

**Direction des bibliothèques**

**AVIS**

Ce document a été numérisé par la Division de la gestion des documents et des archives de l'Université de Montréal.

L'auteur a autorisé l'Université de Montréal à reproduire et diffuser, en totalité ou en partie, par quelque moyen que ce soit et sur quelque support que ce soit, et exclusivement à des fins non lucratives d'enseignement et de recherche, des copies de ce mémoire ou de cette thèse.

L'auteur et les coauteurs le cas échéant conservent la propriété du droit d'auteur et des droits moraux qui protègent ce document. Ni la thèse ou le mémoire, ni des extraits substantiels de ce document, ne doivent être imprimés ou autrement reproduits sans l'autorisation de l'auteur.

Afin de se conformer à la Loi canadienne sur la protection des renseignements personnels, quelques formulaires secondaires, coordonnées ou signatures intégrées au texte ont pu être enlevés de ce document. Bien que cela ait pu affecter la pagination, il n'y a aucun contenu manquant.

**NOTICE**

This document was digitized by the Records Management & Archives Division of Université de Montréal.

The author of this thesis or dissertation has granted a nonexclusive license allowing Université de Montréal to reproduce and publish the document, in part or in whole, and in any format, solely for noncommercial educational and research purposes.

The author and co-authors if applicable retain copyright ownership and moral rights in this document. Neither the whole thesis or dissertation, nor substantial extracts from it, may be printed or otherwise reproduced without the author's permission.

In compliance with the Canadian Privacy Act some supporting forms, contact information or signatures may have been removed from the document. While this may affect the document page count, it does not represent any loss of content from the document.

Université de Montréal

**Caractérisation fonctionnelle des variants génétiques de la région régulatrice (rSNP)  
des gènes du point de contrôle G1/S.**

par

Joëlle Dionne

Département de Biologie Moléculaire  
Faculté de médecine

Mémoire présenté à la Faculté de médecine  
en vue de l'obtention du grade de Maître en science (M.Sc.)  
en Biologie Moléculaire

Avril, 2008

© Joëlle Dionne, 2008



Université de Montréal  
Faculté de médecine

Ce mémoire intitulé :

**Caractérisation fonctionnelle des variants génétiques de la région régulatrice (rSNP)  
des gènes du point de contrôle G1/S.**

présenté par :  
Joëlle Dionne

a été évaluée par un jury composé des personnes suivantes :

Dr. Edward Bradley, président-rapporteur  
Dr. Daniel Sinnett, directeur de recherche  
Dr. Louise R. Simard, membre du jury

## Résumé

Le contrôle du cycle cellulaire est important dans le maintien de l'homéostasie cellulaire et de l'intégrité du matériel génétique. La cellule est réceptive à des signaux antiprolifératifs seulement aux points de contrôle du cycle cellulaire tel que le point G1/S. Le bon fonctionnement de ces points de contrôle est crucial pour contrer les anomalies cellulaires. La dérégulation des gènes codants pour des composantes du point de contrôle de la transition G1/S est fréquemment observée chez divers types de cancers. Nous proposons que des polymorphismes fonctionnels localisés dans la région régulatrice (rSNP) des gènes impliqués dans le point de contrôle G1/S auraient un impact significatif dans l'oncogenèse. Dans cette étude, nous avons validé, à l'aide d'une combinaison d'approches *in silico* et *in vitro*, l'impact fonctionnel de la variabilité génétique sous la forme d'haplotypes (rHAPs), au niveau du promoteur proximal, de 18 gènes codant pour des composantes du point de contrôle de la transition G1/S. Pour ce faire, nous avons sous-cloné les rHAPs majeurs dans un vecteur contenant le gène rapporteur luciférase (pGL3b). Ces constructions furent utilisées dans des essais de transfections transitoires dans 3 lignées cellulaires (Hela, Jeg3 et HepG2). Nous avons observé qu'au moins 12 rHAPs influencent significativement l'activité transcriptionnelle de façon allèle spécifique. Ces rHAPs sont associés aux gènes CDC25a, CDKN1A, CDKN1B, CDKN2A, MADH3, TFDP1 et TGF $\beta$ 1. Les analyses de retard sur gel d'électrophorèse (EMSA) ont permis d'identifier au moins 16 sites de liaison ADN-protéine différentielles dans les rHAPs de ces gènes. La variabilité du niveau d'expression des gènes étudiés pourrait donc contribuer à la susceptibilité interindividuelle aux maladies complexes, comme le cancer.

Le projet est supporté par Génome Québec/Canada et la Fondation Cole.

**Mots-clés** : polymorphisme, région régulatrice, point de contrôle G1/S, transfection cellulaire, facteur de transcription, retard sur gel, maladie complexe.

## Summary

G1/S transition in the cell cycle is a highly regulated biological process. Depending on the context, growth-control genes present in the G1/S checkpoint can stop cell cycle progression and activate survival pathways in the cell, allowing for DNA repair processes or cell death by apoptosis to be initiated. Deregulation of G1/S checkpoint genes is frequently observed in complex diseases, particularly cancers. Indeed genes encoding components of cell cycle control are frequently mutated in human cancers. We hypothesized that functional polymorphisms located in the regulatory region of candidate genes (rSNPs) could lead to variable levels of transcription and thus predispose the individuals carrying these genetic variants to cancer. In this report, we assessed the functional impact of rSNPs located in the proximal promoter region of 18 candidate gene encoding components of the G1/S checkpoint by combining *in silico* analysis and *in vitro* functional assays. Following the subcloning of the major promoter haplotypes (rHap) into luciferase gene reporter vectors (pGL3a), transient transfection assays were performed in 3 cell lines (Hela, Jeg3 and HepG2). We found that at least 12 rHAPs associated with CDC25a, CDKN1A, CDKN1B, CDKN2A, MADH3, TFDPI and TGF $\beta$ 1 significantly influenced transcriptional activity in an allele-specific manner. Further validation by electrophoresis mobility shift assays (EMSA) allowed us to detect 16 differential DNA-protein binding interactions. Although the biological significance of these observations remains to be demonstrated, the expected variability of expression levels in key cell cycle components might influence an individual's risk of cancer. This work was supported by Genome Quebec/Canada and the Cole Foundation.

**Keywords** : polymorphism, promoter region, G1/S checkpoint, transfection assay, transcription factor, EMSA, complex disease.

## Table des matières

RÉSUMÉ	iii
SUMMARY	iv
TABLE DES MATIÈRES	v
LISTE DES TABLEAUX	x
LISTE DES FIGURES	xii
LISTE DES ABRÉVIATIONS	xiv
REMERCIEMENTS	xviii
CHAPITRE 1 – INTRODUCTION	1
1.1 Introduction au cycle cellulaire	2
1.1.1 Le déroulement du cycle cellulaire	2
1.1.2 Régulation du cycle cellulaire par les principaux complexes cycline-cdk	4
1.1.3 La régulation des protéines cyclines et cdk	5
1.2 Le point de contrôle G1/S	7
1.2.1 Fonctionnement du point de contrôle G1/S.	7
1.2.2 Arrêt du cycle cellulaire au point de contrôle G1/S	8
1.3 Les polymorphismes	12
1.3.1 Les types de polymorphismes	12
1.3.2 Les SNPs; la mutation des nucléotides et l'impact de leur localisation	12
1.3.3 Les polymorphismes impliqués dans diverses maladies	15
1.4 La variabilité génétique interindividuelle	15
1.4.1 Définition	15
1.4.2 Les rSNPs impliqués dans la variabilité interindividuelle	16
1.5 Hypothèse	18
1.6 Objectifs du projet	19
CHAPITRE 2 – MATÉRIEL ET MÉTHODES	20
2.1 Identification des SNPs dans la région régulatrice des gènes candidats du point de contrôle G1/S du cycle cellulaire	21
2.1.1 Analyse des sites de liaison des facteurs de transcription	21
2.2 Identification des haplotypes régulateurs (rHap)	21

2.2.1 Population à l'étude	21
2.2.2 Génotypage	22
2.2.3 Construction des haplotypes	27
2.3 Validation fonctionnelle <i>in vitro</i>	27
2.3.1 Construction des vecteurs	27
2.3.2 Transfection	38
2.3.3 Electrophoretic mobility shift assay (EMSA)	40
 CHAPITRE 3 – RÉSULTATS	 45
3.1 Génotypage	46
3.2 Description de l'étude du gène CDC25A	51
3.2.1 Identification des haplotypes régulateurs (rHap) présents dans l'échantillon représentatif de la population mondiale	51
3.2.2 Analyse fonctionnelle d'expression des rHaps, par expériences de transfections transitoires	54
3.2.3 Analyse des modifications de liaison des facteurs de transcription, dû à la présence de polymorphismes régulateurs, dans des expériences de retard sur gel	57
3.3 Description de l'étude du gène CDKN1A	60
3.3.1 Identification des haplotypes régulateurs présents dans l'échantillon représentatif de la population mondiale	60
3.3.2 Analyse fonctionnelle d'expression des rHaps, par expériences de transfections transitoires	62
3.3.3 Analyse des modifications de liaison des facteurs de transcription, dû à la présence de polymorphismes régulateurs, dans des expériences de retard sur gel	64
3.4 Description de l'étude du gène CDKN1B	66
3.4.1 Identification des haplotypes régulateurs présents dans l'échantillon représentatif de la population mondiale	66
3.4.2 Analyse fonctionnelle d'expression des rHaps, par expériences de transfections transitoires	68
3.4.3 Analyse des modifications de liaison des facteurs de transcription, dû à la présence de polymorphismes régulateurs, dans des expériences de retard sur gel	70
3.5 Description de l'étude du gène CDKN2A	72
3.5.1 Identification des haplotypes régulateurs présents dans l'échantillon représentatif de la population mondiale	72
3.5.2 Analyse fonctionnelle d'expression des rHaps, par expériences de transfections transitoires	74
3.5.3 Analyse des modifications de liaison des facteurs de transcription, dû à la présence de polymorphismes régulateurs, dans des expériences de retard sur gel	76
3.6 Description de l'étude du gène MADH3	78
3.6.1 Identification des haplotypes régulateurs présents dans l'échantillon représentatif de la population mondiale	78

3.6.2 Analyse fonctionnelle d'expression des rHaps, par expériences de transfections transitoires	80
3.6.3 Analyse des modifications de liaison des facteurs de transcription, dû à la présence de polymorphismes régulateurs, dans des expériences de retard sur gel	82
3.7 Description de l'étude du gène TFDP1	84
3.7.1 Identification des haplotypes régulateurs présents dans l'échantillon représentatif de la population mondiale	84
3.7.2 Analyse fonctionnelle d'expression des rHaps, par expériences de transfections transitoires	86
3.7.3 Analyse des modifications de liaison des facteurs de transcription, dû à la présence de polymorphismes régulateurs, dans des expériences de retard sur gel	88
3.8 Description de l'étude du gène TGFβ1	90
3.8.1 Identification des haplotypes régulateurs présents dans l'échantillon représentatif de la population mondiale	90
3.8.2 Analyse fonctionnelle d'expression des rHaps, par expériences de transfections transitoires	92
3.8.3 Analyse des modifications de liaison des facteurs de transcription, dû à la présence de polymorphismes régulateurs, dans des expériences de retard sur gel	94
3.9 Description de l'étude du gène SKP2	96
3.9.1 Identification des haplotypes régulateurs présents dans l'échantillon représentatif de la population mondiale	96
3.9.2 Analyse fonctionnelle d'expression des rHaps, par expériences de transfections transitoires	98
3.10 Description de l'étude du gène CCND1	100
3.10.1 Identification des haplotypes régulateurs présents dans l'échantillon représentatif de la population mondiale	100
3.10.2 Analyse fonctionnelle d'expression des rHaps, par expériences de transfections transitoires	102
3.11 Description de l'étude du gène E2F1	104
3.11.1 Identification des haplotypes régulateurs présents dans l'échantillon représentatif de la population mondiale	104
3.11.2 Analyse fonctionnelle d'expression des rHaps, par expériences de transfections transitoires	106
3.12 Description de l'étude du gène HDAC1	108
3.12.1 Identification des haplotypes régulateurs présents dans l'échantillon représentatif de la population mondiale	108
3.12.2 Analyse fonctionnelle d'expression des rHaps, par expériences de transfections transitoires	110
3.13 Description de l'étude du gène RB1	112
3.13.1 Identification des haplotypes régulateurs présents dans l'échantillon représentatif de la population mondiale	112



3.13.2 Analyse fonctionnelle d'expression des rHaps, par expériences de transfections transitoires	114
3.14 Description de l'étude du gène CDKN2B	116
3.14.1 Identification des haplotypes régulateurs présents dans l'échantillon représentatif de la population mondiale	116
3.15 Description de l'étude du gène CCNE	118
3.15.1 Identification des haplotypes régulateurs présents dans l'échantillon représentatif de la population mondiale	118
3.16 Description de l'étude du gène CDK4	120
3.16.1 Identification des haplotypes régulateurs présents dans l'échantillon représentatif de la population mondiale	120
3.17 Description de l'étude du gène CDK6	122
3.17.1 Identification des haplotypes régulateurs présents dans l'échantillon représentatif de la population mondiale	122
3.18 Description de l'étude du gène MADH4	124
3.18.1 Identification des haplotypes régulateurs présents dans l'échantillon représentatif de la population mondiale	124
3.19 Description de l'étude du gène CDK2	126
3.19.1 Identification des haplotypes régulateurs présents dans l'échantillon représentatif de la population mondiale	126
 CHAPITRE 4 – DISCUSSION	 131
4.1 Sommaire des resultants	132
4.2 La dérégulation des gènes du cycle cellulaire et leurs causes possibles	133
4.3 Discussion des résultats d'analyses des gènes du point de contrôle G1/S	134
4.3.1 Le gène CDC25A	134
4.3.2 CDKN1A	136
4.3.3 CDKN1B	138
4.3.4 CDKN2A	140
4.3.5 MADH3	141
4.3.6 TFDPI	143
4.3.7 TGFβ1	144
4.4 Limitations de l'étude	147
4.5 Conclusion	149
 BIBLIOGRAPHIE	 150
 ANNEXE I	 I

ix

ANNEXE II

X

ANNEXE III

XI

CURRICULUM VITAE

XIII

## Liste des tableaux

### Résultats

Tableau I	Liste des rSNPs étudiés pour chacun des gènes régulateurs du point de contrôle G1/S	46
Tableau II	Liste des haplotypes régulateurs retrouvés dans l'échantillon de populations de 40 individus pour le gène CDC25A	53
Tableau III	Liste des haplotypes régulateurs trouvés dans l'échantillon de populations de 40 individus pour le gène CDKN1A	61
Tableau IV	Liste des haplotypes régulateurs trouvés dans l'échantillon de populations de 40 individus pour le gène CDKN1B	67
Tableau V	Liste des haplotypes régulateurs trouvés dans l'échantillon de populations de 40 individus pour le gène CDKN2A	73
Tableau VI	Liste des haplotypes régulateurs trouvés dans l'échantillon de populations de 40 individus pour le gène MADH3	79
Tableau VII	des haplotypes régulateurs trouvés dans l'échantillon de populations de 40 individus pour le gène TFDP1	85
Tableau VIII	Liste des haplotypes régulateurs trouvés dans l'échantillon de populations de 40 individus pour le gène TGF $\beta$ 1	91
Tableau IX	Liste des haplotypes régulateurs trouvés dans l'échantillon de populations de 40 individus pour le gène SKP2	97
Tableau X	Liste des haplotypes régulateurs trouvés dans l'échantillon de populations de 40 individus pour le gène CCND1	101
Tableau XI	Liste des haplotypes régulateurs trouvés dans l'échantillon de populations de 40 individus pour le gène E2F1	105
Tableau XII	Liste des haplotypes régulateurs trouvés dans l'échantillon de populations de 40 individus pour le gène HDAC1	109
Tableau XIII	Liste des haplotypes régulateurs trouvés dans l'échantillon de populations de 40 individus pour le gène RB1	113
Tableau XIV	Liste des haplotypes régulateurs trouvés dans l'échantillon de populations de 40 individus pour le gène CDKN2B	117
Tableau XV	Liste des haplotypes régulateurs trouvés dans l'échantillon de populations de 40 individus pour le gène CCNE	119
Tableau XVI	Liste des haplotypes régulateurs trouvés dans l'échantillon de populations de 40 individus pour le gène CDK4	121
Tableau XVII	Liste des haplotypes régulateurs trouvés dans l'échantillon de populations de 40 individus pour le gène CDK6	123
Tableau XVIII	Liste des haplotypes régulateurs trouvés dans l'échantillon de populations de 40 individus pour le gène MADH4	125
Tableau XIX	Tableau récapitulatif des différents rSNPs étudiés par EMSA pour chacun des gènes	127

**Annexes**

Tableau XX	Liste des sondes utilisées dans le génotypage de type ASO	I
Tableau XXI	Liste des amorces utilisées pour l'amplification PCR des fragment 2Kb	X
Tableau XXII	Liste des sondes utilisées lors des expériences de retard sur gel, EMSA	XI

## Liste des figures

### Introduction

Figure 1.1	Les différentes phases du cycle cellulaire	3
Figure 1.2	Régulation de la progression du cycle cellulaire par les différentes cyclines et leurs partenaires cdk ainsi que leurs protéines inhibitrices	6
Figure 1.3	Interaction des protéines qui régulent le point de contrôle G1/S du cycle cellulaire	10
Figure 1.4	Représentation des mutations présentes dans les régulateurs de la transition G1/S associées aux cancers	11
Figure 1.5	Localisation des SNPs	14

### Matériel et Méthodes

Figure 2.1	Génotypage par la technique d'allèle spécifique oligonucléotides (ASO)	26
Figure 2.2	Clonage par la technique Gateway (Invitrogen)	31
Figure 2.3	Construction des haplotypes régulateur par la technique de mutagenèse dirigée	34
Figure 2.4	Technique de retard sur gel	44

### Résultats

Figure 3.1	Résultat d'analyse d'expression des 4 haplotypes régulateurs testés pour le gène CDC25A par transfection transitoire	56
Figure 3.2	Résultat d'analyse de retard sur gel pour le variant -2030T>G du gène CDC25A et intégration de l'ensemble des résultats	59
Figure 3.3	Résultat d'analyse d'expression des 5 haplotypes régulateurs testés pour le gène CDKN1A par transfections transitoires	63
Figure 3.4	Résultat d'analyse de retard sur gel pour le variant -899T>G du gène CDKN1A et intégration de l'ensemble des résultats	65
Figure 3.5	Résultat d'analyse d'expression des 4 haplotypes régulateurs testés pour le gène CDKN1B par transfections transitoires	69
Figure 3.6	Résultat d'analyse de retard sur gel pour le variant -337C>A du gène CDKN1B et intégration de l'ensemble des résultats	71
Figure 3.7	Résultat d'analyse d'expression des 2 haplotypes régulateurs testés pour le gène CDKN2A par transfections transitoires	75
Figure 3.8	Résultat d'analyse de retard sur gel pour le variant -222T>A du gène CDKN2A et intégration de l'ensemble des résultats	77
Figure 3.9	Résultat d'analyse d'expression des 4 haplotypes régulateurs testés pour le gène MADH3 par transfections transitoires	81

Figure 3.10	Résultat d'analyse de retard sur gel pour le variant -1946C>A du gène MADH3 et intégration de l'ensemble des résultats	83
Figure 3.11	Résultat d'analyse d'expression des 3 haplotypes régulateurs testés pour le gène TFDP1 par transfections transitoires	87
Figure 3.12	Résultat d'analyse de retard sur gel pour le variant -1887C>T du gène TFDP1 et intégration de l'ensemble des résultats	89
Figure 3.13	Résultat d'analyse d'expression des 5 haplotypes régulateurs testés pour le gène TGFβ1 par transfections transitoires	93
Figure 3.14	Résultat d'analyse de retard sur gel pour le variant -1886G>A du gène TGFβ1 et intégration de l'ensemble des résultats	95
Figure 3.15	Résultat d'analyse d'expression des 2 haplotypes régulateurs testés pour le gène SKP2 par transfections transitoires	99
Figure 3.16	Résultat d'analyse d'expression des 2 haplotypes régulateurs testés pour le gène CCND1 par transfections transitoires	103
Figure 3.17	Résultat d'analyse d'expression des 3 haplotypes régulateurs testés pour le gène E2F1 par transfections transitoires	107
Figure 3.18	Résultat d'analyse d'expression des 3 haplotypes régulateurs testés pour le gène HDAC1 par transfections transitoires	111
Figure 3.19	Résultat d'analyse d'expression des 3 haplotypes régulateurs testés pour le gène RB1 par transfections transitoires	115

## Liste des abréviations

°C :	degré Celsius
μCi :	micro Curie
μg :	microgramme
μL :	microlitre
μM :	micromolaire
A :	adénine
ADN :	acide désoxyribonucléique
ARNm :	ARN messenger
ASO :	allèle spécifique oligonucléotides
ATP :	adénosine triphosphate
BCL2:	B-cell CLL/lymphoma 2
C :	cytosine
CDKN1A :	cyclin-dependent kinase inhibitor 1A
CDKN1B :	cyclin-dependent kinase inhibitor 1B
CDKN2A :	cyclin-dependent kinase inhibitor 2A
CDKN2B :	cyclin-dependent kinase inhibitor 2B
CCND1 :	cyclin D1
CCNE1 :	cyclin E1
CDC25 :	cell division cycle 25
CDK :	cycline dépendante kinase
CIP1 :	CDK-interacting protein 1
CKI :	inhibiteur de cycline dépendante kinase
CMHI :	complexe majeur d'histocompatibilité de classe I
cSNP :	coding single nucleotide polymorphism
dbSNP :	single nucleotide polymorphism database
dHPLC :	denaturing high performance liquid chromatography
dNTP :	désoxynucléotide triphosphate
DTT :	Dithiothreitol

E2F :	E2F transcription factor
EDTA :	ethylenediamine-tetraacetic acid
EGTA :	ethylene glycol tetra acetic acid
EMSA :	electrophoretic mobility shift assay
fmol :	fentomol
FOXO3a :	forkhead box o3a
G :	guanine
g:	accélération moyenne de la gravité sur Terre
G0 :	stade quiescent de non division
G1 :	gap 1
G2 :	gap 2
gSNP :	genomic single nucleotide polymorphism
HCl :	acide chloridrique
HDAC :	histone déacétylase
HEPES :	N-2-Hydroxyethylpiperazine-N'-2-ethanesulfonic acid
INK4 :	inhibitors of CDK4
iSNP :	intronic single nucleotide polymorphism
kb :	kilobase
KCl :	chlorure de potassium
KIP :	cycline-dependent kinase inhibitor 1B
LB :	Luria-Bertani
LLA :	leucémie lymphoblastique aiguë
M :	mitose
mA :	milliampère
MADH :	mothers against decapentaplegic
MDM2 :	mouse double minute 2 homolog
Mer :	oligomer
MgCl <sub>2</sub> :	chlorure de magnésium
MgSO <sub>4</sub> :	Sulfate de magnésium
min :	minute



mL :	millilitre
mM :	millimolaire
NaCl :	chlorure de sodium
NaH <sub>2</sub> PO <sub>4</sub> :	sodium phosphate
(NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub> :	sulfate d'ammonium
NaOH :	hydroxyde de sodium
P :	probabilité
PAX :	paired box gene
pb :	paire de base
PCR :	polymerase chain reaction
pmol :	picomol
PMSF :	phenylmethylsulphonyl fluoride
R :	point de restriction dans le point de contrôle G1/S
RB :	retinoblastoma
RBP :	mammalian transcriptional repressor RBP-Jkappa/CBF1
rHap :	haplotype régulateur
rpm :	rotation par minute
rSNP :	single nucleotide polymorphism region régulatrice
S :	synthèse de l'ADN
SDS :	Sodium Dodecyl Sulfate
SKP2 :	S-phase kinase-associated protein 2
SMAD :	voir MADH
SNP :	single nucleotide polymorphism
SP1 :	specificity protein 1
T :	thymine
TF :	transcription factor
TFDP1 :	transcription factor DP1
TGFβ :	transforming growth factor beta
T <sub>m</sub> :	melting temperature
U :	unité enzymatique

UV : ultraviolet  
VNTR : variable number of tandem repeats  
WAF1 : wildtype p53-activated fragment 1  
ZBP : zinc binding protein

## Remerciements

Je remercie mon directeur de recherche, Daniel Sinnett, pour m'avoir donné la chance de faire ma maîtrise dans son laboratoire et spécialement pour son aide tout au long de ma maîtrise. Merci pour m'avoir donné la chance d'assister à tous ces congrès.

Je remercie également tous les gens des laboratoires Krajinovic, Labuda et Sinnett, pour leur aide et conseils au cours de ma maîtrise. Je remercie plus particulièrement les gens qui m'ont aidé dans le projet : Patrick, Alexandru, Robert (merci spécial pour le thé!), Mathieu, Manon, Vincent, Nina, Vania.

J'offre un merci spécial à toutes les personnes qui ont su m'apporter un support moral important par leur merveilleux encouragements, ma famille et amis : Johanne, Pierre, Véronique, Steve, Jasmine et Geneviève.

Enfin, je remercie infiniment mon copain Christopher pour sa patience, ses encouragements et son grand support moral au cours des mois d'écriture du mémoire. Ainsi que pour sa précieuse aide du dernier week-end!

## **Chapitre 1**

**-Introduction-**

# 1. Introduction

## 1.1 Introduction au cycle cellulaire.

### 1.1.1 Le déroulement du cycle cellulaire

La prolifération et la division cellulaire se font par l'intermédiaire d'un processus finement régulé : le cycle cellulaire (figure 1.1). Cette régulation fine du cycle cellulaire, empêche une prolifération infinie et incontrôlée des cellules conduisant au développement de maladies complexes, tel que le cancer. La cellule est apte à arrêter la progression du cycle pour permettre la réparation des altérations présentes au niveau de l'ADN. Ainsi, une transmission fidèle de l'information génétique sous forme d'ADN, sans altération, aux cellules filles est assurée. Le cycle cellulaire comporte 4 principales phases et un point de latence (Swanton 2004). La cellule quiescente se trouve au point G0. En présence de conditions favorables et stimulé par des facteurs de croissance, cette cellule entre en phase G1, de préparation de réplication des chromosomes. La phase G1 du cycle cellulaire est suivie des phases de synthèse (S), où l'ADN est répliquée, la phase de préparation de la mitose (G2) et la phase de ségrégation des chromatides soeurs et de croissance cellulaire, mitose (M) (Schmitt and others 2007). La cellule répond aux signaux de prolifération par les facteurs de croissance seulement jusqu'au deux tiers de la phase G1 du cycle cellulaire; passé ce point de restriction et de non retour « R », la cellule devient indépendante de ces signaux et poursuit indépendamment des conditions du milieu, pour compléter la division cellulaire (Giacinti and Giordano 2006; Orford and Scadden 2008; Swanton 2004). Deux points de contrôles (G1/S et G2/M), positionnés dans le cycle cellulaire, assurent l'ordre et la fidélité lors de la progression du cycle en réponse aux altérations possibles de l'ADN et au stress réplicatif (Schmitt and others 2007).

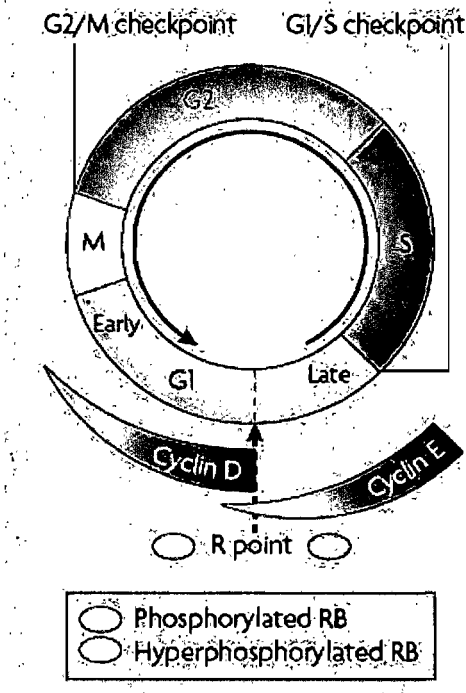


Figure 1.1 : Représentation graphique des différentes phases du cycle cellulaire. Les quatre phases du cycle cellulaire (G1, S, G2, M) ainsi que les deux points de contrôle (G1/S et G2/M) sont montrés dans l'ordre. La concentration des deux principales cyclines (D et E) est montrée en fonction du temps de la gauche vers la droite. La protéine du rétinoblastome (Rb) est montrée sous sa forme phosphorylée (cercle beige) et sous sa forme hyperphosphorylée (cercle orange). Figure tirée de Orford et Scadden 2008.

### **1.1.2 Régulation du cycle cellulaire par les principaux complexes cycline-CDK.**

Le passage d'une cellule à travers les différentes phases du cycle cellulaire est orchestré et régulé par une série de protéines cyclines dépendantes kinases (CDK), dont l'activité spécifique dépend de son association sous forme d'un complexe avec sa sous-unité protéine cycline (Pajalunga and others 2008). La figure 1.2 illustre cette interaction cycline-CDK. La stimulation de la cellule par des facteurs de croissance, induit une accumulation de protéines cycline D au début de la phase G1. La protéine cycline D est à ce moment apte à former des hétérodimères avec les CDK4 et CDK6. Ces complexes actifs vont aller débiter la phosphorylation de la protéine du rétinoblastome (Rb). La protéine Rb phosphorylée commence alors à libérer le facteur de transcription E2F, présent sous la forme d'un complexe avec la protéine TFDP1. Rb phosphorylé libère également les complexes cyclines E-CDK2. Le facteur de transcription E2F libre augmente la transcription de la cycline E et stabilise la protéine cycline E. Cette dernière peut alors former plus de complexes avec la protéine CDK2 et amplifier la phosphorylation de la protéine Rb, qui devient alors hyperphosphorylée. La progression de la cellule à travers le cycle cellulaire devient alors indépendante de la stimulation par les facteurs de croissance, ce qui caractérise la fin de la phase G1. La cycline E et le facteur de transcription E2F établissent alors ensemble les conditions nécessaires pour l'entrée en phase S (Donovan and Slingerland 2000; Orford and Scadden 2008; Pajalunga and others 2008).

En fin de phase G1 et en début de phase S, la protéine CDK2 se trouve de plus en plus liée à son second partenaire, la cycline A, puisque le niveau de cycline E diminue, alors que le niveau de la cycline A augmente. Par la suite, le passage en phase M se caractérise au point G2 par une diminution de la protéine CDK2 libre, à ce moment la cycline A se lie à la CDK1. Le niveau de la cycline B augmente également à ce point. La phase M est contrôlée par la cycline B qui se lie à son partenaire la CDK1 et l'active. La fluctuation de la disponibilité des différentes protéines cyclines et CDK, caractérise le déroulement du cycle cellulaire (Giacinti and Giordano 2006; Schmitt and others 2007).

### 1.1.3 La régulation des protéines cyclines et CDK.

La régulation des protéines cyclines dépendantes kinases se fait principalement de façon post-traductionnelle, car le niveau de protéines demeure sensiblement le même au cours du cycle cellulaire. L'activation des CDK est produite en deux étapes : d'abord une déphosphorylation des résidus thréonines et tyrosines situés à l'extrémité N-terminale par la phosphatase CDC25 et ensuite par une phosphorylation d'un résidu thréonine par le complexe cycline H-CDK7. Également, des familles de protéines inhibitrices des CDK permettent d'inactiver ces protéines lorsqu'une anomalie survient au cours du cycle cellulaire (figure 1.2). Une première famille d'inhibiteurs de cyclines dépendantes kinases est la famille INK4. Cette famille de protéines contient les protéines inhibitrices CDKN2A (p16<sup>INK4a</sup>), CDKN2B (p15<sup>INK4b</sup>), p18<sup>INK4c</sup> et p19<sup>INK4d</sup>. Ces inhibiteurs interagissent avec les protéines CDK4 et CDK6 et bloquent leurs activités de phosphorylation de la protéine Rb. Une deuxième famille importante d'inhibiteurs de CDK est la famille KIP. Cette famille contient les protéines inhibitrices CDKN1A (p21<sup>Cip1/Waf1</sup>), CDKN1B (p27<sup>Kip1</sup>) et p57<sup>Kip2</sup>; et contrairement à la famille des p16<sup>INK4a</sup>, ce groupe d'inhibiteurs peut se lier à la fois à la sous unité cycline ou à la CDK du complexe, pour l'inhiber. La concentration de la protéine CDKN1B augmente en présence d'une stimulation de la cellule par le facteur TGFβ1, mais décroît en présence des facteurs de croissance. La famille d'inhibiteurs p21<sup>Cip1/Waf1</sup> se lie préférentiellement au complexe cycline E-CDK2 ou cycline A-CDK2 pour contrôler leurs activités, mais peuvent également réguler l'activité de la cycline D (Donovan and Slingerland 2000; Swanton 2004).



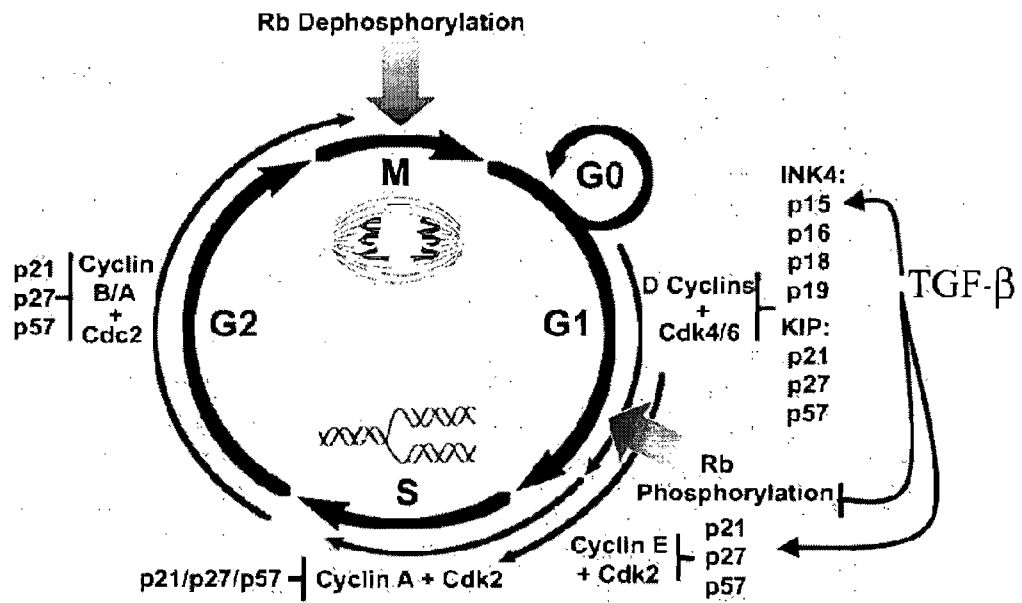


Figure 1.2 : Représentation graphique de la régulation de la progression du cycle cellulaire par les différentes cyclines et leurs partenaires CDK ainsi que leurs protéines inhibitrices. Voir le texte pour les explications détaillées. Figure tirée de (Donovan and Slingerland 2000).

## **1.2 Le point de contrôle G1/S.**

### **1.2.1 Fonctionnement du point de contrôle G1/S.**

Un point de contrôle est défini comme un processus biologique qui permet de ralentir ou arrêter la progression du cycle cellulaire lorsqu'une anomalie est détectée au niveau de l'ADN (Sancar and others 2004). Dans la cellule quiescente en phase G<sub>0</sub>, la cycline D est présente sous une forme instable, sa concentration est maintenue à un niveau faible par protéolyse et son activité est bloquée par la protéine inhibitrice CDKN2a. De plus, la protéine Rb qui se trouve sous une forme hypophosphorylée, recrute la protéine HDAC (une histone désacétylase) pour inhiber l'activité du facteur de transcription E2F. Suite à la stimulation de la cellule par des facteurs de croissance, une cascade de signalisation induit l'accumulation de la cycline D, ce qui permet une augmentation de la formation des complexes cycline D-CDK4 et cycline D-CDK6 (figure 1.3). Ces complexes actifs peuvent alors exercer leur rôle de kinase et débiter la phosphorylation de la protéine Rb. Cette dernière, sous sa forme phosphorylée commence à libérer des facteurs de transcription E2F qui sont présents sous forme d'un complexe avec le facteur TFDPI, ce qui permet de débiter la transcription des gènes nécessaires à la phase S du cycle cellulaire, en plus de la transcription de la protéine cycline E. De plus, la protéine Rb phosphorylée n'est plus apte à se lier à la protéine SKP2, ce qui permet à SKP2 de se lier à la protéine inhibitrice CDKN1B et d'engendrer sa dégradation par ubiquitination. Le niveau de protéine CDKN1B est alors diminué, ce qui permet d'augmenter l'activité des CDK4 et CDK6, mais particulièrement des CDK2. Le fait d'augmenter le niveau de cycline E produite et l'activité de CDK2, permet la formation d'une plus grande quantité de complexes cycline E-CDK2 actifs. Ces complexes cycline E-CDK2 vont alors agir en collaboration avec les complexes cycline D-CDK4 et cycline D-CDK6 pour augmenter la phosphorylation de la protéine Rb. De plus, le complexe cycline E-CDK2 pourra aller phosphoryler son inhibiteur, la protéine CDKN1B, pour induire sa dégradation par ubiquitination. La protéine

Rb hyperphosphorylée libérera totalement le facteur de transcription E2F, se qui entraîne la protéine HDAC, recrutée par Rb, à libérer également le facteur E2F. La transcription des gènes nécessaires à la phase S est alors produite librement par le complexe de transcription E2F-TFDP1. La dimérisation de E2F avec la protéine TFDP1 est nécessaire pour permettre une liaison de haute affinité du facteur E2F avec les promoteurs correspondants (Molinari 2000).

### **1.2.2 Arrêt du cycle cellulaire au point de contrôle G1/S**

Lorsqu'une anomalie est détectée au niveau de l'ADN, des signaux antiprolifératifs sont portés à la cellule. La molécule TGF $\beta$ 1 est un inhibiteur de la prolifération cellulaire qui peut être sécrété lorsqu'une anomalie est détectée. La molécule TGF $\beta$ 1 active ses récepteurs à la surface de la cellule, qui activent à leur tour un mécanisme de signalisation intra-cellulaire Smad2/Smad4 pour permettre la transmission du signal jusqu'au noyau de la cellule. La figure 1.3 illustre le processus par lequel la molécule TGF $\beta$ 1 arrête la progression du cycle cellulaire au point de contrôle G1/S. La stimulation de la cellule par TGF $\beta$ 1 entraîne immédiatement une diminution de la concentration et de l'activité de la protéine CDC25 et une augmentation de l'activité de la protéine inhibitrice CDKN2B ainsi que sa stabilisation. Les protéines CDK ne peuvent alors plus être activées par la phosphatase CDC25. De plus, l'inhibiteur CDKN2B se lie aux protéines kinases CDK4 et CDK6 pour bloquer leurs activités. Les cyclines D et E sont également soumises à une diminution de leur concentration. La protéine cycline D se dissocie des CDK4 et CDK6. Une accumulation de l'inhibiteur CDKN1B permet son association avec le complexe cycline E-CDK2. Puisque les complexes cycline-CDK sont maintenant désactivés et que la haute concentration des inhibiteurs de CDK maintient cette inactivation, la phosphorylation de la protéine Rb ne peut avoir lieu correctement pour produire sa forme hyperphosphorylée. Le facteur de transcription E2F demeure lié et inhibé par la protéine Rb, ce qui empêche la transcription des gènes de la phase S du cycle cellulaire. Les effets

de la stimulation de la cellule par la molécule TGF $\beta$ 1 entraîne un arrêt du cycle cellulaire au point de contrôle G1/S (Donovan and Slingerland 2000).

Les études effectuées au cours des dernières années ont démontrées l'importance des mutations trouvées dans les gènes qui régulent le point de contrôle G1/S. En effet, les gènes codants pour des éléments régulateurs de la transition G1/S sont les plus souvent impliqués dans la formation de cancers (Sherr 2000). Les types d'altérations retrouvés sont, une surexpression de cyclines (D et E) et des CDK (CDK4 et CDK6), ou la perte de l'expression des inhibiteurs de CDK (CDKN1A, CDKN1B, CDKN2A et CDKN2B) et de la protéine Rb (Malumbres and Barbacid 2001). La modification du niveau d'expression de ces régulateurs est souvent associée à une altération chromosomique (amplification de la cycline D et CDK4, translocation de CDK6, délétion des CDKN2A et CDKN2B et de la protéine Rb) (Malumbres and Barbacid 2001). Aussi, des mutations présentes au niveau de la région codante des gènes CDK4 et CDK6 contribuent empêcher leur liaison avec les inhibiteurs de CDK, ce qui supprime l'arrêt de la progression du cycle cellulaire au point de contrôle G1/S (Easton and others 1998; Wolfel and others 1995). Également, la protéine SKP2 qui possède des propriétés oncogéniques, montre un niveau d'expression altéré dans certaines tumeurs (Gstaiger and others 2001; Latres and others 2001). La figure 1.4, résume l'ensemble des cancers humains connus, dont des gènes régulateurs du point de contrôle G1/S sont impliqués dans leurs formations.

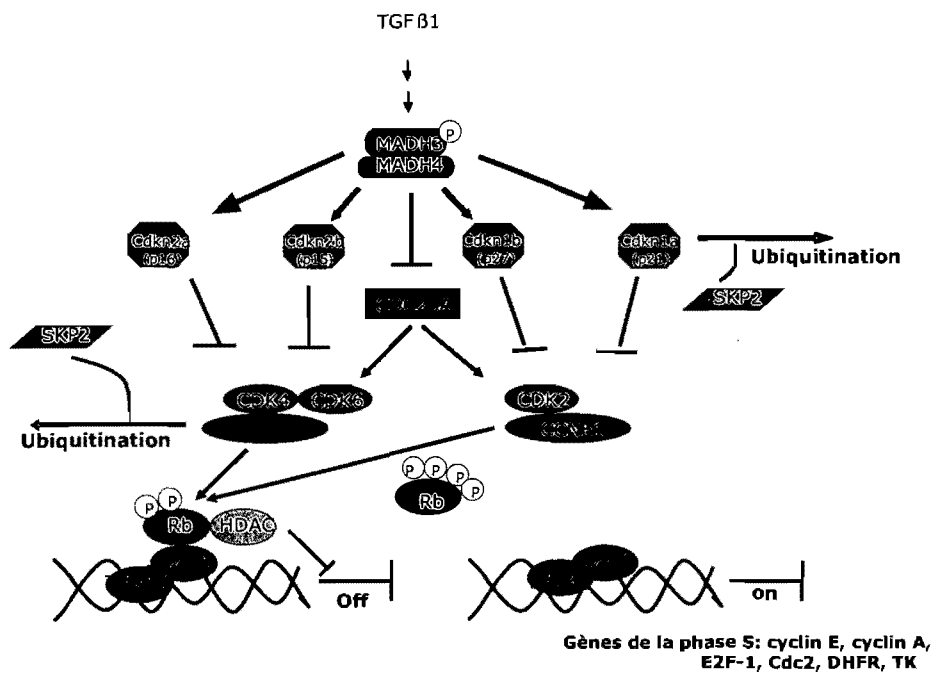


Figure 1.3 : Représentation graphique de l'interaction des protéines qui régulent le point de contrôle G1/S du cycle cellulaire. Voir le texte pour la description détaillée.

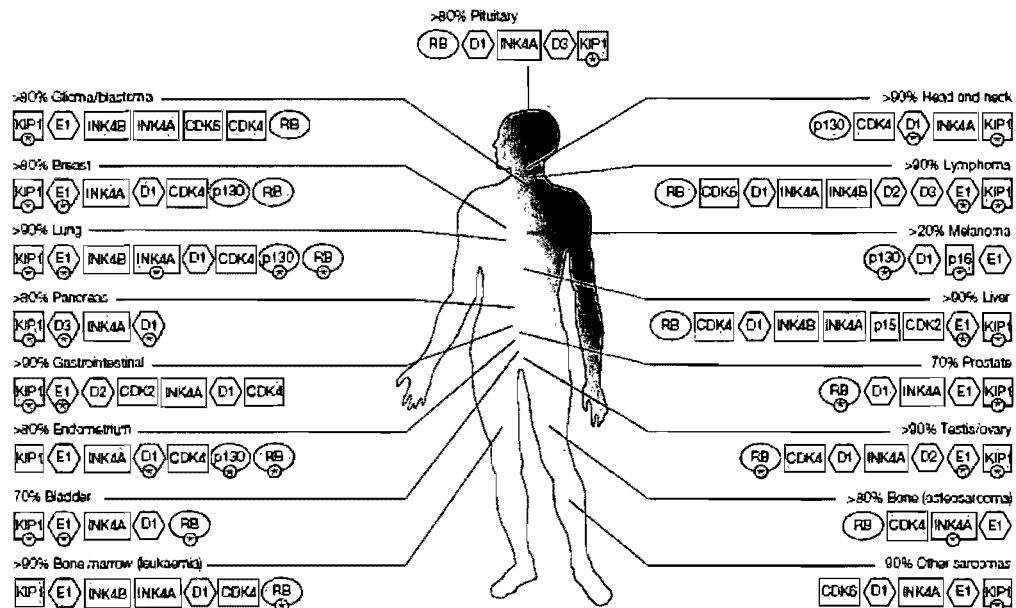


Figure 1.4 : Représentation des mutations présentes dans les régulateurs de la transition G1/S associées aux cancers (Malumbres and Barbacid 2001). Seuls les altérations présentant une fréquence supérieure à 10% dans les tumeurs primaires ont été considérées. Les chiffres indiquent le pourcentage de tumeurs possédant des altérations pour un régulateur du cycle cellulaire donné. Les endroits où la cause génétique ou épigénétique a été identifiée, sont montrés en rose. Les altérations pour lesquelles aucune explication mécanistique n'a été proposée sont en jaune. Les altérations pertinentes dans le pronostic du cancer sont indiquées par une astérisque.

### **1.3 Les polymorphismes**

#### **1.3.1 Les types de polymorphismes**

La définition d'un polymorphisme génétique selon le *Glossaire de génétique médicale et moléculaire* (Dallaire 2006) est : une variation entre individus dans la séquence de gènes. Ces variations qui rendent compte des différents allèles dans une population sont normales et ne sont pas nécessairement pathogènes.

Il existe plusieurs types de polymorphismes, dont les polymorphismes de longueur et les polymorphismes simples. Les polymorphismes de longueur impliquent des délétions ou insertions de plusieurs nucléotides et les séquences répétées en tandem, VNTR (Variations in the Number of Tandem Repeats) (Vignal and others 2002). Les séquences répétées en tandem se distinguent par la longueur du motif de la répétition, comme suit : les satellites (Csink and Henikoff 1998; Mashkova and others 2001), les minisatellites (Jarman and Wells 1989; Jeffreys and others 1985) et les microsatellites (Karlin and Burge 1995). Les SNPs (Single Nucleotide Polymorphism) sont les polymorphismes simples. Ce sont des changements d'un seul nucléotide qui sont très fréquents dans le génome, leur nombre est estimé à au moins 4 millions, c'est-à-dire environ 1 SNP à tous les 500 pb le long du génome (Dallaire 2006).

#### **1.3.2 Les SNPs; la modification des nucléotides et l'impact de leur localisation**

Les SNPs sont issus des changements de nucléotides : insertions, délétions ou des substitutions. Les substitutions de nucléotides sont de deux types : des transitions qui consistent en un changement entre deux purines ou deux pyrimidines; ou des transversions qui sont des changements entre une purine et une pyrimidine.

Dépendamment de leur localisation dans le génome et du changement engendré, le SNP peut avoir un impact différent (figure 1.5). Les SNPs peuvent être localisés dans la région codante d'un gène (cSNP). À ce moment, le SNP peut produire une modification dite silencieuse, qui n'affecte pas le code d'acides aminés de la protéine. Par contre, il peut également produire une mutation faux-sens ou non-sens, qui a un impact fonctionnel sur la protéine produite. La mutation faux-sens est obtenue lorsque le nucléotide muté provoque un changement de codon et que ce nouveau codon code pour un autre acide aminé. Cette mutation modifie la protéine d'un acide aminé et peut ainsi altérer sa fonction, en modifiant par exemple son repliement, un domaine de liaison, un domaine d'activation ou autre. La mutation non-sens survient lorsque le nucléotide substitué transforme le codon normal en un codon d'arrêt; ou lorsqu'une insertion ou délétion d'un nucléotide est produite, alors le cadre de lecture de 3 nucléotides est décalé et ceci résulte pour la plupart des fois en un arrêt prématuré de la synthèse de la protéine (Campbell 1995). Les SNPs situés dans l'intron (iSNP) de la protéine peuvent également avoir un impact fonctionnel sur la protéine. Les iSNPs peuvent modifier l'épissage du transcrit primaire de la protéine et ainsi produire une protéine tronquée, non fonctionnelle (Ye and others 2008). Les gSNPs situés dans la séquence génomique non codante d'un gène n'ont aucun effet sur la protéine. Finalement les rSNPs localisés dans la partie promotrice du gène (région régulatrice) peuvent avoir un impact fonctionnel sur le niveau d'expression du gène, donc le niveau de transcrit de la protéine, en affectant les sites de liaisons des facteurs de transcription.



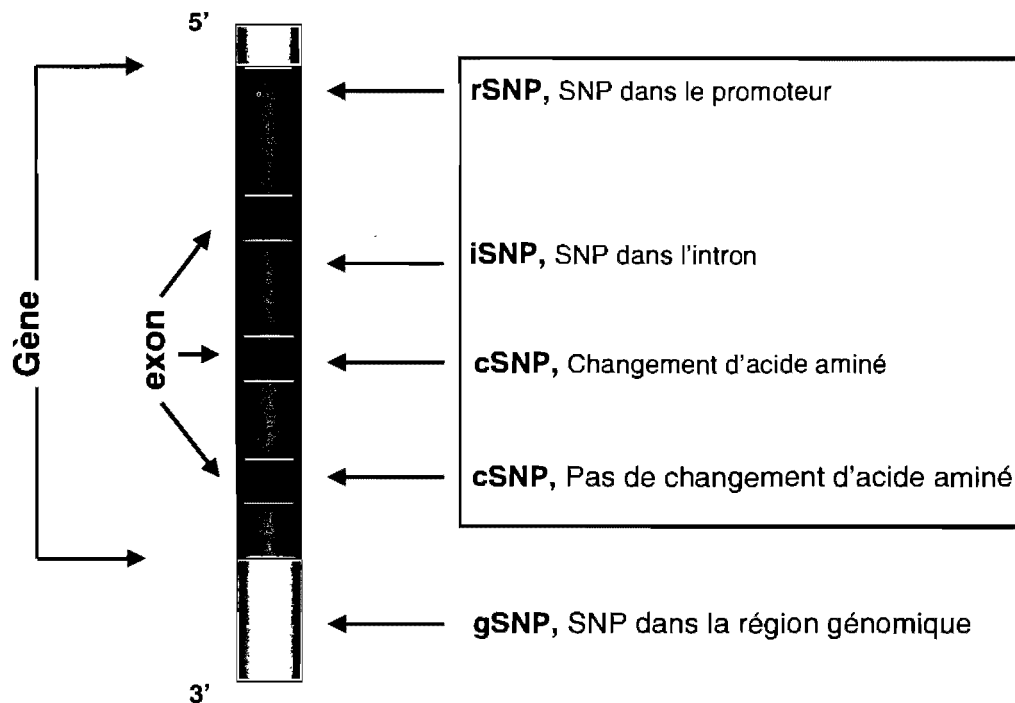


Figure 1.5 : Localisation des SNPs. rSNP, SNP dans la région régulatrice; iSNP, SNP dans la région intronique; cSNP, SNP dans la région codante; gSNP, SNP dans la région génomique. Les SNPs qui sont dans un encadré peuvent mener à un changement fonctionnel.

### **1.3.3 Les polymorphismes régulateurs impliqués dans diverses maladies**

Les rSNPs qui démontrent un impact fonctionnel sont souvent impliqués dans des maladies. Par exemple, une équipe de chercheurs en Allemagne a démontré que le rSNP -938C>A produit une expression différentielle du gène BCL2, qui possède un rôle antiprolifératif et antiapoptotique, la présence de l'allèle A est associée à un effet protecteur sur le cancer du sein (Bachmann and others 2007). Également, une association a été démontrée entre une augmentation du risque de développer une tumeur et la présence du génotype 309GG (rSNP -309T>G) du gène MDM2, qui est un important régulateur de la réponse antitumorale médiée par la protéine p53 (Bougeard and others 2006; Hu and others 2007). Ce n'est que quelques exemples de toutes les maladies auxquelles sont reliées des polymorphismes génétiques.

## **1.4 La variabilité génétique interindividuelle.**

### **1.4.1 Définition**

Une grande variabilité génétique est retrouvée entre les individus, mais cette variabilité est d'autant plus grande entre les différentes populations dans le monde. Cette variabilité est causée principalement par la présence de nombreux polymorphismes à travers tout le génome. Un groupe de scientifiques a analysé 602 gènes de 7 individus, hétérozygotes pour des SNPs contenus dans la région transcrite des gènes, 54% de ces gènes expriment un allèle de façon préférentielle (Lo and others 2003). Ceci démontre que les SNPs sont responsables en grande partie de la variabilité génétique interindividuelle. Cette variabilité produit une susceptibilité propre à chaque individu face à certaines maladies, tel que mentionné précédemment. Également une des grandes préoccupations de l'industrie

pharmaceutique est la façon propre à chaque individu de métaboliser les médicaments et de réagir de façons différentes à ces médicaments. Tous ces effets, propres à chaque individu sont attribués aussi à la variabilité génétique interindividuelle. De nombreuses études de pharmacogénétique sont faites dans le but d'analyser la variabilité génétique en lien avec la façon de métaboliser les médicaments (Oertel and Lotsch 2008; Wilkins and others 2008) (incluant de façon importante, les réponses aux traitements du cancer) (Ansari and others 2007) et également en lien avec la susceptibilité à diverses maladies (Lazarus and others 2002; Sommer and others 2008).

#### **1.4.2 Les rSNPs impliqués dans la variabilité interindividuelle**

Les SNPs positionnés dans la région régulatrice d'un gène peuvent influencer la liaison de facteurs de transcription et ainsi affecter le niveau de transcrits de la protéine (Pastinen and Hudson 2004). La présence de polymorphismes dans le promoteur peut soit produire un gain ou une perte d'un site de liaison d'un facteur de transcription. Cette modification d'un site de liaison explique la différence du niveau de transcrits de la protéine. Les rSNPs pourraient expliquer au moins 25 à 35% de la variabilité interindividuelle (Pastinen and Hudson 2004). La plupart des études réalisées jusqu'à maintenant ont analysées les mutations contenues dans la partie codante de la protéine qui pourraient altérer la fonction de cette protéine (Mitchison 2001). Récemment, l'étude des SNPs contenus dans la région promotrice des gènes tend à gagner de l'importance grâce à leurs impacts connus sur le niveau d'expression des gènes (Gazquez and others 2008; Juhanson and others 2008; Ober and others 2008). Le niveau d'expression d'un gène régule la quantité de transcrits primaires et alors le niveau de protéines traduites et disponibles dans la cellule. Par conséquent, les gènes contenant des polymorphismes régulateurs (rSNPs) dans leur région promotrice et étant impliqués dans la régulation des processus biologiques de haute importance, pourraient contribuer au développement de maladies complexes, tel que le cancer. La susceptibilité interindividuelle aux maladies complexes pourrait être augmentée chez les individus qui possèdent des rSNPs dans des gènes codant pour des protéines de

régulation. Certaines études démontrent effectivement un impact significatif des rSNPs dans la susceptibilité au cancer en affectant des protéines impliquées dans le métabolisme des carcinogènes (Sinnott and others 2006b), des protéines régulant le cycle cellulaire (Healy and others 2007) et autres protéines impliquées dans la progression du cancer, ainsi que la réponse aux médicaments contre le cancer (Milani and others 2007).

### 1.5 Hypothèse

Tel que démontré, les polymorphismes sont retrouvés dans la séquence du génome humain à une très haute fréquence. Ces polymorphismes causent une grande variabilité interindividuelle, qu'ils soient localisés dans la région codante d'un gène (cSNP) ou dans la région promotrice (rSNP). Lorsque ces SNPs affectent des protéines impliquées dans des processus biologiques importants et finement régulés, ils influencent la susceptibilité interindividuelle aux maladies complexes, tel que le cancer. **Nous proposons que des polymorphismes fonctionnels, localisés dans la région régulatrice (rSNPs) des gènes impliqués dans le point de contrôle G1/S, auraient un impact significatif dans l'oncogenèse.** Les rSNPs, localisés dans la région promotrice des gènes, pourraient influencer le niveau de transcrits des protéines régulatrices du point de contrôle G1/S du cycle cellulaire, en modifiant les sites de liaisons des facteurs de transcription.

## 1.6 Objectifs du projet

Les objectifs spécifiques du projet sont:

1. D'évaluer l'impact fonctionnel de 135 variants situés dans la région régulatrice de 2Kb de 18 gènes impliqués dans la régulation du point de contrôle G1/S du cycle cellulaire.
2. Établir par prédictions informatiques, *in silico*, les facteurs de transcription dont le site de liaison est modifié.
3. Valider l'impact fonctionnel des rSNPs à l'aide d'expériences *in vitro* :
  - a. Évaluer l'effet des rSNPs sur la transcription de gènes rapporteurs dans différentes lignées cellulaires.
  - b. Vérifier la présence de liaisons différentielles de facteurs de transcription sur les rSNPs qui démontrent une différence significative d'expression, à l'aide d'expériences de retard sur gel (EMSA).

## **Chapitre 2**

### **-Matériel et Méthodes-**

## **2. Matériel et méthodes**

### **2.1 Identification des SNPs dans la région régulatrice des gènes candidats du point de contrôle G1/S du cycle cellulaire.**

#### **2.1.1 Analyse des sites de liaison des facteurs de transcription.**

Des prédictions informatiques sur toute la région promotrice de 2Kb testée ont été produites à l'aide du programme MatInspector (Genomatix Software GmbH, version 7.0). Pour chacun des polymorphismes étudiés, une portion de 50 pb entourant le SNP a été envoyée à l'analyse. Ce programme a permis de déterminer les sites de liaison des facteurs de transcription connus aux endroits où les SNPs sont présents, par l'intermédiaire de la base de données TRANSFAC (Wingender 2008). Le programme permet également de prévoir les sites perdus, ainsi que les sites créés par la présence d'un polymorphisme. En reconnaissant les sites de liaison présents, les facteurs de transcription qui leur sont liés sont identifiés automatiquement et associés à un score de probabilité de liaison. Uniquement les facteurs possédant un score de similarité optimal (tel que défini par le programme MatInspector) de leur site de liaison, comparativement à la séquence du gène donné, sont incorporés à notre base de données.

### **2.2 Identification des haplotypes régulateurs (rHap).**

#### **2.2.1 Population à l'étude.**

Un échantillon représentatif de la population mondiale a été utilisé pour l'étude des gènes candidats. Cet échantillon est composé de 40 individus, soit un total de 80 chromosomes. Les individus sont répartis selon l'origine ethnique de la façon suivante : 8 Africains (afro-américain et africain sub-saharien), 8 Européens, 8 Amérindiens, 8 Asiatiques, 8 Moyen-



Orientaux. Cet échantillon a servi à détecter la fréquence des polymorphismes à l'étude (préalablement identifiés (Sinnott and others 2006a)), ainsi qu'à la construction des haplotypes présents. Ces polymorphismes sont nommés selon leur position relative au site d'initiation de la transcription +1.

### **2.2.2 Génotypage :**

L'analyse du génotype des individus a été réalisée par la technique d'allèle spécifique oligonucléotides (ASO) (Labuda and others 1999). Sous cette approche, les fragments d'ADN des individus possédant les polymorphismes à l'étude, sont amplifiés par une réaction PCR et transférés sur une membrane de nylon. Ensuite on procède à une hybridation à l'allèle spécifique de l'individu, de sondes oligonucléotidiques radioactives complémentaires (figure 2.1). Un mauvais appariement crée une instabilité. Cette instabilité permet, grâce à des lavages à température près du  $T_m$  de la sonde marquée, de garder associé uniquement les sondes qui ont trouvé un appariement parfait.

### **Amplification PCR.**

Pour chacun des individus de l'échantillon, des fragments de 150pb à 300pb sont amplifiés par PCR, de façon à couvrir toute la portion de 2Kb à tester. Les 30 cycles PCR possèdent plusieurs étapes, soit l'étape de dénaturation de l'ADN qui se fait à une température de 95°C pendant 45 secondes, l'appariement de l'oligonucléotide qui varie entre 50°C à 60°C selon le  $T_m$  de l'amorce pendant 45 secondes et l'élongation des brins d'ADN à 68°C pendant 1 minute. Le tout, est précédé d'une étape de dénaturation de l'ADN et d'activation de l'enzyme à une température de 95°C pendant 5 minutes et se termine par un temps d'élongation supplémentaire de 8 minutes à 68°C. Le mélange réactionnel est composé de 1x de tampon sans  $MgCl_2$  (500 mM Tris/HCl, 100 mM KCl, 50 mM  $(NH_4)_2SO_4$ ) (Roche), 0,75mM de  $MgCl_2$  (Roche), 0,4 $\mu$ M de chacune des amorces (IDT), 0,1mM de dNTP, 1,5 U

de l'enzyme Taq Platinum (Invitrogen), 20 ng d'ADN génomique, dans un volume total de 50 $\mu$ L.

### **Dot Blot.**

Les fragments PCR produits sont ensuite dénaturés dans une solution de dénaturation (NaOH 0,4 N, EDTA 20mM, Orange II dye 0,02%). Un volume de 170  $\mu$ L de solution dénaturante est ajouté directement au 50 $\mu$ L du produit PCR pour chaque fragment et laissé à la température pièce pendant environ 5 minutes. Cette étape permet aux brins d'ADN amplifiés de se séparer pour ensuite s'attacher à la membrane de nylon sous la forme simple brin. Des aliquots de 100 $\mu$ L des mélanges sont par la suite transférés en duplicata sur une membrane de nylon Hybond-N+ (Amersham) par aspiration sous vide à l'aide de l'appareil Minifold I (Schleicher et Schuell). L'ADN est fixé à la membrane par radiation UV : rayonnement ultraviolet à 254nm dans le four UV stratalinker 1800 (Stratagene). Pour éviter leur dégradation avec le temps, les membranes sont rincées à l'eau afin d'éliminer le NaOH restant de la solution dénaturante.

### **Marquage des sondes.**

Les sondes spécifiques à chaque allèle sont marquées par radioactivité ( $\gamma$ -ATP<sup>32</sup>). Les oligonucléotides synthétiques (compagnie IDT) possèdent une longueur totale de 20 nucléotides (20 mers), soit environ 10 mers de part et d'autre du variant à tester. Le mélange réactionnel contient 1,67 $\mu$ M de la sonde, 1x du tampon Kinase *Forward* (350mM Tris-HCl, 50mM MgCl<sub>2</sub>, 500mM KCl, 5mM 2-mercaptoéthanol) (Invitrogen), 10U de l'enzyme T4 Kinase (Invitrogen), 10 $\mu$ Ci de  $\gamma$ -ATP<sup>32</sup>. Le tout est incubé à 37°C pendant 1 heure, puis 15  $\mu$ L de EDTA à 10mM est ajouté pour arrêter la réaction. Les sondes de génotypages sont décrites au tableau x, annexe I.

### Hybridation des membranes.

Pour géotyper un variant, les deux membranes de nylon préalablement préparées, contenant le fragment PCR approprié, sont placées dans des tubes à hybridation distincts. Une solution d'hybridation (SSPE 1x (150 mM NaCl, 10 mM NaH<sub>2</sub>PO<sub>4</sub> et 1mM EDTA), NaCl 0,75mM, SDS 1%, Héparine 100mg, Tris 70mM) à température pièce est ajoutée aux tubes et ces derniers sont placés dans le four à hybridation (*Isotherm*, Fisher Scientific). La température du four est ajustée à la plus haute température du gradient choisi et maintenue pendant 15 minutes. Cette étape de pré-hybridation permet d'amener les membranes à la température supérieure du gradient avant d'ajouter les sondes, afin d'empêcher que celles-ci se collent non-spécifiquement à la membrane. Dans un volume de 20 mL de solution d'hybridation, 5µL de la sonde marquée radioactive et 15 fois plus du compétiteur non-marqué (ratio 15 :1) sont ajoutés. Le compétiteur constitue la sonde froide spécifique à l'autre allèle du variant étudié. Cet excès de compétiteur froid permet de saturer les cibles d'ADN sur la membrane afin de produire une compétition adéquate à la sonde marquée et une association spécifique de celle-ci (Bourgeois and Labuda 2004). Lorsque la sonde et son compétiteur sont ajoutés, les tubes sont placés dans le four à hybridation à une vitesse de rotation moyenne, pour une période d'incubation de 30 minutes. Les tubes sont par la suite soumis à une diminution de température selon le gradient décidé. Le gradient de température est calculé en fonction de la température spécifique d'hybridation des sondes. Le gradient comprend un interval d'environ 10 degrés, soit 5°C au-dessus et en-dessous du T<sub>m</sub> de la sonde. Lorsque la température minimale est atteinte, les tubes sont incubés à cette température pour une durée de 15 minutes. Puis, les membranes sont rincées 1 fois à la température de la pièce avec une solution de lavage, contenant 2x de SSPE et 0,1% de SDS. Ceci permet de retirer l'excédent de sondes non-attachées. Ensuite, on procède à un deuxième rinçage pendant 10 minutes à la température la plus basse du gradient. Ce deuxième rinçage permet de retirer les sondes attachées faiblement suite à une association non-spécifique. Les membranes sont transférées dans un sac de plastique, le sac est ensuite scellé et placé dans une cassette avec un flim X-ray, Super RX (Kodak) entre 2 écrans intensificateurs pour l'exposition du film à la radioactivité. La cassette est placée dans un

congélateur à une température de  $-80^{\circ}\text{C}$  pendant la nuit. Le lendemain, le film est développé afin de révéler le génotype de chaque échantillon testé. Les membranes auxquelles sont attachées les fragments PCR peuvent être réutilisées suite à un décrochage des sondes radioactives. Le détachement des sondes aux membranes est fait par trempage dans une solution bouillante de SDS 1% laissée à la température ambiante pendant 2 heures.

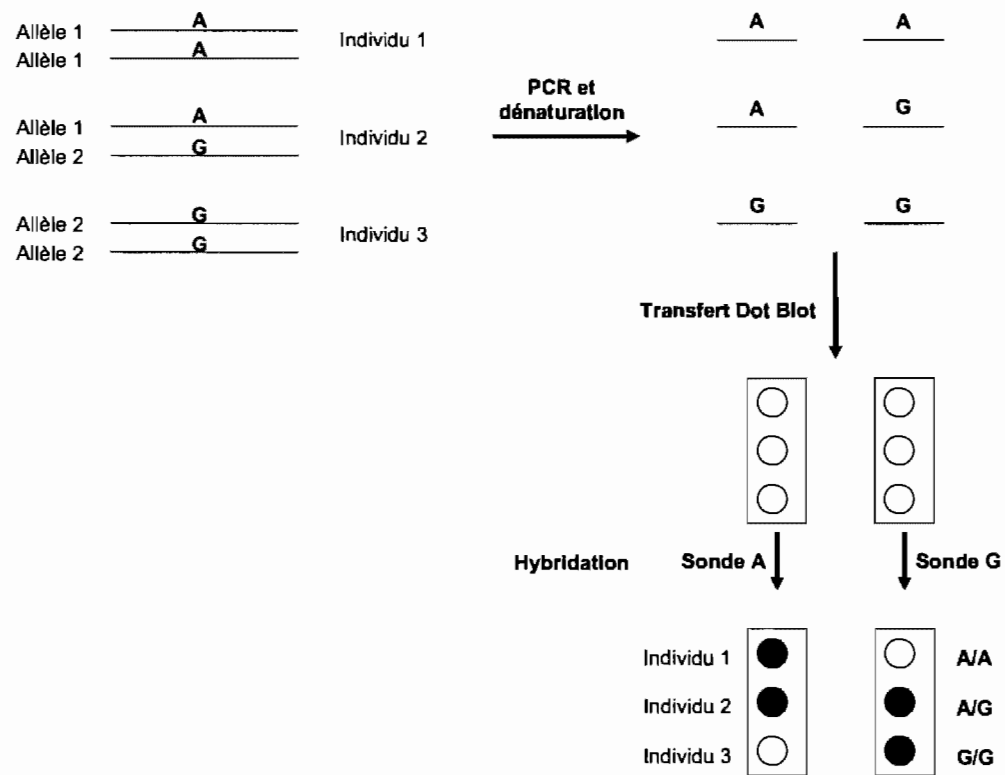


Figure 2.1 : Représentation graphique du génotypage par la technique d'allèle spécifique oligonucléotides (ASO).

### **2.2.3 Construction des haplotypes.**

L'ensemble des rSNPs génotypés dans notre échantillon de population sera étudié sous la forme d'haplotypes régulateurs. Les données de génotypage obtenues ci-dessus sont utilisées pour construire les haplotypes régulateurs. La construction des haplotypes se fait à l'aide du logiciel PHASE, version 2.1 (Université de Washington, Seattle, WA). Les données de génotypage de chacun des variants étudiés pour un gène donné sont entrées dans le logiciel, qui permet ensuite d'identifier les haplotypes présents dans notre échantillon (représentatif de la population mondiale) ainsi que leurs fréquences. L'impact fonctionnel des haplotypes régulateurs fréquents sera déterminé à l'aide d'essais *in vitro*.

## **2.3 Validation fonctionnelle *in vitro*.**

### **2.3.1 Construction des vecteurs.**

Le promoteur proximal est amplifié sous forme d'haplotypes par PCR à partir de l'ADN génomique des individus de l'échantillon. Cette amplification est produite à l'aide d'amorces qui contiennent une partie spécialement conçue pour s'apparier au vecteur de départ du System Gateway (Invitrogen). Le system Gateway permet un clonage rapide et efficace par l'intermédiaire de réactions enzymatiques, sans utilisation d'enzymes de restriction et de ligase (figure 2.2). Le fragment PCR (2 Kb) est inséré par une première réaction enzymatique dans un vecteur d'entrée. Puis ce vecteur est soumis à une seconde réaction enzymatique pour transférer le fragment PCR, dans la bonne orientation, dans le vecteur de destination (par exemple, pour un système gène rapporteur).

### **Amplification PCR.**

L'amplification d'ADN est faite à partir de l'ADN génomique des individus qui possèdent l'haplotype à tester pour un gène donné. L'amplification PCR destinée au clonage Gateway (Invitrogen) est produite en deux réactions PCR distinctes. Lors de la première amplification des amorces, contenant chacune 12 nucléotides complémentaires aux bornes de recombinaison du vecteur d'entrée pDONR Gateway, sont utilisées. Ces séries de nucléotides, appelés queue attB1 et queue attB2, sont respectivement placés sur l'amorce sens et antisens. Les fragments 2 Kb de la région promotrice sont amplifiés par une première réaction PCR spécifique de 40 cycles : dénaturation de l'ADN à une température de 95°C pendant 45 secondes, appariement des amorces (entre 50°C et 60°C selon le T<sub>m</sub> de l'amorce) pendant 45 secondes et élongation des brins d'ADN à une température de 68°C pendant 2 minutes. Les 40 cycles sont précédés, d'un temps de dénaturation de l'ADN et d'activation de l'enzyme à une température de 95°C pendant 10 minutes et se terminent par un temps d'élongation supplémentaire de 8 minutes à 68°C. Le mélange réactionnel PCR est composé de 1x de tampon sans MgCl<sub>2</sub> (500 mM Tris/HCl, 100 mM KCl, 50 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>) (Roche), 3mM de MgCl<sub>2</sub> (Roche), 0,5μM de chacune des amorces (IDT), 0,8mM de dNTP, 1 U de l'enzyme Fast Start (Roche), 15 ng d'ADN génomique, dans un volume final de 25μL. Le second PCR fait suite au premier et il est effectué sur le fragment PCR préalablement amplifié et purifié. Ce second PCR a pour but de compléter les adaptateurs attB1 et attB2 à leur longueur totale de 29 bases. Deux amplifications PCR consécutives avec des amorces d'environ 20 nucléotides sont effectuées; plutôt qu'une seule amplification avec des amorces contenant autour de 40 nucléotides, afin de faciliter l'amplification. Lors du deuxième PCR le programme utilisé est le même que décrit précédemment contenant 40 cycles. Le mélange contient également les même solutions, à l'exception des amorces 0,5μM des adaptateurs Gateway B1 et B2 (Invotrogen) et de l'ADN 3μL du fragment PCR préalablement amplifié (dilué 1 : 50). Les amorces utilisées pour l'amplification des fragments 2Kb sont présentées au tableau x, annexe II.

### **Purification du fragment PCR.**

Afin d'augmenter l'efficacité des étapes ultérieures, le fragment d'ADN amplifié d'une longueur de 2 Kb est purifié. La purification est effectuée par migration de l'ADN sur un gel d'agarose, suivi d'une extraction du gel. Cette étape permet de retirer tous les enzymes et autres réactifs contenus dans le mélange PCR. Un gel d'agarose 1% préparé dans un volume de 100 mL de solution TBE 1X, auquel est ajouté 0,5 $\mu$ g de Bromure d'Éthidium (Sigma), est préparé. Au mélange PCR de 25  $\mu$ L est ajouté 5 $\mu$ L de tampon de migration (40% sucrose, 0,25% bleu de bromophénol, 0,25% xylène cyanol), ensuite le tout (30 $\mu$ L) est chargé sur le gel. On migre également 260ng d'un marqueur de poids moléculaire, *High mass ladder* (Invitrogen), dans un puit voisin afin d'identifier le fragment recherché de 2 Kb. L'électrophorèse est réalisée à 75 Volts (120 mA) pendant 20 minutes. Ensuite, le gel est déposé sur une lampe UV et soumis à 254 nm de longueur d'ondes. Les bandes d'ADN amplifiées et migrées sont visibles sous les rayons UV. La bande correspondante au fragment d'ADN d'une longueur de 2 Kb est découpée du gel. Cette bande est ensuite extraite du gel en utilisant le kit d'extraction de gel *QIAquick gel extraction kit* (QIAGEN), selon les recommandations du fabricant. Puis, l'ADN est resuspendu dans un volume de 30  $\mu$ L de tampon d'élution (QIAGEN). Le fragment est ensuite dosé sur l'appareil *Nano-Drop 1000 Spectrophotometer* (Fisher Scientific).

### **Clonage par la technique Gateway (Invitrogen).**

Le clonage par la technique Gateway consiste en 2 réactions enzymatiques consécutives. La première réaction, effectuée avec l'enzyme BP clonase (Invitrogen), permet d'insérer le fragment PCR dans un vecteur d'entrée pDONR (Gateway). Cette étape facilite ensuite le passage du fragment dans le vecteur désiré, tel que pGL3-basic (Promega). Le mélange de la réaction, comprend 1x de BP clonase (protéine de recombinaison du bactériophage lambda *Integrase* (Int) et protéine *Integration Host Factor* (IHF) de *E. coli*) (Invitrogen), 0,6x de tampon TE pH 8.0, 150ng du vecteur d'entrée pDONR (Invitrogen) et 150ng du



fragment PCR purifié contenant les queues B1 et B2; dans un volume final de 10 $\mu$ L. Le mélange est incubé à la température pièce pendant la nuit. Ensuite, on procède à la deuxième réaction enzymatique avec l'enzyme appelé LR clonase (Invitrogen). Au mélange initial est ajouté 30 mM de NaCl, 1x de LR clonase (protéines de recombinaison *Integrase* (Int) et *Excisionase* (Xis) du bacteriophage lambda et protéine *Integration Host Factor* (IHF) de *E. coli*) (Invitrogen), 225ng du vecteur d'expression pGL3-basic (Promega) modifié, afin d'être compatible avec la technique Gateway. Cette deuxième réaction enzymatique est également incubée à la température pièce pendant la nuit. Puis, pour arrêter la réaction enzymatique, 2 $\mu$ L de la protéinase K (Invitrogen) est ajouté et le mélange final est incubé à 37°C pendant 10 minutes. Ce mélange est conservé à -20°C pour une durée maximale de 2 semaines avant de procéder à l'étape de transformation.

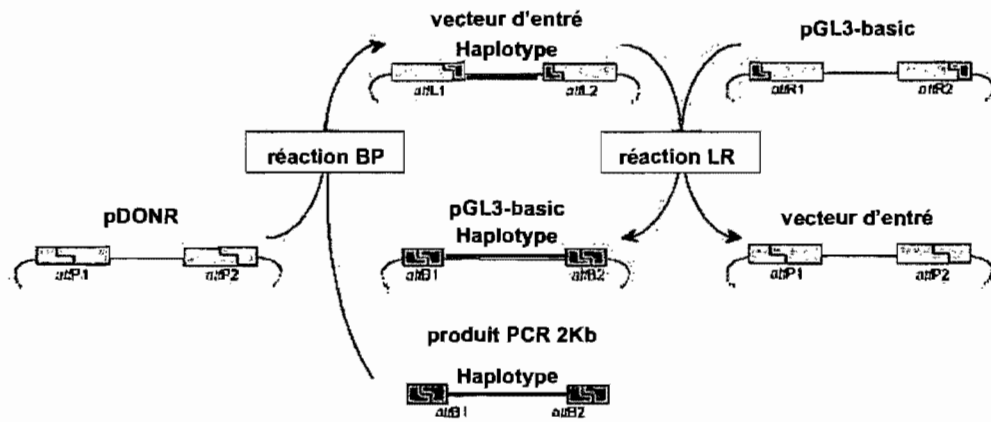


Figure 2.2 : Représentation graphique du clonage par la technique Gateway (Invitrogen).

### **Mutagenèse dirigée.**

La technique de mutagenèse dirigée a été utilisée comme une technique alternative pour certains gènes. Dans certains cas, le fragment de 2 Kb est amplifié plus difficilement chez quelques individus. Puisque le gène testé doit être amplifié chez plusieurs individus afin d'obtenir les haplotypes fréquents, les ADN génomiques plus facile à amplifier peuvent être modifiés pour produire les haplotypes manquants. La trousse de mutagenèse dirigée *QuikChange Multi* (Stratagene) a été utilisée.

Le principe de cette technique est de modifier de façon ponctuelle les sites spécifiques de l'ADN où sont situés les variants (SNPs), à l'aide d'oligonucléotides (figure 2.3). Des oligonucléotides comprenant la base modifiée (SNP), sont construits de façon complémentaire au brin d'ADN gabarit. Les oligonucléotides soumis à des cycles PCR s'hybrident par complémentarité sur un des brins du gabarit d'ADN. Puis, l'augmentation de température active les enzymes pour permettre la synthèse du nouveau brin d'ADN contenant les sites modifiées par les oligonucléotides. En procédant à plusieurs cycles PCR, une grande quantité du nouveau brin d'ADN est obtenue. Lorsque plusieurs nucléotides doivent être modifiés sur un même brin d'ADN, afin d'obtenir l'haplotype à étudier, l'ensemble des oligonucléotides doivent être complémentaires au même brin gabarit.

Afin d'obtenir des résultats optimaux, les oligonucléotides sont construits de la façon suivante : 25 à 45 bases de longueur,  $T_m > 75^\circ\text{C}$ ,  $\%GC > 40\%$ , la base mutée est placée environ au centre de l'oligonucléotide et un ou plusieurs GC sont placés à l'extrémité 3'. Le mélange de la réaction contient : 1x de tampon *QuikChange Multi* (Stratagene), 0,75 $\mu\text{L}$  de *QuikSolution* (Stratagene), 300ng de la matrice d'ADN double brins, 1 $\mu\text{L}$  de dNTP (Stratagene), 2,5 U du mélange d'enzymes *QuikChange Multi* (Stratagene), 100ng de chacun des oligonucléotides (1 à 3 oligonucléotides utilisés) ou 50ng de chacun de chacun des oligonucléotides (4 à 5 oligonucléotides utilisés), dans un volume final de 25  $\mu\text{L}$ . Le mélange est soumis à 30 cycles PCR composés de 3 étapes; 1 minute à  $95^\circ\text{C}$ , 1 minute à  $55^\circ\text{C}$  et 14 minutes à  $65^\circ\text{C}$ . Les cycles sont précédés d'une étape de 1 minute à  $95^\circ\text{C}$ .

Suite aux cycles PCR, un mélange contenant de l'ADN gabarit et le nouvel ADN synthétisé est obtenu. Une digestion enzymatique avec l'enzyme *DpnI* (Stratagene) permet d'éliminer l'ADN gabarit. L'enzyme *DpnI* digère l'ADN méthylé. L'ADN gabarit est méthylé puisqu'il est préalablement amplifié dans des cellules bactériennes, alors cet ADN méthylé est digéré et éliminé du mélange. La digestion contient 15 U de l'enzyme *DpnI* ajoutées directement au mélange et incubé à une température de 37°C pendant 3 heures. Suite à cette digestion seul le nouvel ADN est présent. Le tout est conservé à -20°C jusqu'à la transformation.

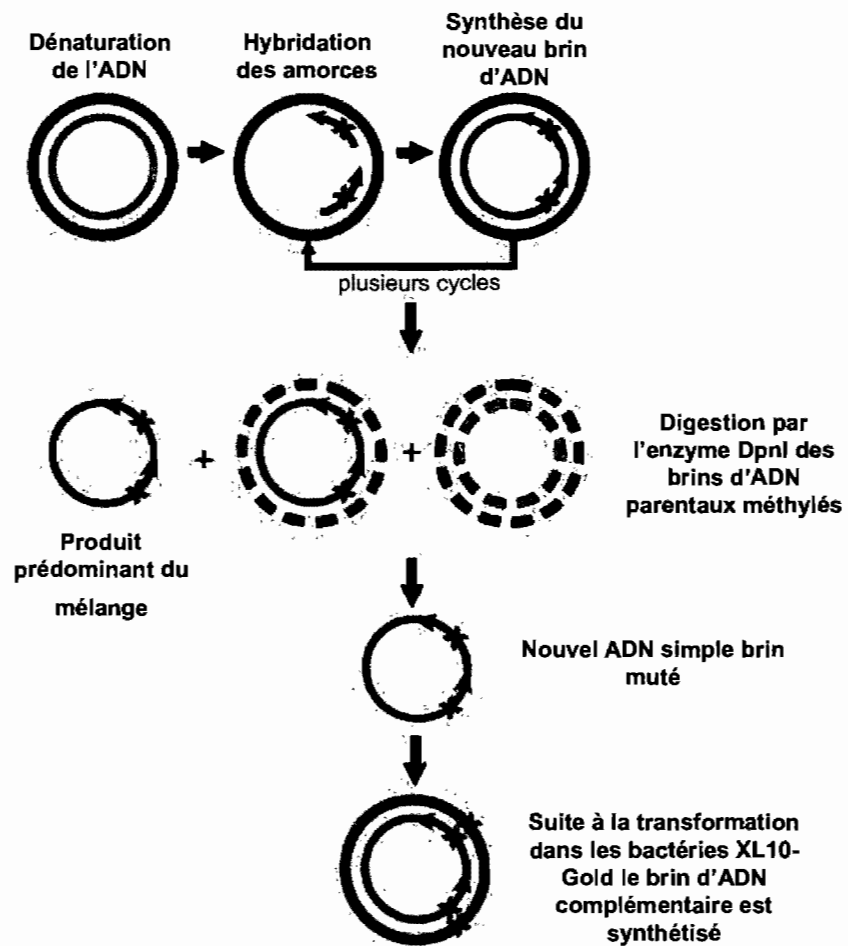


Figure 2.3 : Représentation graphique de la construction des haplotypes régulateur par la technique de mutagenèse dirigée.

**Transformation.**

Le principe de la transformation est de permettre l'entrée de matériel génétique étranger dans une cellule permissive. Pour cette étape, différents niveaux de cellules compétentes sont utilisés. Des cellules bactériennes ultra-compétentes destinées à transformer de long plasmides ou pour des clonages difficiles. Les cellules ultra-compétentes utilisées, XL10-gold (Stratagene) sont de l'ordre de  $5 \times 10^9$  transformants/ $\mu\text{g}$  d'ADN. Pour les transformations moins complexes, les cellules compétentes DH5 $\alpha$  (Invitrogen) sont utilisées. Les DH5 $\alpha$  possèdent un ordre de transformation de  $1 \times 10^6$  transformants/ $\mu\text{g}$  d'ADN.

**Transformation dans les bactéries compétentes DH5 $\alpha$ .** Ces bactéries sont utilisées pour les transformations qui font suite au clonage par le système Gateway. Pour chaque construction d'haplotype à transformer un aliquot de 40 $\mu\text{L}$  de bactéries compétentes est décongelé sur glace. Puis, 2 $\mu\text{L}$  du clonage Gateway (réaction BP/LR) est ajouté au tube. Le tout est ensuite incubé sur glace pour une durée de 30 minutes. Cette incubation est suivie d'un choc thermique infligé aux cellules qui permet aux plasmides d'entrer dans les cellules. Le choc thermique est créé par une courte incubation de 30 secondes à 42°C sur un bloc chauffant. Les cellules sont remises immédiatement sur glace, ce qui complète le choc thermique, pour une durée de 5 minutes. Pour permettre aux bactéries de pousser, 450 $\mu\text{L}$  de milieu SOC (0,5% extrait de levures, 2% Tryptone, 10mM NaCl, 2,5mM KCl, 10mM MgCl<sub>2</sub>, 10mM MgSO<sub>4</sub>, 20mM Glucose) à la température pièce est ajouté.

**Transformation dans les bactéries ultra-compétentes XL10-gold.** Un aliquot de 45 $\mu\text{L}$  de bactéries est décongelé sur glace. Puis, 2 $\mu\text{L}$  de  $\beta$ -mercaptoéthanol (Stratagene) sont ajoutés, le tout est bien mélangé et laissé sur glace pendant 10 minutes. Le  $\beta$ -mercaptoéthanol a pour but d'augmenter l'efficacité de la transformation de 2-3 fois. Au mélange, est ensuite ajouté 1,5 $\mu\text{L}$  de la construction Gateway difficile à transformer ou 4  $\mu\text{L}$  de la digestion *DpnI* de la mutagenèse dirigée, non-purifiée. Le tout est laissé sur glace pour une période de 30 minutes. Un choc thermique est ensuite infligé aux bactéries pour permettre l'entrée du vecteur. Pour produire le choc thermique les cellules sont passées rapidement à 42°C

pendant 30 secondes, puis remises sur glace pour 2 minutes. Un milieu de culture est ajouté pour permettre une croissance efficace des bactéries. Ce milieu appelé NZY<sup>+</sup> contient 10g d'hydrolysate de caséine, 5g d'extrait de levures et 5g de NaCl, dans un volume de 1 litre; à ce volume est ajouté 12,5mM de MgCl<sub>2</sub>, 12,5mM de MgSO<sub>4</sub>, 0,02mM glucose. Un volume de 500μL de milieu NZY<sup>+</sup> est ajouté à chaque transformation.

Les transformations bactériennes sont incubées à 37°C, sous agitation (250 rpm) pendant 1 heure. Lorsque l'incubation est terminée, les tubes sont centrifugés pendant 1 minute à vitesse maximale. Un volume de 400 μL est ensuite retiré, puis le culot bactérien est resuspendu et étalé sur pétri LB ampicilline. Le pétri LB (10g Triptone, 5g. d'extrait de levures, 5g NaCl, 1mL NaOH 1N, 15g d'agar, dans un volume de 1 litre) contient une concentration finale de 50μg/mL d'ampicilline. Les pétris sont incubés pendant la nuit à 37°C. La sélection avec l'ampicilline, permet aux colonies qui ont intégré le plasmide de croître.

### **Mini-culture.**

La mini-culture est une technique qui permet de sélectionner une colonie unique sur pétri pour permettre une croissance encore plus grande de cette colonie dans un bouillon de culture. Afin de trouver des colonies possédant les haplotypes voulus et ne contenant pas d'erreur dans leur séquence, une grande quantité de ces colonies est cultivée en mini-culture. Pour chaque haplotype transformé, environ 20 colonies sont sélectionnées et cultivées en mini-culture. Suite à l'analyse par séquençage, le clone bactérien possédant la bonne séquence est remis en bouillon de culture et une mini-prep est alors produite. Chaque mini-culture est produite dans un volume de 2mL de milieu LB contenant une concentration finale de 50μg/mL d'ampicilline permettant de maintenir la sélection. La culture est placée dans un tube de 13mL (Sarstedt) pour poursuivre avec la mini-prep QIAGEN ou dans un bloc de 96 puits / 2mL (Corning) pour une mini-prep maison. Chaque colonie est piquée sur le pétri à l'aide d'un cure-dents de bois stérilisé et ensuite placé dans le milieu de culture. Tous les tubes sont ensuite placés dans un incubateur *Orbit Environ*

*Shaker* (Labline) à 37°C pendant la nuit avec une agitation de 225 rpm ou dans un agitateur pour bloc *Mini-Grow 384* (Macconnell Research). La mini-culture est toujours suivie d'une étape appelée mini-prep. Les mini-prep servent à lyser les cellules bactériennes pour extraire le matériel génétique désiré. Cette technique permet d'extraire les plasmides pGL3-basic contenant l'haplotype à tester. Deux types de mini-prep sont utilisés. Une mini-prep appelée *maison* est utilisée avant l'étape de séquençage des plasmides et une mini-prep QIAGEN appelée *QIAGEN Plasmid Mini Kit* est utilisée suite au séquençage, avant l'étape de transfection des haplotypes.

**Mini-prep à niveau moyen de pureté.** La plaque de 96 puits est centrifugé (*J2-HC*, Beckman) à 2000 rpm pendant 10 minutes à une température de 4°C, afin d'obtenir un culot bactérien. Le milieu est ensuite retiré par inversion. Puis les culots sont resuspendus dans 100µL de solution (50mM glucose, 10mM EDTA, 25mM Tris-HCl pH 8.0). Ensuite, 200µL de solution de lyse bactérienne (0,2N NaOH, 1% SDS) sont ajoutés. Une fois bien mélangé, le tout est incubé pendant 5 minutes sur glace. Puis, 150µL de solution neutralisante (acétate de potassium 5M, acide acétique glacial pour obtenir une pH 4.8) sont ajoutés, pour permettre d'arrêter la réaction de lyse. Le mélange est incubé sur glace pour 5 minutes. La solution neutralisante ramène le mélange à un pH neutre causant une précipitation des membranes, de l'ADN génomique et de la plupart des protéines. Une centrifugation de la plaque à 2500 rpm pendant 10 minutes suit pour déposer les précipités au fond des puits. Le surnageant est ensuite filtré dans une plaque filtre (3µm verre/0,2 µm Biolnert, Millipore ), puis transféré dans un nouveau bloc. L'isopropanol permet ensuite une précipitation efficace de l'ADN plasmidique. Un volume de 120µL d'isopropanol est ajouté et le tout est centrifugé pour 60 minutes à 2500 rpm. Ensuite, une deuxième précipitation à l'éthanol est effectuée. Étant plus volatile, l'éthanol s'évapore plus rapidement pour permettre de sécher l'ADN. Une quantité de 300µL d'éthanol est ajoutée et le mélange est centrifugé à 2500 rpm pendant 60 minutes. L'ADN séché à l'air est ensuite resuspendu dans 30µL de tampon d'élution (10mM Tris-HCl).

**Mini-prep à haut niveau de pureté. *QIAGEN Plasmid Mini Kit*.** Ce deuxième type de mini-prep est utilisé suite à l'étape de séquençage, lorsque la séquence des clones est



connue. Les clones présentant une séquence génétique du fragment d'ADN parfaite, sont extrait par cette technique. Cette technique permet d'obtenir une qualité supérieure d'ADN extrait. La trousse est utilisée selon les recommandations du manufacturier.

**Séquençage.** Les clones sont envoyés à la plateforme de séquences du Centre d'innovation Technologique Génome Québec, Université McGill. Les ADN plasmidiques extraits par mini-prep sont dilués. Un volume de 5 $\mu$ L de mini-prep est dilué dans 7 $\mu$ L d'eau. Les oligonucléotides sont construits de façon à séquencer tout le promoteur proximal de 2 Kb avec des segments de 700pb se chevauchant. Une concentration d'oligonucléotides de 20 $\mu$ M est envoyée pour le séquençage.

### **2.3.2 Transfection**

La technique de transfection transitoire sert à faire pénétrer, dans les cellules eucaryotes, les vecteurs construits. La transfection transitoire cellulaire permet de mesurer le niveau d'activité du promoteur proximal testé par luminescence. Puisque les promoteurs 2Kb sous forme d'haplotypes sont placés en amont d'un gène rapporteur luciférase dans le vecteur pGL3-basic, les niveaux d'activité des promoteurs résultent en niveaux d'expression de la luciférase. Le niveau d'expression de la luciférase est traduit en intensité lumineuse suite à l'ajout de son substrat. À l'aide d'un appareil luminomètre LmaxII 384 (Molecular Devices) cette intensité lumineuse est mesurée. Les transfections sont produites dans 3 lignées cellulaires différentes, permettant d'assurer que la variabilité d'expression n'est pas un artefact d'une seule lignée. Les 3 lignées cellulaires utilisées sont : HepG2 (hépatocarcinome), Hela (carcinome du col utérin), Jeg3 (carcinome de placenta).

### **Préparation de l'ADN.**

Les constructions d'haplotypes sont cotransfectées avec un second plasmide, contrôle interne, afin de normaliser les expériences. Chaque transfection est produite en 4 réplicats normalisés pour la quantité de cellules et l'efficacité de la transfection par le contrôle interne. Le système commercial de Promega appelé *Dual Glo Luciferase assay System* est utilisé. Les haplotypes d'intérêt sont insérés dans le vecteurs pGL3-basic possédant le gène rapporteur luciférase du luciole (*Firefly*). Le contrôle interne possède le gène rapporteur luciférase de *Renilla*. Leurs origines distinctes confèrent des structures d'enzymes et de substrats différentes permettant de différencier leur bioluminescence. Un ratio de 200 :1, mouche à feu pGL3-basic : *Renilla* pRL-CMV (promega) est préparé pour la transfection. Pour chaque puits à transfecter, 100ng de construction est préparé. Dans un tube 1,5mL, 1200ng de la construction pGL3-basic (40ng/ $\mu$ L) et 8 ng du contrôle interne pRL-CMV sont ajoutés à 30 $\mu$ L de Tris 10mM pH 8.0. En plus du contrôle interne, un contrôle négatif pGL3-basic (Promega) sans insertion et un contrôle positif contenant le promoteur SV40 pGL3-control (Promega) sont transfectés. Les contrôles positif et négatif sont ajoutés à chaque plaque de transfection en quatre réplicats pour chacune des lignées. Dans un volume de 40 $\mu$ L de Tris 10mM, 1600ng de pGL3-basic et 8 ng de pRL-CMV sont ajoutés. La même procédure est suivie pour le pGL3-control (40ng/ $\mu$ L).

### **Transfection.**

Les cellules (Hela, HepG2 et Jeg3) à transfecter sont remises en suspension par aspiration du milieu de culture et ensuite rincées au PBS (Wisent). Puis, le PBS est retiré, 2mL de Trispsine-EDTA 1x (Sigma) est ajouté, puis le mélange est incubé pendant 5 minutes. Ensuite, un volume de 4mL de milieu de culture complet (AMEM (Wisent), 10% Sérum fœtal de bœuf désactivé, 2mM L-glutamine) est ajouté. Les cellules sont décollées par aspiration et comptées dans l'appareil (Beckman coulter) par addition de 1mL de milieu dans l'appareil. Une concentration de  $6 \times 10^3$  cellules/mL (Hela et Jeg3) et  $8 \times 10^3$

cellules/mL (HepG2) est préparée. Pour chaque lignée cellulaire à transfecter (4 réplicats), 10 $\mu$ L des constructions pGL3-basic/pRL-CMV, pGL3-control/pRL-CMV et pGL3-basic/pRL-CMV sont ajoutés à 90 $\mu$ L de milieu sans sérum (AMEM, 2mM L-glutamine). Pour chaque plaque de transfection, 52,5 $\mu$ L de lipofectamine 2000 (Invitrogen) est dilué dans un tube de 15mL de polystyrène (Sarstedt) avec 2572,5 $\mu$ L de milieu de culture sans sérum. Cette dilution est incubée à la température pièce pendant 5 minutes. Ensuite, 100 $\mu$ L de lipofectamine sont ajoutés à 100 $\mu$ L d'ADN préalablement préparé et le tout est incubé pendant 30 minutes à température pièce. Cette incubation permet aux lipides de la lipofectamine, de s'associer à l'ADN afin de faciliter l'entrée du vecteur dans la cellule. Un volume de 425 $\mu$ L de cellules (Hela, Jeg3, HepG2) est ajouté aux mélanges d'ADN. Puis, 150 $\mu$ L de chaque mélange cellules-ADN sont transférés dans un puits de la plaque de transfection opaque, 96 puits (Corning) en 4 réplicats. Les plaques sont placées dans l'incubateur (US Autoflow, Nuair) à 37°C, 5% CO<sub>2</sub> pendant 24 heures.

Les transfections sont ensuite arrêtées. Le milieu est aspiré et les cellules sont rincées au PBS. Le PBS est ensuite aspiré, puis 20 $\mu$ L de solution de lyse 1x PLB (constitution non divulguée) (Proméga) sont ajoutés aux puits. Les plaques sont agitées sur une plateforme agitatrice (Reliable Scientific) pendant 30 minutes à température pièce. Les deux substrats permettant la réaction de bioluminescence de la luciférase *Renilla* et mouche à feu sont préparés et placés dans le luminomètre. Puis, la plaque est placée dans le luminomètre et la lecture de luminescence est effectuée pour chacun des puits. Les mesures des quatre réplicats trouvés sont ensuite normalisées. La luminescence est exprimée sous forme d'un ratio, les mesures de luciférase Luciole sont divisées par les mesures de luciférase *Renilla* (contrôle interne) et le résultat est multiplié par 100.

### **2.3.3 Electrophoretic mobility shift assay (EMSA).**

La technique de retard sur gel (EMSA) sert à détecter les interactions ADN-protéines. Elle permet ainsi de détecter les liaisons différentielles entre les deux allèles d'ADN pour un

variant (SNP) dans la région promotrice des gènes (Buratowski and Chodosh 2001). Les sondes double brins d'ADN marquées par radioactivité sont migrées sur un gel d'acrylamide non-dénaturant en présence d'extraits nucléaires. Les allèles d'ADN du promoteur proximal sont en mesure de lier différents facteurs de transcription présents dans l'extrait nucléaire et ainsi former un complexe qui par sa grosseur est retardé pendant la migration sur gel (figure 2.4). Lorsque le complexe est présent seulement avec un des deux allèles du variant testé, on dit alors qu'il y a liaison différentielle pour le SNP étudié. La spécificité du complexe est vérifiée par l'ajout d'un excès de 50x de compétiteurs non-marqués.

#### **Extraits nucléaires.**

Les extraits nucléaires des 3 lignées cellulaires utilisées pour les transfections sont utilisés pour les expériences de retard sur gel. Les cellules sont récoltées et lavées dans 10 mL de PBS (Wisent). Ensuite, les cellules sont centrifugées pendant 5 minutes à 1500 g. et le culot est remis en suspension dans 1mL de PBS. Le culot resuspendu est ensuite transféré dans un tube de 1,5mL et centrifugé dans une centrifugeuse de table à 3 000 rpm pendant 3 minutes. Le PBS est retiré et le culot est resuspendu délicatement dans 400 $\mu$ L d'un tampon A froid, contenant 10mM HEPES pH 7.9, 10mM KCl, 0,1mM EDTA, 0,1mM EGTA pH 8.0, 1mM DTT, 0,5 mM PMSF, 10 $\mu$ g/mL Aprotinine, 1x capsule Complete mini (3,7mg EDTA, 0,2mg extrait de pancréas, 0,05mg Pronase, 0,005mg thermolysine, 0,03 mg chemotrypsine, 3,3 mg papaïne) (Roche). Le tampon A permet l'inhibition des protéinases. Le tout est placé dans la glace pour 15 minutes, puis 25 $\mu$ L de Nonidet NP-40 10% est ajouté. Le tube est agité vigoureusement pendant 10 secondes. L'homogénat est centrifugé pendant 30 secondes dans une centrifugeuse de table à 13 000 rpm. Puis le culot de noyaux est resuspendu dans 50  $\mu$ L de tampon C froid (20mM HEPES pH 7.9, 0,4M NaCl, 1mM EDTA, 1mM EGTA, 1mM DTT, 1mM PMSF, 20% glycérol, 1x capsule Complete-mini (Roche) et 10 $\mu$ g/mL Aprotinin) et vigoureusement mélangé sur une plateforme agitatrice à 4°C pendant 15 minutes. L'extrait nucléaire est ensuite centrifugé à

13 000 rpm pendant 5 minutes à 4°C. Le surnageant est récupéré et les protéines obtenues sont dosées en utilisant la technique Bradford (Biorad). L'extrait nucléaire est conservé à -80°C.

#### **Appariement des sondes.**

Les oligonucléotides de 20-45 mers qui couvrent le variant (SNP) à tester sont synthétisés par IDT. Les deux sondes simple brin complémentaires (1,75µM) sont incubées dans un tampon d'appariement 10 x (100mM Tris-HCl, 0,5M NaCl, 10mM EDTA) à une température de 95°C pendant 3 minutes. Puis, les sondes sont laissées à température de la pièce durant 1 heure, ce qui permet leur appariement par complémentarité.

#### **Marquage des sondes double brins.**

La sonde double brins (3,5 pmol) est mise en présence de 1µL de tampon d'échange 1x (Invitrogen), 10 U de l'enzyme T4 Kinase (Invitrogen), 10µCi de  $\gamma$ -ATP<sup>32</sup>, dans un volume final de 10µL et le tout est incubé pendant 10 minutes à 37°C. Sont ensuite ajoutés 1µL de EDTA 0,5M pour arrêter la réaction et 89µL de tampon TE 1x (10mM Tris-HCl, 1mM EDTA). Les sondes sont passées sur colonne G-25 (Amersham) pour éliminer la radioactivité non incorporée.

#### **Migration du gel acrylamide.**

D'abord 2 réactions contrôles sont préparées à partir de la trousse de gel de retardation (Promega). Un contrôle négatif, qui ne contient pas d'extrait nucléaire (7µL d'eau et 2µL de tampon de liaison 5x (50 mM Tris-HCl pH 7.5, 5 mM MgCl<sub>2</sub>, 2.5 mM EDTA, 2.5 mM DTT, 250 mM NaCl, 0.25 µg/µL poly déoxyinosinate-déoxycytidylate, 20% glycérol). Le contrôle positif utilisé est la séquence d'ADN connue pour lier le facteur de transcription

Sp1. Le contrôle positif est produit en présence d'un compétiteur spécifique non-marqué (séquence liaison SP1, ATTCGATCGGGGCGGGGCGAGC) et en présence d'un compétiteur non-spécifique non-marqué (séquence SP1 où on a introduit une mutation, ATTCGATCGGTTTCGGGGCGAGC). Les mélanges contiennent chacun 10µg d'extrait nucléaire, 35 fmol de la sonde SP1 double brins marquée par radioactivité, 2µL de tampon de liaison 5x, 1,75 pmol de compétiteur non-marqué, dans un volume final de 10µL. Pour chaque sonde à tester, 4 réactions sont préparées : 1 réaction sans compétiteur contenant 10µg d'extrait nucléaire, 35 fmol de la sonde double brins marquée par radioactivité, 2µL tampon de liaison 5x, dans un volume final de 10µL; 2 réactions de compétition spécifique (sonde non-marquée, sonde de l'allèle opposé non-marquée) et 1 réaction de compétition non-spécifique (sonde NS non-marquée) contenant 10µg d'extrait nucléaire, 35 fmol de la sonde double brins marquée par radioactivité, 2µL tampon de liaison 5x, 1,75 pmol de compétiteur non-marqué, dans un volume final de 10µL. La sonde NS (GATCGAACTGACCGCTTGCGGCCCGT) utilisée a été synthétisée (compagnie IDT) et une vérification de la séquence choisie a été faite pour s'assurer qu'aucun facteur de transcription connu ne peut s'y lier. Lors des premières expériences des gènes de l'ADN de thymus de veau était utilisé comme compétiteur non spécifique. Ce compétiteur c'est révélé inapproprié, puisqu'il semble contenir les sites de liaison de nombreux facteurs de transcription. Toutes les réactions sont ensuite incubées à la température de la pièce pendant 20 minutes. Ensuite, 1µL de tampon de chargement 10x (250 mM Tris-HCl pH 7.5, 0,2% bleu de bromophenole, 40% glycérol) est ajouté. Toutes les réactions sont migrées sur gel d'acrylamide non-dénaturant 6% (acrylamide-bisacrylamide, 37,5 : 1; glycérol 2,5%) dans un tampon 1x TGE (125 mM Tris, 950 mM glycine, 5 mM EDTA). Après avoir déposé les échantillons, l'électrophorèse se déroule à 4°C, 190 V. jusqu'à ce que le bleu de bromophenole soit arrivé au trois-quart du gel. Après démoulage, le gel est séché sur un papier Whatman pour 1 heure à 80°C dans un séchoir à gel (Biorad). Puis, le gel est exposé sur un film X-ray, Super RX (Kodak) pendant la nuit à -80°C en présence d'écrans intensificateurs. Les sondes utilisées lors des expériences de EMSA sont décrites au tableau XXII, annexe III.

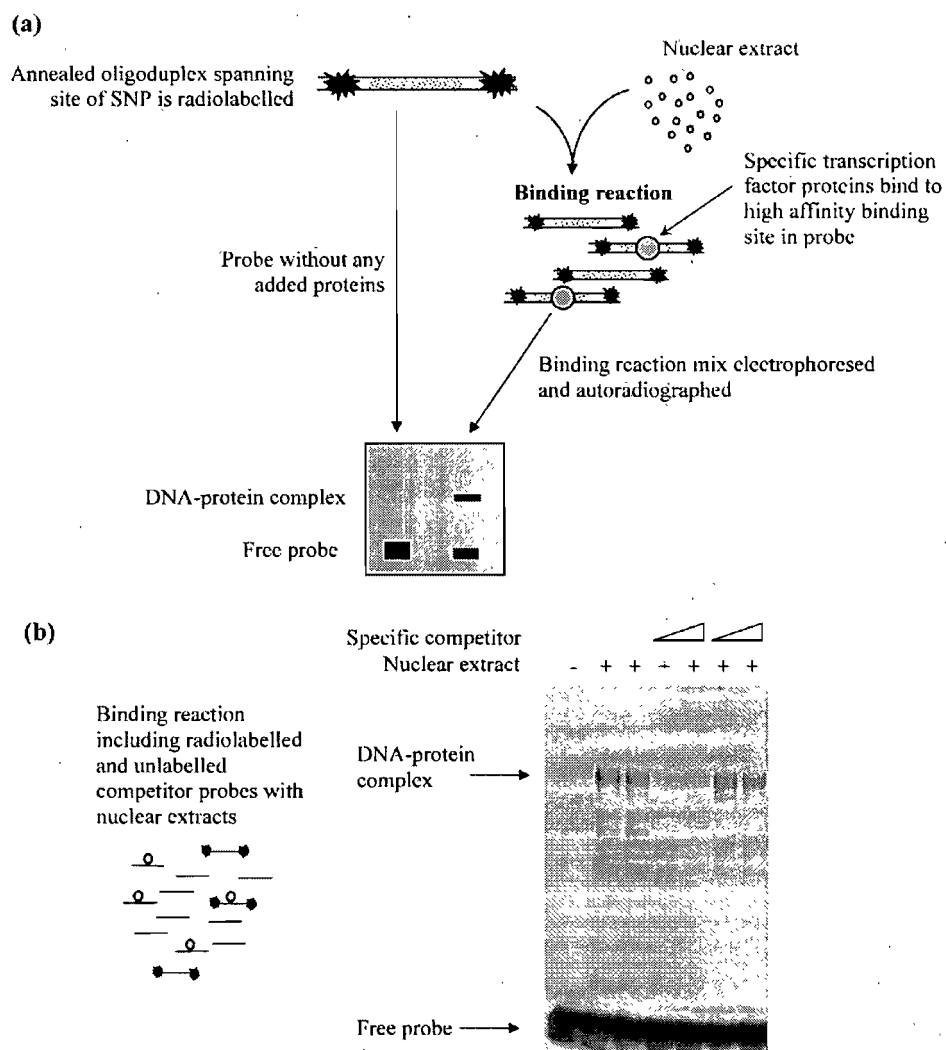


Figure 2.4 : Représentation graphique de la technique de retard sur gel, EMSA (Knight 2003).

## **Chapitre 3**

### **-Résultats-**



### 3. Résultats

#### 3.1 Génotypage.

Les résultats de génotypage de notre échantillon de population pour les 18 gènes du point de contrôle G1/S du cycle cellulaire, ont permis d'identifier la fréquence de chacun des rSNPs à l'étude. Ces résultats sont présentés au tableau I. Les rSNPs sont identifiés selon leur position dans le promoteur, mais également selon leurs numéros d'identification (rs) par la base de données dbSNP (<http://www.ncbi.nlm.nih.gov/projects/SNP/>).

Tableau I : Liste des rSNPs étudiés pour chacun des gènes régulateurs du point de contrôle G1/S.

Gene	SNP	RS	Freq allélique (%)
CCND1	-519 C>T	rs36225064	C (97.5), T (2.5)
	-568 A>C	rs36064048	A (98.7), C (1.3)
	-766 A>T	rs3212860	A (96.2), T (3.8)
	-1537 DEL C	rs36225395	DEL (73.7), C (26.3)
	-1554 G>A	rs36225394	G (97.5), A (2.5)
	-1597 G>A	rs36225073	G (93.6), A (6.4)
	-1701 G>C	rs36225070	G (97.5), C (2.5)
	-1737 A>T	rs36225069	A (98.7), T (1.3)
	-1885 G>T	rs36225068	G (98.7), T (1.3)
	-1888 A>C	rs36225067	A (96.2), C (3.8)
	-1926 INS G	n/d	- (2.5), INS G (97.5)
	-1931 G>C	rs36225071	G (97.5), C (2.5)
	-1938 C>T	rs1944129	C (34.9), T (65.1)

Gene	SNP	RS	Freq allélique (%)
CCNE1	-1366 G>A	rs36225398	G (96.2), A (3.8)
	-1613 T>DEL T	rs36225399	T (98.7), DEL (1.3)
CDC25A	-51 G>C	rs3731485	G (95.0), C (5.0)
	-171 A>T	rs3731484	A (97.5), T (2.5)
	-351 G>A	rs3731482	G (98.7), A (1.3)
	-1433 C>T	rs6809142	C (98.7), T (1.3)
	-1440 G>A	rs34440842	G (97.5), A (2.5)
	-1566 C>T	rs6763242	C (98.7), T (1.3)
	-2030 T>G	rs1903061	T (93.7), G (6.3)
CDK2	-54 C>T	rs2069396	n/d
	-80 A>C	rs2069395	n/d
	-113 A>T	rs2069394	n/d
	-323 C>T	rs2069393	n/d
	-456 C>G	rs2069392	n/d
	-683 C>T	rs2069391	n/d
	-916 A>T	rs2069390	n/d
	-919 A>G	rs2069389	n/d
	-1476 A>T	rs12309895	n/d
	-1856 DELC	rs35283439	n/d
CDK4	-150 C>G	rs36227202	C (98.7), G (1.3)
	-217 C>T	rs238532	C (98.7), T (1.3)
	-290 C>T	rs11172321	C (97.5), T (2.5)
	-612 C>A	rs3211607	C (98.7), A (1.3)
	-615 C>A	rs36227547	C (98.7), A (1.3)
	-724 DEL TC	rs36227545	TC (98.7), DEL (1.3)
	-847 C>T	n/d	C (98.7), T (1.3)
	-937 A>G	rs3211605	A (97.4), G (2.6)
	-1178 A>G	rs36227550	A (98.7), G (1.3)

Gene	SNP	RS	Freq allélique (%)
	-1349 T>A	rs36227551	T (98.7), A (1.3)
	-1887 T>C	rs36227552	T (98.7), C (1.3)
CDK6	-638 C>T	rs36227868	C (97.4), T (2.6)
	-1124 T>A	rs3731258	T (98.7), A (1.3)
	-1270 C>T	rs36227864	C (98.7), T (1.3)
	-1386 A>G	rs36227869	A (98.7), G (1.3)
	-1450 A>T	rs10257231	A (98.7), T (1.3)
	-1789 A>T	rs36227865	A (98.7), T (1.3)
	-1902 DEL CCCCCGCA	rs36227867	CCCCCGCACCCCCGCA (93.7), DEL CCCCCGCA (6.3)
	-1958 C>G	rs36227866	C (98.7), G (1.3)
CDKN1A	-134 G>C	rs4135239	G (80.7), C (19.3)
	-499 G>C	rs4151702	G (48.7), C (51.3)
	-519 G>C	rs730505	G (48.7), C (51.3)
	-791 T>C	rs2395655	T (53.8), C (46.2)
	-899 T>G	rs762624	T (71.7), G (28.3)
	-1021 C>T	rs762623	C (80.7), T (19.3)
	-1284 T>C	rs733590	T (64.1), C (35.9)
	-1466 A>G	rs3829968	A (85.8), G (14.2)
	-1529 T>C	rs3829967	T (85.8), C (14.2)
	-1536 C>T	rs3829968	C (97.4), T (2.6)
	-1731 DEL A	rs36228198	A (52.6), DEL (47.4)
CDKN1B	-221 T>C	rs36228495	T (98.7), C (1.3)
	-337 C>A	rs36228498	C (26.3), A (73.7)
	-373 G>T	rs36228499	G (67.5), T (32.5)
	-522 G>A	rs248743	G (90.0), A (10.0)
	-756 A>C	rs36228497	A (96.2), C (3.8)
	-1084 DEL AGG	rs36228201	AGG (98.7), DEL (1.3)

Gene	SNP	RS	Freq allélique (%)
	-1228 C>T	rs36228202	C (98.7), T (1.3)
	-1608 G>A	rs35756741	G (93.7), A (6.3)
	-1857 C>T	rs3759217	C (86.2), T (13.8)
	-1908 DEL T	rs36228203	T (96.1), DEL (3.9)
CDKN2A	-44 A>T	rs36228836	A (95.0), T (5.0)
	-222 T>A	rs36228834	T (92.5), A (7.5)
	-464 C>T	rs3731238	C (97.5), T (2.5)
	-589 DEL G	rs36228501	G (95.0), DEL (5.0)
	-598 G>A	rs36228502	G (95.0), A (5.0)
	-710 C>A	rs36228503	C (97.5), A (2.5)
	-1600 A>G	n/d	A (98.7), G (1.3)
	-1879 C>T	rs3731236	C (98.7), T (1.3)
CDKN2B	-5 C>G	rs36229156	C (98.7), G (1.3)
	-214 INS T	rs3217975	- (97.5), INS T (2.5)
	-287 C>G	rs2069418	C (22.7), G (77.3)
	-549 C>T	rs3217973	C (98.7), T (1.3)
	-593 A>T, C	rs2069416	A (63.5), T - C (36.5)
	-1001 G>A	rs495490	G (3.8), A (96.2)
	-1164 G>A	rs3808845	G (95.0), A (5.0)
	-1270 C>T	rs36229158	C (92.4), T (7.6)
	-1319 A>G	rs36229155	A (62.3), G (37.7)
E2F1	-187 C>T	rs3213141	C (81.2), T (18.8)
	-243 T>G	rs36231779	T (98.7), G (1.3)
	-300 G>A	rs36231778	G (98.7), A (1.3)
	-897 C>T	rs3213134	C (93.7), T (6.3)
HDAC1	-162 C>T	rs36212114	C (98.7), T (1.3)
	-326 A>T	rs36211753	A (97.5), T (2.5)
	-455 T>C	rs36212119	T (93.7), C (6.3)

Gene	SNP	RS	Freq allélique (%)
	-548 G>A	rs36212118	G (97.5), A (2.5)
	-599 C>T	rs36212117	C (97.5), T (2.5)
	-783 A>G	rs36212116	A (95.0), G (5.0)
	-905 A>C	n/d	A (98.7), C (1.3)
	-1140 G>A	n/d	G (98.7), A (1.3)
	-1193 G>A	rs36212121	G (97.5), A (2.5)
	-1269 T>C	rs1741981	T (55.1), C (44.9)
	-1546 G>T	n/d	G (98.7), T (1.3)
MADH3	-669 C>T	rs11633026	C (77.5), T (22.5)
	-749 G>A	rs36222034	G (87.5), A (12.5)
	-1137 T>G	rs36221699	T (97.5), G (2.5)
	-1938 T>C	rs36221701	T (87.5), C (22.5)
	-1946 C>A	rs34189958	C (88.7), A (21.3)
MADH4	-412 G>T	rs36222036	G (98.7), T (1.3)
	-418 A>C	rs36222037	A (98.7), C (1.3)
RB1	-83 INS CTCGCGGACGT GACGCCGCGGGC	rs36230211	INS CTCGCGGACGT GACGCCGCGGGC (98.7), DEL (1.3)
	-595 T>G	n/d	T (97.5), G (2.5)
	-768 C>A	rs7337056	C (95.0), A (5.0)
	-1554 C>A	rs1573601	C (85.0), A (15.0)
	-1733 INS CTGATA	rs3092879	- (95.0),INS CTGATA (5.0)
	SKP2	-758 G>C	rs36232788
-846 A>G		rs28362652	A (97.4), G (2.6)
TFDP1	-1394 T>C	rs9603847	T (94.9), C (5.1)
	-1515 G>T	rs36203751	G (93.6), T (6.4)
	-1516 C>T	rs36203750	C (93.6), T (6.4)
	-1530 C>T	rs36203748	C (93.6), T (6.4)

Gene	SNP	RS	Freq allélique (%)
	-1655 G>A	rs36204085	G (94.9), A (5.1)
	-1779 A>G	rs36204084	A (96.2), G (3.8)
	-1883 G>C	rs7323179	G (94.9), C (5.1)
	-1887 C>T	rs7324599	C (94.9), T (5.1)
TGFβ1	-448 C>T	rs11466314	C (98.7), T (1.3)
	-508 A>G	rs1800469	A (44.9), G (55.1)
	-799 G>A	rs1800468	G (98.7), A (1.3)
	-1146 C>G	rs3087453	C (98.7), G (1.3)
	-1349 INS C	n/d	- (98.7), INS C (1.3)
	-1550 INS AGG	rs11466313	- (56.4), INS AGG (43.6)
	-1571 A>G	rs4803457	A (52.5), G (47.5)
	-1886 G>A	rs2317130	G (48.7), A (51.3)
	-1960 C>T	rs11466311	C (96.1), T (3.9)

Les numéros rs ont été tirés de la base de données dbSNPs

(<http://www.ncbi.nlm.nih.gov/projects/SNP/>). Pour chacun des allèles la fréquence obtenue dans l'échantillon de population est indiquée (%).

### 3.2 Description de l'étude du gène CDC25A.

#### 3.2.1 Identification des haplotypes régulateurs (rHap) présents dans l'échantillon représentatif de la population mondiale.

La région régulatrice que nous avons étudiée correspond à une portion de 2 Kb en amont du site d'initiation de la transcription. L'haplotype ancestral identifié à partir de l'ADN de chimpanzé constitue la référence nous permettant de déterminer les variants des polymorphismes régulateurs à l'étude (tableau II).

Pour le gène CDC25A, 7 polymorphismes préalablement identifiés (Sinnott and others 2006a), présents dans sa région régulatrice ont été analysés par la technique de génotypage de type A.S.O.. Les polymorphismes, sont nommés selon leur position dans le promoteur du gène relative au site d'initiation de la transcription (+1). En utilisant la fréquence de ces rSNPs, nous avons construit un total de 7 différents haplotypes à l'aide du logiciel PHASE. Les haplotypes sont présents dans la population à des fréquences très variées. Les haplotypes, H1, H2 et H7, montrent une fréquence supérieure aux autres haplotypes. L'haplotype H1, qui représente l'haplotype ancestral, est observé chez plus de 81% de la population et est réparti dans les 5 groupes ethniques. Nous avons voulu valider l'impact fonctionnel des 4 haplotypes les plus fréquents. L'haplotype H3 qui possède la même fréquence (2,5%) que l'haplotype H5, n'a pu être testé pour son niveau d'expression, puisque celui-ci a présenté des difficultés à l'étape du clonage dans le vecteur d'expression pGL3-basic.

Tableau II : Liste des haplotypes régulateurs retrouvés dans l'échantillon de populations de 40 individus pour le gène CDC25A.

Gène	Haplotypes / SNPs							Population							
		-2030	-1566	-1440	-1433	-351	-171	-51	Africain	Amérindien	Européen	Asiatique	Moyent-Orientau x	Total	Freq (%)
CDC25A	Ancestral	G	C	G	C	G	A	G							
	H1	-	-	-	-	-	-	-	13	13	12	13	14	65	81.3
	H2	-	-	-	-	-	-	C		2		2		4	5.0
	H3	-	-	-	-	-	T	-				1	1	2	2.5
	H4	-	-	-	-	A	-	-		1				1	1.3
	H5	-	-	A	-	-	-	-	2					2	2.5
	H6	-	T	-	T	-	-	-	1					1	1.3
	H7	T	-	-	-	-	-	-			4		1	5	6.3

Les zones grisées montrent les haplotypes dont le niveau d'expression a été testé par des analyses de transfections transitoires.



### 3.2.2 Analyse fonctionnelle d'expression des rHaps, par expériences de transfections transitoires.

Nous avons décidé de produire les essais de transfections transitoires sur les haplotypes communs dans la population. Pour permettre l'analyse *in vitro* des haplotypes régulateurs, ceux-ci ont été clonés dans un vecteur contenant le gène rapporteur luciférase de luciole. Les analyses d'expression de la luciférase permettent de mesurer l'activité des promoteurs et de déterminer l'impact des rSNPs sur cette activité. Il est aussi possible de mettre en évidence un déséquilibre, c'est-à-dire une différence d'activité entre les haplotypes régulateurs. Les transfections sont produites dans 3 lignées cellulaires différentes, afin de minimiser les effets cellules-spécifiques. Nous utilisons une définition conservatrice d'un promoteur actif, c'est-à-dire possédant une activité supérieure à 5 fois celle du niveau de base (pGL3-basic sans promoteur). Ensuite nous avons déterminé que la différence d'expression significative entre deux haplotypes doit être supérieure ou égale à 1,5 fois, en plus d'être significative selon le test de Lord (Lord 1957) à  $p \leq 0,05$ . Les haplotypes fréquents, des différents gènes analysés par transfections transitoires, sont comparés à l'haplotype majeur du même gène.

Les 4 haplotypes les plus fréquents du gène CDC25A (H1, H2, H5 et H7) ont été testés par transfections transitoires. Ces expériences de transfections ont permis de valider les différences d'expression significatives en présence des haplotypes H2, H5, H7 et associé aux variants suivants : -51G>C (H2), -1440G>A (H5) et -2030T>G (H7) (tableau II). L'activité promotrice de tous les haplotypes testés du gène CDC25A (figure 3.1). L'activité des promoteurs est de l'ordre de 20 à 30 fois la ligne de base dans les cellules HepG2. Pour les lignées cellulaires Hela et Jeg3 cette différence d'activité est encore beaucoup plus grande. Ce qui classe les promoteurs du gène CDC25A comme étant des promoteurs forts. L'haplotype H2 auquel est associé le polymorphisme -51G>C, démontre une différence allélique très significative ( $P < 0,01$ ) dans les 3 lignées cellulaires, qui se situe autour de 1,7 fois le rHap majeur (H1). Également, l'haplotype H7, qui est du même ordre de fréquence que l'haplotype H2 dans la population, démontre une différence d'expression significative

( $P < 0,01$ ) encore plus importante, qui est d'environ 2 fois le majeur dans les lignées cellulaires Hela et Jeg3. Le SNP -2030T>G est associé à la différence d'expression de l'haplotype H7. L'activité de H7 se retrouve à un niveau inférieur à l'activité des haplotypes H2 et H5 dans la lignée cellulaire HepG2. L'haplotype H5 auquel est associé le variant -1440G>A, montre une différence d'expression significative de 1,7 fois le rHap majeur ( $P < 0,05$ ) dans la lignée cellulaire HepG2. Afin de valider l'effet de ces différences alléliques significatives sur la liaison des facteurs de transcription, nous avons poursuivi l'étude des variants qui leurs sont associés en analyse de retard sur gel.

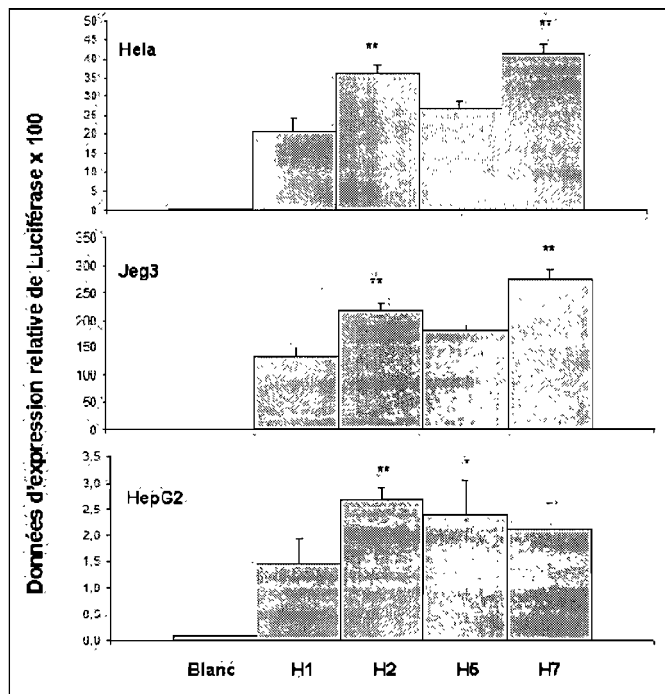


Figure 3.1 : Résultat d'analyse d'expression des 4 haplotypes régulateurs testés pour le gène *CDC25A* par transfection transitoire. Le niveau d'expression de chacun des haplotypes est donné sous la forme d'un ratio de la luciférase mesurée, Luciole/Rénilla, multiplié par 100. Les expériences de transfections sont faites dans trois lignées cellulaires (Hela, Jeg3 et HepG2). Les différences d'expression significatives ont été testées par le test de Lord et les valeurs sont indiquées (\* :  $P < 0,05$ ; \*\* :  $P < 0,01$ ). Les haplotypes fréquents, des différents gènes analysés par transfections transitoires, sont comparés à l'haplotype majeur du même gène. Les haplotypes H2, H5 et H7 démontrent une différence d'expression significative ( $P < 0,01$ ) lorsque comparés à l'haplotype majeur H1. Pour l'haplotype H2 la différence est visible dans les trois lignées cellulaires et pour l'haplotype H7 cette différence est présente seulement dans les cellules HeLa et Jeg3. L'haplotype H5, démontre une différence significative à  $P < 0,05$  dans la lignée cellulaire HepG2 seulement.

### 3.2.3 Analyse des modifications de liaison des facteurs de transcription, dû à la présence de polymorphismes régulateurs, dans des expériences de retard sur gel.

Nous avons voulu tester si les variants contenus dans les haplotypes présentant des différences d'expression significatives affectaient la liaison d'un ou plusieurs facteurs de transcription. À l'aide du logiciel *Mat Inspector*, nous avons effectué des prédictions *In silico* afin d'évaluer l'impact de ces rSNPs sur les sites de liaisons à des facteurs de transcription connus (voir tableau XIX). Les prédictions informatiques ont permis d'identifier 6 facteurs de transcriptions dont le site de liaison est affecté par la présence des variants -1440G>A et -2030T>G. Ensuite, les analyses de retard sur gel ont été produites pour ces rSNPs, y compris ceux dont aucune prédiction de liaison de facteurs de transcription était trouvée (-51G>C).

Les variants -51G>C, -1440G>A et -2030T>G, respectivement associé aux haplotypes H2, H5 et H7 présentant des différences alléliques significatives ( $P < 0,01$ ) ont été étudiés. Les résultats de l'expérience démontrent la présence de liaisons différentielles dans les 3 lignées cellulaires (Hela, Jeg3 et HepG2) pour les variants -51G>C et -2030T>G (tableau XIX). Aucune prédiction de liaison de facteur n'est identifiée pour variant -51G>C, ce qui laisse présumer l'association avec un facteur de transcription inconnu. Le variant -2030T>G a par ailleurs été associé à 4 sites liaisons possibles de facteurs de transcription distincts par la présence de leur site de liaison. La figure 3.2 montre un exemple de l'expérience de retard sur gel pour le polymorphisme -2030T>G en présence de l'extrait nucléaire de cellules HepG2. Cet exemple révèle la formation de 6 complexes spécifiques (complexes 1, 2, 3, 4, 5 et 7) à un allèle du variant -2030T>G. Les complexes 1, 3 et 5 sont associés à l'allèle T. En présence de la sonde chaude T, ces complexes sont compétitionnés lorsque le compétiteur froid 50x sonde T est ajouté (ligne 3) et demeurent présents lorsque le compétiteur froid G (ligne 4) ou compétiteur froid non-spécifique (ligne 5) sont ajoutés. Le complexe 5 est légèrement moins visible que les complexes 1 et 3, mais définitivement présent sur le film original. Les complexes 2 et 7 sont par ailleurs spécifiquement associés à

l'allèle G puisqu'ils sont incapables à lier l'allèle T (ligne 1) et sont clairement visible en présence de la sonde chaude G (ligne 6). De plus, les lignes 7, 8 et 9 révèlent la spécificité de l'association par une absence des complexes lorsqu'ils sont compétitionnés par l'ajout de la sonde froide G (ligne 8) et leur présence lorsque le compétiteur froid T (ligne 7) ou le compétiteur non-spécifique (ligne 9) sont ajoutés. Le complexe 4 est spécifique aux deux sondes T et G (lignes 1, 5, 6 et 9). Ce complexe se trouve compétitionné par tous les compétiteurs froids ajoutés (ligne 3, 4, 7 et 8). Donc, il s'agit d'un facteur qui se lie à la séquence, mais qui n'est pas influencé par la mutation T/G. Le complexe 6 est non spécifique. Pour le variant -1440G>A, aucune liaison différentielle d'un facteur n'a été trouvée.

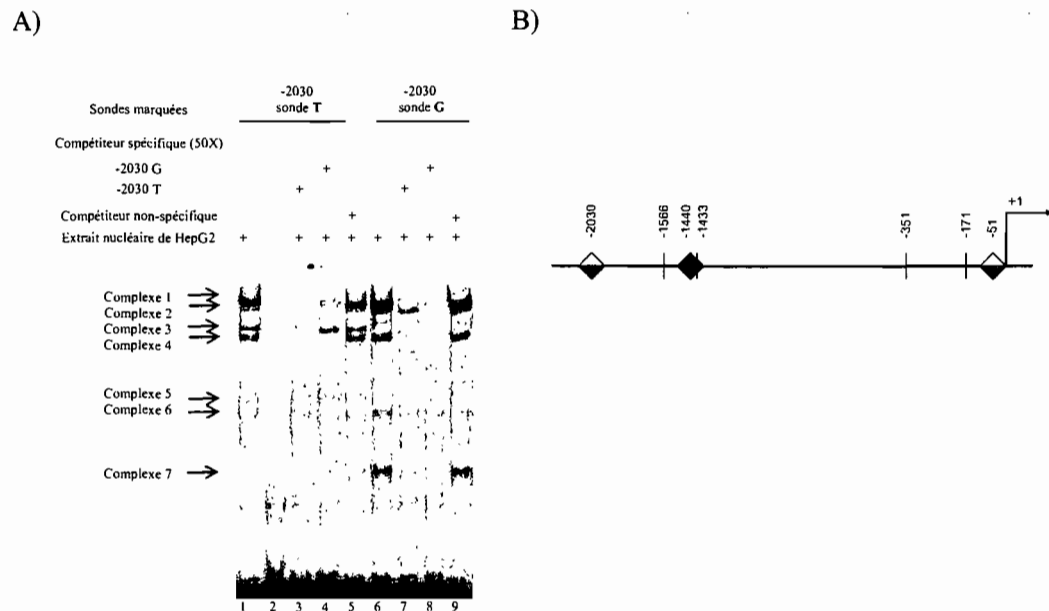


Figure 3.2 : **A)** Résultat d'analyse de retard sur gel pour le variant -2030T>G du gène CDC25A. En présence de l'extrait nucléaire de cellules HepG2, 7 différents complexes sont formés. La ligne 2 représente le contrôle négatif, sans extrait nucléaire. Les complexes 1, 2, 3, 4, 5 et 7 sont spécifiques. Les complexes 1, 3 et 5 sont spécifiques à la sonde T. Ces complexes sont absents totalement ou diminués de beaucoup lorsque le compétiteur spécifique froid 50x T est ajouté (ligne 3) et demeurent présent lorsque les compétiteurs spécifiques sonde opposée, G (ligne 4) et non-spécifique (ligne 5) sont ajoutés. Les complexes 2 et 7 sont à l'opposé spécifique au variant -2030 G. Ils se lient à la sondes G (ligne 6), sont absents lorsque le compétiteur froid spécifique G est ajouté (ligne 8), mais présent lorsque les compétiteurs froid T et non-spécifique sont ajoutés (ligne 7, 9). Le complexe 4 est spécifique au deux sondes T et G. **B)** Intégration des données expérimentales de retard sur gel de la région haplotypique à l'étude. Les losanges démontrent la présence de liaison d'un facteur comme suit : losange bleu, liaison aux deux sondes; losange blanc, absence de liaison aux deux sondes et losange blanc et bleu, présence d'une liaison différentielle. Les résultats ont été observé dans au moins une lignée cellulaire.

### **3.3 Description de l'étude du gène CDKN1A**

#### **3.3.1 Identification des haplotypes régulateurs présents dans l'échantillon représentatif de la population mondiale.**

Pour le gène CDKN1A le génotypage de 11 rSNPs dans notre panel de population a mené à l'identification de 13 rHaps (tableau III). L'haplotype majeur H1, est présent chez près de 50% de la population et est réparti dans les différents groupes ethniques. Les 5 haplotypes les plus fréquents (H1, H2, H3, H8 et H12) ont été étudiés de façon plus approfondie par expériences de transfections transitoires, pour évaluer l'impact fonctionnel au niveau de l'activité du promoteur. Notons que l'haplotype H10 n'a pas été inclus dans la validation fonctionnelle, car il se retrouve uniquement chez les africains. Les rHaps H2, H3, H8 et H12 possèdent les Tag SNPs qui diffèrent de l'haplotype majeur H1 : -134G>C, -499G>C, -519G>C, -791T>C, -899T>G, -1021C>T, -1284T>C, -1466A>G, -1529T>C et -1536C>T (tableau III).

Tableau III : Liste des haplotypes régulateurs trouvés dans l'échantillon de populations de 40 individus pour le gène CDKN1A.

Gène	Haplotypes / SNPs												Population						
		-1731	-1536	-1529	-1466	-1284	-1021	-899	-791	-519	-499	-134	African	Amérindien	Européen	Asiatique	Moyent-Orientaux	Total	Freq (%)
CDKN1A	Ancestral	A	C	T	G	T	C	T	C	G	G	G							
	H1				A	-			T				4	7	11	5	10	37	46,3
	H2				A	-		G	T					1		3		4	5
	H3				A	-		G	-				3	2		2	1	8	10
	H4	-	-	-	A	-	-	G	-	C	C	-	1					1	1,3
	H5	-	-	-	A	C	-	-	-	-	-	-	1					1	1,3
	H6	-	-	-	A	C	-	G	T	-	-	-		1				1	1,3
	H7	-	-	-	A	C	-	G	-	-	-	-	1	1				2	2,5
	H8				A	C	-	G	-	C	C	-		1	2	3	2	8	10
	H9	-	-	-	A	C	T	-	-	-	-	C		2				2	2,5
	H10	-	-	C	-	C	-	-	-	-	-	-	5					5	6,3
	H11	-	-	C	-	C	T	-	-	-	-	C			1			1	1,3
	H12		T	C		G	T								2	3	3	8	10
	H13	DEL	-	-	A	C	-	-	-	C	C	-	1	1				2	2,5

Les zones grisées montrent les haplotypes qui ont été testés pour leur expression, en analyse de transfections transitoire.



### **3.3.2 Analyse fonctionnelle d'expression des rHaps, par expériences de transfections transitoires.**

L'activité promotrice différentielle associée aux 5 haplotypes a été étudiée à l'aide du système gène rapporteur luciférase, dans trois lignées cellulaires (Hela, Jeg3 et HepG2) (figure 3.3). Les haplotypes testés ont été qualifiés de promoteurs forts, puisqu'ils possèdent une activité supérieure à 10 fois la ligne de base. Dans les lignées cellulaires Hela et HepG2, l'activité des promoteurs est de l'ordre de 20 à 60 fois la ligne de base, tandis que dans les cellules Jeg3 l'activité est encore beaucoup plus forte atteignant 1000 fois la ligne de base. Pour l'haplotype H12, présent chez 10% de la population, une différence d'expression très importante a été remarquée dans les 3 lignées cellulaires. En effet, des différences de 2,5 fois; 28 fois et 4 fois l'haplotype majeur H1 ont été observées respectivement dans les lignées cellulaires Hela, Jeg3 et HepG2. Ces différences d'expression sont toutes significatives à  $P < 0,01$ . Également, il est important de mentionner que l'haplotype H12 possède une différence d'expression significativement plus faible que tous les haplotypes (H1, H2, H3 et H5). L'haplotype H2, associé à 5% de la population de notre échantillon, montre une baisse d'expression comparativement à l'haplotype majeur de 1,5 fois ( $P < 0,01$ ) dans les cellules de type Hela. Toutefois lorsque testé dans les deux autres lignées cellulaires (Jeg3 et HepG2), H2 possède une activité promotrice légèrement plus faible, quoique non significative, à l'haplotype H1. Les haplotypes H3 et H8 possèdent également une activité similaire à l'haplotype majeur dans les trois lignées cellulaires.

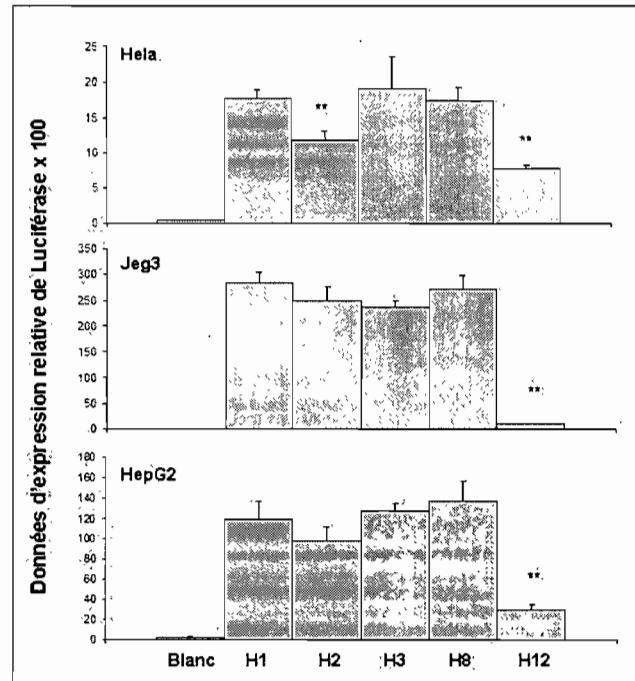


Figure 3.3 : Résultat d'analyse d'expression des 5 haplotypes régulateurs testés pour le gène CDKN1A par transfections transitoires. Le niveau d'expression de chacun des haplotypes est donné sous la forme d'un ratio de la luciférase mesurée Luciole/Rénilla, multiplié par 100. Les haplotypes fréquents, des différents gènes analysés par transfections transitoires, sont comparés à l'haplotype majeur du même gène. Les analyses de transfections sont faites dans trois lignées cellulaires différentes (Hela, Jeg3 et HepG2). Les différences d'expression significatives ont été testées par le test de Lord et les valeurs sont indiquées (\* :  $P < 0,05$ ; \*\* :  $P < 0,01$ ).

### **3.3.3 Analyse des modifications de liaison des facteurs de transcription, dû à la présence de polymorphismes régulateurs, dans des expériences de retard sur gel.**

Nous avons trouvé dans les analyses d'expression du gène CDKN1A que les haplotypes H2 et H12 présentaient des différences significatives d'expression lorsque comparés à l'haplotype majeur H1. Également, nous devons mentionner que les haplotypes H3 et H8 démontrent une expression significativement différente de l'haplotype H12. Afin d'évaluer les modifications de liaisons des facteurs de transcription sur ces haplotypes (H2 et H12), nous avons soumis les polymorphismes: -134G>C, -791T>C, -899T>G, -1021C>T, -1284T>C, -1466A>G, -1529T>C et -1536C>T à des expériences de retard sur gel. Les prédictions de la présence de sites de liaison de facteurs de transcription sont indiquées au tableau XIX pour chacun des variants. Cinq des huit variants testés par expériences de retard sur gel démontrent la présence de liaison différentielle d'un facteur dans au moins 1 des 3 lignées cellulaires (-134G>C, -791T>C, -899T>G, -1021C>T et -1284T>C). Les variants -899T>G et -1021C>T montrent des liaisons différentielles en présence des extraits nucléaires des 3 lignées cellulaire (Hela, Jeg3 et HepG2). Le polymorphisme -899T>G montré dans la figure 3.4, produit 4 complexes spécifiques ADN-protéine. Les complexes 1 et 2 sont associés à l'allèle T (ligne 1). Ces deux complexes sont liés à l'allèle T et sont compétitionnés lorsque la sonde froide compétitrice T est ajoutée (ligne 3), mais demeurent présents et inchangés en présence des compétiteurs froid, G (ligne 4) et non-spécifique (ligne 5). Les complexes 3 et 4 formés, sont par ailleurs associés à l'allèle G du variant -899T>G. Ceux-ci sont compétitionnés seulement en présence du compétiteur froid G (ligne 8). Les prédictions informatiques n'indiquent pas quel facteur possède un site de liaison aux endroits où les polymorphismes -134G>C et -1284T>C sont situés, ce qui laisse place à une possible association avec un nouveau facteur de transcription.

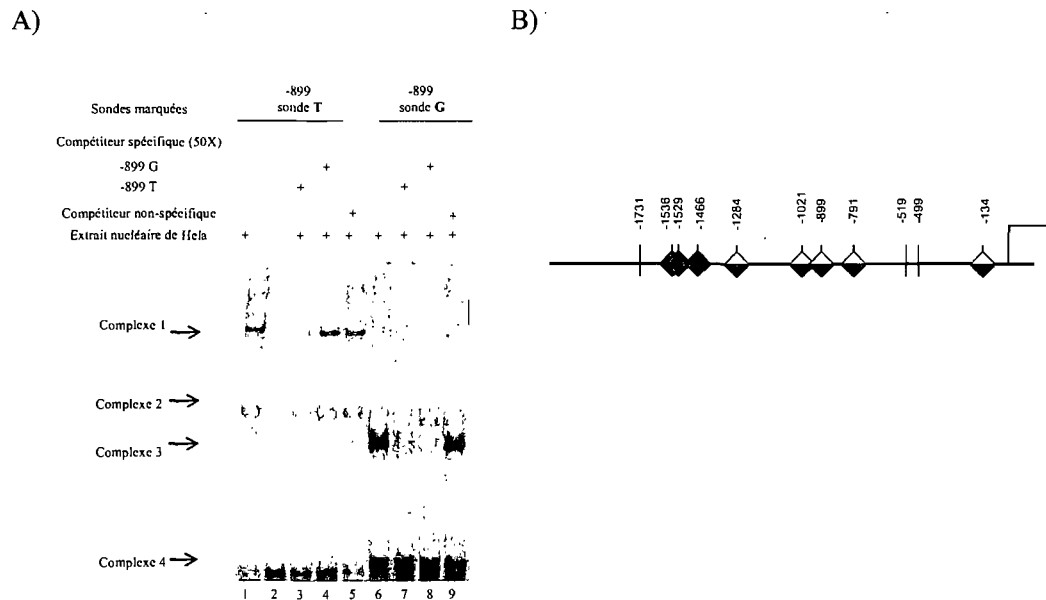


Figure 3.4 : **A)** Résultat d'analyse de retard sur gel pour le variant  $-899T>G$  du gène *CDKN1A*. En présence de l'extrait nucléaire de cellules Hela, 4 différents complexes sont formés. La ligne 2 représente le contrôle négatif, sans extrait nucléaire. Les 4 différents complexes sont spécifiques. Les complexes 1 et 2 sont associés à la sonde T. Ces complexes sont compétitionnés lorsque le compétiteur spécifique froid T 50x est ajouté (ligne 3) et demeurent présents lorsque les compétiteurs spécifiques sonde opposée, G (ligne 4) et non-spécifique (ligne 5) sont ajoutés. Les complexes 3 et 4 sont à l'opposé spécifiques à l'allèle G. Ils se lient à la sonde marquée G (ligne 6), sont beaucoup moins visible lorsque le compétiteur froid spécifique G est ajouté (ligne 8), mais présents lorsque les compétiteurs froid T et non-spécifique sont ajoutés (ligne 7, 9). **B)** Intégration des données expérimentales de retard sur gel de la région haplotypique à l'étude. Les losanges démontrent la présence de liaison d'un facteur comme suit : losange bleu, liaison aux deux sondes; losange blanc, absence de liaison aux deux sondes et losange blanc et bleu, présence d'une liaison différentielle. Les résultats ont été observé dans au moins une lignée cellulaire.

### **3.4 Description de l'étude du gène CDKN1B**

#### **3.4.1 Identification des haplotypes régulateurs présents dans l'échantillon représentatif de la population mondiale.**

Le génotypage de 10 rSNPs a mené à la construction de 11 différents haplotypes, dont les plus fréquents H1 et H3 qui représentent respectivement 31,3% et 32,5% de la population (tableau IV). Ce qui signifie qu'ensemble ces deux haplotypes comptent près de 65% de tous les groupes ethniques. L'haplotype H3 sera considéré comme haplotype majeur. Pour les études de validation fonctionnelle, 4 haplotypes (H1, H3, H7 et H8) ont été clonés dans des vecteurs pGL3-basic contenant le gène rapporteur luciférase de luciole et analysés par transfections transitoires pour déterminer et comparer leur activité promotrice. L'haplotype H4 présent chez 10% de la population n'a pu être analysé par transfections transitoires dû aux difficultés rencontrées lors du clonage.

Tableau IV : Liste des haplotypes régulateurs trouvés dans l'échantillon de populations de 40 individus pour le gène CDKN1B.

Gène	Haplotypes / SNPs											Population						
	-1908	-1857	-1608	-1228	-1084	-756	-522	-373	-337	-221	Africain	Amérindien	Européen	Asiatique	Moyent-Orientaux	Total	Freq (%)	
CDKN1B	Ancestral	T	C	G	C	AAG	A	G	G	A	T							
	H1	-	-	-	-	-	-	-	-	-	-	1	4	6	11	3	25	31.3
	H2	-	-	-	-	-	-	-	-	C	-	1			1	2	2.5	
	H3	-	-	-	-	-	-	T	-	-	-	2	5	8	5	6	26	32.5
	H4	-	-	-	-	-	-	A	-	C	-	4	2	2		8	10.0	
	H5	-	T	-	-	-	-	-	-	C	C	1				1	1.3	
	H6	-	-	A	-	-	-	-	-	-	-	2				2	2.5	
	H7	-	-	A	-	-	-	C	-	-	-		2		1	3	3.8	
	H8	T	-	-	-	-	-	-	-	C	-	2	3		5	10	12.5	
	H9	DEL	-	-	-	-	-	-	-	-	-	1				1	1.3	
	H10	DEL	-	-	T	-	-	-	-	-	-	1				1	1.3	
	H11	DEL	-	-	-	DEL	-	-	-	-	-	1				1	1.3	

Les zones grisées montrent les haplotypes qui ont été testés pour leur expression, en analyse de transfections transitoire.

### **3.4.2 Analyse fonctionnelle d'expression des rHaps, par expériences de transfections transitoires.**

Nous avons validé l'impact fonctionnel des 4 haplotypes (H1, H3, H7 et H8) qui représentent ensemble 80% de la variation interindividuelle présente dans la population (tableau IV). Ces 4 haplotypes régulateurs sont associés à une haute activité promotrice car ils mènent à des niveaux d'expression supérieurs de luciférase à 100 fois la ligne de base (figure 3.5). L'haplotype H8 possède une activité promotrice inférieure aux autres haplotypes. À titre d'exemple, H8 a démontré une différence d'expression significative ( $P < 0,01$ ) de 1,5 fois le majeur (H3) dans la lignée cellulaire Hela et 2 fois dans la lignée Jeg3. La lignée cellulaire HepG2 montre pour l'haplotype H8 une différence d'expression très près de 1,5 fois comparée à l'haplotype majeur, mais toutefois inférieure au seuil d'acceptation requis (1,5 fois). Les haplotypes H1 et H7, possèdent une activité similaire à l'haplotype majeur H3 dans les 3 lignées cellulaires.

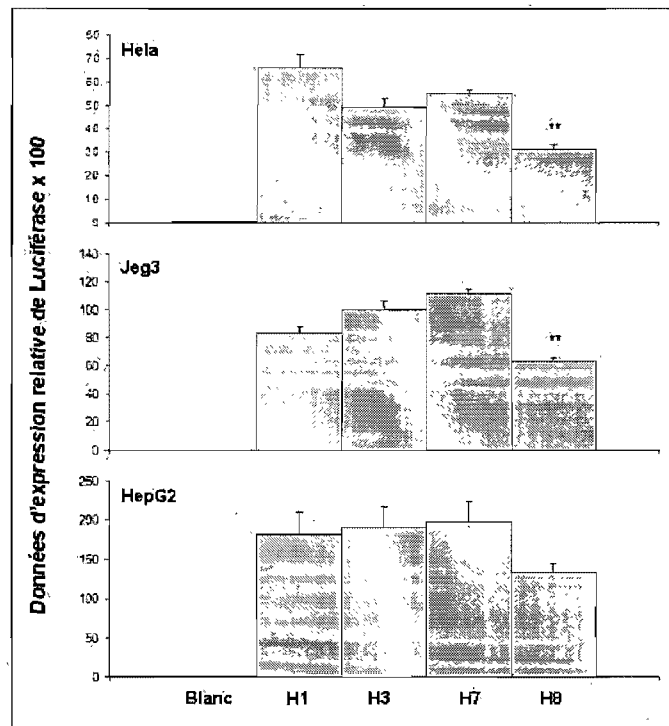


Figure 3.5 : Résultat d'analyse d'expression des 4 haplotypes régulateurs testés pour le gène CDKN1B par transfections transitoires. Le niveau d'expression de chacun des haplotypes est donné sous la forme d'un ratio de la luciférase mesurée Luciole/Rénilla, multiplié par 100. Les haplotypes fréquents, des différents gènes analysés par transfections transitoires, sont comparés à l'haplotype majeur du même gène. Les analyses de transfections sont faites dans trois lignées cellulaires différentes (Hela, Jeg3 et HepG2). Les différences d'expression significatives ont été testées par le test de Lord et les valeurs sont indiquées (\* :  $P < 0,05$ ; \*\* :  $P < 0,01$ ). Seul l'haplotype H8 montre une expression différentielle comparée à l'haplotype majeur H3 ( $P < 0,01$ ) dans les lignées cellulaires Hela et Jeg3.



### **3.4.3 Analyse des modifications de liaison des facteurs de transcription, dû à la présence de polymorphismes régulateurs, dans des expériences de retard sur gel.**

Nous avons vérifié la présence de liaisons différentielles de facteurs de transcription pour les variants présents dans l'haplotype H8 qui possède un niveau d'expression significativement différent aux autres haplotypes notamment H3. Les variants -337C>A, -373G>T et -1857C>T présents uniquement sur l'haplotype H8 comparé au majeur ont été testés (tableau IV). Les prédictions informatiques par le logiciel *Mat Inspector* nous montrent que plusieurs facteurs de transcription possèdent des sites de liaison qui touchent aux variants -373G>T et -1857C>T de la région régulatrice du gène CDKN1B, alors qu'aucun site de liaison n'est prédit à l'endroit où le variant -337C>A est situé (tableau XIX). Dans les expériences de retard sur gel, les rSNPs -373G>T et -1857C>T ne semblent pas affecter la liaison à un facteur de transcription. Aucune liaison différentielle n'est démontrée en présence des 3 extraits nucléaires Hela, Jeg3 et HepG2 par expériences de retard sur gel. Par contre, le rSNP -337C>A montre une liaison différentielle en présence de chacun des 3 extraits nucléaires. En présence de l'extrait nucléaire de Jeg3, 5 complexes ADN-protéine sont formés sur le variant -337C>A (figure 6). Le complexe 1, spécifique à l'allèle A (ligne 6), est compétitionné par l'ajout du compétiteur spécifique froid A (ligne 8) et est partiellement compétitionné par le compétiteur C (ligne 7). Le compétiteur non-spécifique utilisé étant de l'ADN de thymus de veau, il compétitionne totalement tout les complexes formés (ligne 5 et 9). Donc, le complexe 1 est inapte à lier l'allèle C (ligne 1-5). Les complexes 4 et 5 sont également des complexes spécifiques, mais ceux-ci sont associés à l'allèle C (ligne 1 versus 6). Les complexes 2 et 3 sont des complexes non-spécifiques.

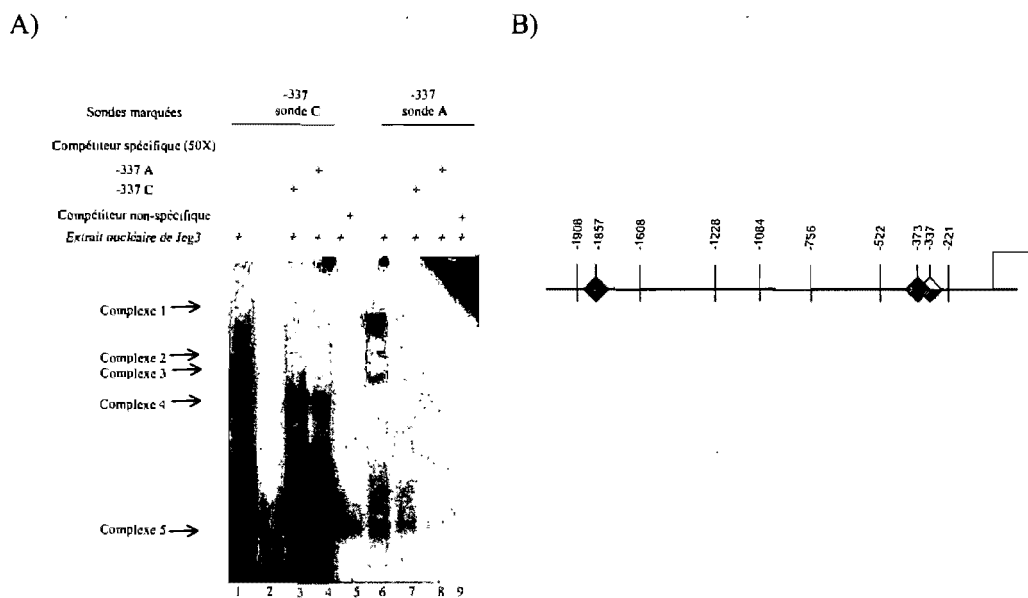


Figure 3.6 : **A)** Résultat d'analyse de retard sur gel pour le variant -337C>A du gène CDKN1B. En présence de l'extrait nucléaire de cellules Jeg3, 5 différents complexes sont formés. La ligne 2 représente le contrôle négatif, sans extrait nucléaire. Les complexes 1, 4 et 5 sont spécifiques. Le complexe 1 est associé à l'allèle A du variant -337 (ligne 6). Ce complexe spécifique est compétitionné lors de l'ajout du compétiteur froid A (ligne 8) et demeure présent, mais diminué partiellement, lorsque le compétiteur froid C est ajouté (ligne 7). Le compétiteur non-spécifique utilisé dans cette expérience (ADN de thymus de veau) compétitionne tous les complexes, ce compétiteur est inapproprié. Les complexes 4 et 5 sont spécifiques à l'allèle C (ligne 1-5) et ne se lient pas à l'allèle A (ligne 6-9). Les complexes 2 et 3 sont non-spécifiques. **B)** Intégration des données expérimentales de retard sur gel de la région haplotypique à l'étude. Les losanges démontrent la présence de liaison d'un facteur comme suit : losange bleu, liaison aux deux sondes; losange blanc, absence de liaison aux deux sondes et losange blanc et bleu, présence d'une liaison différentielle. Les résultats ont été observé dans au moins une lignée cellulaire.

### **3.5 Description de l'étude du gène CDKN2A**

#### **3.5.1 Identification des haplotypes régulateurs présents dans l'échantillon représentatif de la population mondiale.**

Le génotypage de 8 rSNPs a permis d'identifier 7 haplotypes régulateurs. L'haplotype H1 est une copie identique à l'haplotype ancestral identifié à partir de l'ADN de chimpanzé. Cet haplotype (H1) est bien réparti à travers les différents groupes ethniques (tableau V). Seuls les 2 rHaps (H1 et H2) les plus fréquents ont été testés dans les analyses fonctionnelles. L'haplotype H1 représente 80% de la population comparativement à l'haplotype H2 qui représente 7,5% de la population.

Tableau V : Liste des haplotypes régulateurs trouvés dans l'échantillon de populations de 40 individus pour le gène CDKN2A.

Gène	Haplotypes / SNPs									Population						
		-1879	-1600	-710	-598	-589	-464	-222	-44	Africain	Amérindien	Européen	Asiatique	Moyent-Orientaux	Total	Freq (%)
CDKN2A	Ancestral	C	A	C	G	G	C	T	A							
	H1	-	-	-	-	-	-	-	-	15	10	14	12	13	64	80.0
	H2	-	-	-	-	-	-	A	-		4	2			6	7.5
	H3	-	-	-	-	-	T	-	-		1			1	2	2.5
	H4	-	-	-	A	DEL	-	-	T				4		4	5.0
	H5	-	-	A	-	-	-	-	-		1			1	2	2.5
	H6	-	G	-	-	-	-	-	-	1					1	1.3
	H7	T	-	-	-	-	-	-	-					1	1	1.3

Les zones grisées montrent les haplotypes qui ont été testés pour leur expression, en analyse de transfections transitoire.

### **3.5.2 Analyse fonctionnelle d'expression des rHaps, par expériences de transfections transitoires.**

Nous avons testé par transfection transitoire l'expression du gène-rapporteur luciférase placé en aval des rHaps H1 et H2, dans le vecteur pGL3-basic. L'haplotype fréquent H2 a démontré une activité supérieure à celle du niveau de base, de 6 fois dans la lignée cellulaire Hela et de 8 fois dans la lignée cellulaire HepG2 (figure 3.7). L'haplotype majeur H1 démontre une faible activité promotrice telle qu'illustré par une expression de la luciférase seulement 2 fois au dessus du niveau de base dans les lignées cellulaires Hela et HepG2. L'haplotype H2 présente une activité significativement supérieure ( $P < 0,01$ ) à l'haplotype majeur H1 dans les lignées Hela (2,8 fois) et HepG2 (3,7 fois). Dans la lignée Jeg3, toutes les valeurs d'expression trouvées sont basses. En d'autres termes, les haplotypes H1 et H2 sont peu ou pas exprimés dans la lignée cellulaire Jeg3.

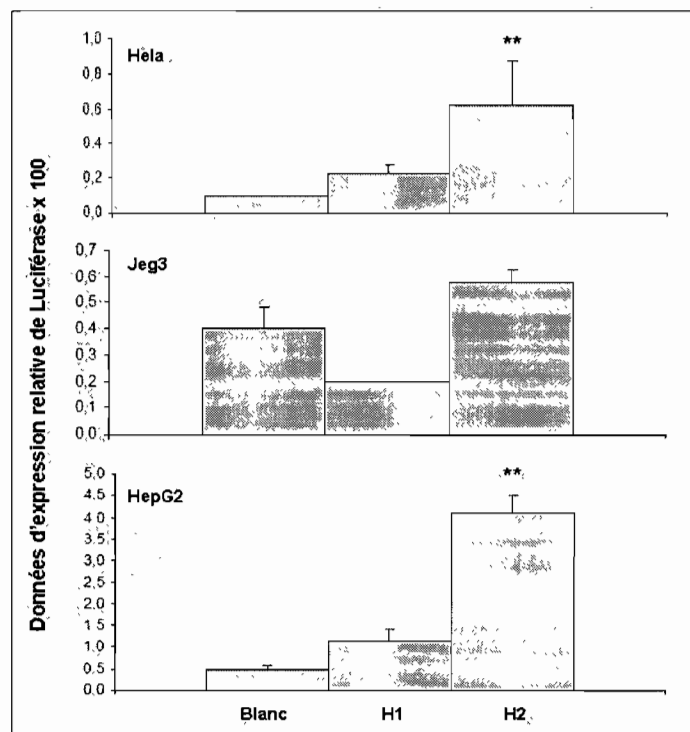


Figure 3.7 : Résultat d'analyse d'expression des 2 haplotypes régulateurs testés pour le gène CDKN2A par transfections transitoires. Le niveau d'expression de chacun des haplotypes est donné sous la forme d'un ratio de la luciférase mesurée Luciole/Rénilla, multiplié par 100. Les haplotypes fréquents, des différents gènes analysés par transfections transitoires, sont comparés à l'haplotype majeur du même gène. Les analyses de transfections sont faites dans trois lignées cellulaires différentes (Hela, Jeg3 et HepG2). Les différences d'expression significatives ont été testées par le test de Lord et les valeurs sont indiquées (\* :  $P < 0,05$ ; \*\* :  $P < 0,01$ ). L'haplotype H2 est comparé pour les résultats d'expression à l'haplotype majeur H1. H2 montre une expression supérieure à l'haplotype H1 de 2,5 fois dans la lignée cellulaire Hela et 3,5 fois dans la lignée cellulaire HepG2 ( $P < 0,01$ ).

### **3.5.3 Analyse des modifications de liaison des facteurs de transcription, dû à la présence de polymorphismes régulateurs, dans des expériences de retard sur gel.**

Nous avons trouvé dans les expériences de transfections transitoires une différence d'expression significative entre l'haplotype majeur H1 et l'haplotype H2. Nous avons donc analysé la liaison des facteurs de transcription aux endroits où les variants, présents dans l'haplotype H2 et différant de l'haplotype H1, se trouvaient. L'unique position qui diffère entre les deux haplotypes étant -222T>A, ce variant est le seul qui est été testé par expériences de retard sur gel (figure 3.8). Les prédictions informatiques ont déterminées une perte de liaison d'un facteur de transcription dû à la présence du rSNP-222 T>A (tableau XIX). Les résultats d'expériences de retard sur gel démontrent une ou plusieurs liaisons différentielles en présence des extraits nucléaires de cellules Hela et HepG2 (tableau XIX). Le figure 3.8 montre le résultat de l'expérience de retard sur gel en présence de l'extrait nucléaire de cellule HepG2. Ce résultat démontre la formation de 5 complexes ADN-protéine. Seuls les complexes 1 et 5 semblent démontrer une tendance à être spécifiques. Cette tendance est associée à l'allèle A (lignes 6-9). Nous mentionnons une tendance, car l'intensité de la présence du complexe est plus grande lorsqu'il est associé à la sonde marquée A pour laquelle il montre une tendance, cependant cette intensité varie peu lorsque les compétiteurs sont ajoutés (lignes 7 et 8). La ligne 7 qui contient le compétiteur froid T et la ligne 8 qui contient le compétiteur spécifique A froid varient peu pour ces complexes, alors que ceux-ci devraient être compétitionnés presque totalement lorsque le compétiteur A (ligne 8) est ajouté. Le compétiteur non spécifique utilisé étant de l'ADN de thymus de veau, celui-ci fait disparaître les complexes 1 et 5 en le compétitionnant (ligne 9). Les complexes 2, 3 et 4 sont non spécifiques, car les complexes 2 et 4 sont présents dans tous les puits, à l'exception du puit contrôle négatif (ligne 2). Le complexe 3 est à l'inverse compétitionné par tous les compétiteurs ajoutés, c'est-à-dire compétiteurs spécifiques (lignes 3-4 et 7-8) et non spécifiques (lignes 5 et 9), en présence des deux différentes sondes chaudes.

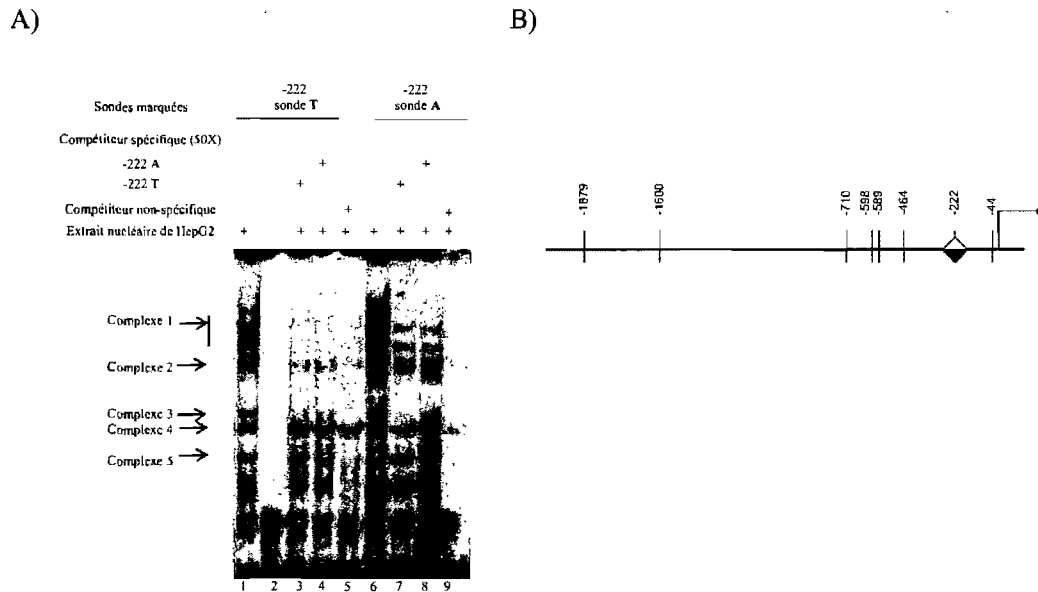


Figure 3.8 : **A)** Résultat d'analyse de retard sur gel pour le variant -222T>A du gène CDKN2A. En présence de l'extrait nucléaire de cellule HepG2, 5 différents complexes sont formés. La ligne 2 représente le contrôle négatif, sans extrait nucléaire. Le compétiteur non-spécifique utilisé dans cette expérience (ADN de thymus de veau) compétitionne tous les complexes, ce compétiteur est inapproprié. **B)** Intégration des données expérimentales de retard sur gel de la région haplotypique à l'étude. Les losanges démontrent la présence de liaison d'un facteur comme suit : losange bleu, liaison aux deux sondes; losange blanc, absence de liaison aux deux sondes et losange blanc et bleu, présence d'une liaison différentielle. Les résultats ont été observé dans au moins une lignée cellulaire.



### **3.6 Description de l'étude du gène MADH3**

#### **3.6.1 Identification des haplotypes régulateurs présents dans l'échantillon représentatif de la population mondiale.**

Le génotypage de 5 rSNPs, a permis d'identifier 5 haplotypes régulateurs dans notre échantillon de population. L'haplotype majeur H1 représente environ 40% de la population et les haplotypes H2, H3 et H5 se répartissent près de 50% du reste de la population de façon assez uniforme (tableau VI). Seul l'haplotype H4, peu fréquent à 2,5%, n'a pas été analysé par les expériences de transfections transitoires. D'ailleurs, ce dernier n'est observé que chez les asiatiques.

Tableau VI : Liste des haplotypes régulateurs trouvés dans l'échantillon de populations de 40 individus pour le gène MADH3.

Gène	Haplotypes / SNPs					Population						Freq (%)	
	-1946	-1938	-1137	-749	-669	Africain	Amérindien	Européen	Asiatique	Moyent-Orientaux	Total		
MADH3	Ancestral	C	T	T	G	C							
	H1	-	-	-	-	-	9	8	6	6	6	35	43.8
	H2	-	-	-	A	-		4	4	2	10	12.5	
	H3	-	C	-	-	T	4	2	4	2	4	16	20.0
	H4	-	C	G	-	T			2		2	2.5	
	H5	A	-	-	-	-	3	2	2	6	4	17	21.3

Les zones grisées montrent les haplotypes qui ont été testés pour leur expression, en analyse de transfections transitoire

### **3.6.2 Analyse fonctionnelle d'expression des rHaps, par expériences de transfections transitoires.**

Nous avons analysé par transfections transitoires les 4 haplotypes régulateurs fréquents du gène MADH3 (H1, H2, H3 et H5). Les rHaps H1, H2 et H3 sont associés à une activité promotrice similaire (figure 3.9). Par contre, dans les 3 lignées cellulaires testées on note une plus grande expression des gènes analysés, plus particulièrement dans Jeg3 et Hela. L'haplotype H5 montre en effet une différence significative à  $P < 0,01$  dans les lignées cellulaires Hela et Jeg3. H5 montre des données d'expression relative du gène luciférase 2 fois supérieure à l'haplotype majeur H1 dans les cellules Hela; 1,6 fois supérieure dans les cellules Jeg3 et 1,3 fois supérieure dans les cellules HepG2. Les niveaux d'activité promotrice varient beaucoup entre les trois lignées cellulaires pour l'ensemble des rHaps. Les cellules Jeg3 montrent des promoteurs très forts, car le niveau d'expression pour l'ensemble des haplotypes est de l'ordre de 20 à 30 fois supérieur au niveau de base. La lignée HepG2 montre un ensemble de promoteurs forts (entre 5 et 10 fois le niveau d'expression de la ligne de base). Par contre, l'activité promotrice présentée dans la lignée cellulaire Hela est faible et les haplotypes H1, H2 et H3 sont inactifs, puisqu'ils présentent une activité promotrice de 4 fois l'activité de la ligne de base. L'étude de la liaison possible de facteurs de transcription a été effectuée sur l'haplotype H5 qui présente des données d'expression significativement différentes des autres haplotypes.

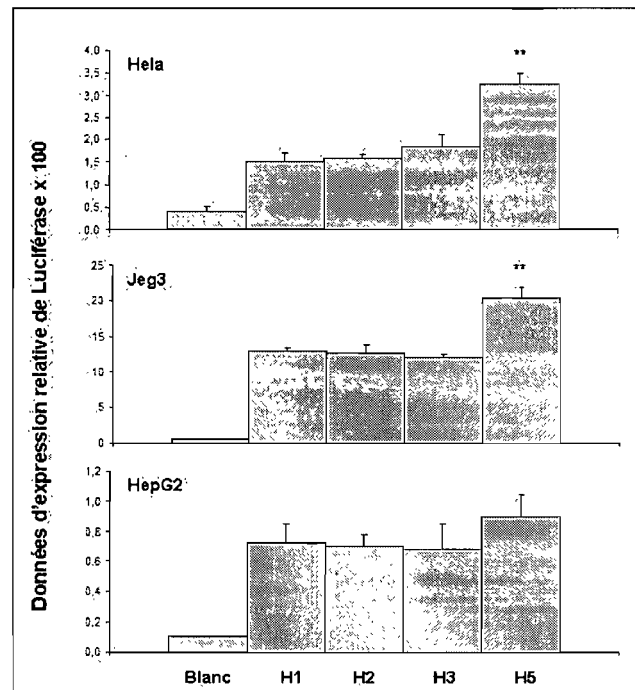


Figure 3.9 : Résultat d'analyse d'expression des 4 haplotypes régulateurs testés pour le gène MADH3 par transfections transitoires. Le niveau d'expression de chacun des haplotypes est donné sous la forme d'un ratio de la luciférase mesurée Luciole/Rénilla, multiplié par 100. Les haplotypes fréquents, des différents gènes analysés par transfections transitoires, sont comparés à l'haplotype majeur du même gène. Les analyses de transfections sont faites dans trois lignées cellulaires différentes (Hela, Jeg3 et HepG2). Les différences d'expression significatives ont été testées par le test de Lord et les valeurs sont indiquées (\* :  $P < 0,05$ ; \*\* :  $P < 0,01$ ). Les haplotypes fréquents, des différents gènes analysés par transfections transitoires, sont comparés à l'haplotype majeur du même gène.

### **3.6.3 Analyse des modifications de liaison des facteurs de transcription, dû à la présence de polymorphismes régulateurs, dans des expériences de retard sur gel.**

Nous avons analysé la présence de possibles liaisons différentielles des facteurs de transcription aux sites où les variants présents dans l'haplotype H5 diffèrent de ceux présents dans l'haplotype majeur H1. Un seul variant diffère entre l'haplotype majeur et H5 (-1946C>A). Les prédictions informatiques prévoient que le facteur de transcription Ets1 peut se lier à l'allèle A du variant. Le tableau XIX montre qu'une liaison différentielle est observée dans chacune des trois lignées cellulaires testées. La figure 3.10 montre qu'en présence de l'extrait nucléaire de cellules HepG2, 3 complexes globaux sont formés. Les lignes de complexes, qui apparaissent sur le film de retard sur gel; qui sont rapprochées et qui suivent la même tendance dans chacun des puits, sont considérées comme un seul complexe large. Les complexes 2 et 3 sont des complexes larges (globaux). Les complexes 1 et 2 sont spécifiquement associés à l'allèle A (ligne 6 versus 1). Ces complexes sont compétitionnés lors de l'ajout du compétiteur spécifique froid A (ligne 8) et demeurent présents lorsque les compétiteurs froids, T (ligne 7) et non spécifique (ligne 9) sont ajoutés.

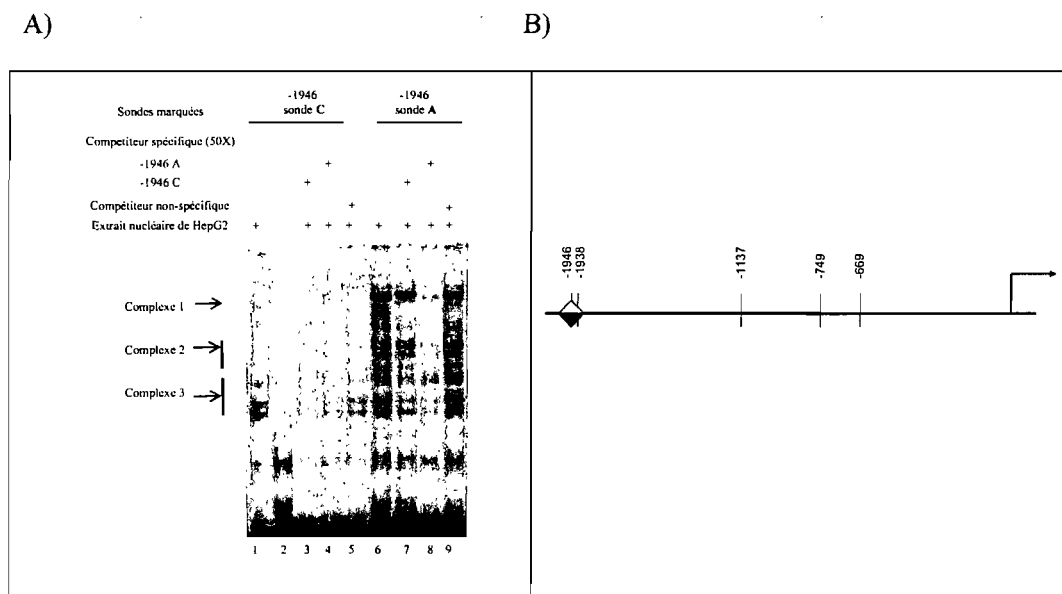


Figure 3.10 : **A)** Résultat d'analyse de retard sur gel pour le variant -1946C>A du gène MADH3. En présence de l'extrait nucléaire de cellule HepG2, 3 complexes globaux sont formés. La ligne 2 représente le contrôle négatif, sans extrait nucléaire. Un contrôle positif (non montré) est effectué pour assurer la validité de l'expérience. Deux des trois complexes formés démontrent de façon claire une spécificité envers la sonde A, complexes 1 et 2. Le complexe 3 présente une intensité similaire dans tous les puits, ce qui montre sa non-spécificité. **B)** Intégration des données expérimentales de retard sur gel de la région haplotypique à l'étude. Les losanges démontrent la présence de liaison d'un facteur comme suit : losange bleu, liaison aux deux sondes; losange blanc, absence de liaison aux deux sondes et losange blanc et bleu, présence d'une liaison différentielle. Les résultats ont été observé dans au moins une lignée cellulaire.

### **3.7 Description de l'étude du gène TFDP1**

#### **3.7.1 Identification des haplotypes régulateurs présents dans l'échantillon représentatif de la population mondiale.**

Suite à l'analyse des génotypes de 8 rSNPs présents dans la région promotrice de 2 Kb, 4 haplotypes régulateurs ont pu être identifiés (tableau VII). L'haplotype le plus fréquent H1, est présent chez 82,5% de la population. Cet haplotype est identique, pour chacun des variants du gène TFDP1, à l'haplotype ancestral. Les sous-populations amérindienne et européenne contiennent uniquement l'haplotype majeur. Les haplotypes H2, H3 et H4, sont répartis dans la population à des fréquences très similaires et ces fréquences sont de beaucoup inférieures à la fréquence de l'haplotype H1. Les haplotypes fréquents ont été évalués par analyse fonctionnelle, cependant l'haplotype H2 qui est le second plus fréquent avec une fréquence de 6% dans la population, n'a pu être étudié au cours des analyses fonctionnelles, puisqu'il a présenté des difficultés lors du clonage. Les différences d'expression des haplotypes H1, H3 et H4 ont été étudiées par expériences de transfections transitoires.

Tableau VII : Liste des haplotypes régulateurs trouvés dans l'échantillon de populations de 40 individus pour le gène TFDP1.

Gène	Haplotypes / SNPs									Population						
		-1887	-1883	-1779	-1655	-1530	-1516	-1515	-1394	Africain	Amérindien	Européen	Asiatique	Moyent-Orientaux	Total	Freq (%)
	Ancestral	C	G	A	G	C	C	G	T							
	H1	-	-	-	-	-	-	-	-	9	16	16	11	14	66	84.6
	H2	-	-	-	-	T	T	T	-	2			3		5	6.4
	H3	-	-	G	-	-	-	-	-				1	2	3	3.8
	H4	T	C	-	A	-	-	-	C	3			1		4	5.1

Les zones grisées montrent les haplotypes qui ont été testés pour leur expression, en analyse de transfections transitoire



### **3.7.2 Analyse fonctionnelle d'expression des rHaps, par expériences de transfections transitoires.**

Les haplotypes régulateurs H1, H3 et H4 ont été analysés pour détecter les différences significatives d'expression, comparativement à l'haplotype majeur H1 du gène TFDPI. Les données d'expression relative (figure 3.11) montrent que le gène TFDPI possède des haplotypes régulateurs actifs, les haplotypes H3 et H4 (activité supérieure entre 5 et 10 fois au niveau de base), et que l'haplotype majeur H1 est un promoteur fort (expression supérieure à 10 fois la ligne de base). La tendance d'expression reste la même dans les trois lignées cellulaires pour l'ensemble des haplotypes. Les deux haplotypes testés, H3 et H4, présentent une différence d'expression significative ( $P < 0,01$ ) dans les lignées cellulaires Hela et Jég3, mais également dans la lignée HepG2 pour l'haplotype H4. Cette différence significative ce chiffre à environ 2 fois l'haplotype majeur pour les deux haplotypes régulateurs H3 et H4.

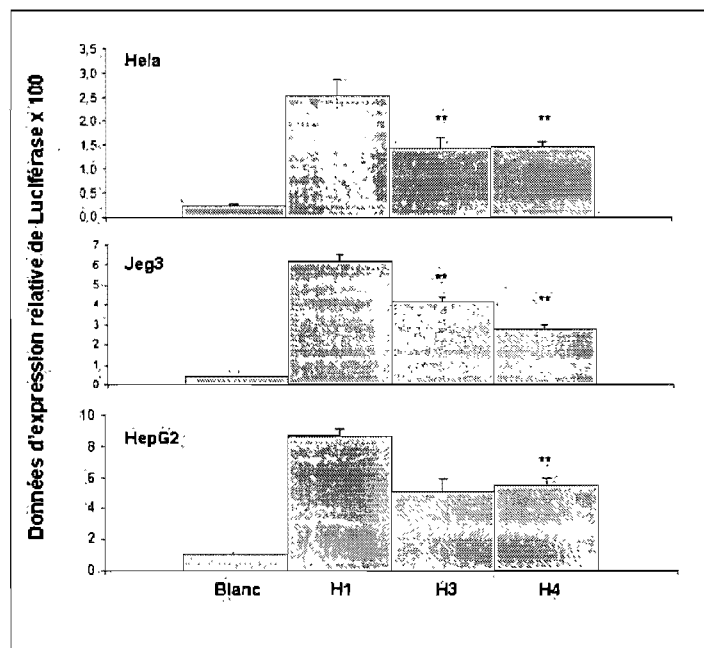


Figure 3.11 : Résultat d'analyse d'expression des 3 haplotypes régulateurs testés pour le gène TFDP1 par transfections transitoires. Le niveau d'expression de chacun des haplotypes est donné sous la forme d'un ratio de la luciférase mesurée Luciole/Rénilla, multiplié par 100. Les haplotypes fréquents, des différents gènes analysés par transfections transitoires, sont comparés à l'haplotype majeur du même gène. Les analyses de transfections sont faites dans trois lignées cellulaires différentes (Hela, Jeg3 et HepG2). Les différences d'expression significatives ont été testées par le test de Lord et les valeurs sont indiquées (\* :  $P < 0,05$ ; \*\* :  $P < 0,01$ ).

### **3.7.3 Analyse des modifications de liaison des facteurs de transcription, dû à la présence de polymorphismes régulateurs, dans des expériences de retard sur gel.**

Les analyses d'expression ont démontrées des différences significatives entre l'haplotype majeur et les deux haplotypes régulateurs H3 et H4. Pour évaluer la présence de liaisons différentielles de facteurs de transcription reliées à ces deux haplotypes, les variants -1394T>C, -1655G>A, -1779A>G, -1883G>C et -1887C>T ont été étudiés par expériences de retard sur gel. Le tableau XIX montre les prédictions informatiques de liaisons différentielles de facteurs de transcription pour les rSNPs étudiés. Tous les variants testés du gène TFDPI ont démontrés une liaison différentielle dans au moins une des trois lignées cellulaires testées, à l'exception du polymorphisme -1779A>G. Le rSNP -1779A>G qui n'a obtenu aucune liaison différentielle de facteur de transcription, ne possédait aucune prédiction informatique de liaison d'un facteur. Par ailleurs, le rSNP -1394T>C qui ne possède également aucune prédiction informatique de liaison d'un facteur de transcription, démontre en présence des extraits nucléaires des trois lignées cellulaires une ou plusieurs liaisons différentielles. La figure 3.12 montre le résultat de l'expérience de retard sur gel pour le variant -1887C>T, sur lequel deux facteurs de transcription sont aptes à lier l'allèle T (tableau XIX). En présence de l'extrait nucléaire de cellules HepG2, 7 complexes ADN-protéine sont formés. Les complexes spécifiques 1, 2, 3 et 5 sont associés à l'allèle T (ligne 6-9). Ces complexes sont compétitionnés lorsque que le compétiteur froid T (ligne 8) est ajouté et demeurent présents lors de l'ajout des compétiteurs C (ligne 7) et non-spécifique (ligne 9). Les complexes 1, 2, 3 sont aptes à lier l'allèle C (ligne 1), mais sont compétitionnés par l'ajout des compétiteurs froids C et T (ligne 3 et 4). Les complexes 4, 6 et 7 sont non spécifiques.

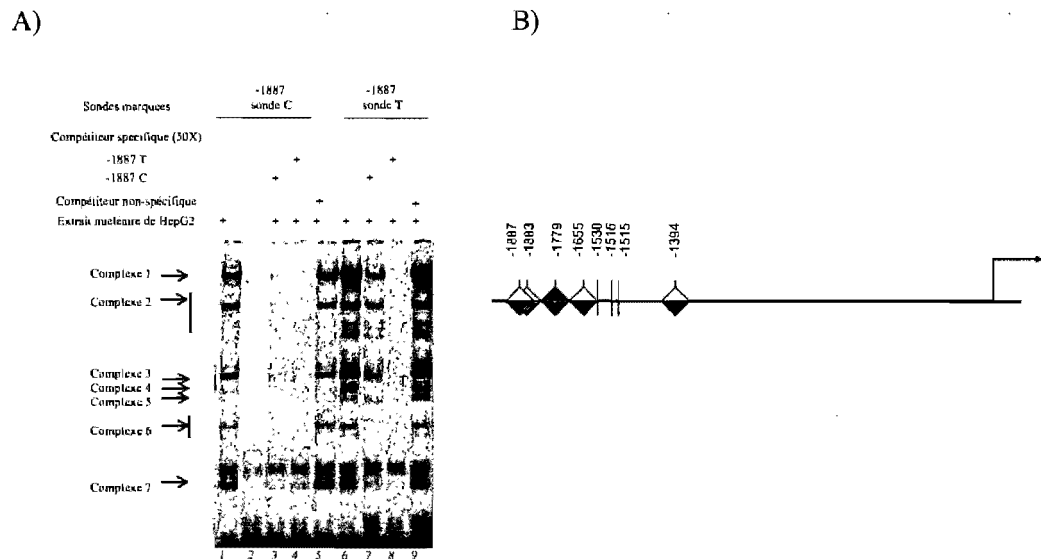


Figure 3.12 : **A)** Résultat d'analyse de retard sur gel pour le variant  $-1887C>T$  du gène *TFDP1*. En présence de l'extrait nucléaire de cellule HepG2, 7 différents complexes sont formés. La ligne 2 représente le contrôle négatif, sans extrait nucléaire. Un contrôle positif (non montré) est effectué pour assurer la validité de l'expérience. Les complexes 1, 2, 3 et 5 sont spécifiques à l'allèle T. Les complexes 4, 6 et 7 sont non spécifiques. Le complexe formé au dessus du complexe 7 n'est pas un complexe spécifique, puisqu'il demeure présent même en absence d'extrait nucléaire (ligne 2). **B)** Intégration des données expérimentales de retard sur gel de la région haplotypique à l'étude. Les losanges démontrent la présence de liaison d'un facteur comme suit : losange bleu, liaison aux deux sondes; losange blanc, absence de liaison aux deux sondes et losange blanc et bleu, présence d'une liaison différentielle. Les résultats ont été observé dans au moins une lignée cellulaire.

### **3.8 Description de l'étude du gène TGF $\beta$ 1**

#### **3.8.1 Identification des haplotypes régulateurs présents dans l'échantillon représentatif de la population mondiale.**

Le génotypage de 9 rSNPs a mené à l'identification de 9 haplotypes régulateurs (tableau VIII). Deux des 9 rHaps peuvent être considérés comme haplotypes majeurs. Les haplotypes H1 et H6 possèdent des fréquences élevées. Cependant, l'haplotype H6 possède une fréquence dans la population de 45%, qui est légèrement supérieure à celle de l'haplotype H1 (40%). Ce qui détermine l'haplotype H6 comme étant l'haplotype majeur pour le gène TGF $\beta$ 1. L'ensemble des 7 autres haplotypes se partagent une fréquence de totale restante de 15%. Les deux haplotypes fréquents sont bien répartis dans les différents groupes ethniques. Puisque les fréquences des haplotypes autres que H1 et H6 sont faibles dans la population, il a été difficile de déterminer les haplotypes régulateurs importants à étudier. Les haplotypes H3 et H8 ayant présentés des difficultés à l'étape du clonage, n'ont pu être étudiées de façon fonctionnelle.

Tableau VIII : Liste des haplotypes régulateurs trouvés dans l'échantillon de populations de 40 individus pour le gène TGFβ1.

Gène	Haplotypes / SNPs										Population					
	-1960	-1886	-1571	-1550	-1349	-1146	-799	-508	-448	Africain	Amérindien	Européen	Asiatique	Moyent-Orientaux	Total	Freq (%)
	Ancestral	C	G	A	INS AGG	-	C	G	G	C						
	H1	-	-	-	DEL AGG	-	-	A	-	-	2	8	5	9	8	32 41.0
	H2	-	-	-	DEL AGG	-	G	-	A	-	1				1	1.3
	H3	-	-	-	DEL AGG	INS C	-	-	A	-		1			1	1.3
	H4	-	-	-	-	-	-	-	A	-	1				1	1.3
	H5	-	A	-	-	-	-	-	-	-		1		2	3	3.8
	H6	-	A	G	-	-	-	-	-	-	7	6	10	7	6	36 46.2
	H7	-	A	G	-	-	-	A	-	-			1		1	1.3
	H8	T	-	-	-	-	-	-	-	-	2				2	2.6
	H9	T	-	-	-	-	-	-	-	T	1				1	1.3

Les zones grisées montrent les haplotypes qui ont été testés pour leur expression, en analyse de transfections transitoire.

### **3.8.2 Analyse fonctionnelle d'expression des rHaps, par expériences de transfections transitoires.**

Les résultats d'analyses d'expression des haplotypes régulateurs du gène TGF $\beta$ 1 démontrent que les expériences produites dans la lignée cellulaire Hela présentent des activités promotrices faibles (différence d'expression entre les rHaps et la ligne de base est inférieure à 5 fois). Les haplotypes régulateurs H2 et H5 sont faiblement actifs dans la lignée Hela, lorsque l'on compare leur expression au niveau de base (figure 3.13). Cependant, ces haplotypes H2 et H5 démontrent une différence d'expression significative ( $P < 0,01$ ) comparés à l'haplotype majeur H6, dans les lignées cellulaires Jeg3 et HepG2. Les différences d'expression des haplotypes H2 et H5 trouvées dans les résultats de transfections de la lignée Jeg3, sont de 2 fois supérieures à l'haplotype majeur. Dans la lignée cellulaire HepG2, l'haplotype H6 montre un niveau d'expression 11 fois supérieur à la ligne de base, ce qui contraste avec les résultats observés dans les deux autres lignées cellulaires. Les résultats de transfections de la lignée cellulaire HepG2 présentent un différence significative d'expression de H2 et H5 à l'inverse de celle observée dans les cellules Jeg3, soit de 2 fois inférieure au niveau d'expression de l'haplotype majeur.

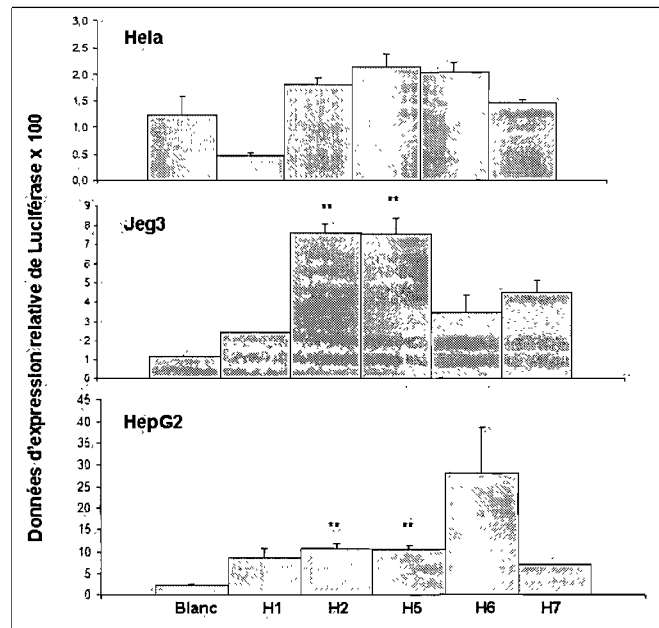


Figure 3.13 : Résultat d'analyse d'expression des 5 haplotypes régulateurs testés pour le gène TGF $\beta$ 1 par transfections transitoires. Le niveau d'expression de chacun des haplotypes est donné sous la forme d'un ratio de la luciférase mesurée Luciole/Rénilla, multiplié par 100. Les haplotypes fréquents, des différents gènes analysés par transfections transitoires, sont comparés à l'haplotype majeur du même gène. Les analyses de transfections sont faites dans trois lignées cellulaires différentes (Hela, Jeg3 et HepG2). Les différences d'expression significatives ont été testées par le test de Lord et les valeurs sont indiquées (\* :  $P < 0,05$ ; \*\* :  $P < 0,01$ ).



### **3.8.3 Analyse des modifications de liaison des facteurs de transcription, dû à la présence de polymorphismes régulateurs, dans des expériences de retard sur gel.**

Les haplotypes régulateurs H2 et H5 ont démontrés des différences d'expression significatives ( $P < 0,01$ ) comparés à l'haplotype majeur H6, dans deux lignées cellulaires. Les variants suivant diffèrent entre les haplotypes régulateurs H2, H7 et l'haplotype majeur H6 : -508A>G, -1146C>G, -1550INSAGG, -1571A>G, -1886G>A. Ces variants ont été étudiés dans des expériences de retard sur gel. Les prédictions informatiques proposent la présence de sites de liaison de facteurs de transcription affectés par chacun des variants étudiés (tableau XIX). Les résultats des expériences de retard sur gel montrent que des liaisons différentielles de facteurs de transcription ont été trouvées pour les variants -1550 INSAGG et -1886G>A. La figure 3.14, nous montre un exemple de liaison différentielle ADN-protéine trouvée pour le rSNP -1886G>A en présence de l'extrait nucléaire de cellules Hela. La figure, présente 4 complexes différents formés en présence de l'extrait nucléaire. Le complexe 1 est spécifique à l'allèle G (ligne 1 versus 6). Ce complexe est compétitionné totalement lorsque le compétiteur spécifique froid G est ajouté (ligne 3), mais demeure présent lors de l'ajout du compétiteur froid A (ligne 4).

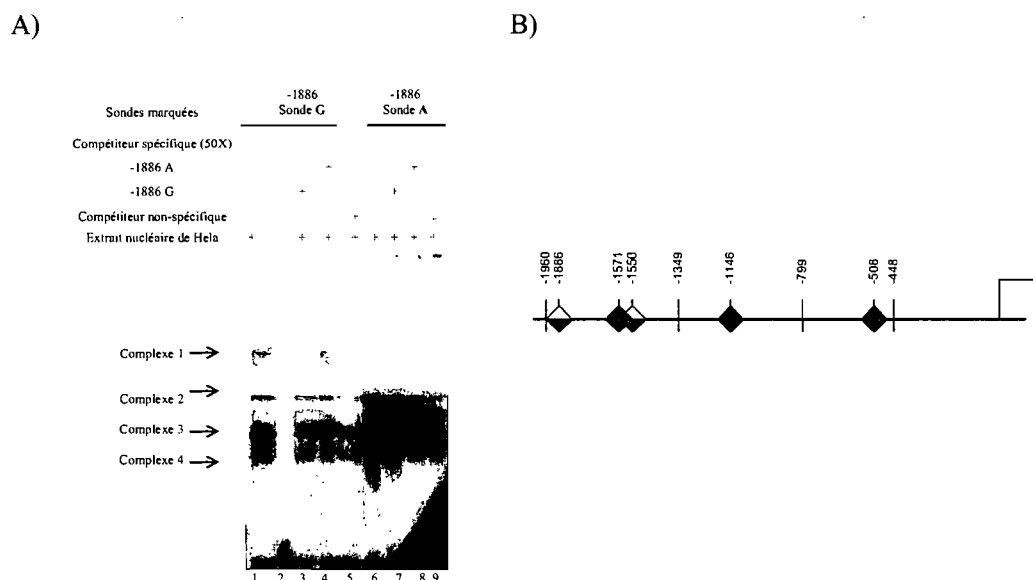


Figure 3.14 : **A)** Résultat d'analyse de retard sur gel pour le variant -1886G>A du gène TGF $\beta$ 1. En présence de l'extrait nucléaire de cellule Hela, 4 différents complexes sont formés. La ligne 2 représente le contrôle négatif, sans extrait nucléaire. Un contrôle positif (non montré) est effectué pour assurer la validité de l'expérience. Le complexe 1 est spécifique à l'allèle G (ligne 1 versus 6). Ce complexe est compétitionné lors de l'ajout du compétiteur spécifique froid G (ligne 3) et demeure présent lorsque le compétiteur A froid est ajouté (ligne 4). Le compétiteur non spécifique (ADN de thymus de veau) compétitionne tous les complexes, ce compétiteur est non approprié. Les complexes 2, 3 et 4 sont non spécifiques. **B)** Intégration des données expérimentales de retard sur gel de la région haplotypique à l'étude. Les losanges démontrent la présence de liaison d'un facteur comme suit : losange bleu, liaison aux deux sondes; losange blanc, absence de liaison aux deux sondes et losange blanc et bleu, présence d'une liaison différentielle. Les résultats ont été observé dans au moins une lignée cellulaire.

### **3.9 Description de l'étude du gène SKP2**

#### **3.9.1 Identification des haplotypes régulateurs présents dans l'échantillon représentatif de la population mondiale.**

L'analyse du génotype de 2 rSNPs de la population, a permis l'identification de trois haplotypes régulateurs (tableau IX). De ces trois rHaps un est de façon très marquée le plus fréquent dans la population. L'haplotype H1 représente 92,5% de la population, ce qui le place en position d'haplotype majeur.

Tableau IX : Liste des haplotypes régulateurs trouvés dans l'échantillon de populations de 40 individus pour le gène SKP2.

Gène	Haplotypes / SNPs			Population						Freq (%)
				Africain	Amérindien	Européen	Asiatique	Moyent-Orientaux	Total	
SKP2	-846	-758								
	Ancestral	A	G							
	H1			14	15	16	13	16	74	94.9
	H2	-	C	2					2	2.6
	H3	G			1		1		2	2.6

Les zones grisées montrent les haplotypes qui ont été testés pour leur expression, en analyse de transfections transitoire.

### **3.9.2 Analyse fonctionnelle d'expression des rHaps, par expériences de transfections transitoires.**

Nous avons décidé d'étudier les haplotypes régulateurs H1 et H3. Les haplotypes H2 et H3 possèdent des fréquences identiques, mais H2 n'est observé que chez les africains. Les analyses fonctionnelles de transfections transitoires nous montrent que les deux haplotypes régulateurs testés possèdent une activité promotrice forte (figure 3.15). Le niveau d'expression des haplotypes est de l'ordre de 60 fois la ligne de base dans les cellules Hela, 200 fois dans les cellules Jeg3 et de 12 fois dans les cellules HepG2. Lorsque les niveaux d'expression des haplotypes régulateurs H1 et H3 sont comparés, ceux-ci démontrent une expression très similaire, qui ne présente aucune différence significative. Par conséquent, aucune analyse de retard sur gel n'a été produite pour le gène SKP2.

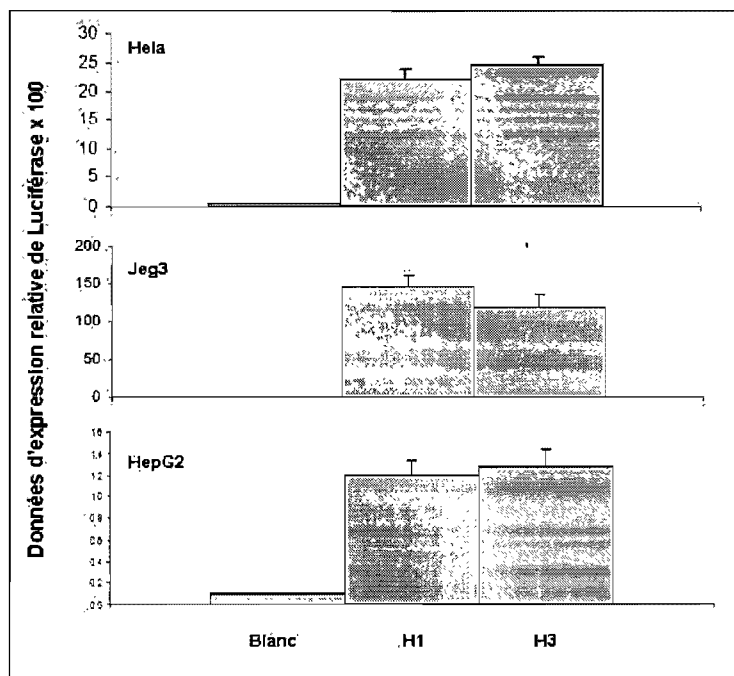


Figure 3.15 : Résultat d'analyse d'expression des 2 haplotypes régulateurs testés pour le gène SKP2 par transfections transitoires. Le niveau d'expression de chacun des haplotypes est donné sous la forme d'un ratio de la luciférase mesurée Luciole/Rénilla, multiplié par 100. Les haplotypes fréquents, des différents gènes analysés par transfections transitoires, sont comparés à l'haplotype majeur du même gène. Les analyses de transfections sont faites dans trois lignées cellulaires différentes (Hela, Jeg3 et HepG2).

### **3.10 Description de l'étude du gène CCND1**

#### **3.10.1 Identification des haplotypes régulateurs présents dans l'échantillon représentatif de la population mondiale.**

Le génotypage de 13 rSNPs a mené à l'identification de 11 rHaps (tableau X). Dans les 11 haplotypes régulateurs identifiés, deux sont très fréquents (H1 et H7). Ces deux haplotypes représentent ensemble 81% de la population et leur fréquence respective sont pour l'haplotype H1 56% et pour l'haplotype H7 25%.

Tableau X : Liste des haplotypes régulateurs trouvés dans l'échantillon de populations de 40 individus pour le gène CCND1.

Gène	Haplotypes / SNPs														Population						
		-1938	-1931	-1926	-1888	-1885	-1737	-1701	-1597	-1554	-1537	-766	-568	-519	Africain	Amérindien	Européen	Asiatique	Moyent-Orientaux	Total	Freq (%)
CCND1	Ancestral	C	G	-	A	G	A	G	G	G	DEL C	A	A	C							
	H1	T	-	-	-	-	-	-	-	-	-	-	-	-	6	12	7	12	8	45	56.3
	H2	T	-	-	-	-	-	-	-	-	-	-	-	T	2					2	2.5
	H3	-	-	-	-	-	-	A	-	-	T	C	-	1						1	1.3
	H4	T	-	-	-	T	-	-	-	-	-	-	-				1			1	1.3
	H5	T	-	INS G	-	-	-	-	-	-	T	-	-	2						2	2.5
	H6	T	C	-	-	-	-	-	-	-	-	-	-			1			1	2	2.5
	H7	-	-	-	-	-	-	-	-	C	-	-	-	3	3	6	3	5	20	25.0	
	H8	-	-	-	-	-	-	A	-	-	-	-	-	1						1	1.3
	H9	-	-	-	-	-	C	-	A	-	-	-	-	1					1	2	2.5
	H10	-	-	-	-	T	-	-	-	C	-	-	-			1				1	1.3
	H11	-	-	-	C	-	-	A	-	-	-	-	-		1	1			1	3	3.8

Les zones grisées montrent les haplotypes qui ont été testés pour leur expression, en analyse de transfections transitoire.



### **3.10.2 Analyse fonctionnelle d'expression des rHaps, par expériences de transfections transitoires.**

Nous avons décidé d'analyser les deux haplotypes fréquents obtenus pour le gène CCND1, soit les haplotypes H1 et H7. Puisque tous les autres haplotypes démontrent des fréquences semblables, nous avons limité les analyses fonctionnelles à ces deux haplotypes régulateurs. Le figure 3.16, montre les résultats des expériences de transfections transitoires produites dans les trois lignées cellulaires Hela, Jeg3 et HepG2. Les haplotypes H1 et H7 possèdent une région régulatrice active, car le niveau d'expression de ces haplotypes se situe entre 5 et 10 fois le niveau d'expression de base. La tendance d'expression pour l'ensemble des haplotypes testés reste la même dans les lignées cellulaires Hela et Jeg3, mais diffère dans la lignée HepG2. C'est-à-dire que l'haplotype H1 démontre un niveau d'expression légèrement plus bas que l'haplotype H7 dans les lignées cellulaires Hela et Jeg3, alors que dans la lignée HepG2 l'expression de H7 se retrouve inférieure à l'expression de H1. Toutefois, les données d'expression relative de H1 et H7 demeurent similaires dans les trois lignées cellulaires étudiées et aucune différence significative d'expression n'est observée entre les haplotypes fréquents du gène CCND1. Par conséquent, aucune étude ultérieure n'a été produite pour ce gène.

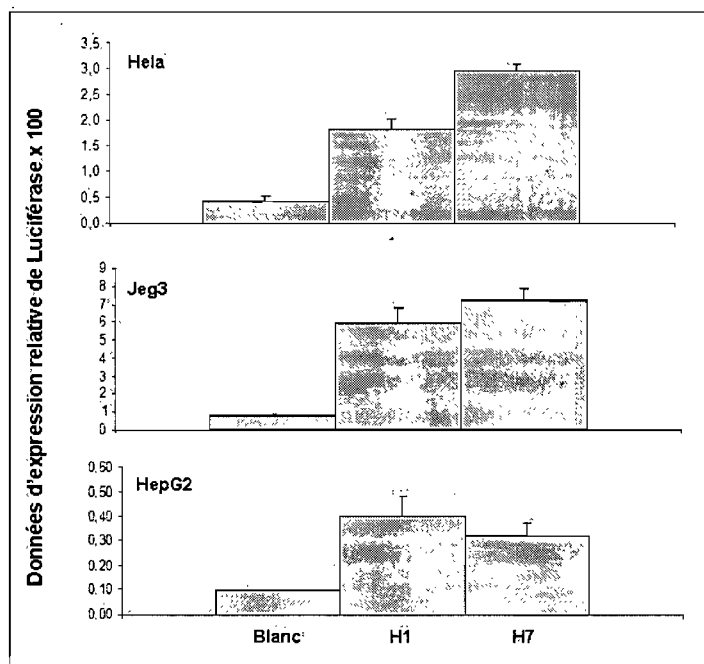


Figure 3.16 : Résultat d'analyse d'expression des 2 haplotypes régulateurs testés pour le gène CCND1 par transfections transitoires. Le niveau d'expression de chacun des haplotypes est donné sous la forme d'un ratio de la luciférase mesurée Luciole/Rénilla, multiplié par 100. Les haplotypes fréquents, des différents gènes analysés par transfections transitoires, sont comparés à l'haplotype majeur du même gène. Les analyses de transfections sont faites dans trois lignées cellulaires différentes (Hela, Jeg3 et HepG2).

### **3.11 Description de l'étude du gène E2F1**

#### **3.11.1 Identification des haplotypes régulateurs présents dans l'échantillon représentatif de la population mondiale.**

Les données de génotypage pour 4 rSNPs ont permis d'identifier cinq haplotypes régulateurs (tableau XI). Dans l'ensemble des haplotypes régulateurs trouvés, deux (H1 et H2) sont bien répartis dans tous les groupes ethniques constituant l'échantillon de populations. Ces deux haplotypes sont également les rHaps les plus représentatifs de l'ensemble de la population pour le gène E2F1. L'haplotype H1 est présent chez 73,8% de la population, ce qui le désigne comme haplotype majeur. L'haplotype H2, le second plus fréquent, représente 17,5% de la population. Les trois haplotypes régulateurs H1, H2 et H3, les plus retrouvés dans l'échantillon de populations ont été analysés de façon fonctionnelle.

Tableau XI : Liste des haplotypes régulateurs trouvés dans l'échantillon de populations de 40 individus pour le gène E2F1.

Gène	Haplotypes / SNPs					Population						
		-897	-300	-243	-187	Africain	Amérindien	Européen	Asiatique	Moyent-Orientaux	Total	Freq (%)
	Ancestral	C	G	T	C							
	H1	-	-	-	-	13	13	13	9	11	59	73.8
	H2	-	-	-	T	3	3	3	3	2	14	17.5
	H3	T	-	-	-				2	3	5	6.3
	H4	-	-	G	T				1		1	1.3
	H5	-	A	-	-				1		1	1.3

Les zones grisées montrent les haplotypes qui ont été testés pour leur expression, en analyse de transfections transitoire.

### **3.11.2 Analyse fonctionnelle d'expression des rHaps, par expériences de transfections transitoires.**

Nous avons étudié les haplotypes régulateurs : H1, H2 et H3. Les résultats d'analyse d'expression par expériences de transfections transitoires nous montrent que le patron d'expression pour l'ensemble de ces haplotypes régulateurs est maintenu dans les trois lignées cellulaires (Hela, Jeg3 et HepG2), tel que démontre la figure 3.17. Les haplotypes régulateurs H1, H2 et H3 du gène E2F1 démontrent par les analyses d'expression qu'ils possèdent des promoteurs forts. Des différences d'expression observées entre les rHaps et la ligne de base sont supérieures à 10 fois. La lignée cellulaire Hela présente une expression des haplotypes régulateurs d'environ 150 fois supérieure au niveau de base. Dans les cellules Jeg3, ces haplotypes sont en moyenne 250 fois supérieurs au niveau de base et dans la lignée cellulaire HepG2, les haplotypes montrent une expression autour de 15 fois supérieure au niveau de base. Toutefois, aucune différence d'expression significative n'est observée entre les haplotypes régulateurs (H2 et H3) et l'haplotype majeur H1. Alors aucune expérience de retard sur gel n'a été produite pour ce gène.

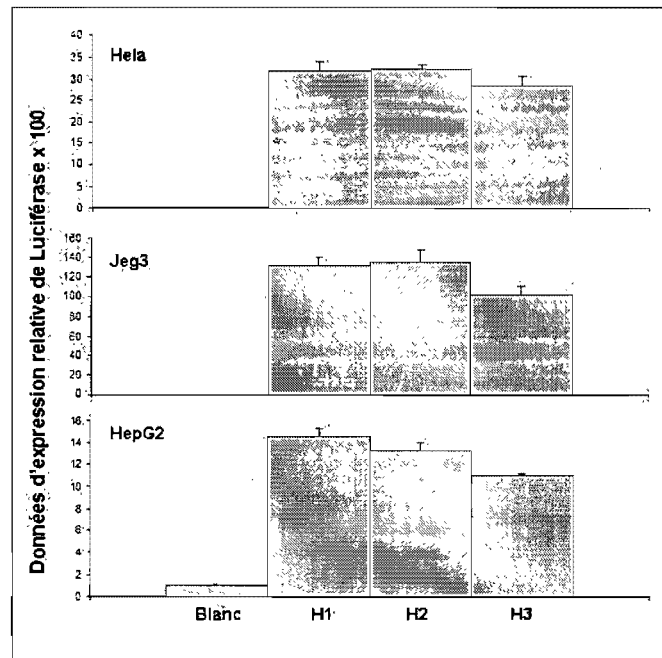


Figure 3.17 : Résultat d'analyse d'expression des 3 haplotypes régulateurs testés pour le gène E2F1 par transfections transitoires. Le niveau d'expression de chacun des haplotypes est donné sous la forme d'un ratio de la luciférase mesurée Luciole/Rénilla, multiplié par 100. Les haplotypes fréquents, des différents gènes analysés par transfections transitoires, sont comparés à l'haplotype majeur du même gène. Les analyses de transfections sont faites dans trois lignées cellulaires différentes (Hela, Jeg3 et HepG2).

### **3.12 Description de l'étude du gène HDAC1**

#### **3.12.1 Identification des haplotypes régulateurs présents dans l'échantillon représentatif de la population mondiale.**

Le génotypage de 12 rSNPs a permis l'identification de 11 haplotypes régulateurs (tableau XII). Deux de ces 12 rHaps peuvent être considérés comme haplotype majeur. Les haplotypes H1 et H2 possèdent des fréquences élevées. Cependant, l'haplotype H1 est légèrement plus fréquent avec une représentation de la population de 46%. Le second haplotype est présent chez 35% de la population. Les haplotypes H1 et H2 sont bien répartis sur l'ensemble des groupes ethniques qui composent l'échantillon, toutefois, H1 est peu fréquent chez les africains. Les trois haplotypes les plus fréquents H1, H2 et H3 ont été analysés de façon fonctionnelle, afin de déterminer leur patron d'expression.

Tableau XII : Liste des haplotypes régulateurs trouvés dans l'échantillon de populations de 40 individus pour le gène HDAC1.

Gène	Haplotypes / SNPs													Population						
		-1546	-1269	-1193	-1140	-905	-786	-783	-599	-548	-455	-326	-162	Africain	Amérindien	Européen	Asiatique	Moyent-Orientaux	Total	Freq (%)
HDAC1	Ancestral	G	C	G	G	A	T	G	C	G	T	A	C							
	H1	-	T	-	-	-	-	A	-	-	-	-	-	1	12	8	6	10	37	46.3
	H2	-	-	-	-	-	-	A	-	-	-	-	-	8	4	5	7	4	28	35.0
	H3	-	T	-	-	-	-	A	-	-	C	-	-			2		2	4	5.0
	H4	-	-	-	-	-	-	-	T	A	-	T	-	2					2	2.5
	H5	-	-	-	-	-	-	-	-	-	-	-	-	2					2	2.5
	H6	-	-	-	A	-	-	A	-	-	-	-	-	1					1	1.3
	H7	-	T	-	-	-	-	A	-	-	C	-	T			1			1	1.3
	H8	T	-	-	-	-	-	A	-	-	-	-	-				1		1	1.3
	H9	-	T	A	-	-	-	A	-	-	-	-	-				2		2	2.5
	H10	-	-	-	-	-	C	A	-	-	-	-	-	1					1	1.3
	H11	-	-	-	-	C	-	A	-	-	-	-	-	1					1	1.3

Les zones grisées montrent les haplotypes qui ont été testés pour leur expression, en analyse de transfections transitoire.



### **3.12.2 Analyse fonctionnelle d'expression des rHaps, par expériences de transfections transitoires.**

Nous avons analysé par transfections transitoires les 3 haplotypes régulateurs fréquents du gène HDAC1 (H1, H2 et H3). La figure 3.18 montre une tendance d'expression qui se maintient dans les trois lignées cellulaires (Hela, Jeg3 et HepG2). Les haplotypes démontrent une activité promotrice forte dans les trois lignées cellulaires utilisées. Les lignées cellulaires Hela et Jeg3 montrent une différence des données d'expression relative entre les haplotypes régulateurs et la ligne de base, d'environ 200 fois et les données d'expression relative trouvées dans la lignée cellulaire HepG2 sont autour de 60 fois supérieures au niveau de base. Cette tendance démontre que les rHaps H2 et H3 suivent bien le niveau d'expression de l'haplotype majeur H1. Aucune différence d'expression significative n'est observée pour le gène HDAC1. Alors l'analyse du gène HDAC1 n'a pas été poursuivie. Aucune expérience de retard sur gel n'a été effectuée.

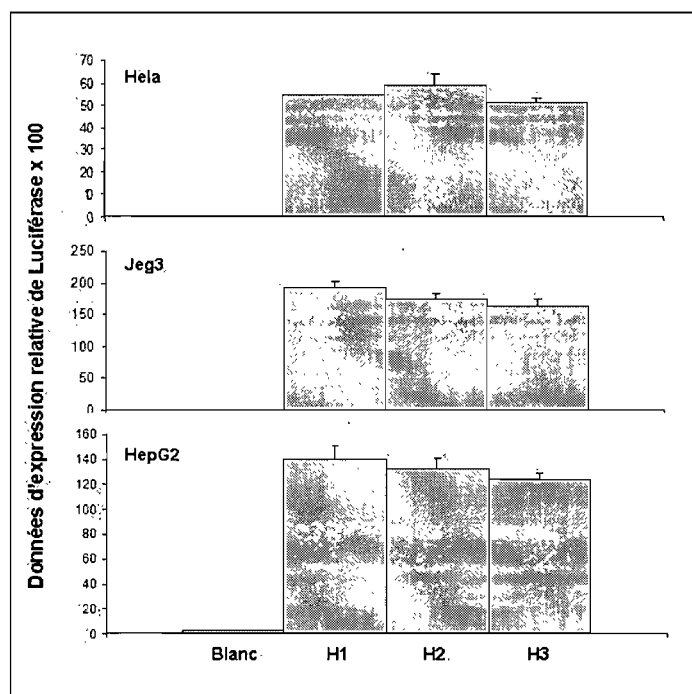


Figure 3.18 : Résultat d'analyse d'expression des 3 haplotypes régulateurs testés pour le gène HDAC1 par transfections transitoires. Le niveau d'expression de chacun des haplotypes est donné sous la forme d'un ratio de la luciférase mesurée Luciole/Rénilla, multiplié par 100. Les haplotypes fréquents, des différents gènes analysés par transfections transitoires, sont comparés à l'haplotype majeur du même gène. Les analyses de transfections sont faites dans trois lignées cellulaires différentes (HeLa, Jeg3 et HepG2).

### **3.13 Description de l'étude du gène RB1**

#### **3.13.1 Identification des haplotypes régulateurs présents dans l'échantillon représentatif de la population mondiale.**

Le génotypage de 5 rSNPs a mené à l'identification de 6 haplotypes régulateurs. L'haplotype H1 est une copie identique à l'haplotype ancestral identifié à partir de l'ADN de chimpanzé. Cet haplotype est bien réparti à travers les différents groupes ethniques (tableau XIII). Seul les deux haplotypes régulateurs fréquents (H1 et H4) ont été testés dans les analyses fonctionnelles par transfections transitoires. L'haplotype H1 représente 73,8% de la population comparativement à l'haplotype H4 qui représente que 12,5% de la population.

Tableau XIII : Liste des haplotypes régulateurs trouvés dans l'échantillon de populations de 40 individus pour le gène RB1.

Gène	Haplotypes / SNPs					Population						
	-1733	-1554	-768	-595	-83	Africain	Amérindien	Européen	Asiatique	Moyent-Orientaux	Total	Freq (%)
	INS CTCGCGGACGT											
Ancestral	-	C	C	T	GACGCCGCGGGC							
H1	-	-	-	-	-	8	12	13	12	14	59	73.8
H2	-	-	-	-	DEL			1			1	1.3
H3	-	-	A	-	-	4					4	5.0
H4	-	A	-	-	-		2	2	4	2	10	12.5
H5	-	A	-	G	-		2				2	2.5
H6	INS CTGATA					4					4	5.0

Les zones grisées montrent les haplotypes qui ont été testés pour leur expression, en analyse de transfections transitoire.

### **3.13.2 Analyse fonctionnelle d'expression des rHaps, par expériences de transfections transitoires.**

Les 2 haplotypes les plus fréquents du gène RB1, H1 et H4 ont été testés par transfections transitoires. Ces analyses fonctionnelles d'expression ont permis de valider que la présence du rSNP -1554C>A ne cause aucune différence d'expression entre les haplotypes régulateurs (figure 3.19). L'activité promotrice, des haplotypes testés du gène RB1, est très élevée lorsque comparée au niveau de base (pGL3-basic, sans promoteur). L'activité des promoteurs est de l'ordre de 700 fois supérieure à la ligne de base dans les cellules Hela, près de 300 fois supérieure dans la lignée cellulaire Jeg3 et 50 fois supérieure dans les cellules HepG2, ce qui classe les promoteurs du gènes RB1 comme étant des promoteurs forts. L'expression de l'haplotype H4, auquel est associé le polymorphisme -1554C>A, démontre aucune différence allélique significative comparée à l'haplotype majeur H1 dans les 3 lignées cellulaires. Par conséquent, puisque aucune différence allélique significative n'a été trouvée, aucune analyse subséquente du gène RB1 n'a été produite.

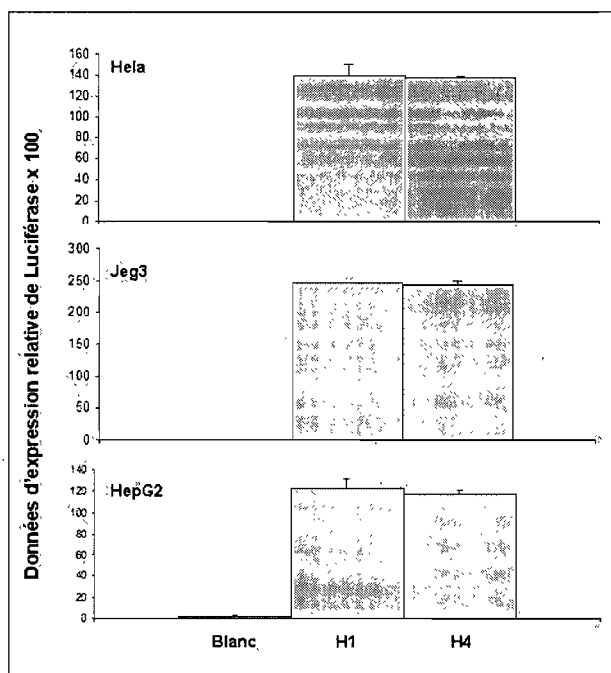


Figure 3.19 : Résultat d'analyse d'expression des 3 haplotypes régulateurs testés pour le gène RB1 par transfections transitoires. Le niveau d'expression de chacun des haplotypes est donné sous la forme d'un ratio de la luciférase mesurée Luciole/Rénilla, multiplié par 100. Les haplotypes fréquents, des différents gènes analysés par transfections transitoires, sont comparés à l'haplotype majeur du même gène. Les analyses de transfections sont faites dans trois lignées cellulaires différentes (Hela, Jeg3 et HepG2).

### **3.14 Description de l'étude du gène CDKN2B**

#### **3.14.1 Identification des haplotypes régulateurs présents dans l'échantillon représentatif de la population mondiale.**

Le génotypage de 9 rSNPs a permis d'identifier 15 différents haplotypes régulateurs (tableau XIV). Les cinq haplotypes fréquents (H1, H2, H3, H11 et H14) ont été sélectionnés pour être évalués dans les analyses fonctionnelles (tableau XIV). Ces 5 haplotypes représentent ensemble 77,5% de la population. Les 10 autres haplotypes régulateurs du gène CDKN2B se partagent le reste. L'haplotype majeur H11 est présent chez 25% de la population. Les rHaps fréquents H3, H11 et H14 sont bien répartis dans les différents groupes ethniques. Les expériences de transfections transitoires sont présentement en cours dans le laboratoire pour valider les différences d'expression entre les haplotypes fréquents.

Tableau XIV : Liste des haplotypes régulateurs trouvés dans l'échantillon de populations de 40 individus pour le gène CDKN2B.

Gène	Haplotypes / SNPs										Population						
	-1319	-1270	-1164	-1001	-593	-549	-287	-214	-5		Africain	Amérindien	Européen	Asiatique	Moyent-Orientaux	Total	Freq (%)
	Ancestral	A	C	G	A	A	C	G	-	C							
	H1	-	-	-	-	-	-	-	-	-	4	1	2			7	8.8
	H2	-	-	-	-	-	C	-	-	-	1	1	5		4	11	13.8
	H3	-	-	-	T	-	-	-	-	-	2	1	3	8	3	17	21.3
	H4	-	-	-	-	T	-	-	-	G					1	1	1.3
	H5	-	-	-	-	C	-	-	-	-					1	1	1.3
	H6	-	-	-	G	-	-	C	-	-	1	1			1	3	3.8
	H7	-	-	A	-	-	-	-	INS	T				2		2	2.5
	H8	-	-	A	-	T	-	-	-	-				2		2	2.5
	H9	-	T	-	-	-	-	-	-	-		2	1			3	3.8
	H10	-	T	-	-	-	-	C	-	-		2	1			3	3.8
	H11	G	-	-	-	-	-	-	-	-	6	4	3	3	4	20	25.0
	H12	G	-	-	-	-	-	C	-	-		1				1	1.3
	H13	G	-	-	-	-	T	-	-	-	1					1	1.3
	H14	G	-	-	-	T	-	-	-	-	1	2	1	1	2	7	8.8
	H15	G	-	-	-	C	-	-	-	-		1				1	1.3

Les zones grisées montrent les haplotypes qui ont été testés pour leur expression, en analyse de transfections transitoire.



### **3.15 Description de l'étude du gène CCNE**

#### **3.15.1 Identification des haplotypes régulateurs présents dans l'échantillon représentatif de la population mondiale.**

Le génotypage de 2 rSNPs a permis d'identifier 3 haplotypes régulateurs. L'haplotype H1 est une copie identique de l'haplotype ancestral montré au tableau XV. Un seul haplotype régulateur (H1) du gène CCNE est fréquent dans la population, possédant une fréquence de 92,5%. Le second haplotype (H2), que nous avons décidé de tester dans les analyses fonctionnelles, est présent chez moins de 4% de la population. Les essais de transfections transitoires sont présentement en cours dans le laboratoire.

Tableau XV : Liste des haplotypes régulateurs trouvés dans l'échantillon de populations de 40 individus pour le gène CCNE.

Gène	Haplotypes / SNPs			Population						
	-1938	-1931		Africain	Amérindien	Européen	Asiatique	Moyent-Orientaux	Total	Freq (%)
	Ancestral	T	G							
	H1			15	15	16	12	16	74	94.9
	H2	-	A		1		2		3	3.8
	H3	DEL T	-	1					1	1.3

Les zones grisées montrent les haplotypes qui ont été testés pour leur expression, en analyse de transfections transitoire.

### **3.16 Description de l'étude du gène CDK4**

#### **3.16.1 Identification des haplotypes régulateurs présents dans l'échantillon représentatif de la population mondiale.**

Au total 11 rSNPs ont été génotypés (tableau XVI), ce qui nous a permis d'identifier 10 haplotypes régulateurs. L'haplotype H1, identique à l'haplotype ancestral, représente 87,5% de la population. Les analyses fonctionnelles présentement en cours permettront de valider le niveau d'expression de l'haplotype majeur, comparé au second haplotype le plus fréquent H4, qui est présent chez 2,5% de la population.

Tableau XVI : Liste des haplotypes régulateurs trouvés dans l'échantillon de populations de 40 individus pour le gène CDK4.

Gène	Haplotypes / SNPs												Population					
	-1887	-1349	-1178	-937	-847	-724	-615	-612	-290	-217	-150	Africain	Amérindien	Européen	Asiatique	Moyent-Orientaux	Total	Freq (%)
	Ancestral	T	T	A	A	C	TC	C	C	C	C							
	H1	-	-	-	-	-	-	-	-	-	-	12	13	14	15	16	70	87.5
	H2	-	-	-	-	-	-	-	-	-	G	1					1	1.3
	H3	-	-	-	-	-	-	-	-	T	-			1			1	1.3
	H4	-	-	-	-	-	-	T	-	-	-		2				2	2.5
	H5	-	-	-	-	-	DEL	-	-	-	-		1				1	1.3
	H6	-	-	-	-	T	-	-	-	-	-			1			1	1.3
	H7	-	-	-	G	-	-	-	A	-	-				1		1	1.3
	H8	-	-	G	-	-	-	-	-	-	-	1					1	1.3
	H9	-	A	-	G	-	-	-	-	-	-	1					1	1.3
	H10	C	-	-	-	-	-	A	-	-	-	1					1	1.3

Les zones grisées montrent les haplotypes qui ont été testés pour leur expression, en analyse de transfections transitoire.

### **3.17 Description de l'étude du gène CDK6**

#### **3.17.1 Identification des haplotypes régulateurs présents dans l'échantillon représentatif de la population mondiale.**

L'identification des 9 haplotypes régulateurs présents dans l'échantillon de populations a été produite à partir du génotypage de 8 rSNPs (tableau XVII). Un seul haplotype H1 est présent à 85%, alors que les 8 autres rHaps se partagent 15% de la population. L'haplotype majeur H1, est bien réparti dans les différents groupes ethniques et est identique à l'haplotype ancestral. Les analyses fonctionnelles en cours sur le gène CDK6 nous informerons sur une possible expression différentielle entre les deux haplotypes H1 et H8.

Tableau XVII : Liste des haplotypes régulateurs trouvés dans l'échantillon de populations de 40 individus pour le gène CDK6.

Gène	Haplotypes / SNPs									Population						
	-1958	-1902	-1789	-1450	-1386	-1270	-1124	-638	Africain	Amérindien	Européen	Asiatique	Moyent-Orientaux	Total	Freq (%)	
	Ancestral	C	CCCCCGCA	A	A	A	C	T	C							
	H1	-	-	-	-	-	-	-	-	13	13	14	14	14	68	85.0
	H2	-	-	-	-	-	-	-	T	1				1	1.3	
	H3	-	-	-	-	-	-	A	-				1	1	1.3	
	H4	-	-	-	-	-	T	-	-	1				1	1.3	
	H5	-	-	-	-	G	-	-	-				1	1	1.3	
	H6	-	-	-	T	-	-	-	T	1				1	1.3	
	H7	-	-	T	-	-	-	-	-		1			1	1.3	
	H8	-	DEL CCCCCGCA	-	-	-	-	-	-		3	1	1	5	6.3	
	H9	G	-	-	-	-	-	-	-			1		1	1.3	

Les zones grisées montrent les haplotypes qui ont été testés pour leur expression, en analyse de transfections transitoire.

### **3.18 Description de l'étude du gène MADH4**

#### **3.18.1 Identification des haplotypes régulateurs présents dans l'échantillon représentatif de la population mondiale.**

Les données de génotypage pour 2 rSNPs ont permis d'identifier 2 haplotypes régulateurs (tableau XVIII). Pour les 2 haplotypes identifiés, un représente l'haplotype ancestral (H1) et est présent chez tous les individus de l'échantillon de populations analysés. Le second haplotype (H2) est présent uniquement chez un individu africain. Puisque la variabilité interindividuelle trouvée est insuffisante, aucune analyse fonctionnelle ne sera produite pour ce gène.

Tableau XVIII : Liste des haplotypes régulateurs trouvés dans l'échantillon de populations de 40 individus pour le gène MADH4.

Gène	Haplotypes / SNPs		Population							
			Africain	Amérindien	Européen	Asiatique	Moyent-Orientaux	Total	Freq (%)	
MADH4	-418	-412								
	Ancestral	A	G							
	H1	-	-	15	16	16	14	16	77	98.7
	H2	C	T	1					1	1.3

Les zones grisées montrent les haplotypes qui ont été testés pour leur expression, en analyse de transfections transitoire.



### **3.19 Description de l'étude du gène CDK2**

#### **3.19.1 Identification des haplotypes régulateurs présents dans l'échantillon représentatif de la population mondiale.**

Les analyses de génotypage de l'échantillon représentatif de la population mondiale sont présentement en cours pour le gène CDK2. Les 10 rSNPs connus, identifiés dans la littérature et présents dans la région régulatrice de 2 Kb à l'étude, seront testés. Ces polymorphismes sont les suivants : -54C>T, -80A>C, -113A>T, -323C>T, -456C>G, -683C>T, -916A>T, -919A>G, -1476A>T et -1856DEL. Une fois les différents haplotypes régulateurs identifiés, des essais fonctionnels d'expression par transfections transitoires seront produits sur les haplotypes fréquents.

Tableau XIX : Tableau récapitulatif des différents rSNPs étudiés par EMSA pour chacun des gènes.

Gènes	rSNPs	Prédictions informatiques		Résultats d'analyses EMSA		
		Gain	Perte	Hela	Jeg3	HepG2
CDC25A	-2030T>G	- Zinc binding protein factors (0.91) - Basic and erythroid Krueppel like factors (0.91) - Myeloid zinc finger 1 factors (0.99)	- RBPJ-kappa (0.94)	+	+	+
	<del>-1566C&gt;T</del>					
	-1440G>A	- Metal induced transcription factor (0.88)	- AHR-arnt heterodimers and AHR-related factors (0.90)	-	-	-
	<del>-1433C&gt;T</del>					
	<del>-351G&gt;A</del>					
	<del>-171A&gt;T</del>					
	-51G>C	-	-	+	+	+
CDKN1A (P21)	<del>-1731DEL A</del>					
	-1536C>T	- Nuclear receptor subfamily 2 factors (0.83) - Peroxisome proliferation-activated receptor (0.70)		-	-	-
	-1529T>C		- NGFI-B response elements, nur subfamily of nuclear receptors (0.86) - Insulinoma associated factors (0.90)	-	-	-
	-1466A>G	-	-	-	-	-
	-1284T>C	-	-	+	+	-
	-1021C>T	- Human acute myelogenous leukemia factors (0.93)	- E2F-myc activator/cell cycle regulator (0.84) - Zinc finger with interaction domain factors (0.85)	+	+	+

Gènes	rSNPs	Prédictions informatiques		Résultats d'analyses EMSA		
		Gain	Perte	Hela	Jeg3	HepG2
	-899T>G	- Cellular and viral myb-like transcriptional regulators (0.90)	- Positive regulatory domain I binding factor (0.81)	+	+	+
	-791T>C		- Growth factor independence transcriptional repressor (0.90)	-	+	-
	-519G>C					
	-499G>C					
	-134G>C	-	-	-	+	+
CDKN1B (P27)	-1908 DEL T					
	-1857C>T	- Evl1-myeloid transforming protein (0.83)	- Myoblast determining factors (0.92) - Lim homeodomain factors (0.81) - Tata-binding protein factor (0.78)	-	-	-
	-1608G>A					
	-1228C>T					
	-1084 DEL AGG					
	-756A>C					
	-522G>A					
	-373G>T	- GA-Boxes (0.78) - PAX-1 binding sites (0.62) - Signal transducer and activator of transcription 1 (0.77)	- Myc associated zinc fingers protein (0.90)	-	-	-
	-337C>A	-	-	+	+	+
-221T>C						
CDKN2A (P16)	-1879C>T					
	-1600A>G					
	-710C>A					
	-598G>A					
	-589 DEL G					
	-464C>T					
	-222T>A		- Cellular and viral myb-like transcriptional regulators (0.96)	+	-	+
	-44A>T					

Gènes	rSNPs	Prédictions informatiques		Résultats d'analyses EMSA		
		Gain	Perte	Hela	Jeg3	HepG2
MADH3	-1946C>A	- Human and murine ETS1 factors (0.96)		+	+	+
	-1938T>C					
	-1137T>G					
	-749G>A					
	-669C>T					
TFDP1	-1887C>T	- GC-Box factors SP1/GC (0.83) - Basic and erythroid krueppel like factors (0.95)		-	-	+
	-1883G>C	- Basic and erythroid krueppel like factors (0.95)		+	+	-
	-1779A>G	-	-	-	-	-
	-1655G>A	- PAX-4/PAX-6 paired domain binding sites (0.87) - Cellular and viral myb-like transcriptional regulators (0.90)	- GC-Box factors SP1/GC (0.88) - Myc associated zinc fingers (0.88)	-	+	+
	-1530C>T					
	-1516C>T					
	-1515G>T					
-1394T>C	-	-	+	+	+	
TGFβ1	-1960C>T					
	-1886G>A	- Factors with moderate activity to homeo domain consensus sequence (0.87)	- AP4 and related proteins (0.86) - NeuroD, Beta2, HLH domain (0.92) - GATA binding factors (0.91)	+	-	+
	-1571A>G	- Nuclear receptor subfamily 2 factors (0.82) - Neuron-specific-olfactory factor (0.88)	- Gut-enriched Krueppel like binding factor (0.86)	-	-	-
	-1550 INS AGG	- PAX-4/PAX-6 paired domain binding sites (0.83) - Myc associated zinc	- Chorion-specific transcription factors with a GCM DNA binding domain (0.85)	+	+	+

Gènes	rSNPs	Prédictions informatiques		Résultats d'analyses EMSA		
		Gain	Perte	Hela	Jeg3	HepG2
		fingers (0.90)				
	-1349INS,C					
	-1146C>G	- Pancreas transcription factor 1, heterotrimeric transcription factor (0.76)	- GA-Boxes (0.78)	-	-	-
	-799G>A					
	-508A>G	- Neuron-specific-olfactory factor (0.88)	- HOX-PBX complexes (0.81) - Gut-enriched Krueppel like binding factor (0.86)	-	-	-
	-448C>T					

Les rSNPs dont on a mesuré une différence d'expression allélique significative, ont été analysés par expérience de retard sur gel et sont surlignés en rose. Les rSNPs qui ont été analysés par transfections transitoire mais dont aucune différence allélique significative a été trouvée sont surlignés en gris.

## **Chapitre 4**

### **-Discussion-**

## 4. Discussion

### 4.1 Sommaire des résultats

Nous avons étudié la région régulatrice de 18 gènes codant pour des composantes du point de contrôle de la transition G1/S du cycle cellulaire. Puisque les rSNPs situés dans la portion promotrice influenceraient la liaison de différents facteurs de transcription en modifiant leur site de liaison, nous avons voulu vérifier cette hypothèse de façon fonctionnelle. Nous avons défini arbitrairement pour le projet, la région régulatrice comme étant une portion de 2 Kb située en amont du site d'initiation de la transcription. Nous sommes conscients que des régions régulatrices peuvent être localisées à une grande distance du site d'initiation de la transcription. Par contre, nous croyons que la présence de SNPs au niveau du promoteur procure un impact important sur le niveau d'efficacité des promoteurs. Cette régulation est appuyée par plusieurs études récentes (Bachmann and others 2007; Bougeard and others 2006; Gazquez and others 2008; Hu and others 2007; Juhanson and others 2008; Ober and others 2008; Pastinen and Hudson 2004).

Dans cette étude, nous avons trouvé 7 gènes montrant une différence d'expression allélique significative : CDC25A, CDKN1A, CDKN1B, CDKN2A, MADH3, TGF $\beta$ 1 et TFDP1. Ces différences alléliques du niveau d'expression mesuré peuvent sembler drastique, cependant des différences d'expression importantes entre haplotypes ont déjà été observées dans au moins une autre étude expérimentale, (Joosten and others 2001). Ces 7 gènes impliquent au total 26 variants à analyser par expériences de retard sur gel, afin de trouver les liaisons différentielles de facteurs de transcription qui pourraient possiblement expliquer cette différence allélique d'expression.

Les 5 autres gènes (SKP2, CCND1, E2F1, HDAC1 et RB1) qui ont été testés de façon fonctionnelle par expériences de transfections transitoires, n'ont démontré aucune différence d'expression allélique significative. Nous avons préalablement déterminé un seuil d'activité promotrice qui doit être supérieur à 5 fois le niveau de base. Ceci est une

définition conservatrice d'un promoteur actif. Certaines études ont démontré que lorsque les promoteurs sont classifiés dans les bases de données comme tel, il ne nécessite pas de prendre un seuil de signification de l'activité promotrice aussi élevé (Hoogendoorn and others 2003) et qu'une activité supérieure à 2 fois le niveau de base est suffisante. En utilisant cette dernière définition du seuil de signification (2 fois le niveau de base), l'haplotype H7 du gène CCND1 démontre alors une différence d'expression significative ( $P < 0,01$ ) dans la lignée cellulaire Hela comparativement à l'haplotype majeur H1.

#### **4.2 La dérégulation des gènes du cycle cellulaire et leurs causes possibles.**

Il a déjà été établi par de nombreuses études, que les différentes contraintes auxquelles sont soumises les cellules telles que les radiations ultra-violet et les radicaux oxygène, sont sujet à causer des dommages à l'ADN (Boutros and others 2007). Si la cellule continue son cycle cellulaire, à ce moment les lésions présentes au niveau de l'ADN seront transmises aux cellules filles. Cette perte d'intégrité de la séquence d'ADN provoque une instabilité génétique qui conduit très souvent au cancer. Même si tous les types de cancers surviennent à partir de causes diverses et que chaque cas individuel provient d'une cause différente, tous les cancers ont un point en commun : une progression désordonnée du cycle cellulaire et la présence d'irrégularités telles que des délétions, mutations ou une surexpression des molécules régulant le cycle cellulaire (Kristjansdottir and Rudolph 2004). Des polymorphismes trouvés dans les gènes régulant le point de contrôle G1/S du cycle cellulaire et démontrant des différences significatives d'expression, peuvent être impliqués dans la susceptibilité interindividuelle au cancer. En effet, ces différences d'expression peuvent altérer le déroulement normal du point de contrôle G1/S et pousser la cellule à poursuivre sa progression dans le cycle cellulaire, même si celle-ci présente des anomalies au niveau de son ADN. Une surexpression des molécules conduisant la cellule à poursuivre le cycle vers la phase S ou une diminution de l'expression des molécules provoquant l'arrêt au point de contrôle G1/S, conduira vers une prolifération désordonnée et à la formation de tumeurs (Malumbres and Barbacid 2001).



### 4.3 Discussion des résultats d'analyses des gènes du point de contrôle G1/S.

Nous discuterons uniquement des gènes qui ont démontré une activité promotrice différentielle.

#### 4.3.1 Le gène CDC25A

Le gène CDC25 code pour une protéine phosphatase impliquée dans la régulation du cycle cellulaire. Cette phosphatase permet la déphosphorylation des résidus thréonine 14 et tyrosine 15 des protéines CDK pour activer les complexes cycline-cdk au cours du cycle cellulaire (Kristjansdottir and Rudolph 2004). Il existe trois différentes isoformes de CDC25 chez l'humain : CDC25A, CDC25B et CDC25C. La CDC25A, qui est celle étudiée dans le projet, active principalement la protéine CDK2 dans les complexes CDK2-cycline E et CDK2-cycline A au cours de la transition du point de contrôle G1/S, mais elle possède aussi un rôle dans la régulation du point de contrôle G2/M. Les CDC25B et CDC25C sont principalement requises pour l'entrée en phase de mitose (M) (Boutros and others 2007).

Nos analyses de transfections pour le gène CDC25A ont démontrées une différence d'expression allélique significative pour trois haplotypes H2, H5 et H7. Ces trois haplotypes sont surexprimés par rapport à l'haplotype majeur H1. La surexpression du gène CDC25A pourrait mener à une suractivation des CDK et ainsi conduire la cellule à passer outre le point de contrôle G1/S et continuer la progression du cycle cellulaire, ce qui est un trait commun du cancer. Plusieurs études démontrent qu'une surexpression du gène CDC25A est souvent impliquée dans plusieurs types de cancers, tel que : le cancer du sein, le cancer du colon et dans les lymphomes non hodgkinien (Boutros and others 2007; Kristjansdottir and Rudolph 2004). De plus, la surexpression de la protéine CDC25A est impliquée dans la formation de lymphomes (œsophage, thyroïdien et poumons), du cancer du sein ou du cancer du foie. La surexpression du gène entraîne un développement de la maladie plus agressif et les chances de survie sont plus faibles.

La différence allélique trouvée chez les trois haplotypes régulateurs H2, H5 et H7, nous a mené à l'analyse des liaisons possibles des facteurs de transcription des trois variants suivants : -2030T<C, -1440G<A et -51G<C. Les variants -2030 et -51 causent des liaisons différentielles de facteurs de transcription (tableau XIX). Nos analyses de retard sur gel démontrent clairement une liaison différentielle présente dans les trois lignées cellulaires (HepG2, Hela et Jeg3) aux endroits où sont situés les rSNPs -2030 et -51 (figure 3.2). Nous pouvons alors supposer que ces liaisons différentielles expliqueraient la différence d'expression allélique observée pour les haplotypes H2 et H7; et que ces surexpressions pourraient augmenter la susceptibilité interindividuelle au cancer. Aucun des variants situés dans la région régulatrice du gène CDC25A n'ont été étudiés dans la littérature jusqu'à présent. Ces nouvelles données sont particulièrement intéressantes car certaines études (Boutros and others 2007; Kristjansdottir and Rudolph 2004) n'ont pas réussi à démontrer l'évidence d'une surexpression causée par une mutation ou amplification au niveau génétique, mais ont notés que la surexpression était plutôt causée par un événement qui survient suite à la transcription. Ces résultats apportent une nouvelle connaissance sur le gène CDC25A.

Pour le rSNP -51, aucun facteur de transcription n'a pu être prédit pour expliquer la liaison différentielle, ce qui signifie qu'un nouveau facteur de transcription dont le site de liaison est toujours inconnu, se lie possiblement à cet endroit. Des analyses plus approfondies pourraient permettre l'identification de ce facteur de transcription qui semble être présent dans la plupart des tissus. Le rSNP -2030 affecte la séquence des sites de liaison de plusieurs facteurs de transcription et 6 complexes spécifiques ont été trouvés dans les expériences de retard sur gel. Quatre complexes spécifiques à l'allèle mineur (-2030T) (figure 3.2) ont été trouvés et le seul facteur prédit pouvant être en cause est le facteur RBP J-kappa. Peu d'études ont été faites sur ce facteur de transcription. Cette protéine a d'abord été identifiée comme étant impliquée dans la recombinaison de la chaîne légère  $\kappa$  du CMH I (Waltzer and others 1994). Par la suite, on lui a reconnu un possible rôle de facteur de transcription (Tun and others 1994), mais encore à ce jour on ne connaît pas les gènes qui sont régulés par ce facteur. L'allèle majeur (-2030G) est retrouvé chez 93,8% de la population et est associé à 2 complexes ADN-protéine. Ces complexes pourraient impliquer

les facteurs prédits Zinc, Krueppel et un facteur Doigt de zinc myéloïde. Pour les facteurs se liant à l'allèle majeur (-2030G), la liaison possible au facteur Doigt de zinc myéloïde qui est présent uniquement dans des cellule myéloïde (Morris and others 1994), est éliminée, puisqu'aucune des lignées cellulaires utilisées pour les analyses de retard sur gel n'est différenciée en lignée myéloïde. Les facteurs de la famille Krüepfel, tel que ZBP-89 sont associés au cycle cellulaire pour leur régulation de protéines importantes comme p53 (Bai and Merchant 2001; Okada and others 2006). On peut alors supposer qu'il est possible que ce soit la famille de facteurs de transcription Krüepfel qui affecte l'expression du gène CDC25A. Cependant, une étude récente (Ray and Kiyokawa 2008) démontre, par l'intermédiaire de souris transgéniques; qu'une surexpression du gène CDC25A est insuffisante pour conduire au développement du cancer, mais que cette surexpression couplée à une surexpression du gène RAS conduit au développement rapide de tumeurs chez la souris.

#### 4.3.2 CDKN1A

Ce régulateur de CDK dans le cycle cellulaire possède plusieurs noms : CDKN1A (Cycline dépendante kinase inhibiteur 1A), CIP1 (CDK-interacting protein 1), WAF1 (wildtype p53-activated fragment 1) et p21. Son nom WAF1 lui vient de son rôle dans le processus biologique de suppression de tumeurs, WAF1 est activée par la protéine p53 (el-Deiry and others 1993). Puis, son nom CIP1, pour son rôle de régulateur des CDK dans la progression du cycle cellulaire (Harper and others 1993). La transcription du gène CDKN1A est finement régulée par la protéine p53. CDKN1A agit ensuite comme molécule médiatrice de p53, pour réguler l'activité des complexes cyclines E-CDK2, cycline A-CDK2 ou de la cycline D. Une étude démontre que CDKN1A est une molécule modulatrice de la fonction de suppression de tumeur par p53 (Zakut and Givol 1995). En effet, la surexpression de CDKN1A dans une lignée tumorale provoque l'arrêt de la formation de colonies, de façon similaire à un arrêt produit par la molécule p53. Par contre CDKN1A peut également agir seule comme suppresseur de tumeurs. Une étude (Gartel 2007) a démontrée avec des souris contenant le gène p53 muté et le gène p21 aboli (p53R172P+/+, p21-/-) comparativement

aux souris contenant p53 muté seulement; que la perte de la protéine p21 empêche totalement l'arrêt du cycle cellulaire et accélère le développement de tumeurs. La protéine inhibitrice CDKN1A est connue pour être impliquée dans le développement du cancer, mais indirectement par l'intermédiaire de d'autres molécules. En effet, CDKN1A est plutôt sous-réglée par les protéines et facteurs qui contrôlent son niveau d'expression (Arima and others 2004; Gonzalez-Fernandez and others 2008).

Nos analyses d'expression par expériences de transfections transitoires ont montré une baisse d'expression très importante lorsque l'haplotype régulateur H12 est transfecté par rapport à l'haplotype majeur H1 (figure 3.3). Cette baisse d'expression est présente dans les trois lignées cellulaires. L'haplotype H12 par sa différence d'expression, fait ressortir l'effet de un ou plusieurs des variants suivants : -1536, -1529, -1466, -1284, -1021, -791 et -134. Lorsque ces variants ont été testés pour la liaison de facteurs de transcription, on a pu trouver des liaisons différentielles dans les variants -1284, -1021, -791 et -134 (tableau XIX). Le variant -1021 montre la formation d'un complexe qui se lie spécifiquement à l'allèle T, ce qui signifie que le seul facteur de transcription que nous avons identifié par prédictions informatiques pouvant être en cause est *Human acute myelogenous leukemia*, car il est le seul qui se lie à l'allèle T. De plus, on peut facilement déduire une possible association entre ce facteur impliqué dans le cancer de type leucémie et la sous-expression de CDKN1A. Cependant, aucune étude ne démontre une association entre CDKN1A et le facteur *Human acute myelogenous leukemia*. Pour le variant -791, un seul complexe dans la lignée cellulaire Jeg3 est retrouvé sur les analyses de retard sur gel et ce complexe est associé à l'allèle C. Le facteur de transcription prédit par les analyses informatiques pour ce variant, se trouve associé à l'allèle T, donc il s'agit d'un facteur inconnu qui crée le complexe de liaison différentielle à la position -791. Les liaisons différentielles de facteurs de transcription trouvées aux rSNPs -1284 et -134 n'ont pu être associées à aucun facteur de transcription par les prédictions informatiques. Par conséquent, des analyses plus approfondies de ces complexes devraient être produites pour possiblement identifier ces facteurs. La sous-régulation de l'haplotype H12 de CDKN1A peut alors être associée à une combinaison des effets des variants -1021 et -791.

Lors des expériences de transfections transitoires, un second haplotype régulateur (H2) a montré une différence d'expression significative. L'expression différentielle démontrée par cet haplotype peut être associée à un seul variant distinct de l'haplotype majeur (H1), -899T>G. Plusieurs liaisons différentielles ont été trouvées pour le rSNP -899 (figure 3.4). De plus, des complexes se sont formés sur les deux allèles du variant. L'allèle mineur (-899G) présent dans l'haplotype H2 est prédit pour se lier au facteur C-myb. Ce facteur de transcription (C-myb), connu comme oncogène pour son rôle dans le développement de la leucémie lymphoblastique aigüe de type T (Lahortiga and others 2007), pourrait être relié à la diminution d'expression de l'haplotype H2 de CDKN1A .

#### 4.3.3 CDKN1B

La protéine CDKN1B est un inhibiteur de l'activité des CDK qui agit en cas de dommages à l'ADN. Cette protéine permet à l'aide des autres inhibiteurs de CDK d'assurer le contrôle de la progression du cycle cellulaire. Une étude a démontrée une régulation réciproque entre CDKN1B et le complexe cycline E-CDK2 (Sheaff and others 1997). Puisque la cellule a besoin d'énergie pour se répliquer, elle se trouve également dépendante du niveau d'ATP disponible pour faire progresser le cycle cellulaire. Lorsque la concentration du complexe cycline E-CDK2 est près du niveau physiologique de l'ATP (> 1mM), à ce moment le complexe phosphoryle CDKN1B sur le résidu thréonine 187, ce qui conduit à son élimination de la cellule et à la progression du cycle cellulaire en phase S. Dès que la concentration de l'ATP dans la cellule se retrouve à un bas niveau, les fonctions inhibitrices de CDKN1B sont augmentées et la cellule arrête sa prolifération. La protéine CDKN1B est connue pour être impliquée dans divers types de cancers. Par exemple, la perte de CDKN1B dans la différenciation cellulaire ostéogénique conduit à la formation d'ostéosarcomes (Thomas and others 2004). Également, une étude a démontrée une coopération des facteurs C-myc et FOXO3a qui se lient à la région promotrice de CDKN1B et régulent son expression (Chandramohan and others 2008). Cependant, une expression anormalement élevée de C-myc empêche l'activation de la transcription de CDKN1B par FOXO3a, ce qui conduit la cellule au développement de cancers tels que la leucémie, le

cancer de la prostate et le cancer de la langue. Une signalisation oncogénique de la cellule peut mener à la protéolyse de p27 de façon accélérée, ce qui conduit également à la formation de plusieurs types de cancers : du cerveau, de la cavité buccale, thyroïdien, du larynx, de l'œsophage, des poumons, de l'estomac, du sein, du foie, du colon, des ovaires, de la vessie, de la prostate, de l'endomètre, des lymphomes, mélanomes ou leucémie (Besson and others 2007; Chu and others 2008; Philipp-Staheli and others 2001; Slingerland and Pagano 2000). Il est également proposé que p27 pourrait contribuer activement au développement de tumeurs; par sa localisation dans le cytoplasme (Besson and others 2007). Des études ont démontrées que des souris p27<sup>-/-</sup> sont moins susceptibles de développer une tumeur que les souris p27<sup>+/-</sup> (Besson and others 2007; Sicinski and others 2007).

Nous avons démontré lors des expériences de transfections transitoires, une diminution significative de l'expression de l'haplotype H8 (figure 3.5). Cette baisse d'expression pourrait être expliquée par une liaison différentielle de facteurs de transcription sur un ou plusieurs des trois variants suivants qui diffèrent de l'haplotype majeur (H3) : -337, -373 et -1857 (tableau IV). L'allèle -373G est l'allèle majeur, présent à 67,5% de la population et chez tous les haplotypes, à l'exception de l'haplotype majeur H3. Il est peut probable que ce variant cause la diminution d'expression présente seulement avec l'haplotype H8, car lors des expériences de retard sur gel, aucune liaison différentielle n'a été observée sur ce variant (tableau XIX). À l'opposé, les allèles mineurs -337C et -1857T sont présents uniquement sur l'haplotype H8 et pourraient possiblement expliquer l'expression différentielle observée. En effet, le variant -337 démontre une liaison différentielle lors des expériences de retard sur gel (figure 3.6). Par contre, aucun facteur de transcription n'est prédit pour expliquer cette liaison différentielle. Le variant -1857 ne démontre pas de liaison de facteur qui diffère entre les deux allèles. On peut donc affirmer que la baisse d'expression est causée par la liaison de facteurs de transcription, qui est affectée par la présence du variant -337C>A. Des analyses plus approfondies devraient être faites pour essayer d'identifier ce facteur, par exemple à l'aide d'anticorps contre les facteurs connus se liant au promoteur CDKN1B discutés dans les études précédemment mentionnées.

#### 4.3.4 CDKN2A

Le gène CDKN2A a été grandement étudié au cours des dernières années. Son rôle dans l'arrêt du cycle cellulaire au point de contrôle G1/S est bien défini. Cet inhibiteur de CDK se lie aux CDK4 et CDK6, qui forment toutes deux un complexe avec la cycline D, les empêchant ainsi de produire leur activité de phosphorylation de la protéine Rb (Serrano and others 1993; Stott and others 1998). La protéine Rb qui demeure alors hypophosphorylée reste attachée au facteur de transcription E2F pour l'empêcher de transcrire les gènes essentiels à la phase S. Donc, la cellule ne peut poursuivre sa progression dans le cycle cellulaire et alors les mécanismes de réparation de l'ADN peuvent agir. La délétion de façon homozygote ou hétérozygote du gène CDKN2A (Cairns and others 1995; Hebert and others 1994; Igaki and others 1994; Kamb and others 1994; Liu and others 1995; Nobori and others 1994; Ogawa and others 1994; Okuda and others 1995; Petty and others 1993; Serrano and others 1996; Tsutsumi and others 1998; Williamson and others 1995), des mutations entraînant la formation d'une protéine p16 tronquée (Della Torre and others 2001; FitzGerald and others 1996; Goldstein 2004; Honoki and others 2007; Koh and others 1995; Lukas and others 1995; Murphy and others 2004; Quelle and others 1997; Ranade and others 1995; Smith-Sorensen and Hovig 1996) ou une inactivation du gène CDKN2A (Kratzke and others 1995), ont été bien défini comme étant souvent impliqué dans l'oncogenèse.

Nos résultats d'analyses d'expression ont démontrés pour le gène CDKN2A (figure 3.7) un très haut niveau de transcrits de l'haplotype H2, de l'ordre de 2,8 fois et plus l'haplotype majeur H1, dépendamment des lignées cellulaires. Cet haplotype présent chez 7,5% de la population pourrait avoir un effet protecteur quant au développement de maladies complexes, tel que discuté précédemment par de nombreuses études. Il a été démontré que CDKN2A joue un rôle de gène supresseur de tumeurs et qu'il est d'ailleurs impliqué, en lien avec se rôle, dans l'arrêt permanent de prolifération des cellules par l'induction de la sénescence (Ohtani and others 2001). La liaison du facteur de transcription Ets1, qui s'accumule dans la cellule, au promoteur de CDKN2A, entraîne une augmentation de la

transcription de CDKN2A et un arrêt de la prolifération cellulaire (sénescence). Cependant, cette explication impliquant le facteur Ets1, ne corrèle pas avec nos données d'analyses de facteurs de transcription. L'allèle -222A, présent sur l'haplotype H2, est le seul variant qui diffère de l'haplotype majeur H1. Les analyses de retard sur gel de ce site, démontre la présence d'un complexe spécifique à l'allèle -222A (figure 3.8). Les prédictions informatiques obtenues prédisent la perte du site de liaison du facteur myb en présence de l'allèle -222A, mais aucun gain de liaison de facteur en présence de cet allèle (tableau XIX). Par conséquent, nous pouvons supposer que le facteur myb n'est pas la cause de cette liaison différentielle. De plus, aucun site de liaison ETS n'est reconnu à cet endroit dans la séquence du gène, alors le facteur Ets1 n'est pas non plus en cause. Des analyses plus approfondies seraient nécessaire pour déterminer le facteur de transcription responsable de cette augmentation significative d'expression de l'haplotype H2. À l'inverse, on peut également prédire que l'haplotype majeur produit une susceptibilité accrue au développement de tumeurs. Il a été démontré par la plupart des études effectuées sur le gène CDKN2A, tel que mentionné précédemment, qu'une délétion du gène entraîne dans la plupart des cas le développement de tumeurs, tel que le cancer du cerveau, des poumons, du sein, des os, de la peau, du foie, des ovaires, un lymphome, une leucémie ou la formation de mélanomes.

#### **4.3.5 MADH3**

Les gènes MADH3 et MADH4 agissent de façon synergique en réponse à la signalisation anti-tumorale du gène TGF $\beta$ 1 (Zawel and others 1998; Zhang and others 1996). MADH3 est un médiateur direct du signal anti-prolifératif produit par TGF $\beta$ 1 (Chen and others 2002). Les récepteurs cellulaires de TGF $\beta$ 1 vont activer MADH3, localisé dans le cytoplasme, par phosphorylation (Inman and others 2002). Par la suite, MADH3 se lie et forme un complexe avec MADH4 dans le cytoplasme et ce complexe va ensuite migrer au noyau pour agir comme facteur de transcription et activer l'expression des inhibiteurs de CDK, ce qui va entraîner l'arrêt de la progression du cycle cellulaire au point de contrôle G1/S (Blobe and others 2000; Matsuura and others 2004). De plus, MADH3 possède



également la capacité de sous-réguler l'expression de l'oncogène C-myc. Puisque MADH3 est un important gène impliqué dans la réponse aux signaux antiprolifératifs par TGF $\beta$ 1, on peut supposer une possible implication de ce gène dans le développement de cancers, lorsque celui-ci est délété ou que sa protéine transcrite est tronquée. En effet, une étude a montrée que la perte de la protéine MADH3 fonctionnelle est reliée au développement de la leucémie de type T (ALL) (Sekimoto and others 2007; Wolfram and others 2004).

Nous avons trouvé dans les analyses d'expression, une surexpression de l'haplotype H5 dans les trois lignées cellulaires (Hela, Jeg3 et HepG2) comparativement au niveau moyen fonctionnel de tous les autres haplotypes testés (H1, H2 et H3) (figure 3.9). L'haplotype H5 présent chez 21% de la population, diffère de l'haplotype majeur H1 et de tous les autres haplotypes testés par la présence de l'allèle mineur A du variant -1946 (tableau VI). Par les résultats d'expérience de retard sur gel pour le variant -1946, on peut distinguer deux complexes se liant spécifiquement à l'allèle A (figure 3.10). Les prédictions informatiques, indiquent la gain d'un site de liaison du facteur Ets1 sur l'allèle -1946A. Aucune étude ne semble démontrer une régulation de l'expression de MADH3 par le facteur Ets1. Cependant, quelques études démontrent que les facteurs MADH3, Ets1 et souvent SP1 agissent ensemble de façon synergique pour augmenter l'expression de la molécule *parathyroid hormone-related protein* (PTHrP) dans les cellules cancéreuses du sein en réponse au signal antiprolifératif médié par TGF $\beta$ 1 (Lindemann and others 2003). Il est alors probable que les facteurs MADH3 et Ets1 qui agissent en synergie, pourrait également agir de façon réciproque pour augmenter l'expression de leur collaborateur et ainsi Ets1 augmenterait l'expression de MADH3. Le second facteur apte à se lier au promoteur MADH3 lorsque l'allèle -1946A est présent demeure inconnu et aucune prédiction informatique n'est ressortie pour aider à détecter de quel facteur il s'agit. Le gain de liaison d'un ou des deux facteurs confèrent à l'haplotype H5 une augmentation de l'expression de la protéine MADH3 et ainsi, un effet protecteur quant au développement de tumeurs.

#### 4.3.6 TFDPI

Le facteur de transcription DP1 est un joueur important dans la régulation du point de contrôle G1/S du cycle cellulaire. Le facteur DP1 est un partenaire connu pour former une hétérodimérisation avec le facteur E2F (Yasui and others 2003). La formation de l'hétérodimère DP1/E2F est essentielle pour permettre au facteur E2F de se lier avec une haute affinité aux promoteurs des gènes dont il régule l'expression (Attwooll and others 2004; Bandara and others 1993). Le facteur DP1 est donc requis pour permettre la transcription des gènes de la phase S, par le facteur E2F, au point de contrôle G1/S; suite à la libération du facteur E2F par la protéine Rb hyperphosphorylée. Par contre, DP1 ne peut non plus induire la transcription seul, puisqu'il ne possède pas de domaine d'activation de la transcription, ni de domaine de liaison à la protéine Rb qui sont tous deux positionnés sur le facteur E2F (Wang and others 2001). Le facteur DP1 a été identifié comme étant impliqué dans l'oncogenèse. La surexpression du facteur DP1 contribue à la progression des carcinomes de cellules hépatiques, en augmentant la croissance des tumeurs (Yasui and others 2003). Aussi, une autre étude a conclu que la surexpression du facteur DP1 est une cible de la progression du cancer du sein (Abba and others 2007). Toutefois le facteur DP1 demeure peu étudié au cours des dernières années.

Nos résultats d'analyses d'expression par transfection transitoire démontrent une sous-régulation des haplotypes H3 et H4, comparativement à l'haplotype majeur H1 (figure 3.11). Cette baisse d'expression est significative pour les deux haplotypes H3 et H4. Puisqu'il a été mentionné précédemment que la surexpression du facteur DP1 contribue à la progression des tumeurs, on peut supposer que les haplotypes H3 et H4 possèdent un effet protecteur contre la progression de tumeurs. Par contre, à l'inverse une concentration trop faible du facteur DP1 disponible, pourrait diminuer le nombre de complexes DP1/E2F formés et ainsi ralentir la progression du cycle cellulaire en phase S, ce qui nuirait au développement normal de la cellule et pourrait entraîner possiblement des malformations ou autres anomalies du développement. Notons qu'une étude (Attwooll and others 2004) faite sur des cellules embryonnaires DP1 -/- n'a pas trouvée de diminution de la prolifération cellulaire. Par contre, il a été observé dans cette étude que l'embryon formé

meurt rapidement dû aux nombreuses altérations présentes dans les cellules du trophoblaste. Des expériences de modèles animaux de génotype DP1<sup>-/-</sup> serait requises pour éclaircir la nécessité de la formation de l'hétérodimaire DP1/E2F, tel que mentionné par Attwooll.

La sous-régulation des haplotypes H3 et H4 pourrait être associée au variant -1394, -1655, -1779, -1883 ou -1887. Les variants -1394, -1655, -1883 et -1887 ont démontrés, lors des expériences de retard sur gel, une liaison différentielle dans une ou plusieurs des lignées cellulaires testées (tableau XIX). Les complexes ADN-protéine qui se sont formés lors de l'expérience sont associés aux sondes suivantes : -1394T, -1655GA, -1883C et -1887T. Ces variants pourraient expliquer la baisse d'expression de l'haplotype régulateur H4. Plusieurs facteurs de transcription ont été prédits pour avoir leur site de liaison affecté par ces variants. Cependant, puisque peu d'études ont été effectuées sur le gène TFDPI, il nous est impossible de prédire de façon adéquate si un facteur prédit est de façon plus probable en cause. Des études futures concernant ce facteur pourront aider à éclaircir ce point. Le faible niveau d'expression de l'haplotype H3 ne peut être expliqué par notre étude, car le variant présent sur cet haplotype et différant de l'haplotype majeur, ne présente aucune liaison différentielle de facteurs de transcription. Le gène DP1, nécessite certainement des analyses futures pour permettre de tirer des conclusions de façon adéquate sur les effets et les causes de ces différences d'expression.

#### **4.3.7 TGFβ1**

TGFβ1, qui est exprimé de façon ubiquitaire, est une cytokine multifonctionnelle connue pour jouer un rôle crucial dans les processus normaux d'une cellule, mais également dans le développement de maladies complexes (Blobe and others 2000). La molécule TGFβ1, qui est diffusée via le plasma, agit à différent niveaux dans plusieurs maladies, telles que : le cancer (Derynck and others 2001), la fibrose cardiaque (Zeisberg and others 2007), l'obésité (Long and others 2003), l'oedème pulmonaire (Pittet and others 2001), la Sclérodémie (Dong and others 2002), la néphropathie diabétique (Reeves and Andreoli 2000), la dystrophie musculaire de Duchenne (Bernasconi and others 1995), dans les

maladies de type auto-immune, dans la prolifération tumorale et la formation de métastases (Shah and others 2006b). La molécule TGF $\beta$ 1 agit lorsqu'une anomalie est détectée au niveau de l'ADN, en transmettant un signal antiprolifératif à la cellule. La cellule répond à ce signal lors du point de contrôle G1/S. La molécule TGF $\beta$ 1 se rend à la cellule, puis elle se lie à son récepteur membranaire de type II ou III. La formation de ce complexe entraîne le récepteur de type I à s'y joindre, celui-ci se fait phosphoryler et est alors activé. L'activation des protéines kinases du récepteur I, leurs permettra de phosphoryler ensuite les protéines MADH2 ou MADH3 présentes dans le cytoplasme. La protéine MADH2 ou MADH3 phosphorylée se lie à la protéine MADH4 et le complexe migre vers le noyau. Ce complexe agit dans le noyau comme facteur de transcription pour activer la transcription des inhibiteurs de CDK (Blobe and others 2000). Une augmentation de la concentration de TGF $\beta$ 1 dans le plasma est associée à l'oncogenèse (Blobe and others 2000).

Le promoteur du gène TGF $\beta$ 1 a été analysé dans de nombreuses études. Plusieurs rSNPs sont bien caractérisés par ces études. Notamment le rSNP -509T>C (nommé -508A>G dans notre étude), dont le génotype TT est significativement associé à une incidence accrue du cancer du sein, mais toutefois sans affecter le niveau d'expression de TGF $\beta$ 1 (Dunning and others 2003). Par contre, ce rSNP est connu pour être en déséquilibre de liaison avec le SNP +29T>C, qui affecte la sécrétion de TGF $\beta$ 1. Le génotype -509 TT a également été associé à une augmentation du risque de développer l'asthme (Salam and others 2007). Plusieurs des études ayant trouvées une association entre le génotype TT et une augmentation du risque de développer une maladie supposait qu'un facteur de transcription se liant à l'allèle T, augmentait l'expression de TGF $\beta$ 1. Cependant, une étude plus récente a démontrée qu'il s'agit plutôt d'un facteur de transcription qui se lie à l'allèle -509C qui régule de façon négative l'expression de TGF $\beta$ 1 (Shah and others 2006a). L'allèle -509A est présent chez deux des haplotypes testés (H1 et H2) par transfections transitoires dans notre étude (tableau VIII). Nos résultats d'expression de ces deux haplotypes montrent une augmentation significative de l'expression seulement de l'haplotype H2. De plus, le second haplotype démontrant une surexpression, est l'haplotype H5, qui est associé à l'allèle -509G. Alors, l'explication de la liaison du facteur de transcription négatif apportée par

Shah, ne soutient pas nos résultats. Par contre dans l'étude de Shah, une lignée cellulaire (A549, adénocarcinome de poumon) différentes de celles utilisées dans le projet est employée. Aussi, nos analyses de liaisons de facteurs de transcription pour le variant -509, ne montrent aucune liaison différentielle (tableau XIX). Nous devons donc supposer que la surexpression trouvée avec les haplotypes H2 et H5 est causée par un autre variant.

Les résultats d'expériences de retard sur gel montrent que les variants -1550 INSAGG et -1886G>A possèdent une liaison différentielle de facteurs de transcription. Ces deux variants diffèrent entre l'haplotype H2 et l'haplotype majeur. Le complexe présent au variant -1550 se lie sur l'allèle sans INSAGG, qui est associé à l'haplotype H2. Les prédictions informatiques indiquent qu'il peut s'agir des facteurs PAX4/PAX6 ou Myc. Le facteur c-Myc, connu comme étant une proto-oncogène, agit dans la régulation de l'expression de certains gènes pour entraîner la formation de tumeurs (Wang and others 2007). Ce facteur pourrait possiblement expliquer la haute expression de l'haplotype H2. Les rôles des facteurs PAX4 et PAX6 sont relativement peu définis, alors il est difficile des les associer à la régulation de l'haplotype H2. Pour le variant -1886, les résultats de retard sur gel montrent des complexes en liaisons différentielles, spécifiques aux deux allèles. Pour toutes les prédictions informatiques possible pour ce variant, une étude (Shah and others 2006b) a déjà démontrée la liaison des facteurs de transcription GATA1 et AP4 au promoteur de TGF $\beta$ 1 et que cette liaison entraîne une augmentation du niveau d'expression. Ceci coïncide avec nos résultats de surexpression de l'haplotype H2. Aucun des deux variants (-1550 et -1886), démontrant une liaison différentielle de facteurs de transcription, ne peut expliquer le haut niveau d'expression trouvé sur l'haplotype H5 par rapport à l'haplotype majeur H6; car seul le variant -1571 diffère entre ces haplotypes. Il serait difficile de se prononcer sur une possibilité entraînant cette différence d'expression, car les hypothèses sont nombreuses, mais aucune évidence nous pousse vers l'une d'elle. Les haplotypes H2 et H5 peuvent toutefois être associés à une augmentation de la susceptibilité interindividuelle au cancer.

#### 4.4 Limitations de l'étude

Nous sommes conscients que certaines limitations sont présentes dans les études *In vitro*. En effet, ces études démontrent des rSNPs présents dans les gènes candidats qui pourraient, potentiellement avoir un impact fonctionnel. Toutefois, ces études ne reflètent pas exactement le potentiel que ces rSNPs pourraient avoir *In vivo*, puisque le contexte génomique n'est pas présent. Le contexte génomique naturel étant très sophistiqué, il est possible que le niveau d'expression ne soit pas influencé de la même façon dans un contexte naturel. De plus, la machinerie transcriptionnelle des cellules non transformées, en relation avec les processus de traduction et de post-traduction, peuvent possiblement contrer l'effet du niveau de transcrite observé *In vitro*. Il se peut que des changements de la concentration du niveau de transcrits d'ARNm, ne résultent pas en un changement au niveau de l'abondance des protéines. Également, lors des expériences de transfections transitoires, les constructions pénètrent dans les cellules en de nombreuses copies et ne sont pas insérées dans le génome de la cellule. Ceci influence beaucoup l'effet de la construction haplotypique comparativement au contexte naturel. Une nouvelle approche a été mise au point dans le laboratoire pour permettre de contrer cette limitation. Les constructions seront alors intégrées dans le génome de la cellule toujours au même endroit et en une seule copie, par ce nouveau système. Ceci permettra de réduire une grande partie des limitations des études *In vitro*. L'utilisation du logiciel Mat Inspector a apporté d'autres limitations. Ce logiciel permet de prédire la présence du site de liaison d'une famille de facteurs de transcription, mais il ne permet pas l'identification exacte du facteur membre de la famille possédant le site de liaison. De plus, la base de données de Mat Inspector contient les sites de liaison des facteurs de transcription communs, ce qui veut dire que plusieurs facteurs de transcription moins connus à ce jour, qui pourraient possiblement se lier à la séquence que l'on soumet à l'étude, ne sont pas prédit. Pour les expériences de retard, les résultats des complexes ADN-protéine obtenus sont dépendants des lignées cellulaires utilisées. Plusieurs facteurs de transcription sont spécifiques à un type de tissu, ce qui explique que les complexes ADN-protéine formés peuvent différer selon la lignée cellulaire utilisée pour obtenir l'extrait nucléaire. Afin de contrer cet effet, nous avons utilisé trois lignées cellulaires provenant de tissus très différents : HepG2 (hépatocarcinome), Hela (carcinome

du col utérin), Jeg3 (carcinome de placenta). Les expériences de retard sur gel en présence d'anticorps, *supershift*, auraient pu permettre d'identifier exactement les facteurs de transcription présents dans les différents complexes ADN-protéine. Cependant, le but de l'étude était de trouver si une liaison ADN-protéine pourrait possiblement être la cause de la différence de niveau d'expression observé entre les haplotypes et non de trouver de quel facteur il s'agit.

#### 4.5 Conclusion

En conclusion, nous avons analysé la portion promotrice de 2Kb en amont du site d'initiation de la transcription de 18 gènes impliqués dans la régulation du point de contrôle G1/S du cycle cellulaire. De ces gènes, 7 ont montrés une différence allélique significative : CDC25A, CDKN1A, CDKN1B, CDKN2A, MADH3, TGFβ1 et TFDP1. Ces 7 gènes démontrant une différence allélique significative impliquent ces mêmes 26 variants, dont 61,5% présentaient une liaison différentielle lors des expériences de retard sur gel. Les prédictions *In silico* nous ont permis d'associer 34,6% de ces liaisons différentielles à un facteur de transcription dont le site de liaison est modifié par la présence d'un variant. Cinq des 18 gènes candidats (SKP2, CCND1, E2F1, HDAC1 et RB1) qui ont pu être testés de façon fonctionnelle par expériences de transfections transitoires, n'ont démontrés aucune différence d'expression allélique significative selon nos critères. Un gène (MADH4) présentant seulement un singleton, n'a pas été soumis aux analyses pour cause de variabilité interindividuelle insuffisante. Les analyses à compléter des 5 autres gènes (CDKN2B, CCNE, CDK4, CDK6, CDK2) sont présentement en cours dans le laboratoire. Cette étude a permis de mettre en évidence plusieurs rSNPs potentiellement impliqués dans l'oncogenèse. La validation fonctionnelle de 13 des 18 gènes impliqués dans le fonctionnement du point de contrôle G1/S facilitera l'étude ultérieure de ceux-ci, quant à leur implication dans la susceptibilité interindividuelle aux maladies complexes. Le laboratoire a également produit des études d'association concernant certains de ces gènes candidats (Healy and others 2007). Des analyses *In vivo*, par exemple dans des modèles de souris, de ces rSNPs pourraient permettre de valider l'impact réel dans un contexte naturel de ces mutations. Finalement, cette étude a permis d'augmenter les connaissances de la science quant aux variants génétiques présents dans la région promotrice des gènes impliqués dans la régulation du point de contrôle G1/S du cycle cellulaire et d'évaluer l'impact des ces variants dans la susceptibilité interindividuelle au cancer.



## Bibliographie

- Abba MC, Fabris VT, Hu Y, Kittrell FS, Cai WW, Donehower LA, Sahin A, Medina D, Aldaz CM. 2007. Identification of novel amplification gene targets in mouse and human breast cancer at a syntenic cluster mapping to mouse ch8A1 and human ch13q34. *Cancer Res* 67(9):4104-12.
- Ansari M, St-Onge G, Krajcinovic M. 2007. [Pharmacogenomics of acute lymphoblastic leukemia]. *Med Sci (Paris)* 23(11):961-7.
- Arima Y, Hirota T, Bronner C, Mousli M, Fujiwara T, Niwa S, Ishikawa H, Saya H. 2004. Down-regulation of nuclear protein ICBP90 by p53/p21Cip1/WAF1-dependent DNA-damage checkpoint signals contributes to cell cycle arrest at G1/S transition. *Genes Cells* 9(2):131-42.
- Attwooll C, Lazzerini Denchi E, Helin K. 2004. The E2F family: specific functions and overlapping interests. *EMBO J* 23(24):4709-16.
- Bachmann HS, Otterbach F, Callies R, Nuckel H, Bau M, Schmid KW, Siffert W, Kimmig R. 2007. The AA genotype of the regulatory BCL2 promoter polymorphism ( 938C>A) is associated with a favorable outcome in lymph node negative invasive breast cancer patients. *Clin Cancer Res* 13(19):5790-7.
- Bai L, Merchant JL. 2001. ZBP-89 promotes growth arrest through stabilization of p53. *Mol Cell Biol* 21(14):4670-83.
- Bandara LR, Buck VM, Zamanian M, Johnston LH, La Thangue NB. 1993. Functional synergy between DP-1 and E2F-1 in the cell cycle-regulating transcription factor DRTF1/E2F. *EMBO J* 12(11):4317-24.
- Bernasconi P, Torchiana E, Confalonieri P, Brugnoli R, Barresi R, Mora M, Cornelio F, Morandi L, Mantegazza R. 1995. Expression of transforming growth factor-beta 1 in dystrophic patient muscles correlates with fibrosis. Pathogenetic role of a fibrogenic cytokine. *J Clin Invest* 96(2):1137-44.
- Besson A, Hwang HC, Cicero S, Donovan SL, Gurian-West M, Johnson D, Clurman BE, Dyer MA, Roberts JM. 2007. Discovery of an oncogenic activity in p27Kip1 that causes stem cell expansion and a multiple tumor phenotype. *Genes Dev* 21(14):1731-46.
- Blobe GC, Schiemann WP, Lodish HF. 2000. Role of transforming growth factor beta in human disease. *N Engl J Med* 342(18):1350-8.
- Bougeard G, Baert-Desurmont S, Tournier I, Vasseur S, Martin C, Brugieres L, Chompret A, Bressac-de Paillerets B, Stoppa-Lyonnet D, Bonaiti-Pellie C and others. 2006. Impact of the MDM2 SNP309 and p53 Arg72Pro polymorphism on age of tumour onset in Li-Fraumeni syndrome. *J Med Genet* 43(6):531-3.
- Bourgeois S, Labuda D. 2004. Dynamic allele-specific oligonucleotide hybridization on solid support. *Anal Biochem* 324(2):309-11.
- Boutros R, Lobjois V, Ducommun B. 2007. CDC25 phosphatases in cancer cells: key players? Good targets? *Nat Rev Cancer* 7(7):495-507.
- Buratowski S, Chodosh LA. 2001. Mobility shift DNA-binding assay using gel electrophoresis. *Curr Protoc Mol Biol* Chapter 12:Unit 12 2.
- Cairns P, Polascik TJ, Eby Y, Tokino K, Califano J, Merlo A, Mao L, Herath J, Jenkins R, Westra W and others. 1995. Frequency of homozygous deletion at p16/CDKN2 in primary human tumours. *Nat Genet* 11(2):210-2.

- Campbell NA. 1995. *Biology*. Mathieu R, translator. Inc Édrrp, editor: The Benjamin/Cummings Publishing company, Inc. 1190 p.
- Chandramohan V, Mineva ND, Burke B, Jeay S, Wu M, Shen J, Yang W, Hann SR, Sonenshein GE. 2008. c-Myc represses FOXO3a-mediated transcription of the gene encoding the p27(Kip1) cyclin dependent kinase inhibitor. *J Cell Biochem*.
- Chen CR, Kang Y, Siegel PM, Massague J. 2002. E2F4/5 and p107 as Smad cofactors linking the TGFbeta receptor to c-myc repression. *Cell* 110(1):19-32.
- Chu IM, Hengst L, Slingerland JM. 2008. The Cdk inhibitor p27 in human cancer: prognostic potential and relevance to anticancer therapy. *Nat Rev Cancer* 8(4):253-67.
- Csink AK, Henikoff S. 1998. Something from nothing: the evolution and utility of satellite repeats. *Trends Genet* 14(5):200-4.
- Dallaire L. 2006. *Glosaire de génétique médicale et moléculaire Sainte-Justine ÉdC*, editor. Montréal: Marquis. 127 p.
- Della Torre G, Pasini B, Frigerio S, Donghi R, Rovini D, Delia D, Peters G, Huot TJ, Bianchi-Scarra G, Lantieri F and others. 2001. CDKN2A and CDK4 mutation analysis in Italian melanoma-prone families: functional characterization of a novel CDKN2A germ line mutation. *Br J Cancer* 85(6):836-44.
- Derynck R, Akhurst RJ, Balmain A. 2001. TGF-beta signaling in tumor suppression and cancer progression. *Nat Genet* 29(2):117-29.
- Dong C, Zhu S, Wang T, Yoon W, Li Z, Alvarez RJ, ten Dijke P, White B, Wigley FM, Goldschmidt-Clermont PJ. 2002. Deficient Smad7 expression: a putative molecular defect in scleroderma. *Proc Natl Acad Sci U S A* 99(6):3908-13.
- Donovan J, Slingerland J. 2000. Transforming growth factor-beta and breast cancer: Cell cycle arrest by transforming growth factor-beta and its disruption in cancer. *Breast Cancer Res* 2(2):116-24.
- Dunning AM, Ellis PD, McBride S, Kirschenlohr HL, Healey CS, Kemp PR, Luben RN, Chang-Claude J, Mannermaa A, Kataja V and others. 2003. A transforming growth factorbeta1 signal peptide variant increases secretion in vitro and is associated with increased incidence of invasive breast cancer. *Cancer Res* 63(10):2610-5.
- Easton J, Wei T, Lahti JM, Kidd VJ. 1998. Disruption of the cyclin D/cyclin-dependent kinase/INK4/retinoblastoma protein regulatory pathway in human neuroblastoma. *Cancer Res* 58(12):2624-32.
- el-Deiry WS, Tokino T, Velculescu VE, Levy DB, Parsons R, Trent JM, Lin D, Mercer WE, Kinzler KW, Vogelstein B. 1993. WAF1, a potential mediator of p53 tumor suppression. *Cell* 75(4):817-25.
- FitzGerald MG, Harkin DP, Silva-Arrieta S, MacDonald DJ, Lucchina LC, Unsal H, O'Neill E, Koh J, Finkelstein DM, Isselbacher KJ and others. 1996. Prevalence of germline mutations in p16, p19ARF, and CDK4 in familial melanoma: analysis of a clinic-based population. *Proc Natl Acad Sci U S A* 93(16):8541-5.
- Gartel AL. 2007. P21(WAF1/CIP1) may be a tumor suppressor after all. *Cancer Biol Ther* 6(8):1171-2.
- Gazquez C, Oriola J, de Mateo S, Vidal-Taboada JM, Balleca JL, Oliva R. 2008. A Common Protamine 1 Promoter Polymorphism (-190 C to A) Correlates with Abnormal Sperm Morphology and Increased Protamine P1/P2 Ratio in Infertile Patients. *J Androl*.
- Giacinti C, Giordano A. 2006. RB and cell cycle progression. *Oncogene* 25(38):5220-7.
- Goldstein AM. 2004. Familial melanoma, pancreatic cancer and germline CDKN2A mutations. *Hum Mutat* 23(6):630.

- Gonzalez-Fernandez O, Jimenez A, Villalobo A. 2008. Differential p38 mitogen-activated protein kinase-controlled hypophosphorylation of the retinoblastoma protein induced by nitric oxide in neuroblastoma cells. *Free Radic Biol Med* 44(3):353-66.
- Gstaiger M, Jordan R, Lim M, Catzavelos C, Mestan J, Slingerland J, Krek W. 2001. Skp2 is oncogenic and overexpressed in human cancers. *Proc Natl Acad Sci U S A* 98(9):5043-8.
- Harper JW, Adami GR, Wei N, Keyomarsi K, Elledge SJ. 1993. The p21 Cdk-interacting protein Cip1 is a potent inhibitor of G1 cyclin-dependent kinases. *Cell* 75(4):805-16.
- Healy J, Belanger H, Beaulieu P, Lariviere M, Labuda D, Sinnett D. 2007. Promoter SNPs in G1/S checkpoint regulators and their impact on the susceptibility to childhood leukemia. *Blood* 109(2):683-92.
- Hebert J, Cayuela JM, Berkeley J, Sigaux F. 1994. Candidate tumor-suppressor genes MTS1 (p16INK4A) and MTS2 (p15INK4B) display frequent homozygous deletions in primary cells from T- but not from B-cell lineage acute lymphoblastic leukemias. *Blood* 84(12):4038-44.
- Honoki K, Stojanovski E, McEvoy M, Fujii H, Tsujiuchi T, Kido A, Takakura Y, Attia J. 2007. Prognostic significance of p16 INK4a alteration for Ewing sarcoma: a meta-analysis. *Cancer* 110(6):1351-60.
- Hoogendoorn B, Coleman SL, Guy CA, Smith K, Bowen T, Buckland PR, O'Donovan MC. 2003. Functional analysis of human promoter polymorphisms. *Hum Mol Genet* 12(18):2249-54.
- Hu Z, Jin G, Wang L, Chen F, Wang X, Shen H. 2007. MDM2 promoter polymorphism SNP309 contributes to tumor susceptibility: evidence from 21 case-control studies. *Cancer Epidemiol Biomarkers Prev* 16(12):2717-23.
- Igaki H, Sasaki H, Kishi T, Sakamoto H, Tachimori Y, Kato H, Watanabe H, Sugimura T, Terada M. 1994. Highly frequent homozygous deletion of the p16 gene in esophageal cancer cell lines. *Biochem Biophys Res Commun* 203(2):1090-5.
- Inman GJ, Nicolas FJ, Hill CS. 2002. Nucleocytoplasmic shuttling of Smads 2, 3, and 4 permits sensing of TGF-beta receptor activity. *Mol Cell* 10(2):283-94.
- Jarman AP, Wells RA. 1989. Hypervariable minisatellites: recombinators or innocent bystanders? *Trends Genet* 5(11):367-71.
- Jeffreys AJ, Wilson V, Thein SL. 1985. Hypervariable 'minisatellite' regions in human DNA. *Nature* 314(6006):67-73.
- Joosten PH, Toepoel M, Mariman EC, Van Zoelen EJ. 2001. Promoter haplotype combinations of the platelet-derived growth factor alpha-receptor gene predispose to human neural tube defects. *Nat Genet* 27(2):215-7.
- Juhanson P, Kepp K, Org E, Veldre G, Kelgo P, Rosenberg M, Viigimaa M, Laan M. 2008. N-acetyltransferase 8, a positional candidate for blood pressure and renal regulation: resequencing, association and in silico study. *BMC Med Genet* 9(1):25.
- Kamb A, Gruis NA, Weaver-Feldhaus J, Liu Q, Harshman K, Tavtigian SV, Stockert E, Day RS, 3rd, Johnson BE, Skolnick MH. 1994. A cell cycle regulator potentially involved in genesis of many tumor types. *Science* 264(5157):436-40.
- Karlin S, Burge C. 1995. Dinucleotide relative abundance extremes: a genomic signature. *Trends Genet* 11(7):283-90.
- Knight JC. 2003. Functional implications of genetic variation in non-coding DNA for disease susceptibility and gene regulation. *Clin Sci (Lond)* 104(5):493-501.

- Koh J, Enders GH, Dynlacht BD, Harlow E. 1995. Tumour-derived p16 alleles encoding proteins defective in cell-cycle inhibition. *Nature* 375(6531):506-10.
- Kratzke RA, Otterson GA, Lincoln CE, Ewing S, Oie H, Geradts J, Kaye FJ. 1995. Immunohistochemical analysis of the p16INK4 cyclin-dependent kinase inhibitor in malignant mesothelioma. *J Natl Cancer Inst* 87(24):1870-5.
- Kristjansdottir K, Rudolph J. 2004. Cdc25 phosphatases and cancer. *Chem Biol* 11(8):1043-51.
- Labuda D, Krajcinovic M, Richer C, Skoll A, Sinnott H, Yotova V, Sinnott D. 1999. Rapid detection of CYP1A1, CYP2D6, and NAT variants by multiplex polymerase chain reaction and allele-specific oligonucleotide assay. *Anal Biochem* 275(1):84-92.
- Lahortiga I, De Keersmaecker K, Van Vlierberghe P, Graux C, Cauwelier B, Lambert F, Mentens N, Beverloo HB, Pieters R, Speleman F and others. 2007. Duplication of the MYB oncogene in T cell acute lymphoblastic leukemia. *Nat Genet* 39(5):593-5.
- Latres E, Chiarle R, Schulman BA, Pavletich NP, Pellicer A, Inghirami G, Pagano M. 2001. Role of the F-box protein Skp2 in lymphomagenesis. *Proc Natl Acad Sci U S A* 98(5):2515-20.
- Lazarus R, Vercelli D, Palmer LJ, Klimecki WJ, Silverman EK, Richter B, Riva A, Ramoni M, Martinez FD, Weiss ST and others. 2002. Single nucleotide polymorphisms in innate immunity genes: abundant variation and potential role in complex human disease. *Immunol Rev* 190:9-25.
- Lindemann RK, Braig M, Ballschmieter P, Guise TA, Nordheim A, Dittmer J. 2003. Protein kinase C $\alpha$  regulates Ets1 transcriptional activity in invasive breast cancer cells. *Int J Oncol* 22(4):799-805.
- Liu Q, Yan YX, McClure M, Nakagawa H, Fujimura F, Rustgi AK. 1995. MTS-1 (CDKN2) tumor suppressor gene deletions are a frequent event in esophagus squamous cancer and pancreatic adenocarcinoma cell lines. *Oncogene* 10(3):619-22.
- Lo HS, Wang Z, Hu Y, Yang HH, Gere S, Buetow KH, Lee MP. 2003. Allelic variation in gene expression is common in the human genome. *Genome Res* 13(8):1855-62.
- Long JR, Liu PY, Liu YJ, Lu Y, Xiong DH, Elze L, Recker RR, Deng HW. 2003. APOE and TGF- $\beta$ 1 genes are associated with obesity phenotypes. *J Med Genet* 40(12):918-24.
- Lord FM. 1957. A significance test for the hypothesis that two variables measure the same trait except for errors of measurement. *Psychometrika* 22(3):207-220.
- Lukas J, Parry D, Aagaard L, Mann DJ, Bartkova J, Strauss M, Peters G, Bartek J. 1995. Retinoblastoma-protein-dependent cell-cycle inhibition by the tumour suppressor p16. *Nature* 375(6531):503-6.
- Malumbres M, Barbacid M. 2001. To cycle or not to cycle: a critical decision in cancer. *Nat Rev Cancer* 1(3):222-31.
- Mashkova TD, Oparina NY, Lacroix MH, Fedorova LI, IGT, Zinovieva OL, Kisselev LL. 2001. Structural rearrangements and insertions of dispersed elements in pericentromeric alpha satellites occur preferably at kinkable DNA sites. *J Mol Biol* 305(1):33-48.
- Matsuura I, Denissova NG, Wang G, He D, Long J, Liu F. 2004. Cyclin-dependent kinases regulate the antiproliferative function of Smads. *Nature* 430(6996):226-31.
- Milani L, Gupta M, Andersen M, Dhar S, Fryknas M, Isaksson A, Larsson R, Syvanen AC. 2007. Allelic imbalance in gene expression as a guide to cis-acting regulatory single nucleotide polymorphisms in cancer cells. *Nucleic Acids Res* 35(5):e34.

- Mitchison NA. 2001. Polymorphism in regulatory gene sequences. *Genome Biol* 2(1):COMMENT2001.
- Molinari M. 2000. Cell cycle checkpoints and their inactivation in human cancer. *Cell Prolif* 33(5):261-74.
- Morris JF, Hromas R, Rauscher FJ. 1994. Characterization of the DNA-binding properties of the myeloid zinc finger protein MZF1: two independent DNA-binding domains recognize two DNA consensus sequences with a common G-rich core. *Molecular and Cellular Biology* 14(3):1786.
- Murphy JA, Barrantes-Reynolds R, Kocherlakota R, Bond JP, Greenblatt MS. 2004. The CDKN2A database: Integrating allelic variants with evolution, structure, function, and disease association. *Hum Mutat* 24(4):296-304.
- Nobori T, Miura K, Wu DJ, Lois A, Takabayashi K, Carson DA. 1994. Deletions of the cyclin-dependent kinase-4 inhibitor gene in multiple human cancers. *Nature* 368(6473):753-6.
- Ober C, Tan Z, Sun Y, Possick JD, Pan L, Nicolae R, Radford S, Parry RR, Heinzmann A, Deichmann KA and others. 2008. Effect of Variation in CHI3L1 on Serum YKL-40 Level, Risk of Asthma, and Lung Function. *N Engl J Med*.
- Oertel B, Lotsch J. 2008. Genetic mutations that prevent pain: implications for future pain medication. *Pharmacogenomics* 9(2):179-94.
- Ogawa S, Hirano N, Sato N, Takahashi T, Hangaishi A, Tanaka K, Kurokawa M, Tanaka T, Mitani K, Yazaki Y and others. 1994. Homozygous loss of the cyclin-dependent kinase 4-inhibitor (p16) gene in human leukemias. *Blood* 84(8):2431-5.
- Ohtani N, Zebedee Z, Huot TJ, Stinson JA, Sugimoto M, Ohashi Y, Sharrocks AD, Peters G, Hara E. 2001. Opposing effects of Ets and Id proteins on p16INK4a expression during cellular senescence. *Nature* 409(6823):1067-70.
- Okada M, Tessier A, Bai L, Merchant JL. 2006. P53 mutants suppress ZBP-89 function. *Anticancer Res* 26(3A):2023-8.
- Okuda T, Shurtleff SA, Valentine MB, Raimondi SC, Head DR, Behm F, Curcio-Brint AM, Liu Q, Pui CH, Sherr CJ and others. 1995. Frequent deletion of p16INK4a/MTS1 and p15INK4b/MTS2 in pediatric acute lymphoblastic leukemia. *Blood* 85(9):2321-30.
- Orford KW, Scadden DT. 2008. Deconstructing stem cell self-renewal: genetic insights into cell-cycle regulation. *Nat Rev Genet* 9(2):115-28.
- Pajalunga D, Mazzola A, Franchitto A, Puggioni E, Crescenzi M. 2008. The logic and regulation of cell cycle exit and reentry. *Cell Mol Life Sci* 65(1):8-15.
- Pastinen T, Hudson TJ. 2004. Cis-acting regulatory variation in the human genome. *Science* 306(5696):647-50.
- Petty EM, Gibson LH, Fountain JW, Bologna JL, Yang-Feng TL, Housman DE, Bale AE. 1993. Molecular definition of a chromosome 9p21 germ-line deletion in a woman with multiple melanomas and a plexiform neurofibroma: implications for 9p tumor-suppressor gene(s). *Am J Hum Genet* 53(1):96-104.
- Philipp-Staheli J, Payne SR, Kemp CJ. 2001. p27(Kip1): regulation and function of a haploinsufficient tumor suppressor and its misregulation in cancer. *Exp Cell Res* 264(1):148-68.
- Pittet JF, Griffiths MJ, Geiser T, Kaminski N, Dalton SL, Huang X, Brown LA, Gotwals PJ, Kotliansky VE, Matthay MA and others. 2001. TGF-beta is a critical mediator of acute lung injury. *J Clin Invest* 107(12):1537-44.

- Quelle DE, Cheng M, Ashmun RA, Sherr CJ. 1997. Cancer-associated mutations at the INK4a locus cancel cell cycle arrest by p16INK4a but not by the alternative reading frame protein p19ARF. *Proc Natl Acad Sci U S A* 94(2):669-73.
- Ranade K, Hussussian CJ, Sikorski RS, Varmus HE, Goldstein AM, Tucker MA, Serrano M, Hannon GJ, Beach D, Dracopoli NC. 1995. Mutations associated with familial melanoma impair p16INK4 function. *Nat Genet* 10(1):114-6.
- Ray D, Kiyokawa H. 2008. CDC25A phosphatase: a rate-limiting oncogene that determines genomic stability. *Cancer Res* 68(5):1251-3.
- Reeves WB, Andreoli TE. 2000. Transforming growth factor beta contributes to progressive diabetic nephropathy. *Proc Natl Acad Sci U S A* 97(14):7667-9.
- Salam MT, Gauderman WJ, McConnell R, Lin PC, Gilliland FD. 2007. Transforming growth factor-1 C-509T polymorphism, oxidant stress, and early-onset childhood asthma. *Am J Respir Crit Care Med* 176(12):1192-9.
- Sancar A, Lindsey-Boltz LA, Unsal-Kacmaz K, Linn S. 2004. Molecular mechanisms of mammalian DNA repair and the DNA damage checkpoints. *Annu Rev Biochem* 73:39-85.
- Schmitt E, Paquet C, Beauchemin M, Bertrand R. 2007. DNA-damage response network at the crossroads of cell-cycle checkpoints, cellular senescence and apoptosis. *J Zhejiang Univ Sci B* 8(6):377-97.
- Sekimoto G, Matsuzaki K, Yoshida K, Mori S, Murata M, Seki T, Matsui H, Fujisawa J, Okazaki K. 2007. Reversible Smad-dependent signaling between tumor suppression and oncogenesis. *Cancer Res* 67(11):5090-6.
- Serrano M, Hannon GJ, Beach D. 1993. A new regulatory motif in cell-cycle control causing specific inhibition of cyclin D/CDK4. *Nature* 366(6456):704-7.
- Serrano M, Lee H, Chin L, Cordon-Cardo C, Beach D, DePinho RA. 1996. Role of the INK4a locus in tumor suppression and cell mortality. *Cell* 85(1):27-37.
- Shah R, Hurley CK, Posch PE. 2006a. A molecular mechanism for the differential regulation of TGF-beta1 expression due to the common SNP -509C-T (c. -1347C > T). *Hum Genet* 120(4):461-9.
- Shah R, Rahaman B, Hurley CK, Posch PE. 2006b. Allelic diversity in the TGFBI regulatory region: characterization of novel functional single nucleotide polymorphisms. *Hum Genet* 119(1-2):61-74.
- Sheaff RJ, Groudine M, Gordon M, Roberts JM, Clurman BE. 1997. Cyclin E-CDK2 is a regulator of p27Kip1. *Genes Dev* 11(11):1464-78.
- Sherr CJ. 2000. The Pezcoller lecture: cancer cell cycles revisited. *Cancer Res* 60(14):3689-95.
- Sicinski P, Zacharek S, Kim C. 2007. Duality of p27Kip1 function in tumorigenesis. *Genes Dev* 21(14):1703-6.
- Sinnett D, Beaulieu P, Belanger H, Lefebvre JF, Langlois S, Theberge MC, Drouin S, Zotti C, Hudson TJ, Labuda D. 2006a. Detection and characterization of DNA variants in the promoter regions of hundreds of human disease candidate genes. *Genomics* 87(6):704-10.
- Sinnett D, N'Diaye N, Labuda D, Krajcinovic M. 2006b. [Genetic determinants of childhood leukemia]. *Bull Cancer* 93(9):857-65.
- Slingerland J, Pagano M. 2000. Regulation of the cdk inhibitor p27 and its deregulation in cancer. *J Cell Physiol* 183(1):10-7.

- Smith-Sorensen B, Hovig E. 1996. CDKN2A (p16INK4A) somatic and germline mutations. *Hum Mutat* 7(4):294-303.
- Sommer CA, Pavarino-Bertelli EC, Goloni-Bertollo EM, Henrique-Silva F. 2008. Identification of dysregulated genes in lymphocytes from children with Down syndrome. *Genome* 51(1):19-29.
- Stott FJ, Bates S, James MC, McConnell BB, Starborg M, Brookes S, Palmero I, Ryan K, Hara E, Vousden KH and others. 1998. The alternative product from the human CDKN2A locus, p14(ARF), participates in a regulatory feedback loop with p53 and MDM2. *EMBO J* 17(17):5001-14.
- Swanton C. 2004. Cell-cycle targeted therapies. *Lancet Oncol* 5(1):27-36.
- Thomas DM, Johnson SA, Sims NA, Trivett MK, Slavin JL, Rubin BP, Waring P, McArthur GA, Walkley CR, Holloway AJ and others. 2004. Terminal osteoblast differentiation, mediated by runx2 and p27KIP1, is disrupted in osteosarcoma. *J Cell Biol* 167(5):925-34.
- Tsutsumi M, Tsai YC, Gonzalgo ML, Nichols PW, Jones PA. 1998. Early acquisition of homozygous deletions of p16/p19 during squamous cell carcinogenesis and genetic mosaicism in bladder cancer. *Oncogene* 17(23):3021-7.
- Tun T, Hamaguchi Y, Matsunami N, Furukawa T, Honjo T, Kawaichi M. 1994. Recognition sequence of a highly conserved DNA binding protein RBP-J kappa. *Nucleic Acids Res* 22(6):965-71.
- Vignal A, Milan D, SanCristobal M, Eggen A. 2002. A review on SNP and other types of molecular markers and their use in animal genetics. *Genet Sel Evol* 34(3):275-305.
- Waltzer L, Logeat F, Brou C, Israel A, Sergeant A, Manet E. 1994. The human J kappa recombination signal sequence binding protein (RBP-J kappa) targets the Epstein-Barr virus EBNA2 protein to its DNA responsive elements. *EMBO J* 13(23):5633-8.
- Wang D, Russell J, Xu H, Johnson DG. 2001. Deregulated expression of DP1 induces epidermal proliferation and enhances skin carcinogenesis. *Mol Carcinog* 31(2):90-100.
- Wang X, Southard RC, Allred CD, Talbert DR, Wilson ME, Kilgore MW. 2007. MAZ drives tumor-specific expression of PPAR gamma 1 in breast cancer cells. *Breast Cancer Res Treat*.
- Wilkins JJ, Savic RM, Karlsson MO, Langdon G, McIlleron H, Pillai GC, Smith PJ, Simonsson US. 2008. Population pharmacokinetics of rifampin in pulmonary tuberculosis patients including a semi-mechanistic model to describe variable absorption. *Antimicrob Agents Chemother*.
- Williamson MP, Elder PA, Shaw ME, Devlin J, Knowles MA. 1995. p16 (CDKN2) is a major deletion target at 9p21 in bladder cancer. *Hum Mol Genet* 4(9):1569-77.
- Wingender E. 2008. The TRANSFAC project as an example of framework technology that supports the analysis of genomic regulation. *Brief Bioinform*.
- Wolfel T, Hauer M, Schneider J, Serrano M, Wolfel C, Klehmann-Hieb E, De Plaen E, Hankeln T, Meyer zum Buschenfelde KH, Beach D. 1995. A p16INK4a-insensitive CDK4 mutant targeted by cytolytic T lymphocytes in a human melanoma. *Science* 269(5228):1281-4.
- Wolfrain LA, Fernandez TM, Mamura M, Fuller WL, Kumar R, Cole DE, Byfield S, Felici A, Flanders KC, Walz TM and others. 2004. Loss of Smad3 in acute T-cell lymphoblastic leukemia. *N Engl J Med* 351(6):552-9.
- Yasui K, Okamoto H, Arii S, Inazawa J. 2003. Association of over-expressed TFDPI with progression of hepatocellular carcinomas. *J Hum Genet* 48(12):609-13.

- Ye Y, Yang H, Grossman HB, Dinney C, Wu X, Gu J. 2008. Genetic variants in cell cycle control pathway confer susceptibility to bladder cancer. *Cancer*.
- Zakut R, Givol D. 1995. The tumor suppression function of p21 Waf is contained in its N-terminal half ('half-WAF'). *Oncogene* 11(2):393-5.
- Zawel L, Dai JL, Buckhaults P, Zhou S, Kinzler KW, Vogelstein B, Kern SE. 1998. Human Smad3 and Smad4 are sequence-specific transcription activators. *Mol Cell* 1(4):611-7.
- Zeisberg EM, Tarnavski O, Zeisberg M, Dorfman AL, McMullen JR, Gustafsson E, Chandraker A, Yuan X, Pu WT, Roberts AB and others. 2007. Endothelial-to-mesenchymal transition contributes to cardiac fibrosis. *Nat Med* 13(8):952-61.
- Zhang Y, Feng X, We R, Derynck R. 1996. Receptor-associated Mad homologues synergize as effectors of the TGF-beta response. *Nature* 383(6596):168-72.



## Annexe I

Tableau XX : Liste des sondes utilisées dans le géotypage de type ASO.

Gène	rSNP	Séquence sonde (5'-3')
CCND1	-519 C	TTAGTTCTTGCAATT
	-519 T	TTAGTTTTTGCAATT
	-568 A	AAACTTGACAGGGGTT
	-568 C	AAACTTGCCCAGGGGTT
	-766 A	AGGAAGGCAGCCCGAAG
	-766 T	AGGAAGGCTGCCCGAAG
	-1537 DELC	GTCCTGGGGGGTAAC
	-1537 C	GTCCTGGGGGGGTAA
	-1554 G	GGAATCCGGGGGTAA
	-1554 A	GGAATCCAGGGGTAA
	-1597 G	AGGGAGCACCTACGGTC
	-1597 A	AGGGAGCATCTACGGTC
	-1701 G	GAATGAAAAAGCAG
	-1701 C	GAATGAAAAACCAG
	-1737 A	CATCCAGGCGGCCGC
	-1737 T	CATCCTGGCGGCCGC
	-1885 G	CGGAGTGGCGGAGC
	-1885 T	CGGAGTTGGCGGAGC
	-1888 A	GCCGGCGTTGGCG
	-1888 C	GCCGGCGTGGCG
	-1926 INSG	GATGGGGGGGTGAAGC T
	-1926 -	GATGGGGGGGTGAAGCT
	-1931 G	GGATGGGGGGGTGAAG
-1931 C	GATGCGGGGGGTGAAGCT	

Gène	rSNP	Séquence sonde (5'-3')
	-1938 C	TTGCGCCCGGGATGG
	-1938 T	TTGCGCCTGGGATGG
CCNE1	-1366 G	ATAGTGCCTGGACCA
	-1366 A	ATAGTGCTTGGACCA
	-1613T	GCTCTAACTCAAGAAA
	-1613 DELT	GCTCTAACTCAAGAAA
CDC25A	-51 G	CTTTCCTAGTTGGCGC
	-51 C	CTTTCCTACTTGGCGC
	-171 A	AGAGCCGAAGACCTGGC
	-171 T	AGAGCCGATGACCTGGC
	-351 G	CCAGTTAGCTTCGGC
	-351 A	CCAGTTAACTTCGGC
	-1433 C	GCTCAACAGGCCGGG
	-1433 T	GCTCAACAGGCTGGG
	-1440 G	GGGAGCGGTGGCTC
	-1440 A	GGGAGCAGTGGCTC
	-1566 C	TTACCGGGCGTGGTGG
	-1566 T	TTACTGGGCGTGGTGG
	-2030 T	TGGGGCTGGTGGGAAGT
	-2030 G	TGGGGCTGGGGGAAGT
CDK4	-150 C	CGTGACCGCTTGACATTGC
	-150 G	CGTGACCGGTTGACATTGC
	-217 C	CCACAACCCCGCTGCA
	-217 T	CCACAACCCCTGCTGCA
	-290 C	CTAGGCCACCCAGAGCC
	-290 T	CTAGGCCATCCAGAGCC
	-612 C	AATGGGGTCTTCTACAA
	-612 A	AATGGGGTATTCTACAA
	-615 C	GGGTCTTCTACAATAGG

Gène	rSNP	Séquence sonde (5'-3')
	-615 A	GGGTCTTATACAATAGG
	-724 DELTC	GAAAAGTGAGGATT
	-724 TC	GAAAAGTGAGAGGATT
	-847 C	TTTTTCCAGTAGGAAGG
	-847 T	TTTTTTCAGTAGGAAGG
	-937 A	TTATCACATCATATATAAG
	-937 G	TTATCACGTCATATATAAG
	-1178 A	GTAGACACAAGTAGATT
	-1178 G	GTAGACACAGGTAGATT
	-1349 T	GTTAAGGATAACAGAAG
	-1349 A	GTTAAGGAAAACAGAAG
	-1887 T	CTTCTTCTGTTAATTAC
	-1887 C	CTTCTCCGTTAATTAC
CDK6	-638 C	GTGTTGCGCCGCTACCCTCCCC
	-638 T	GTGTTGCGCTGCTACCCTCCCC
	-1124 T	CGCCACAATTCGCTTGTCG
	-1124 A	CGCCACAATTCGCTTGTCG
	-1270 C	CCTGCGCCTCCAGGGGTGAG
	-1270 T	CCTGCGCCTCAGGGGTGAG
	-1386 A	CCCCTCCGGCTAAAGGCCCGG
	-1386 G	CCCCTCCGGCTGAAGGCCCGG
	-1450 A	GCGCCAGAGTAGGGTACCC
	-1450 T	GCGCCAGAGTTGGGTACCC
	-1789 A	CCACACAGGTAGCCGGGGAGC
	-1789 T	CCACACAGGTTGCCGGGGAGC
	-1902 DELCCCCGCA	CGTCCCACCCCGCAGGGAAGTCCG
	-1902 CCCCCGCA CCCCCGCA	CGTCCCACCCCGCACCCCGCAGGG

Gène	rSNP	Séquence sonde (5'-3')
	-1958 C	CCCTGGCTCTCCCCCACTCG
	-1958 G	CCCTGGCTCTGCCCCCACTCG
CDKN1A	-134 G	CTGGCTCGGCGCTGG
	-134 C	GCTGGCTCCGCGCTGG
	-499 G	GGGAATAGAGGTGAT
	-499 C	GGGAATACAGGTGAT
	-519 G	TGACCTAGTGAGGGA
	-519 C	TGACCTACTGAGGGA
	-791 T	GAAATCCCTGTGG
	-791 C	GAAGAAACCCCTGTG
	-899 T	GGGGAAACTGGGGCTCA
	-899 G	GGGGAAACGGGGGCTCA
	-1021 C	ATTTTGGAGCCACAGAA
	-1021 T	ATTTTGGAAACCACAGAA
	-1284 T	ATCCCACTAAAAACAG
	-1284 C	ATCCCACTGAAAAACAG
	-1466 A	GACACAGTAGCAGAC
	-1466 G	GACACAGCAGCAGAC
	-1529 T	GTGGCCCTTCTGG
	-1529 C	GTGGCCCTTCTGG
	-1536 C	GGCCACGGCACAGC
	-1536 T	GGCCACAGCACAGC
-1731 DELA	GGAGAAATTTTTTTTA	
-1731 A	GGAGAAAATTTTTTTTA	
CDKN1B	-221 T	AGAGCAGGTTTGTGGC
	-221 C	AGAGCAGGCTTGTGGC
	-337 C	GGGTTCGCCGCGGCG
	-337 A	GGGTTCGACGCGGCG
	-373 G	TTCCCGGGCGAGGAGC

Gène	rSNP	Séquence sonde (5'-3')
	-373 T	TTCCCGGGAGAGGAGC
	-522 G	CTGGGGGGGCTGGGC
	-522 A	CTGGGGGAGCTGGGC
	-756 A	TTAAAAAAAAAAAAAC
	-756 C	TTAAAAAACAAAAAC
	-1084 DELAGG	AGAGGGCTGAGTGAAA
	-1084 AGG	AGAGGGCTGAAGGGTG
	-1228 C	AGTGTGGGCTTGAGC
	-1228 T	AGTGTGGACTTGAGC
	-1608 G	TGTCCAGGGACATGCAT
	-1608 A	TGTCCAGGAACATGCAT
	-1857 C	CTCAAGACAGCTGCA
	-1857 T	CTCAAGATAGCTGCA
	-1908 DELT	GAAGGACGGGGTTGA
	-1908 T	GAAGGACAGGGGTTG
CDKN2A	-44 A	CAGAGCCAGCGTTGGCA
	-44 T	CAGAGCCTGCGTTGGCA
	-222 T	GGCAGTTTGGGAAGGT
	-222 A	GGCAGTTAGGAAGGT
	-464 C	AGGGGAAGGAGAGAGCA
	-464 T	AGGGGAAGAAGAGAGCA
	-589 DELG	GGACGGGGAGAATTC
	-589 G	GACGGGGGAGAATTC
	-598 G	GAGAATTCTGCCTGTAG
	-598 A	GAGAATTCTACCTGTAG
	-710 C	AGAGTTGCACAGTGA
	-710 A	AGAGTTGAACAGTGA
	-1600 A	CTGGGAGACAAGAGC
	-1600 G	CTGGGAGGCAAGAGC

Gène	rSNP	Séquence sonde (5'-3')
	-1879 C	GGATTATCCAGTAGGCA
	-1879 T	GGATTATCTAGTAGGCA
CDKN2B	-5 C	CGCGTCGGCCCCGGGCCTGGCC
	-5 G	CGCGTCGGCCGGGGCCTGGCC
	-214 INST	AACCAACTGGTCTGGCTGAAG
	-214 -	AACCAACTGGCTGGCTGAAG
	-287 C	CTTAAGAAACACGGAGTTAT
	-287 G	CTTAAGAAAGACGGAGTTAT
	-549 C	TTGCCACTCTCAATCTCGAAC
	-549 T	TTGCCACTCTTAATCTCGAAC
	-593 A	CAGATTCTTAAAGTATAAT
	-593 T	CAGATTCTTATAGTATAAT
	-593 C	CAGATTCTTACAGTATAAT
	-1001 G	GTAAACAAAAACTGTTCACT
	-1001 A	GTAAACAAAAATTGTTCACT
	-1164 G	AAGAAAGAGCTTTTTTTAATTC
	-1164 A	AAGAAAGAGTTTTTTTAATTC
	-1270 C	CGGTTCCCTTCCCTGTCCAG
	-1270 T	CGGTTCCCTTACCTGTCCAG
-1319 A	GAAATGGGGGTTGGATGGGAA	
-1319 G	GAAATGGGGGCTGGATGGGAA	
E2F1	-187 C	CGCGCCCCGCCCCG
	-187 T	CGCGCCTCGCCCCG
	-243 T	CCGGGCTGGGGGC
	-243 G	CCGGGCGGGGGGC
	-300 G	GCGTCGGGGCGGGG
	-300 A	GCGTCGGAGCGGGG
	-897 C	GAGATGGGTCTTCAC
	-897 T	GAGATGGATCTTCAC

Gène	rSNP	Séquence sonde (5'-3')
HDAC1	-162 C	AGAGGCCGCCGAGGG
	-162 T	AGAGGCCACCGAGGG
	-326 A	TCTGATTTAACGGAGAA
	-326 T	TCTGATTTTACGGAGAA
	-455 T	TAATTA AAAAGAAAAGC
	-455 C	TAATTAAGAAGAAAAGC
	-548 G	CCAAGTACTGTGCGA
	-548 A	CCAAGTATTGTGCGA
	-599 C	AGCTCTCGGCCAAGG
	-599 T	AGCTCTCAGCCAAGG
	-783 A	GGAGGTAGATTTTGG
	-783 G	GGAGGTAGGTTTGG
	-905 A	CAAAAAAAAAAAAAAAAAAAAAAC
	-905 C	CAAAAAAAAAAAAAACAAAAAAAAAC
	-1140 G	GGTGGGCGGATCCTG
	-1140 A	GGTGGGCAGATCCTG
	-1193 G	CTAGGCCGGGCGCTG
	-1193 A	CTAGGCCAGGCGCTG
	-1269 T	GGAAGCTATTCTGTG
	-1269 C	GGAAGCTGTTCTGTG
-1546 G	GATAATTGATGATCTGG	
-1546 T	GATAATTGATTATCTGG	
MADH3	-669 C	GCGACCCGGGCGGCG
	-669 T	GCGACCCAGGCGGCG
	-749 G	GGAAGTCCTGGGAAA
	-749 A	GGAAGCTTGGGAAA
	-1137 T	TGGGGACACCAGTGA
	-1137 G	TGGGGACCCAGTGA
	-1938 T	AGCGCGTTAACAGGAG

Gène	rSNP	Séquence sonde (5'-3')
	-1938 C	AGCGCGTCAACAGGAG
	-1946 C	ACAGCGGCAGCGCGT
	-1946 A	ACAGCGGAAGCGCGT
MADH4	-412 G	CAAAAATTGGCCTTCAA
	-412 T	CAAAAATTGTCTTCAA
	-418 A	GCCTTCAACAGCCAGTCG
	-418 C	GCCTTCCACAGCCAGTCG
RB1	-768 C	TTAATCCCCAGAAAC
	-768 A	TTAATCACCAGAAAC
	-1554 C	GGTAGGACTGGCTGT
	-1554 A	GGTAGGAATGGCTGT
	-1733 INSCTGATA	CCAGCCTGATAAGAC
	-1733 -	CCAGCAGACTACTCC
SKP2	-758 G	AGGGCATGGGCAGCCTT
	-758 C	AGGGCATGCGCAGCCTT
	-846 A	ACCCGTGCACGCAG
	-846 G	ACCCGTGCGCGCAG
TFDP1	-1394 T	TGTGGACACACGGAG
	-1394 C	TGTGGACGCACGGAG
	-1515 G	GGGCACGTAGGACG
	-1515 T	GGGCACTTAGGACG
	-1516 C	GGGCACCGAGGACG
	-1516 T	GGGCACTGAGGACG
	-1530 C	CAGGCTGCGTCCGGT
	-1530 T	CAGGCTGTGTCCGGT
	-1655 G	GGCCCTAGCGGCGG
	-1655 A	GGCCCTAACGGGCGG
	-1779 A	GAGGACTTGAGGAGC
	-1779 G	GAGGACTCGAGGAGC



Gène	rSNP	Séquence sonde (5'-3')
	-1883 G	GTGCGGCGTTGCCTG
	-1883 C	GTGCGGCCTTGCCTG
	-1887 C	GTGCGGCGTTGCCTG
	-1887 T	GTGTGGCGTTGCCTG
TGFβ1	-448 C	AGAGGGTCTGTCAAC
	-448 T	AGAGGGTTTGTCAAC
	-508 A	CACCTGAAGGATGGA
	-508 G	CACCTGAGGGATGGA
	-799 G	GTGGTGACGTTGGAG
	-799 A	GTGGTGATGTTGGAG
	-1146 C	AAAGGAGACAGAGGGGA
	-1146 G	AAAGGAGAGAGAGGGGA
	-1349 INSC	CCTTAGCAGGGGAGTA
	-1349 -	CCTTAGCAGGGGAGTA
	-1550 INSAGG	CATGAGGAGGGAAGGCA
	-1550 -	CATGAGGGAAGGCA
	-1571 A	CCCAAAGAGAGCAGG
	-1571 G	CCCAAAGGAGCAGG
	-1886 G	GGTAGATCAGATGGT
	-1886 A	GGTAGATTAGATGGT
-1960 C	CAAGAAGGGGATCAA	
-1960 T	CAAGAAGAGGATCAA	

## Annexe II

Tableau XXI : Liste des amorces utilisées pour l'amplification PCR des fragment 2Kb.

Gène	Nom de l'amorce	Séquence (5'-3')	Longueur (pb)
CCND1	ccnd1.B1.F	AAAAAGCAGGCTCTTGCAATTTTGGCGCTCGT	2103
	ccnd1.B2.R	AGAAAGCTGGGTGCAACTTCAACAAAACCTCCC	
CDC25A	cdc25a.B2.F	AGAAAGCTGGGTCCGACCTACACCTCTTACCC	2230
	cdc25a.B1.R	AAAAAGCAGGCTAGGCCAAACTCTCAGCAAA	
CDKN1A	cdkn1a.attB1.F	AAAAAGCAGGCTGGATCTGATGCATGTGTGCT	2092
	cdkn1a.attB2.R	AGAAAGCTGGGTGGCTCCACAAGGAACTGACT	
CDKN1B	cdkn1b.attB2.F	AGAAAGCTGGGTAAAACACCCCGAAAAGACG	1992
	cdkn1b.attB1.R	AAAAAGCAGGCTCACTGCCCCCTTACTGTTC	
CDKN2A	cdkn2a.attB2.F	AGAAAGCTGGGTCTGTCCCTCAAATCCTCTG	2139
	cdkn2a.attB1.R	AAAAAGCAGGCTTCCCAAATTGTTTTTCATTGG	
E2F1	e2f1.B2.F	AGAAAGCTGGGTAAAGTCCCGGCCACTTTTA	2100
	e2f1.B1.R	AAAAAGCAGGCTGAATGCAGTGTAGGCCCAAT	
HDAC	hdac.attB1.F	AAAAAGCAGGCTGACAAGGCCTGGCTCTATTG	2037
	hdac.attB2.R	AGAAAGCTGGGTCCGTCCCTACCGTCAGTC	
MADH3	madh3.B1.F	AAAAAGCAGGCTCTTGTCTCTCTTGACGGACG	2186
	madh3.B2.R	AGAAAGCTGGGTCTTGACCAGGCTCTTGACC	
RB1	rb1.attB1.F	AAAAAGCAGGCTTGCAAACAAGTAAAACCTCA	2028
	rb1.attB2.R	AGAAAGCTGGGTCAACGTCCCTGAGAAAAAC	
SKP2	skp2.B2.F	AGAAAGCTGGGTCTCTACAGCCCGCTCTGC	2015
	skp2.B1.R	AAAAAGCAGGCTAAGGCAATAGTCATGATTAGAA GG	
TFDP1	tfdp1.B1.F	AAAAAGCAGGCTGGGTGTCACTGTTGCCATC	2003
	tfdp1.B2.R	AGAAAGCTGGGTCTCCTCTTCTGCTGCTCCTC	
TGFβ1	tgfb1.+31.R	CTCGCTGTCTGGCTGCTC	2093
	tgfb1.-2062.F	TGGCTAGGAGCACTCTTCC	

## Annexe III

Tableau XXII : Liste des sondes utilisées lors des expériences de retard sur gel, EMSA.

Gène	rSNP	Séquence sonde (5'-3')
CDC25A	-51 G	TGCTGCCCCGCCCCCTTTCTAGTTGGCGCCAAACGGAATC
	-51 C	TGCTGCCCCGCCCCCTTTCTACTTGGCGCCAAACGGAATC
	-1440 G	AACAGGCCGGGAGCGGTGGCTCACGCCTGTAATCCCAGC
	-1440 A	AACAGGCCGGGAGCAGTGGCTCACGCCTGTAATCCCAGC
	-2030 T	AAGGGGAGCTTGGGGCTGGTGGGAAGTGCTTTCTAAG
	-2030 G	AAGGGGAGCTTGGGGCTGGGGGGAAGTGCTTTCTAAG
CDKN1A	-134 G	TCGGCCAGGCTCAGCTGGCTCGGCGCTG
	-134 C	TCGGCCAGGCTCAGCTGGCTCCGCGCTG
	-791 T	ACAGAAGAAATCCCTGTGGTT
	-791 C	ACAGAAGAAACCCCTGTGGTT
	-899 T	ACTGCTGGGGAAACTGGGGCTCAGA
	-899 G	ACTGCTGGGGAAACGGGGGCTCAGA
	-1021 C	TATTTCTGTGGCTCCAAAATGACAAAAAT
	-1021 T	TATTTCTGTGGTTCCAAAATGACAAAAAT
	-1284 T	TTCTGTTTTTTAGTGGGATTT
	-1284 C	TTCTGTTTTTCAGTGGGATTT
	-1466 A	GCAGGCCAAGGGGGTCTGCTACTGTGTCCTCCCACCCCTACCTGGG
	-1466 G	GCAGGCCAAGGGGGTCTGCTGCTGTGTCCTCCCACCCCTACCTGGG
	-1529 T	GCCTTTCTGGGGTTTAGCCACAATCCT
	-1529 C	GCCCTTCTGGGGTTTAGCCACAATCCT
	-1536 C	CCTGAGGCTGTGCCGTGGCCTTTCTGGGGTTTAGCC
-1536 T	CCTGAGGCTGTGCTGTGGCCTTTCTGGGGTTTAGCC	
CDKN1B	-337 C	CTGCGTTGGCGGGTTTCGCCGCGGCGCCTAAGCC
	-337 A	CTGCGTTGGCGGGTTTCGACGCGGCGCCTAAGCC

	-373 G	TAAGCCCCGACCTCCCTCCCGCTCCTCGCCCGGAAGCCGGGAC
	-373 T	TAAGCCCCGACCTCCCTCCCGCTCCTCTCCCGGAAGCCGGGAC
	-1857 C	TACCACAGGCTCAAGACAGCTGCATTTAA
	-1857 T	TACCACAGGCTCAAGATAGCTGCATTTAA
CDKN2A	-222 T	ACAACCTTCCAACTGCCAAATTGAATCGGGGTGT
	-222 A	ACAACCTTCCAACTGCCAAATTGAATCGGGGTGT
MADH3	-1946 C	AAACCTAAGACCACAGCGGCAGCGGTAAACAGGA
	-1946 A	AAACCTAAGACCACAGCGGAAGCGGTAAACAGGA
TFDP1	-1394 T	CTCCTCCGTGTGCCACATAC
	-1394 C	CTCCTCCGTGCGTCCACATAC
	-1655 G	TGCTCGGGCCTGGCCCTAGCGGCGGGGCCCTCC
	-1655 A	TGCTCGGGCCTGGCCCTAACGGCGGGGCCCTCC
	-1779 A	CAGCTCCTCAAGTCTCTGTA
	-1779 G	CAGCTCCTCGAGTCTCTGTA
	-1883 G	CCAGAGCTTTAGGGGGTGCGGCCTGCCTGAGGGGCAGGA
	-1883 C	CCAGAGCTTTAGGGGGTGCGGCCTGCCTGAGGGGCAGGA
	-1887 C	CCAGAGCTTTAGGGGGTGCGGCCTGCCTGA
	-1887 T	CCAGAGCTTTAGGGGGTGTGGCCTGCCTGA
TGFβ1	-508 A	ACAGGACACCTGAAGGATGGAAGGGTCAG
	-508 G	ACAGGACACCTGAGGGATGGAAGGGTCAG
	-1146 C	CACCCATCCCTCCTTTCCCCTCTCTCCTTTCTGCAGGCTGGC
	-1146 G	CACCCATCCCTCCTTTCCCCTCTGTCTCCTTTCTGCAGGCTGGC
	-1550	GGGACATGAGGAGGGAAGGCAG
	INSAGG	
	-1550 -	GGGACATGAGGGAAGGCAG
	-1571 A	GAAAGAAGGCCCTGGGCCCAAAGAGAGCAGGGCAG
	-1571 G	GAAAGAAGGCCCTGGGCCCAAAGGGAGCAGGGCAG
	-1886 G	AATTACCACCATCTGATCTACCCTATA
	-1886 A	AATTACCACCATCTAATCTACCCTATA

## Curriculum Vitae

---

Joëlle Dionne

---

**Molecular biology researcher with excellent technical and organizational skills.  
Seeking exciting job with continuous learning opportunities.**

### Education

---

2006-2008	M.Sc. Molecular biology, University of Montreal
2002-2005	B.Sc. Biological sciences, option Microbiology, University of Montreal
1999-2002	DEC, Chemical procedures technologies, Collège Maisonneuve
1998-1999	(50 credits) Transportation logistics technologies, Cégep Lionel-Groulx
Summer 1998	English immersion, University of Dalhousie, Halifax (governmental program)

### Research experience

---

2006-2008	<p><b>M.Sc. Molecular biology</b>, University of Montreal  Faculty of Graduate Studies  Sainte-Justine Hospital, University Research Center  Department of Haematology-Oncology  Laboratory of Genome Quebec, Cancer genetics  Daniel Sinnett, Ph.D., director</p> <ul style="list-style-type: none"> <li>• Research project: Functional impact of regulatory polymorphisms (rSNPs) in G1/S cell cycle checkpoint genes.</li> <li>• Other project: Transforming Flp-In vector system (Invitrogen) with the Gateway (Invitrogen) cloning system.</li> </ul>
-----------	--

- Weekly laboratory meeting and journal club presentations.
- Analysed, interpreted and discussed results

Summer 2004-2005 **Laboratory research training, University of Montreal**

Sainte-Justine Hospital, University Research center

Department of Neurology

Laboratory Neuro-genetics

Louise R. Simard, Ph.D., supervisor

Project: Study the transcriptional regulation of the SMN gene in mouse P19  
EC cells *in vitro*.

### Bursaries, Awards and Distinctions

---

#### Bursaries:

- Master Research Fellowship from the Cole Foundation, March 2007 to September 2008
- Master Research Bursary from the Foundation for Research into Children's Diseases and the Sainte-Justine Hospital Foundation, September 2006 to August 2007

#### Awards:

- Student Travel Award from Genome Canada, *HUGO's 12<sup>th</sup> Human Genome Meeting*, Montréal (2007)
- Prize for best poster from Boehringer Ingelheim, Research training in Dr. Louise R. Simard laboratory, *COOPSE Student Conference*, Montreal (2006)
- Excellence award from Department of microbiology and immunology of the University of Montreal (2003)

#### Distinctions:

- Graduated First of Class from the Collège de Maisonneuve, Montreal (2002)

### Scientific Contributions

---

#### Article in preparation:

- **Dionne, J\***, Healy, J\*, Bélanger, H, Beaulieu, P, Sinnett, D (2007) Impact of promoter SNPs on *TGFβ1* expression and association with childhood leukemia. *Cancer Research*.

#### Oral presentations:

- **Dionne, J**, Healy, J, Sinnett, D (2007) Déterminants génétiques de la leucémie de l'enfant. *The Sainte-Justine Hospital University Research Centre's Viral, Immune Disorders and Cancer Research Axis' Annual Retreat*. Val- David, QC, Canada.
- **Dionne, J**, Ouimet, M, Gagné, V, Larivière, M, Labuda, D, Sinnett, D (2007) Functional impact of regulatory polymorphisms (rSNP) in G1/S cell cycle checkpoint genes. *HUGO's 12<sup>th</sup> Human Genome Meeting*. May, Montreal, QC, Canada.

## Poster presentations:

- **Dionne, J**, Ouimet, M, Gagné, V, Larivière, M, Labuda, D, Sinnett, D (2007) Functional impact of regulatory polymorphisms (rSNP) in G1/S cell cycle checkpoint genes. *The Sainte-Justine Hospital University Research Centre's XXIst Annual Student Conference*. June, Montreal, QC, Canada.
- **Dionne, J**, Ouimet, M, Gagné, V, Larivière, M, Labuda, D, Sinnett, D (2007) Functional impact of regulatory polymorphisms (rSNP) in G1/S cell cycle checkpoint genes. *HUGO's 12<sup>th</sup> Human Genome Meeting*. May, Montreal, QC, Canada.
- **Dionne, J**, Ouimet, M, Gagné, V, Larivière, M, Labuda, D, Sinnett, D (2007) Functional impact of regulatory polymorphisms (rSNP) in G1/S cell cycle checkpoint genes. *The XIII<sup>th</sup> Annual University of Montréal, Molecular Biology Student Conference*. May, Montreal, QC, Canada.
- **Dionne, J**, Ouimet, M, Gagné, V, Larivière, M, Labuda, D, Sinnett, D (2007) Functional impact of regulatory polymorphisms (rSNP) in G1/S cell cycle checkpoint genes. *The American Association for Cancer Research Annual Meeting*. April, Los Angeles, CA, USA.
- **Dionne, J**, Ouimet, M, Gagné, V, Larivière, M, Labuda, D, Sinnett, D (2006) Functional impact of regulatory polymorphisms (rSNP) in G1/S cell cycle checkpoint genes. *Genome Québec, Gene Regulators in Disease Retreat*. November, Montreal, QC, Canada.
- **Dionne, J**, Ouimet, M, Gagné, V, Larivière, M, Labuda, D, Sinnett, D (2006) Functional impact of regulatory polymorphisms (rSNP) in G1/S cell cycle checkpoint genes. *Genome Canada's International Conference*. October, Quebec, QC, Canada.
- **Dionne J.**, Boli F., Simard L.R., (2006) Étude de l'activité transcriptionnelle du gène *SMN* *in vitro* dans des cellules neuronales murines. *COOPSE Student Conference*, January, Montreal, Qc, Canada.

### Technical areas of experience

---

- **Cellular biology :**
  - Cell culture
  - Transfection
  - Luminescence analysis
- **Molecular biology :**
  - Southern Blot
  - Perfecting long range PCR
  - A.S.O. Genotyping with radioactive isotopes
  - DNA sequencing
  - Analysing EMSA
  - Cloning
- **Work in a Level 2 biocontainment laboratory**
- **Software often used :** Word, Power Point, Excel, Internet Explorer
- **Bioinformatics :** Experienced at accessing information from databases (NCBI)

### Other experience

---

Summer 2002-2003 **Laboratory Technician**, quality control  
Groupe Avmor, Laval, Qc.

1998-2002

**Part-time job at Canadian Tire,**

Cash and customer services, St-Eustache, Qc.

**Skills**

---

- Excellent planning and organizational skills
- Ability to work well independently and within a team
- Conscientious and fast learner
- Persevering and flexible

Languages: French and English

References available upon request