas être utilisée dans le génome humain. De plus,

RLGS est techniquement complexe et les signaux identifiés comme méthylés différemment sont difficiles à détecter sans un appareil de balayage automatisé. Cependant, cette technique a permis d'identifier un nouveau gène, *Gnas*, avec empreinte génomique chez l'humain en utilisant de l'ADN d'un patient dont les leucocytes périphériques étaient entièrement parthénogénotes (Hayward et coll. 1998).

Nous décrivons dans cette thèse une méthode alternative qui pourrait être utilisée pour cribler des changements de méthylation à la grandeur du génome d'une manière simple et rapide afin d'isoler des fragments spécifiques d'ADN potentiellement associés à des régions avec empreinte parentale. Elle pourrait être appliquée à n'importe quelle source d'ADN génomique de mammifères et puisqu'il y a 25 % d'ICGs de plus dans le génome humain que dans celui de la souris (Antequera et Bird 1993), FP-CGI pourrait permettre la détection de ces RMDs qui ne seraient pas détectées par d'autres techniques qui utilisent le génome murin. Ni FP-CGI ni RLGS n'ont été conçues pour un isolement direct de gènes. Ces deux méthodes permettent d'identifier des séquences dans des ICGs ou près des ICGs qui sont associés aux extrémités 5' des gènes. Comme une bonne partie des séquences (~25%) que nous avons isolées par FP-CGI est constituée de ESTs, il est probable qu'une proportion des clones isolés de cette façon sera constituée de séquences exprimées.

Finalement, nous nous permettons de croire que le système de cellules somatiques hybrides pourrait être une addition intéressante à l'éventail d'applications de FP-CGI. Nous pourrions envisager, par exemple, de chercher par FP-CGI de séquences méthylées différemment au niveau d'un chromosome en particulier.

#### 4.4 IMPORTANCE DES "RMDs"

Lorsqu'on confirme qu'une séquence isolée par FP-CGI est méthylée différemment, il faut faire face à trois questions principales. D'abord, est-ce que c'est une méthylation différentielle allélique?; bien que les profils globaux de méthylation différentielle varient plutôt d'une façon histospécifique (Heinke et coll. 1987; Silva et White 1988), nous ne pouvons pas écarter la possibilité de l'existence de

variations de méthylation au niveau de certains sites de restriction. Ensuite, si la méthylation différentielle d'un clone donné est allélique, est-ce qu'elle est aussi parentale-spécifique?. Question qui n'est pas facile à répondre, parce que bien que pour une séquence donnée ce soit toujours un seul allèle qui est méthylé, nous ne pouvons pas ignorer la possibilité que cette méthylation allélique soit aléatoire. Cela nous mène à la troisième question, qui est incontournable, est-ce que le profil de méthylation différentielle se traduit en une expression monoallélique?. La réponse à la troisième question va nous donner des indices sur la deuxième question, car il y a des gènes qui ont une expression monoallélique qui n'est pas parentale-spécifique, l'expression allélique étant aléatoire (Chess 1998; Holländer et coll. 1998). En fait, les gènes sur le chr X chez les femmes sont un bon exemple de méthylation et expression différentielle monoallélique mais sans caractère parental-spécifique. Toutefois il n'y a pas beaucoup de cas de gènes autosomiques ayant une expression monoallélique aléatoire et encore moins avec une méthylation différentielle monoallélique aléatoire.

Il est intéressant de noter que les clones méthylés différentiellement se retrouvent dans des régions impliquées dans des translocations ou délétions. Peut-être la méthylation différentielle allélique d'un gène ou d'une région quelconque pourrait être associée à des réarrangements chromosomiques. Elle pourrait avoir une influence dans l'instabilité génomique et la probabilité d'avoir des délétions. La différence de méthylation dans les deux allèles pourrait être "détectée" par un système quelconque et rendre l'un des deux allèles plus vulnérable à certains types de réarrangements génomiques. Ça expliquerait, par exemple, le fait qu'on ait trouvé deux de nos clones (L53 et L88) exactement dans la même région, le chr 22q11 étant une région très instable (Puech et coll. 1997). Il existe d'autres évidences suggérant que la méthylation est associée à des phénomènes de mutabilité (Andrew et coll. 1998) et d'instabilité génétique dans les cancers colorectaux (Lengauer et coll. 1997), à des mécanismes de recombinaison homologue (Colot et coll. 1996), à des maladies associées à des trinuécléotides répétées (Knight et coll. 1993; Sandberg et Schalling 1997), au syndrome de Down (Kuromitsu et coll. 1997) et d'autres. Il serait donc intéressant d'étudier le profil de méthylation de nos séquences chez des patients de

DiGeorge ou de patients leucémiques et d'individus normaux afin de détecter des changements dans le profil de méthylation différentielle.

De l'ADN de granulocytes femelles a été utilisé comme source d'une grande majorité des clones obtenus par FP-CGI, mais nous n'en avons isolé aucun jusqu'à maintenant provenant du chromosome X, ce qui est étonnant vu le grand nombre de séquences méthylées différentiellement qui se trouvent sur ce chromosome et vu le petit nombre de domaines qui ont une empreinte parentale. Cela pourrait être attribué au hasard, quelques clones sur le chromosome X pouvant en effet être parmi ceux pour lesquels nous n'avons pas encore de données de localisation chromosomique ou des données de méthylation différentielle par hybridations Southern. Mais cela pourrait notamment suggérer que les chromosomes de mammifères portent beaucoup plus de RMDs de ce qui est couramment accepté. Des régions qui pourraient être potentiellement associées à des réarrangements chromosomiques impliqués dans certaines maladies telles que des leucémies ou des maladies neurologiques.

#### **4.5 FP-CGI ET LES ÉTUDES DE MÉTHYLATION DANS LE CANCER.**

Finalement, il faut souligner la flexibilité de la technique que nous avons développée. Nous pouvons appliquer le FP-CGI à des phénomènes où l'on soupçonne la présence d'une méthylation différentielle particulièrement celui de la méthylation des promoteurs de gènes suppresseurs qu'on observe dans certains types de cancer. Dans ce cas, la méthylation différentielle ne serait pas allélique mais plutôt dûe à la progression tumorale.

Il existe déjà quelques techniques bien établies destinées à isoler ce type de séquences en utilisant de modèles murins de progression tumorale (Akama et coll. 1997), (Gonzalzo et coll. 1997), (Ushijima et coll. 1997) mais étant donné que les techniques ne sont pas identiques, nous estimons que FP-CGI pourrait fournir une population de clones différente à celles obtenues par les autres techniques.

**CHAPITRE 5**

**CONCLUSIONS**

À l'aide de la méthode FP-CGI décrite dans les sections précédentes, nous avons identifié cinq séquences: L3, L59, L62, L88 et L211. La présence de certaines caractéristiques nous permet de croire que ces séquences sont localisées dans des régions ayant une empreinte parentale.

## 5.1 PERSPECTIVES

Quelques stratégies pourraient être envisagées pour mieux caractériser ces régions. Par exemple: nous pourrions déterminer avec certitude si L59 est différenciellement méthylé d'une façon allélique ou s'il a une empreinte parentale. Il serait intéressant d'étudier la méthylation différentielle et l'expression monoallélique de gènes connus (de gènes de maladies, par exemple) se trouvant sur les régions que nous avons identifiées. Nous suggérons aussi de chercher des gènes (e.g. par "exon trapping" en utilisant des BACs ou des YACs) dans les régions que nous avons identifiées et déterminer leur expression monoallélique et/ou parental-spécifique.

## 5.2 RÉGIONS MÉTHYLÉES DIFFÉRENTIELLEMENT

FP-CGI pourrait servir à décélérer des régions importantes dans des phénomènes autres que l'empreinte parentale et dont le caractère d'instabilité chromosomique serait souligné par une méthylation différentielle allélique mais non pas nécessairement parentale spécifique. Cette cartographie des RMDs des clones obtenus ou qui seront obtenus par FP-CGI pourrait grandement faciliter la compréhension de ces phénomènes lors de la fin du séquençage du génome humain, vers l'an 2003.

**CHAPITRE 6**

**RÉFÉRENCES**



- Akama, T. O. *et coll.* (1997). "Restriction landmark genomic scanning (RLGS-M)-based genome-wide scanning of mouse liver tumors for alterations in DNA methylation status." *Cancer Research* **57**: 3294-3299.
- Altschul, S. F. *et coll.* (1990). "Basic local alignment search tool." *Journal of Molecular Biology* **215**: 403-410.
- Altschul, S. F. *et coll.* (1997). "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs." *Nucleic Acids Research* **25**(17): 3389-3402.
- Andrew, S. E. *et coll.* (1998). "Tissues of MSH2-deficient mice demonstrate hypermutability on exposure to a DNA methylating agent." *Proc. Natl. Acad. Sci. USA* **95**: 1126-1130q.
- Antequera, F. et Bird, A. (1993)a. "Number of CpG islands and genes in human and mouse." *Proc. Natl. Acad. Sci. USA* **90**: 11995-11999.
- Antequera, F. et Bird, A. (1993)b. CpG Islands. DNA Methylation: Molecular biology and biological significance. J. P. Jost et H. P. Saluz, Birkhäuser Verlag.
- Araujo, F. D. *et coll.* (1998). "Concurrent replication and methylation at mammalian origins of replication." *Molecular and Cellular Biology* **18**(6): 3475-3482.
- Bakin, A. V. et Curran, T. (1999). "Role of DNA 5-methylcytosine transferase in cell transformation by *fos*." *Science* **283**: 387-390.
- Balmain, A. (1995). "Exploring the bowels of DNA methylation." *Current Biology* **5**(9): 1013-1016.
- Barlow, D.P. *et coll.* (1991). "The mouse insulin-like growth factor type-2 receptor is imprinted and closely linked to the *Tme* locus." *Nature* **49**: 84-87

- Barlow, D. P. (1997). "Competition-a common motif for the imprinting mechanism?" *The EMBO Journal* **16**(23): 6899-6905.
- Bartolomei, M. S. et Tilghman, S. M. (1997). "Genomic imprinting in mammals." *Annu. Rev. Genet.* **31**: 493-525.
- Bartolomei, M. S. *et coll.* (1993). "Epigenetic mechanisms underlying the imprinting of the mouse *H19* gene." *Genes & Development* **7**: 1663-1673.
- Battacharya, S. K. *et coll.* (1999). "A mammalian protein with specific demethylase activity for mCpG DNA." *Nature* **397**: 579.
- Bestor, T. H. et Tycko, B. (1996). "Creation of genomic methylation patterns." *Nature Genetics* **12**: 363-367.
- Bestor, T. H. et Verdine, G. L. (1994). "DNA methyltransferases." *Current Biology* **6**: 380-389.
- Bird, A. *et coll.* (1995). "Studies of DNA methylation in animals." *Journal of Cell Science* **19**: 37-39.
- Bird, A. P. et Taggart, M. H. (1980). "Variable patterns of total DNA and rDNA methylation in animals." *Nucleic Acids Research* **8**(7): 1485-1497.
- Birger, Y. *et coll.* (1999). "The imprinting box of the mouse *Igf2r* gene." *Nature* **397**: 84-88.
- Blagitko, N. *et coll.* (1999). " $\gamma$ 2-COP, a novel imprinted gene on chromosome 7q32, defines a new imprinting cluster in the human genome." *Human Molecular Genetics* **8**: 2387-2396.
- Boccaccio, I. *et coll.* (1999). "The human *MAGEL2* gene and its mouse homologue are paternally expressed and mapped to the Prader-Willi region." *Human Molecular Genetics* **8**: 2497-2505.

Bork, P. et Koonin, E. V. (1998). "Predicting functions from protein sequences-where are the bottlenecks?" *Nature Genetics* **18**: 313-318.

Brown, A. *et coll.* (1994). "Human homolog of a mouse sequence from the dystonia musculorum locus is on Chromosome 6p12." *Mammalian Genome* **5**: 434-437.

Brown, C. J. *et coll.* (1992). "The human *XIST* gene: an analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus." *Cell* **71**: 527-542.

Cameron, E. E. *et coll.* (1999). "Synergy of demethylation and histone deacetylase inhibition in the re-expression of genes silenced in cancer." *Nature Genetics* **21**(1): 103.

Casna, N. J. *et coll.* (1986). "Genomic analysis II: isolation of high molecular weight heteroduplex DNA following differential methylase protection and formamide-PERT hybridization." *Nucleic Acids Research* **14**(18): 7285-7303.

Caspary, T. *et coll.* (1998). "Multiple mechanisms regulate imprinting of the mouse distal chromosome 7 gene cluster." *Molecular and Cellular Biology* **18**(6): 3466-3474.

Cattanach, B. M. et Beechey, C. V. (1990). "Autosomal and X-chromosome imprinting." *Development Supplement*: 63-72.

Cedar, H. et Verdine, G. L. (1999). "The amazing demethylase." *Nature* **397**: 568-569.

Chen, R. Z. *et coll.* (1998). "DNA hypomethylation leads to elevated mutation rates." *Nature* **395**: 89-93.

Chess, A. (1998). "Expansion of the allelic exclusion principle?" *Science* **279**: 2067-2068.

Chieffo, C. *et coll.* (1997). "Isolation and characterization of a gene from the DiGeorge chromosomal region homologous to the mouse *Tbx1* gene." *Genomics* **43**: 267-277.

Chuang, L. S.-H. *et coll.* (1997). "Human DNA-(Cytosine-5) methyltransferase-PCNA complex as a target for p21<sup>WAF1</sup>." *Science* **277**: 1996-2000.

Church, D. M. *et coll.* (1994). "Isolation of genes from complex sources of mammalian genomic DNA using exon amplification." *Nature Genetics* **6**: 98-105.

Clark, S. J. *et coll.* (1995). "CpNpG methylation in mammalian cells." *Nature Genetics* **10**: 20-27.

Clark, S. J. *et coll.* (1994). "High sensitivity mapping of methylated cytosines." *Nucleic Acids Research* **22**(15): 2990-2997.

Claverie, J.-M. (1997). "Computational methods for the identification of genes in vertebrate genomic sequences." *Human Molecular Genetics* **6**(10): 1735-1744.

Colot, V. *et coll.* (1996). "Interchromosomal transfer of epigenetic states in *Ascomobolus*: transfer of DNA methylation is mechanistically related to homologous recombination." *Cell* **86**: 855-864.

Côté, S. et Momparler, R. L. (1997). "Activation of the retinoic acid receptor  $\beta$  gene by 5-aza-2'- deoxycytidine in human DLD-1 colon carcinoma cells." *Anticancer Drugs* **8**: 56-61.

Craig, J. M. et Bickmore, W. A. (1994). "The distribution of CpG islands in mammalian chromosomes." *Nature Genetics* **7**(3): 376-82.

Cross, S. H. et Bird, A. P. (1995). "CpG islands and genes." *Current Opinion in Genetics and Development* **5**: 309-314.

Cross, S. H. *et coll.* (1997). "A component of the transcriptional repressor MeCP1 shares a motif with DNA methyltransferase and HRX proteins." *Nature Genetics* **16**(3): 256-259.

Dao, D. *et coll.* (1998). "IMPT1, an imprinted gene similar to polyspecific transporter and multi-drug resistance genes." *Human Molecular Genetics* **7**: 597-608.

DeChiara, T. M. *et coll.* (1991). "Parental imprinting of the mouse insulin-like growth factor II gene." *Cell* **64**: 849-859.

Demczuk, S. *et coll.* (1995). "Excess of deletions of maternal origin in the DiGeorge/velo-cardio-facial syndromes. A study of 22 new patients and review of the literature." *Hum Genet* **96**: 9-13.

Denissenko, M. F. *et coll.* (1997). "Cytosine methylation determines hot spots of DNA damage in the human P53 gene." *Proc. Natl. Acad. Sci. USA* **94**: 3893-3898.

Dittrich, B. *et coll.* (1996). "Imprint switching on human chromosome 15 may involve alternative transcripts of the SNRPN gene." *Nature Genetics* **14**: 163-170.

Dobrovic, A. et Simpendorfer, D. (1997). "Methylation of the *BRCA1* gene in sporadic breast cancer." *Cancer Research* **57**: 3347-3350.

Dracopoli, N. C. *et coll.* (1998). Current protocols in human genetics, John Wiley & Sons, Inc.

Dubois, B. L. et Naylor, S. L. (1993). "Characterization of NIGMS human/rodent somatic cell hybrid mapping panel 2 by PCR." *Genomics* **16**: 315-319.

Eggenchwiler, J. *et coll.* (1997). "Mouse mutant embryos overexpressing IGF-II exhibit phenotypic features of the Beckwith-Wiedemann and Simpson-Golabi-Behmel syndromes." *Genes & Development* **11**(23).

El, A. *et coll.* (1997). "Transient neonatal diabetes mellitus in a child with invdup(6)(q22q23) of paternal origin." *Eur. J. Hum. Genet.* **5**(6): 417-419.

Feinberg, A. P. (1993). "Genomic imprinting and gene activation in cancer." *Nature Genetics* **4**: 110-113.

Gabriel, J. M. *et coll.* (1998). "A model system to study genomic imprinting of human genes." *Proc. Natl. Acad. Sci. USA* **95**.

Garrick, D. *et coll.* (1998). "Repeat-induced gene silencing in mammals." *Nature Genetics* **18**(1): 56-59.

Giddings, S. J. *et coll.* (1994). "Allele-specific inactivation of insulin 1 and 2, in the mouse yolk sac, indicates imprinting." *Nature Genetics* **6**: 310-313.

Gonzalzo, M. L. et Jones, P. A. (1997). "Rapid quantitation of methylation differences at specific sites using methylation-sensitive single nucleotide primer extension (Ms-SNuPE)." *Nucleic Acids Research* **25**(12): 2529-2531.

Gonzalzo, M. L. *et coll.* (1997). "Identification and characterization of differentially methylated regions of genomic DNA by methylation-sensitive arbitrarily primed PCR." *Cancer Research* **57**: 594-599.

Gould, T. D. et Pfeifer, K. (1998). "Imprinting of mouse *Kvlqt1* is developmentally regulated." *Human Molecular Genetics* **7**: 483-487.

Graff, J. R. *et coll.* (1995). "E-Cadherin expression is silenced by DNA hypermethylation in human breast and prostate carcinomas." *Cancer Research* **55**: 5195-5199.

Greally, J. M. *et coll.* (1998). "The mouse *H19* locus mediates a transition between imprinted and non-imprinted DNA replication patterns." *Human Molecular Genetics* **7**(1): 91-95.

Greenwood, A. D. et Burke, D. T. (1996). "Single nucleotide primer extension: quantitative range, variability, and multiplex analysis." *Genome Research* **6**: 336-348.

Grondin, B. *et coll.* (1996). "The KRAB zinc finger gene *ZNF74* encodes an RNA-binding protein tightly associated with the nuclear matrix." *The Journal of Biological Chemistry* **271**(26): 15458-15467.

Guillemot, F. *et coll.* (1995) "Genomic imprinting of *Mash-2*, a mouse gene required for trophoblast development." *Nature Genetics* **9**: 235-241.

Haas, O. A. *et coll.* (1992). "Parental origin of chromosomes involved in the translocation t(9;22)." *Nature* **359**: 414-416.

Hagiwara, Y. *et coll.* (1997). "Screening for imprinted genes by allelic message display: Identification of a paternally expressed gene *Impact* on mouse chromosome 18." *Proc. Natl. Acad. Sci. USA* **94**: 9249-9254.

Hashimoto, K. *et coll.* (1995). "Loss of imprinting in choriocarcinoma." *Nature Genetics* **9**: 109-110.

Hatada, I. *et coll.* (1991). "A genomic scanning method for higher organisms using restriction sites as landmarks." *Proc. Natl. Acad. Sci. USA* **88**: 9523-9527.

Hatada, I. *et coll.* (1993). "A new imprinted gene by a methylation-sensitive genome scanning method." *Nucleic Acids Research* **21**(24): 5577-5582.

Hatada, I. et Mukai, T. (1995). "Genomic imprinting of p57/KIP2, a cyclin-dependent kinase inhibitor, in mouse." *Nature Genetics* **11**: 204-206.

Hayward, B. E. *et coll.* (1998). "The human *GNAS1* gene is imprinted and encodes distinct paternally and biallelically expressed G proteins." *Proc. Natl. Acad. Sci. USA* **95**: 10038-10043.

Heard, E. *et coll.* (1997). "X-chromosome inactivation in mammals." *Annu. Rev. Genet.* **31**: 571-610.

Heinke, C. *et coll.* (1987). "Tissue specific methylation of C-MYC in adult chickens." *Biochemical and Biophysical Research Communications* **149**(1): 313-317.

Hendrich, B. et Bird, A. (1998). "Identification and characterization of a family of mammalian methyl-CpG binding proteins." *Molecular and Cellular Biology* **18**(11): 6538-6547.

Hendriks, R. W. (1992). "An X chromosome inactivation assay based on differential methylation of a CpG island coupled to a VNTR polymorphism at the 5' end of the monoamine oxidase A gene." *Human Molecular Genetics* **1**(3): 187-194.

Henikoff, S. et Matzke, M. A. (1997). "Exploring and explaining epigenetic effects." *Trends in Genetics* **13**(8): 293-295.

Herman, J. G. *et coll.* (1994). "Silencing of the *VHL* tumor-suppressor gene by DNA methylation in renal carcinoma." *Proc. Natl. Acad. Sci. USA* **91**: 9700-9704.

Herman, J. G. *et coll.* (1996). "Hypermethylation-associated inactivation indicates a tumor suppressor role for *p15<sup>INK4B1</sup>*." *Cancer Research* **56**: 722-727.

Hiltunen, M. O. *et coll.* (1997). "Hypermethylation of the *APC* (adenomatous polyposis coli) gene promoter region in human colorectal carcinoma." *Int. J. Cancer* **70**: 644-648.

Holländer, G. A. *et coll.* (1998). "Monoallelic expression of the interleukin-2 locus." *Science* **279**: 2118-2121.

Hollick, J. B. *et coll.* (1997). "Paramutation and related allelic interactions." *Trends in Genetics* **13**(8): 302-308.



Hong, R. (1998). "The DiGeorge anomaly (Catch 22, Digeorge/Velocardiofacial syndrome." *Seminars in Hematology* **35**(4): 282-290.

Hu, J.-F. *et coll.* (1998). "The role of histone acetylation in the allelic expression of the imprinted human insulin-like growth factor II gene." *Biochemical and Biophysical Research Communications* **251**: 403-408.

Hurst, L. D. *et coll.* (1996). "Imprinted genes have few and small introns." *Nature Genetics* **12**: 234-237.

Issa, J.-P. J. *et* Baylin, S. B. (1996). "Epigenetics and human disease." *Nature Medicine* **2**(3): 281-282.

Jackson, A. *et coll.* (1998). "The human homologue of the *Drosophila tailless* gene (*TLX*): characterization and mapping to a region of common deletion in human lymphoid leukemia on chromosome 6q21." *Genomics* **50**: 34-43.

Jaenisch, R. (1997). "DNA methylation and imprinting: why bother?" *Trends in Genetics* **13**(8): 323-329.

Jinno, Y. *et coll.* (1994). "Mosaic and polymorphic imprinting of the WT1 gene in humans." *Nature Genetics* **6**: 305-309.

Jones, P. A. *et* Laird, P. W. (1999). "Cancer epigenetics comes of age." *Nature Genetics* **21**(2): 163.

Jones, P. L. *et coll.* (1998). "Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription." *Nature Genetics* **19**: 187-191.

Jong, M. T. C. *et coll.* (1999). "A novel imprinted gene, encoding a RING zinc-finger protein, and overlapping antisense transcript in the Prader-Willi syndrome critical region." *Human Molecular Genetics* **8**: 783-793.

Kadonaga, J. T. (1998). "Eukaryotic transcription: an interlaced network of transcription factors and chromatin-modifying machines." *Cell* **92**: 307-313.

Kaghad, M. *et coll.* (1997). "Monoallelically expressed gene related to p53 at 1p36, a region frequently deleted in neuroblastoma and other human cancers." *Cell* **90**: 809-819.

Kagitani, F. *et coll.* (1997). "*Peg5/Neuronatin* is an imprinted gene located on sub-distal chromosome 2 in the mouse." *Nucleic Acids Research* **25**(17): 3428-3432.

Kane, M. F. *et coll.* (1997). "Methylation of the *hMLH1* promoter correlates with lack of expression of *hMLH1* in sporadic colon tumors and mismatch repair-defective human tumor cell lines". *Cancer Research* **57**(5): 808-811.

Kaneko-Ishino, T. *et coll.* (1995). "*Peg/Mest* imprinted gene on chromosome 6 identified by cDNA subtraction hybridization." *Nature Genetics* **11**: 52-59.

Kass, S. U. *et coll.* (1997). "How does DNA methylation repress transcription?" *TIG* **13**(11): 444-449.

Kato, M. V. *et coll.* (1998). "Paternal imprinting of the mouse serotonin receptor 2A gene *Htr2* in embryonic eye: a conserved imprinting regulation on the *RB/Rb* locus." *Genomics* **47**: 146-148.

Keshet, I. *et coll.* (1986). "DNA methylation affects the formation of active chromatin." *Cell* **44**: 535.

Kim, J. *et coll.* (1999). "*Zim1*, a maternally expressed mouse Kruppel-type zinc-finger gene located in proximal chromosome 7." *Human Molecular Genetics* **8**: 847-854.

Kitsberg, D. *et coll.* (1993). "Allele-specific replication timing of imprinted gene regions." *Nature* **364**: 459-463.

Knight, S. J. L. *et coll.* (1993). "Trinucleotide repeat amplification and hypermethylation of a CpG island in *FRAXE* mental retardation." *Cell* **74**: 127-134.

Knoll, J. H. M. *et coll.* (1994). "Allele specificity of DNA replication timing in the Angelman/Prader-Willi syndrome imprinted chromosomal region." *Nature Genetics* **6**: 41-46.

Knudson, A. G. (1971). "Mutation and cancer: statistical study of retinoblastoma." *Proc. Natl. Acad. Sci. USA* **68**: 820-823.

Kochanek, S. *et coll.* (1993). "DNA methylation in the Alu sequences of diploid and haploid primary human cells." *The EMBO Journal* **12**(3): 1141-1151.

Kohne, D. E. *et coll.* (1977). "Room temperature method for increasing the rate of DNA reassociation by many thousandfold: the phenol emulsion reassociation technique." *Biochemistry* **16**(24): 5329-5341.

Kubota, T. *et coll.* (1997). "Methylation-specific PCR simplifies imprinting analysis." *Nature Genetics* **16**: 16-17.

Kuromitsu, J. *et coll.* (1997). "A unique downregulation of h2-calponin gene expression in Down Syndrome: a possible attenuation mechanism for fetal survival by methylation at the CpG island in the trisomic chromosome 21." *Molecular and Cellular Biology* **17**(2): 707-712.

Laird, P. W. *et coll.* (1995). "Suppression of intestinal neoplasia by DNA hypomethylation." *Cell* **81**: 197-205.

Lamour, V. *et coll.* (1995). "A human homolog of the *S. cerevisiae* *HIR1* and *HIR2* transcriptional repressor cloned from the DiGeorge syndrome critical region." *Human Molecular Genetics* **4**(5): 791-799.

LaSalle, J. M. et Lalande, M. (1996). "Homologous association of oppositely imprinted chromosomal domains." *Science* **272**: 725-728.

Lefebvre, L. *et coll.* (1997). "Genomic structure and parent-of-origin-specific methylation of *Peg1*." *Human Molecular Genetics* **6**(11): 1907-1915.

Leff, S. E. *et coll.* (1992). "Maternal imprinting of the mouse *Snrpn* gene and conserved linkage homology with the human Prader-Willi syndrome region." *Nature Genetics* **2**: 259-264.

Lei, H. *et coll.* (1996). "De novo DNA cytosine methyltransferase activities in mouse embryonic stem cells." *Development* **122**: 3195-3205.

Lemieux, N. *et coll.* (1992). "A simple method for simultaneous R- or G-banding and fluorescence in situ hybridization of small single-copy genes." *Cytogenetics and Cell Genetics* **59**: 311-312.

Lengauer, C. *et coll.* (1997). "DNA methylation and genetic instability in colorectal cancer cells." *Proc. Natl. Acad. Sci. USA* **94**: 2545-2550.

Lewin, B. (1998). "The mystique of epigenetics." *Cell* **93**: 301-303.

Li, E. *et coll.* (1993). "Role for DNA methylation in genomic imprinting." *Nature* **366**: 362-366.

Li, E. *et coll.* (1992). "Targeted mutation of the DNA methyltransferase gene results in embryonic lethality." *Cell* **69**: 915-926.

Lighten, A. D. *et coll.* (1997). "*IGF2* is parentally imprinted in human preimplantation embryos." *Nature Genetics* **15**: 122-123.

Lindsay, E. A. *et coll.* (1998). "*ES2*, a gene deleted in DiGeorge syndrome, encodes a nuclear protein and is expressed during early mouse development, where it shares an expression domain with a Goosecoid-like gene." *Human Molecular Genetics* **7**(4): 629-635.

- Lock, L. F. *et coll.* (1986). "Methylation of the mouse *hprt* gene differs on the active and inactive X chromosomes." *Molecular and Cellular Biology* **6**(3): 914-924.
- MacDonald, H. R. et Wevrick, R. (1997). "The *necdin* gene is deleted in Prader-Willi syndrome and is imprinted in human and mouse." *Human Molecular Genetics* **6**: 1873-1878.
- MacLeod, A. R. *et coll.* (1995). "Regulation of DNA methylation by the Ras signaling pathway." *The Journal of Biological Chemistry* **270**(19): 11327-11337.
- Macleod, D. *et coll.* (1994). "Sp1 sites in the mouse *aprt* gene promoter are required to prevent methylation of the CpG island." *Genes & Development* **8**: 2282-2292.
- Malagnac, F. *et coll.* (1997). "A gene essential for De Novo methylation and development in *Ascobolus* reveals a novel type of eukaryotic DNA methyltransferase structure." *Cell* **91**: 281-290.
- Maloisel, L. et Rossignol, J.-L. (1998). "Suppression of crossing-over by DNA methylation in *Ascobolus*." *Genes & Development* **12**: 1381-1389.
- Mancini, D. *et coll.* (1997). "Constitutively methylated CpG dinucleotides as mutation hot spots in the retinoblastoma gene (RB1)." *Am. J. Hum. Genet.* **61**: 80-87.
- Marra, M. A. *et coll.* (1998). "Expressed sequence tags-ESTablishing bridges between genomes." *Trends in Genetics* **14**(1): 4-7.
- Matsuo, K. *et coll.* (1993). "Evidence for erosion of mouse CpG islands during mammalian evolution." *Somatic Cell and Molecular Genetics* **19**(6): 543-555.
- McCarthy, L. C. *et coll.* (1997). "A first-generation whole genome-radiation hybrid map spanning the mouse genome." *Genome Research* **7**: 1153-1161.
- McQueen, H. *et coll.* (1998). "Chicken microchromosomes are hyperacetylated, early replicating, and gene rich." *Genome Research* **8**: 621-630.

Meehan, R. R. *et coll.* (1992). "Characterization of MeCP2, a vertebrate DNA binding protein with affinity for methylated DNA." *Nucleic Acids Research* **20**(19): 5085-5092.

Melo, J. V. et Goldman, J. M. (1995). "Reply to Haas: are ABL and BCR imprinted?" *Leukemia* **9**: 1407-1408.

Melo, J. V. *et coll.* (1995). "Balanced parental contribution to the ABL component of the BCR-ABL gene in chronic myeloid leukemia." *Leukemia* **9**: 734-739.

Merlo, A. *et coll.* (1995). "5' CpG island methylation is associated with transcriptional silencing of the tumour suppressor *p16/CDKN2/MTS1* in human cancers." *Nature Medicine* **1**(7): 686-692.

Miller, A. P. et Willard, H. F. (1998). "Chromosomal basis of X chromosome inactivation: identification of a multigene domain in Xp11.21-p11.22 that escapes X inactivation." *Proc. Natl. Acad. Sci. USA* **95**: 8709-8714.

Miller, N. *et coll.* (1997). "The *MAS* proto-oncogene is imprinted in human breast tissue." *Genomics* **46**: 509-512.

Miyoshi, N. *et coll.* (1998). "Identification of the *Meg1/Grb10* imprinted gene on mouse proximal chromosome 11, a candidate for the Silver-Russell syndrome gene." *Proc. Natl. Acad. Sci. USA* **95**: 1102-1107.

Moore, T. *et coll.* (1997). "Multiple imprinted sense and antisense transcripts, differential methylation and tandem repeats in a putative imprinting control region upstream of mouse *Igf2*." *Proc. Natl. Acad. Sci. USA* **94**: 12509-12514.

Moore, T. et Haig, D. (1991). "Genomic imprinting in mammalian development: a parental tug-of-war." *TIG* **7**(2): 45-49.

Mori, N. *et coll.* (1998). "Chromosome band 1p36 contains a putative tumor suppressor gene important in the evolution of chronic myelocytic leukemia." *Blood* **92**(9): 3405-3409.

Mummaneni, P. *et coll.* (1998). "The primary function of a redundant Sp1 binding site in the mouse *aprt* gene promoter is to block epigenetic gene inactivation." *Nucleic Acids Research* **26**(22): 5163-5169.

Nan, X. *et coll.* (1997). "MeCP2 is a transcriptional repressor with abundant binding sites in genomic chromatin." *Cell* **88**: 471-481.

Nan, X. *et coll.* (1993). "Dissection of the methyl-CpG binding domain from the chromosomal protein MeCP2." *Nucleic Acids Research* **21**(21): 4886-4892.

Nan, X. *et coll.* (1998). "Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex." *Nature* **393**: 386-389.

Nelson, M. et McClelland, M. (1991). "Site-specific methylation: effect on DNA modification methyltransferases and restriction endonucleases." *Nucleic Acids Research* **19**(supplement): 2045-2071.

Nelson, S. F. *et coll.* (1993). "Genomic mismatch scanning: a new approach to genetic linkage mapping." *Nature Genetics* **4**: 11-17.

Neumann, B. *et coll.* (1995). "Characteristics of imprinted genes." *Nature Genetics* **9**: 12-13.

Neumann, B. et Barlow, D.P. (1996). "Multiple roles for DNA methylation in gametic imprinting." *Current Opinion in Genetics and Development* **6**: 159-163.

Nezer, C. *et coll.* (1999). "An imprinted QTL with major effect on muscle mass and fat deposition maps to the *IGF2* locus in pigs." *Nature Genetics* **21**(2): 155.

Nishigaki, K. *et coll.* (1985). "Type II restriction endonucleases cleave single-stranded DNAs in general." *Nucleic Acids Research* **13**(16): 5747-5760.

Ohyashiki, J. H. *et coll.* (1993). "The methylation status of the major breakpoint cluster region in human leukemia cells, including Philadelphia chromosome-positive cells, is linked to the lineage of hematopoietic cells." *Leukemia* **7**(6): 801-807.

Okano, M. *et coll.* (1998). "Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases." *Nature Genetics* **19**: 219-220.

Ottaviano, Y. L. *et coll.* (1994). "Methylation of the estrogen receptor gene CpG island marks loss of estrogen receptor expression in human breast cancer cells." *Cancer Research* **54**: 2552-2555.

Pàldi, A. *et coll.* (1995). "Imprinted chromosomal regions of the human genome display sex-specific meiotic recombination frequencies." *Current Biology* **5**: 1030-1035.

Paulsen, M. *et coll.* (1998). "Syntenic organization of the mouse distal chromosome 7 imprinting cluster and the Beckwith-Wiedemann syndrome region in chromosome 11p15.5." *Human Molecular Genetics* **7**(7).

Pazin, M. J. et Kadonaga, J. T. (1997). "What's up and down with histone deacetylation and transcription." *Cell* **89**: 325-328.

Penny, G. D. *et coll.* (1996). "Requirement for *Xist* in X chromosome inactivation." *Nature* **379**: 131-137.

Pfeifer, G. P. *et coll.* (1990). "In vivo footprint and methylation analysis by PCR-aided genomic sequencing: comparison of active and inactive X chromosomal DNA at the CpG island and promoter of human PGK-1." *Genes & Development* **4**: 1277-1287.

Pfeifer, K. et Tilghman, S. M. (1994). "Allele-specific gene expression in mammals: the curious case of the imprinted RNAs." *Genes & Development* **8**: 1867-1874.



Pikaart, M. J. *et coll.* (1998). "Loss of transcriptional activity of a transgene accompanied by DNA methylation and histone deacetylation and is prevented by insulators." *Genes & Development* **12**(18): 2852.

Pirrotta, V. (1998). "Polycomb the genome: PcG, trxG, and chromatin silencing." *Cell* **93**: 333-336.

Plass, C. *et coll.* (1996). "Identification of *Grfl* on mouse chromosome 9 as an imprinted gene by RLGS-M." *Nature Genetics* **14**: 106-109.

Pradhan, S. *et coll.* (1997). "Baculovirus-mediated expression and characterization of the full-length murine DNA methyltransferase." *Nucleic Acids Research* **25**(22): 4666-4673.

Prusiner, S. B. *et coll.* (1998). "Prion protein biology." *Cell* **93**: 337-348.

Puech, A. *et coll.* (1997). "Comparative mapping of the human 22q11 chromosomal region and the orthologous region in mice reveals complex changes in gene organization." *Proc. Natl. Acad. Sci. USA* **94**: 14608-14613.

Qian, N. *et coll.* (1997). "The *IPL* gene on chromosome 11p15.5 is imprinted in humans and mice and is similar to *TDAG51*, implicated in Fas expression and apoptosis." *Human Molecular Genetics* **6**: 2021-2029.

Ramchandani, S. *et coll.* (1997). "Inhibition of tumorigenesis by a cytosine-DNA methyltransferase, antisense oligodeoxynucleotide." *Proc. Natl. Acad. Sci. USA* **94**: 684-689.

Rastan, S. et Beeley, L. J. (1997). "Functional genomics: going forwards from the databases." *Current Opinion in Genetics & Development* **7**: 777-783.

Reik, W. et Walter, J. (1998). "Imprinting mechanisms in mammals." *Current Opinion in Genetics & Development* **8**: 154-164.

Rein, T. *et coll.* (1998). "Identifying 5-methylcytosine and related modifications in DNA genomes." *Nucleic Acids Research* **26**(10): 2255-2264.

Rein, T. *et coll.* (1997). "Active mammalian replication origins are associated with a high-density cluster of <sup>m</sup>CpG dinucleotides." *Molecular and Cellular Biology* **17**(1): 416-426.

Relaix, F. *et coll.* (1998). "*Peg3/Pw1* is an imprinted gene involved in the TNF-NFκB signal transduction pathway." *Nature Genetics* **18**: 287-291.

Riesewijk, A. M. *et coll.* (1996). "The *MAS* proto-oncogene is not imprinted in humans." *Genomics* **35**: 380-382.

Riesewijk, A. M. *et coll.* (1996). "Maternal-specific methylation of the human *IGF2R* gene is not accompanied by allele-specific transcription." *Genomics* **31**: 158-166.

Riggs, A. D. (1975). "X inactivation, differentiation and DNA methylation." *Cytogenet. Cell Genet.* **14**: 9-11.

Riggs, A. D. *et coll.* (1996). Introduction. Epigenetic mechanisms of gene regulation. V. E. A. Russo, R. A. Martienssen et A. D. Riggs, Cold Spring Harbor Laboratory Press.

Rougeulle, C. *et coll.* (1998). "An imprinted antisense RNA overlaps *UBE3A* and a second maternally expressed transcript." *Nature Genetics* **19**(1): 15-16.

Rougeulle, C. *et coll.* (1997). "The Angelman syndrome candidate gene, *UB3A/E6-AP*, is imprinted in brain." *Nature Genetics* **17**(1): 14-15.

Rubin, C. M. *et coll.* (1994). "*Alu* repeated DNAs are differentially methylated in primate germ cells." *Nucleic Acids Research* **22**(23): 5121-5127.

Ryan, A. K. *et coll.* (1997). "Spectrum of clinical features associated with interstitial chromosome 22q11 deletions: a European collaborative study." *J Med Genet.* **34**: 798-804.

Saitoh, S. *et coll.* (1996). "Minimal definition of the imprinting center and fixation of chromosome 15q11-q13 epigenotype by imprinting mutations." *Proc. Natl. Acad. Sci. USA* **93**: 7811-7815.

Sambrook, J. *et coll.* (1989). Molecular cloning: a laboratory manual, Cold Spring Harbor.

Sandberg, G. et Schalling, M. (1997). "Effect of *in vitro* promoter methylation and CGG repeat expansion on *FMR-1* expression." *Nucleic Acids Research* **25**(14): 2883-2887.

Schorderet, D. F. et Gartler, S. M. (1992). "Analysis of CpG suppression in methylated and nonmethylated species." *Proc. Natl. Acad. Sci. USA* **89**: 957-961.

Sheardown, S. A. *et coll.* (1997). "Stabilization of *Xist* RNA mediates initiation of X chromosome inactivation." *Cell* **91**: 99-107.

Shizuya, H. *et coll.* (1992). "Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector." *Proc. Natl. Acad. Sci. USA* **89**: 8794-8797.

Silva, A. J. et White, R. (1988). "Inheritance of allelic blueprints for methylation patterns." *Cell* **54**: 145-152.

Singer-Sam, J. et Riggs, A. D. (1993). X chromosome inactivation and DNA methylation. DNA Methylation: molecular biology and biological significance. J. P. Jost and H. P. Saluz. Basel, Birkhäuser Verlag: 358-384.

Stirzaker, C. *et coll.* (1997). "Extensive DNA methylation spanning the *Rb* promoter in retinoblastoma tumors." *Cancer Research* **57**: 2229-2237.

Su, L.-K. *et coll.* (1992). "Multiple intestinal neoplasia caused by a mutation in the murine homolog of the APC gene." *Science* **256**: 668-670.

Surani, M. A. *et coll.* (1990). "Genome imprinting and development in the mouse." *Development Supplement*: 89-98.

Sutcliffe, J. S. *et coll.* (1994). "Deletions of a differentially methylated CpG island at the *SNRPN* gene define a putative imprinting control region." *Nature Genetics* **8**(1): 52-58.

Szyf, M. (1996). "The DNA methylation machinery as a target for anticancer therapy." *Pharmacol. Ther.* **70**(1): 1-37.

Takeuchi, S. *et coll.* (1998). "Frequent loss of heterozygosity on the long arm of chromosome 6: identification of two distinct regions of deletion in childhood acute lymphoblastic leukemia." *Cancer Research* **58**: 2618-2623.

Tate, P. *et coll.* (1996). "The methyl-CpG binding protein MeCP2 is essential for embryonic development in the mouse." *Nature Genetics* **12**: 205-208.

Tate, P. H. et Bird, A. P. (1993). "Effects of DNA methylation on DNA-binding proteins and gene expression." *Current Opinion in Genetics and Development* **3**: 226-231.

Temple, I. K. *et coll.* (1996). "Further evidence for an imprinted gene for neonatal diabetes localised to chromosome 6q22-q23." *Human Molecular Genetics* **5**(8): 1117-1121.

Thomas, J. H. (1995). "Genomic imprinting proposed as a surveillance mechanism for chromosome loss." *Proc. Natl. Acad. Sci. USA* **92**: 480-482.

Thorvaldsen, J. L. *et coll.* (1998). "Deletion of the *H19* differentially methylated domain results in loss of imprinted expression of *H19* and *Igf2*." *Genes & Development* **12**.

Tilghman, S. (1993). "DNA methylation: a phoenix rises." *Proc. Natl. Acad. Sci. USA* **90**: 8761-8762.

Toniolo, D. *et coll.* (1988). "Expression of the G6PD locus on the human X chromosome is associated with demethylation of three CpG islands within 100 kb of DNA." *The EMBO Journal* **7**(2): 401-406.

Tremblay, K. D. *et coll.* (1995). "A paternal-specific methylation imprint marks the alleles of the mouse *H19* gene." *Nature Genetics* **9**.

Tucker, K. L. *et coll.* (1996). "Germ-line passage is required for establishment of methylation and expression patterns of imprinted but not of nonimprinted genes." *Genes & Development* **10**: 1008-1020.

Ushijima, T. *et coll.* (1997). "Establishment of methylation-sensitive-representational difference analysis and isolation of hypo- and hypermethylated genomic fragments in mouse liver tumors." *Proc. Natl. Acad. Sci. USA* **94**: 2284-2289.

Varmuza, S. et Mann, M. (1994). "Genomic imprinting-defusing the ovarian time bomb." *TIG* **10**(4): 118-123.

Veigl, M. L. *et coll.* (1998). "Biallelic inactivation of *hMLH1* by epigenetic gene silencing, a novel mechanism causing human MSI cancers." *Proc. Natl. Acad. Sci. USA* **95**: 8698-8702.

Verdine, G. L. (1994). "The flip side of DNA methylation." *Cell* **76**: 197-200.

Villar, A. J. et Pedersen, R. A. (1994). "Parental imprinting of the *Mas* protooncogene in mouse." *Nature Genetics* **8**: 373-379.

Wakefield, M. J. *et coll.* (1997). "Histone underacetylation is an ancient component of mammalian X chromosome inactivation." *Proc. Natl. Acad. Sci. USA* **94**: 9665-9668.

Wales, M. K. *et coll.* (1995). "p53 activates expression of *HIC-1*, a new candidate tumour suppressor gene on 17p13.3." *Nature Medicine* **1**(6): 570-577.

Walsh, C. P. et Bestor, T. H. (1999). "Cytosine methylation and mammalian development." *Genes & Development* **13**: 26-34.

Weisbach, A. (1993). A chronicle of DNA methylation (1948-1975). DNA Methylation: molecular biology and biological significance. J. P. Jost and H. P. Saluz. Basel, Birhäuser Verlag: 1-10.

Weitzel, J. M. *et coll.* (1997). "Chicken MAR-binding protein ARBP is homologous to rat methyl-CpG-binding protein MeCP2". *Molecular and Cellular Biology* **17**(9): 5656-5666.

Wevrick, R. et Francke, U. (1997). "An imprinted mouse transcript homologous to the human imprinted in the Prader-Willi syndrome region." *Human Molecular Genetics* **6**: 325-332.

Wiens, G. R. et Sorger, P. K. (1998). "Centromeric chromatin and epigenetic effects in kinetochore assembly." *Cell* **93**: 313-316.

Woodcock, D. M. *et coll.* (1997). "Asymmetric methylation in the hypermethylated CpG promoter region of the human L1 retrotransposon." *The Journal of Biological Chemistry* **272**(12): 7810-7816.

Wu, Y.-Q. *et coll.* (1999). "Molecular refinement of the 1p36 deletion syndrome reveals size diversity and a preponderance of maternally derived deletions." *Human Molecular Genetics* **8**(2): 313-321.

Wutz, A. *et coll.* (1997). "Imprinted expression of the *Igf2r* gene depends on an intronic CpG island." *Nature* **389**: 745-749.

Xiong, Z. *et coll.* (1998). "Differential replication timing of X-linked genes measured by a novel method using single-nucleotide primer extension." *Nucleic Acids Research* **26**(2): 684-686.

Yamagishi, H. *et coll.* (1999). "A molecular pathway revealing a genetic basis for human cardiac and craniofacial defects." *Science* **283**: 1158-1161.

Yoder, J. A. *et coll.* (1997). "Cytosine methylation and the ecology of intragenomic parasites." *Trends in Genetics* **13**(8): 335-340.

Zhang, X. et Verdine, G. L. (1996). "Mammalian DNA cytosine-5 methyltransferase interacts with p23 protein." *FEBS Letters* **392**: 179-183.

Zingg, J.-M. et Jones, P. A. (1997). "Genetic and epigenetic aspects of DNA methylation on genome expression, evolution, mutation and carcinogenesis". *Carcinogenesis* **18**(5): 869-882.

Zion, M. *et coll.* (1994). "Progressive *de novo* DNA methylation at the *brc-abl* locus in the course of chronic myelogenous leukemia." *Proc. Natl. Acad. Sci. USA* **91**: 10722-10726.

Zneimer, S. M. *et coll.* (1998). "Partial trisomy of chromosome 6q: an interstitial duplication of the long arm." *American Journal of Medical Genetics* **80**: 133-135.

Zuccotti, M. et Monk, M. (1995). "Methylation of the mouse *Xist* gene in sperm and eggs correlates with imprinted *Xist* expression and paternal X-inactivation." *Nature Genetics* **9**: 316-320.