

2m11.3311.1

Université de Montréal

Variabilité génétique dans les régions régulatrices de gènes impliqués dans le
contrôle du cycle cellulaire

par
Hélène Bélanger

Programme de biologie moléculaire
Faculté des études supérieures

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

Avril 2005

© Hélène Bélanger, 2005



QH
506

U54

2005

v.031

AVIS

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

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
Faculté des études supérieures

Ce mémoire intitulé:
Variabilité génétique dans les régions régulatrices de gènes impliqués dans le
contrôle du cycle cellulaire

présenté par :
Hélène Bélanger

a été évalué par un jury composé des personnes suivantes :

Dr Richard Bertrand

.....
président-rapporteur

Dr Daniel Sinnett

.....
directeur de recherche

Dr François Robert

.....
membre du jury

Sommaire

Le contrôle du cycle cellulaire est essentiel au maintien de l'homéostasie cellulaire et de l'intégrité du matériel génétique. Nous proposons que des polymorphismes fonctionnels dans des gènes candidats mèneraient à une variabilité dans les mécanismes contrôlant le cycle cellulaire, d'où un risque variable de développer certaines maladies, tel que le cancer. L'objectif principal de mon projet de maîtrise est d'identifier et de valider des polymorphismes fonctionnels dans le promoteur (pSNPs), arbitrairement définis comme un segment de 2 kb en amont de l'exon 1. À l'aide de HPLC dénaturant, nous avons identifié des variants d'ADN dans la région régulatrice de 17 gènes candidats régulant la transition G1/S du cycle cellulaire, et ce dans un panel de 40 individus provenant de différents groupes ethniques. Nous avons ainsi détecté 136 pSNPs. Notre attention s'est portée particulièrement sur 13 polymorphismes fonctionnels qui, selon des prédictions *in silico*, suggèrent l'apparition de nouveaux sites de liaison pour des facteurs de transcription. Les expériences de retards sur gel ont permis de démontrer que 12 de ces pSNPs mènent à une modification de la liaison à des facteurs de transcription. L'impact fonctionnel a été évalué à l'aide d'expérience de gène rapporteur en comparant l'effet de la présence des variants sur l'efficacité de transcrire le gène de la luciférase. Des différences significatives ont été détectées pour des polymorphismes situés dans le promoteur des gènes TGF β 1, E2F1, HDAC1, RB1 et CCND1. L'identification et la caractérisation des polymorphismes dans les gènes impliqués dans le contrôle du cycle cellulaire permettront d'évaluer l'importance de ceux-ci dans la susceptibilité au cancer. Le projet est supporté par Génome Québec/Canada.

Mots clés : polymorphisme, transition G1/S, facteur de transcription, retard sur gel, gène rapporteur, maladie complexe.

Summary

Cell cycle checkpoints are essential in order to maintain homeostasis and preserve genome integrity. We propose that functional polymorphisms in candidate genes could bring variability in the mechanisms that regulate the cell cycle which could give rise to differences in the risk of developing complex diseases such as cancer. The main objectives are to identify and validate functional polymorphisms in the promoter (pSNPs), arbitrary defined as a 2kb genomic segment upstream of exon 1. Mutation analysis was carried out on 17 genes involved in the G1/S checkpoint. Polymorphisms were detected using denaturing HPLC in a population panel of 40 individuals from different ethnic backgrounds. A total of 136 pSNPs were identified. In this study we selected 13 polymorphisms that were predicted using *in silico* tool, to lead to gains of binding sites for transcription factors. EMSA showed that nuclear proteins from different cell lines exhibited differences in the binding affinity for 12 of the pSNPs tested. Expression analyses were performed by cloning 2kb haplotypes corresponding to different construction of the selected polymorphism into a luciferase reporter gene vector (pGL3) and were tested for differences in their ability to drive transcription in different cell lines. Differences in expression greater than 1.5-fold were detected between haplotypes of TGFB1, E2F1, RB1, HDAC1 and CCND1. The detection and validation of polymorphisms in the regulatory regions of genes implicated in cell cycle checkpoints will provide a better understanding of their involvement in susceptibility to complex diseases. This work is supported by Genome Quebec/Canada.

Key words: polymorphism, G1/S checkpoint, transcription factor, EMSA, reporter gene assay, complex disease.

Table des matières

| | |
|---|-----------|
| Sommaire | iii |
| Summary | iv |
| Table des matières | v |
| Liste des tableaux | viii |
| Liste des figures | x |
| Liste des abréviations | xii |
| Remerciements | xv |
| | |
| Chapitre 1-Introduction | 1 |
| | |
| 1.1 Le cycle cellulaire | 2 |
| 1.1.1 Les phases du cycle cellulaire | 2 |
| 1.1.2 La régulation du cycle cellulaire | 4 |
| 1.1.2.1 Régulation par les kinases cycline-dépendantes (CDK) | 4 |
| 1.1.2.2 Régulation par les points de contrôle du cycle cellulaire | 6 |
| 1.2 Variabilité interindividuelle, cycle cellulaire et maladies complexes | 15 |
| 1.3 Nature d'un polymorphisme | 16 |
| 1.3.1 Type | 16 |
| 1.3.2 Impact d'un polymorphisme | 17 |
| 1.3.3 Détection d'un polymorphisme | 18 |
| 1.3.4 Annotation d'un polymorphisme | 21 |
| 1.3.4.1 Approche <i>in silico</i> | 23 |
| 1.3.4.2 Approche <i>in vitro</i> | 23 |
| 1.4 Régulation de l'expression génique | 27 |
| 1.4.1 Architecture générale d'un promoteur | 29 |
| 1.4.2 La génétique régulatrice | 31 |
| 1.5 Hypothèse | 32 |
| 1.6 Objectifs du projet | 33 |
| | |
| Chapitre 2- Article 1: Detection and characterization of DNA variants in the promoter regions of hundreds of human diseases candidate genes. | 34 |
| | |
| Abstract | 36 |

| | |
|---|-----------|
| Introduction | 37 |
| Material and methods | 38 |
| Results | 39 |
| Discussion | 42 |
| Acknowledgements | 44 |
| References | 45 |
| | |
| Chapitre 3- Article 2: Functional promoter SNPs in the cell cycle checkpoint genes | 58 |
| | |
| Abstract | 60 |
| Introduction | 61 |
| Material and methods | 62 |
| Results | 63 |
| Discussion | 66 |
| Acknowledgements | 69 |
| References | 70 |
| | |
| Chapitre 4- Résultats | 83 |
| | |
| 4.1 Description des polymorphismes identifiés pour TGF β 1. | 84 |
| 4.2 Étude <i>in vitro</i> de l'impact des pSNPs du gène codant pour TGF β 1. | 84 |
| 4.2.1 Retard sur gel | 84 |
| 4.2.2 Gène rapporteur | 89 |
| | |
| Chapitre 5- Discussion | 92 |
| | |
| 5.1 État de la question | 93 |
| 5.2 Identification de polymorphismes | 93 |
| 5.3 Prédications <i>in silico</i> | 95 |
| 5.4 Génomique comparative | 96 |
| 5.5 Retard sur gel | 96 |
| 5.6 Gène Rapporteur | 97 |
| 5.7 TGF β 1 | 98 |

| | |
|--------------------------------------|--------|
| 5.8 E2F1 | 100 |
| 5.9 RB1 | 101 |
| 5.10 HDAC1 | 103 |
| 5.11 CCND1 | 104 |
| 5.12 CDKN1A | 105 |
| 5.13 Limitations des études in vitro | 106 |
| 5.14 Conclusion | 107 |
| | |
| Références | 108 |
| | |
| Annexe I | xvi |
| Annexe II | xx |
| Annexe III | xxi |
| Annexe IV | xxii |
| Annexe V | xxiv |
| Annexe VI | xxviii |
| | |
| Curriculum vitae | lxxxii |

Liste des tableaux

Article 1

| | |
|---|----|
| Table 1: Distribution of pSNPs according to the type of nucleotide substitution. | 50 |
| Table 2: Distribution of pSNPs among repetitive elements. | 51 |
| Table 3: List of Gene Ontology groups with significant differences in the promoter genetic diversity. | 52 |

Article 2

| | |
|---|----|
| Table 1 : List of the promoter SNPs associated with predicted gains of transcription factor binding sites and the corresponding oligonucleotides used in EMSAs. | 77 |
| Table 2: Summary of the electrophoretic mobility shift assay results for the chosen pSNP. | 78 |
| Table 3: Promoter haplotypes studied in gene reporter assays | 79 |
| Table 4: Results of the gene reporter assays for the promoter haplotype tested | 80 |

Résultats

| | |
|--|----|
| Tableau IV-I. Description détaillée des polymorphismes identifiés dans le promoteur du gène codant pour TGF β 1. | 87 |
|--|----|

Annexes

| | |
|---|---------|
| Tableau I : Liste des amorces utilisées pour l'amplification d'ADN génomique pour l'analyse au dHPLC ainsi que les conditions PCR utilisées. | xvi |
| Tableau II : Liste des amorces utilisées pour l'amplification d'ADN génomique pour les constructions de gène rapporteur ainsi que les conditions PCR utilisées. | xx |
| Tableau III: Conditions de PCR | xxi |
| Tableau IV: Protocole pour l'amplification PCR | xxii |
| Tableau V: Température d'analyse au dHPLC | xxiv |
| Tableau VI: Description détaillée des polymorphismes identifiés dans le promoteur du gène codant pour CCND1. | xxix |
| Tableau VII: Description détaillée des polymorphismes identifiés dans le promoteur du gène codant pour CCNE1. | xxxiii |
| Tableau VIII: Description détaillée des polymorphismes identifiés dans le promoteur du gène codant pour CDC25A. | xxxv |
| Tableau IX: Description détaillée des polymorphismes identifiés dans le promoteur du gène codant pour CDK4. | xxxviii |
| Tableau X: Description détaillée des polymorphismes identifiés dans le promoteur du gène codant pour CDK6. | xlii |

| | |
|---|--------|
| Tableau XI: Description détaillée des polymorphismes identifiés dans le promoteur du gène codant pour CDKN1A. | xliv |
| Tableau XII: Description détaillée des polymorphismes identifiés dans le promoteur du gène codant pour CDKN1B. | xlviii |
| Tableau XII: Description détaillée des polymorphismes identifiés dans le promoteur du gène codant pour CDKN2A. | li |
| Tableau XIV: Description détaillée des polymorphismes identifiés dans le promoteur du gène codant pour CDKN2B. | liii |
| Tableau XV: Description détaillée des polymorphismes identifiés dans le promoteur du gène codant pour E2F1. | lvi |
| Tableau XVI: Description détaillée des polymorphismes identifiés dans le promoteur du gène codant pour HDAC1. | lx |
| Tableau XVII: Description détaillée des polymorphismes identifiés dans le promoteur du gène codant pour MADH3. | lxiv |
| Tableau XVIII: Description détaillée des polymorphismes identifiés dans le promoteur du gène codant pour MADH4. | lxvii |
| Tableau XIX: Description détaillée des polymorphismes identifiés dans le promoteur du gène codant pour RB1. | lxix |
| Tableau XX: Description détaillée des polymorphismes identifiés dans le promoteur du gène codant pour SKP2. | lxxii |
| Tableau XXI: Description détaillée des polymorphismes identifiés dans le promoteur du gène codant pour TFDP1. | lxxiv |
| Table XXII: 197 genes screened by dHPLC for the presence of pSNPs and their predicted impact on TFBSs. | lxxvi |

Liste des figures

Introduction

| | |
|---|----|
| Figure 1-1: Les phases du cycle cellulaire. | 3 |
| Figure 1-2: Mécanismes qui régulent les CDKs. | 5 |
| Figure 1-3: Le point de contrôle G1/S. | 8 |
| Figure 1-4: Mécanismes d'arrêt cellulaire par la voie de TGF β et la dérégulation de cette voie dans le cancer. | 9 |
| Figure 1-5: Mutations dans les régulateurs de la transition G1/S associées aux cancers humains. | 11 |
| Figure 1-6: Le point de contrôle G2/M. | 14 |
| Figure 1-7 : Type de substitutions: transitions et transversions. | 19 |
| Figure 1-8 : Localisation des SNPs. | 20 |
| Figure 1-9 : Résolution des hétéroduplexes. | 22 |
| Figure 1-10 : Profile de dénaturation à 3 températures d'un fragment de 185 pb tel qu'évaluer par WAVEMAKER. | 22 |
| Figure 1-11 : Principe de EMSA. | 25 |
| Figure 1-12 : Principe de l'empreinte à la DNase I. | 26 |
| Figure 1-13 : A. Modèle typique d'un gène avec les composantes jouant un rôle dans l'activation et l'inactivation de ce gène. B. Activation d'un gène et assemblage du complexe de pré-initiation de la Pol II. | 28 |
| Figure 1-14 : Éléments de séquence composant le cœur de promoteur | 30 |

Article 1

| | |
|---|----|
| Figure 1: Distribution of pSNPs among different dinucleotide subtypes. | 53 |
| Figure 2: Correlation between the prevalence of pSNPs and the physical distance. | 54 |
| Figure 3: Distribution of normalized mutation rates in the selected genes. | 55 |
| Figure 4: Correlation between the prevalence of pSNPs and the Gene Ontology groups. | 56 |

Article 2

| | |
|--|----|
| Figure 1: Representative EMSA illustrating allelic DNA-protein interactions in the promoter of CCND1. | 81 |
| Figure 2: Assessment of the functional impact of the HDAC1 -905A<C variant on the promoter activity using a gene reporter assay. | 82 |

Résultats

| | |
|---|----|
| Figure 4-1: Légende pour les figures des représentations graphiques des promoteurs des gènes étudiés. | 85 |
|---|----|

| | |
|---|----|
| Figure 4-2: Représentation graphique du promoteur du gène TGFβ1. | 86 |
| Figure 4-3: Résultats de retard sur gel pour le polymorphisme à la position – 1550 du gène TGFβ1. | 88 |
| Figure 4-4: Haplotypes majeurs de TGFβ1. | 90 |
| Figure 4-5: Résultats de gène rapporteur pour les haplotypes A et B du gène TGFβ1. | 91 |

Discussion

| | |
|---|----|
| Figure 5-1: Effet d'un SNP sur le risque de développer une maladie. | 93 |
|---|----|

Annexes

| | |
|---|--------|
| Figure 1: Représentation graphique du promoteur du gène CCND1. | xxviii |
| Figure 2: Représentation graphique du promoteur du gène CCNE1. | xxxii |
| Figure 3: Représentation graphique du promoteur du gène CDC25A. | xxxiv |
| Figure 4: Représentation graphique du promoteur du gène CDK4. | xxxvii |
| Figure 5: Représentation graphique du promoteur du gène CDK6. | xli |
| Figure 6: Représentation graphique du promoteur du gène CDKN1A. | xliv |
| Figure 7: Représentation graphique du promoteur du gène CDKN1B. | xlvii |
| Figure 8: Représentation graphique du promoteur du gène CDKN2A. | l |
| Figure 9: Représentation graphique du promoteur du gène CDKN2B. | lii |
| Figure 10: Représentation graphique du promoteur du gène E2F1. | lv |
| Figure 11: Représentation graphique du promoteur du gène HDAC1. | lix |
| Figure 12: Représentation graphique du promoteur du gène MADH3. | lxiii |
| Figure 13: Représentation graphique du promoteur du gène MADH4. | lxvi |
| Figure 14: Représentation graphique du promoteur du gène RB1. | lxviii |
| Figure 15: Représentation graphique du promoteur du gène SKP2. | lxxi |
| Figure 16: Représentation graphique du promoteur du gène TFDP1. | lxxiii |

Liste des abréviations

| | |
|----------|---|
| A : | adénine |
| ADN : | acide désoxyribonucléique |
| ARN : | acide ribonucléique |
| ARNm : | ARN messenger |
| ATM : | ataxie-telangiectasie mutated |
| ATP : | adénosine triphosphate |
| ATR : | ataxia-telangiectasia and rad3-related |
| BRCA : | breast cancer |
| BRE : | TFIIB recognition element |
| C : | cytosine |
| c-Abl : | abelson murine leukemia viral oncogene homolog 1 |
| CAK : | cdk-activating kinase |
| 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 kinase-dépendante |
| CHK : | choline kinase |
| CKI : | inhibiteur de cycline kinase-dépendante |
| CLC-K1 : | chloride channel-K1 |
| cSNP : | coding single nucleotide polymorphism |
| CYP2A6 : | cytochrome P450 2A6 |
| D1 : | cycline D1 |
| DBSNP : | single nucleotide polymorphism database |
| dHPLC : | denaturing high performance liquid chromatography |
| DNA-PK : | DNA-dependent protein kinase |

| | |
|-------------|---|
| DPE : | downstream promoter element |
| DTT : | Dithiothreitol |
| E1 : | cycline E1 |
| E2F : | E2F transcription factor |
| ECE : | endothelin-converting enzyme |
| EDTA : | ethylenediamine-tetraacetic acid |
| EMSA : | electrophoretic mobility shift assay |
| EVI-1 : | ecotropic viral integration site 1 encoded factor |
| F-SSCP : | fluorescent single-stranded conformation polymorphism |
| G : | guanine |
| G0 : | stade quiescent de non division |
| G1 : | gap 1 |
| G2 : | gap 2 |
| GADD45 : | growth arrest- and dna damage-inducible |
| gSNP : | genomic single nucleotide polymorphism |
| GSK-3beta : | glycogen synthase kinase 3-beta |
| HDAC : | histone déacétylase |
| HFH-2 : | hepatocyte nuclear factor 3/forkhead homolog 2 |
| HGBASE : | human genic bi-allelic sequences |
| HGMD : | human gene mutation database |
| HPLC : | high performance liquid chromatography |
| INK4 : | inhibitors of CDK4 |
| iSNP : | intronic single nucleotide polymorphism |
| kb : | kilobase |
| kDa : | kilodalton |
| M : | mitose |
| MADH : | mothers against decapentaplegic |
| MAPs : | microtubule associated proteins |
| MAR : | matrix attachment region |
| MAZ : | myc-associated zinc finger protein |
| MDM2 : | mouse double minute 2 homolog |

| | |
|---------------------|---|
| MgCl ₂ : | chlorure de magnésium |
| MYT1 : | myelin transcription factor 1 |
| pb : | paire de base |
| PCR : | polymerase chain reaction |
| PLK : | polo-like kinase |
| Pol : | polymérase |
| pSNP : | promoter single nucleotide polymorphism |
| Py : | pyrimidine |
| RB : | retinoblastoma |
| RBP : | mammalian transcriptional repressor RBP-Jkappa/CBF1 |
| S : | synthèse de l'ADN |
| SKP2 : | S-phase kinase-associated protein 2 |
| SMAD: | voir MADH |
| SNP : | single nucleotide polymorphism |
| SRY: | sex-determining region Y |
| SSCP : | single-stranded conformation polymorphism |
| T : | thymine |
| TβR : | transforming growth factor beta receptor |
| TAF : | TBP associating factor |
| TBP : | TATA-binding protein |
| TFBSs: | Transcription factor binding sites |
| TFDP1 : | transcription factor DP1 |
| TGFβ : | transforming growth factor beta |
| Thr : | thréonine |
| TMHA : | temperature-modulated heteroduplex analysis |
| UTR : | untranslated region |
| VNTR : | variable number of tandem repeats |
| WEE1 : | wee 1 tyrosine kinase |
| YY1 : | Yin Yang-1 |

Remerciements

Je tiens à remercier en tout premier lieu mon directeur de recherche, Daniel Sinnett pour sa confiance et sa générosité ainsi que pour m'avoir encadré et fait bénéficier de ses conseils dans la réalisation de ce projet de maîtrise.

Je souhaite également remercier toutes les personnes à l'étage B du pavillon vidéo-tron, pour leur accueil chaleureux, pour les nombreuses discussions et surtout pour l'ambiance agréable au laboratoire et pendant les partys! Je tiens à saluer ici les personnes qui, de près ou de loin, ont contribué à la concrétisation de ce travail : Patrick, Caroline, Sylvie, Claudia, Gino, Josiane, Marie-Christine, Nina, Rocio, Jasmine, Stéphane, Dominika, Claude.

Finalement, ma gratitude s'adresse à mes parents pour leur encouragement et leur soutien au cours de mes études et à Danny pour m'avoir remotivé quand le besoin s'en faisait sentir et pour sa patience.

Chapitre 1

-Introduction-

1.1 Le cycle cellulaire

1.1.1 Les phases du cycle cellulaire

La division cellulaire est un processus fondamental par lequel une cellule-mère donne deux cellules-filles. Le cycle cellulaire des eucaryotes est divisé en 4 phases : G1, S, G2 et M (Figure 1-1). La phase G1 est une intervalle dans laquelle la cellule effectue sa croissance, intègre les signaux mitogènes ou anti-mitogènes et se prépare pour entrer correctement en phase S (synthèse). Selon les stimuli extracellulaires, la cellule décidera de passer à la phase suivante pour répliquer l'ADN ou de quitter le cycle cellulaire et entrer en phase G0, stade quiescent de non division. Lorsque la cellule passe le point de non-retour, c'est-à-dire le point de restriction R, le cycle cellulaire est irréversiblement engagé et l'entrée en division ne dépend plus de la présence de signaux mitogènes. Ensuite, il y a réplication de l'ADN au cours de la phase S. La phase suivante, l'intervalle G2, permet la préparation pour l'entrée en phase M (mitose), où les chromosomes dédoublés seront répartis dans les deux cellules-filles grâce au fuseau de division. La phase M est divisée en cinq étapes : la prophase, la prométaphase, la métaphase, l'anaphase, et la télophase. Finalement, la cytokinèse achève la division de la cellule.

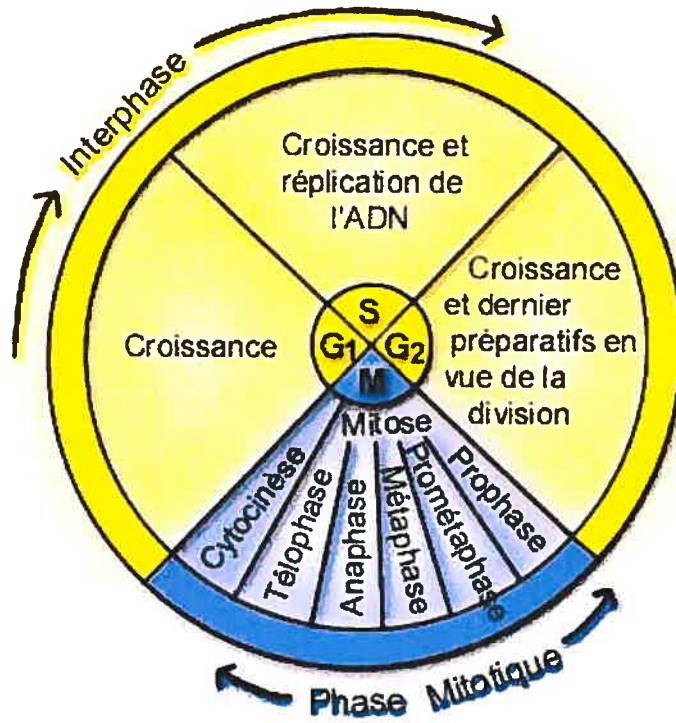


Figure 1-1: Les phases du cycle cellulaire. (voir texte pour détails)

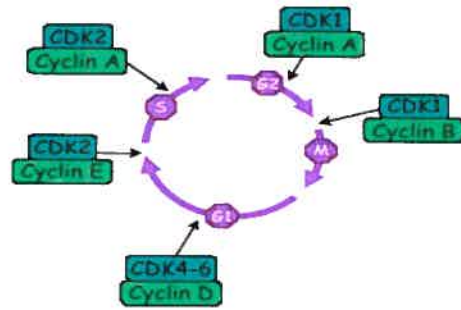
(<http://www.ustboniface.mb.ca/cusb/abernier/Biologie/Cellule/mitose-2.html>)

1.1.2. La régulation du cycle cellulaire

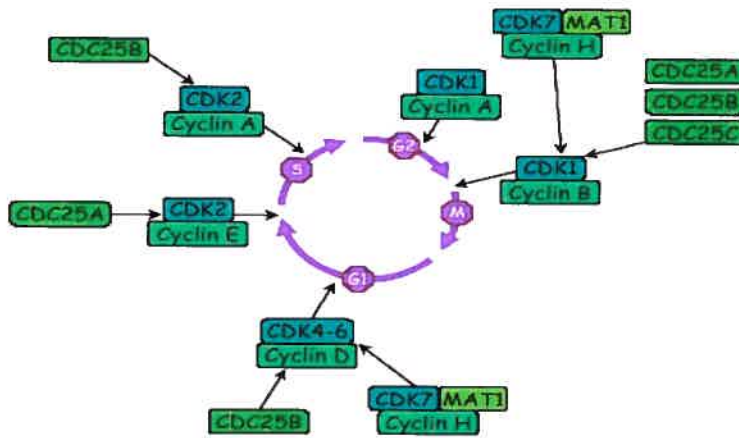
1.1.2.1 Régulation par les kinases cycline-dépendantes (CDK)

Tous les jours un milliard de cellules doivent être renouvelées pour remplacer les cellules perdues de façon continue, plus particulièrement au niveau de la peau, du tube digestif et du système hématopoïétique. D'où l'importance d'avoir un mécanisme de division cellulaire complexe et hautement régulé à différents niveaux (Figure 1-2). Premièrement, au niveau de la régulation du cycle, ce sont essentiellement des kinases cycline-dépendantes (CDKs) qui interviennent. Les CDKs forment une famille de protéines kinases de petites tailles (35 à 40 kDa) qui régulent l'activité de nombreuses protéines par phosphorylation de résidus sérine/thréonine. Le taux d'expression des CDKs est à peu près constant durant le cycle cellulaire, cependant les CDKs sont inactives. Pour être activées, elles doivent se lier avec leurs cyclines correspondantes. Il s'agit d'une étape clé puisque les cyclines sont hautement régulées par la transcription et la dégradation par la voie de protéolyse ubiquitine-dépendante à des moments précis du cycle cellulaire[1]. Ensuite, l'association d'une CDK avec sa cycline permettra la phosphorylation d'un résidu thréonine (thréonine 160/161), commun à toute les CDKs, par la CAK (cdk-activating kinase)[2]. Dans certain cas, l'activation des CDKs nécessite la déphosphorylation des résidus thréonine 14 et tyrosine 15 par une phosphatase de la famille cdc25[1]. Les CDKs activés, ainsi que le processus d'activation, peuvent être contrebalancé par deux familles d'inhibiteur des CDKs (CKI). Les membres de la famille CIP/KIP (p21, p27, p57) se lient aux complexes cycline-CDK et les membres de la famille INK4 (p15 p16, p18, p19) entrent en compétition avec les cyclines pour la liaison aux CDK4/CDK6[3].

A)



B)



C)

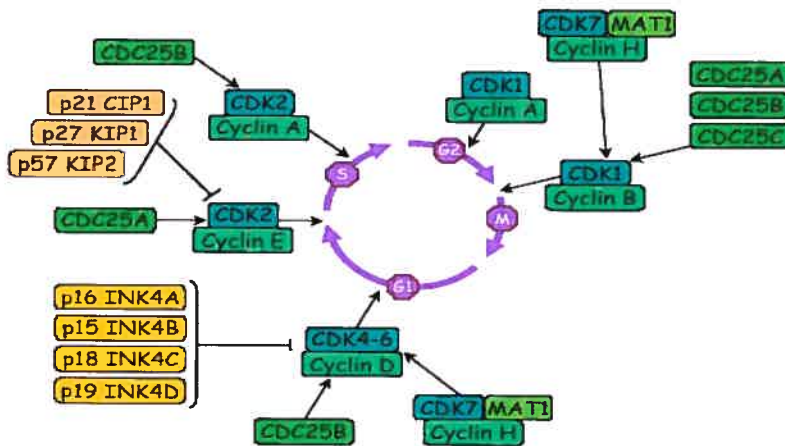


Figure 1-2: Mécanismes qui régulent les CDKs. (voir texte pour détails)

A. Les CDKs contrôlent le cycle cellulaire. B. Phosphorylations et déphosphorylations régulent les CDKs. C. Les CDKs sont régulées par des inhibiteurs protéiques.

(<http://www.sb-roscoff.fr/CyCell/Articles/140-Meijer-Oncologie.pdf>)

1.1.2.2 Régulation par les points de contrôle du cycle cellulaire

Si seules les CDKs intervenaient, l'enchaînement des phases du cycle pourrait continuer à avoir lieu, même si l'ADN est endommagé, ce qui conduirait finalement à des anomalies génétiques ou chromosomiques graves pour les cellules, comme la perte d'un chromosome. Par conséquent, les cellules ont développé des mécanismes de surveillance afin de s'assurer de l'exécution complète des événements avant de procéder à l'étape suivante, ce qui permet de s'assurer de l'intégrité du génome. Les cellules sont donc en mesure d'arrêter la progression à travers le cycle cellulaire en phase G1, S ou G2 lorsqu'ils détectent que l'ADN n'est pas répliqué correctement ou lorsqu'il est endommagé. De plus, la cellule intervient aussi lorsque les chromatides-sœurs ne sont pas équitablement réparties le long du fuseau chromatique. Lorsque le dommage est trop important la cellule peut décider d'entrer en apoptose. La perte de ses points de contrôle pourrait mener à un excès de prolifération et entraîner des effets sévères sur la cellule tel que des mutations, l'amplification de gènes, erreurs de réplication et aberrations chromosomiques. La compréhension des mécanismes de régulations de ses points de contrôle est donc essentielle pour l'étude de maladies complexes.

1.1.2.2.1 Point de contrôle G1/S

Le point de contrôle G1/S est essentiel au maintien de l'intégrité du génome. Durant la phase G1 (Figure 1-3), la progression de la cellule vers la phase S dépend de facteurs de croissance. Lorsque les facteurs de croissance stimulent l'expression de la cycline D, celle-ci va former des complexes avec les CDK4 et CDK6. Les complexes activés vont phosphoryler la protéine RB qui, sous sa forme active (hypophosphorylée), a la capacité de séquestrer le facteur de transcription E2F. RB est un répresseur actif qui recrute HDAC1, une histone déacétylase, ce qui va réprimer la transcription[4]. E2F est un hétérodimère formé d'un membre de la famille E2F (E2F-1/6) et un membre de la famille TFDP (TFDP-1/2)[5], qui sont régulés par différents membres de la famille RB[6]. RB se lie principalement au complexe formé de E2F-1, -2 ou -3 avec TFDP-1. La phosphorylation de Rb en fin de phase G1 va permettre la libération de E2F qui va

activer la transcription de gènes essentiels à l'entrée en phase S (par exemple, thymidylate synthase, thymidine kinase, ribonucléotide réductase, l'ADN polymérase- α et cdk1)[7]. Le processus de phosphorylation de RB est initié par les complexes cycline D-CDK4/6 et accéléré par le complexe cycline E-CDK2. En absence de facteurs de croissances, en présence de signaux antiprolifératifs (comme TGF β 1) ou lors de dommage à l'ADN, la cellule peut arrêter en phase G1 si elle n'a pas atteint le point de restriction R. Les deux derniers stimuli peuvent induire différents membres des familles INK4 et CIP/KIP. TGF β 1 peut aussi inhiber la transcription de CDC25A, une phosphatase nécessaire à l'activation des CDK (Figure 1-4). En absence de facteurs de croissance, il y a activation de GSK-3 β , qui phosphoryle la cycline D qui sera alors dégradée par ubiquitination[8].

Depuis les dix dernières années, les études montrent que les gènes qui sont le plus souvent mutés dans les cancers humains sont ceux qui codent pour les éléments de la machinerie de la transition G1/S (Figure 1-4, 1-5) [9, 10]. En effet, les mutations dans ces régulateurs du cycle cellulaire peuvent donner un avantage prolifératif pour la cellule. Cependant, les cellules doivent accumuler les mutations pour développer des tumeurs. Récemment, des études ont montré que les cellules qui ont une dérégulation au niveau du contrôle G1/S, ont une phase S sous optimale[11]. Par conséquence, ces cellules n'ont pas une réplication optimale et sont prédisposées à accumuler des mutations. En général, un dérèglement du cycle cellulaire implique une mutation ou une surexpression d'un régulateur positif qui mène à son activation permanente ou l'inactivation d'un répresseur causé par une mutation, délétion, ou une hyperméthylation du promoteur. Par exemple, la surexpression de la cycline D1 est retrouvée dans plusieurs types de cancers : cancer du sein, cancer de la prostate, cancer du poumon, cancer colorectal[12, 13]. De plus, la surexpression de la cycline E est associée avec différentes tumeurs[14]. Après p53, l'inactivation de p16 est l'aberration génétique la plus souvent associée aux cancers humains. Elle est retrouvée surtout dans les cancers hématologiques et épithéliaux[10, 15]. D'autres maladies complexes ont aussi été associées avec un dérèglement de certains joueurs clé de la transition G1/S. Par exemple,

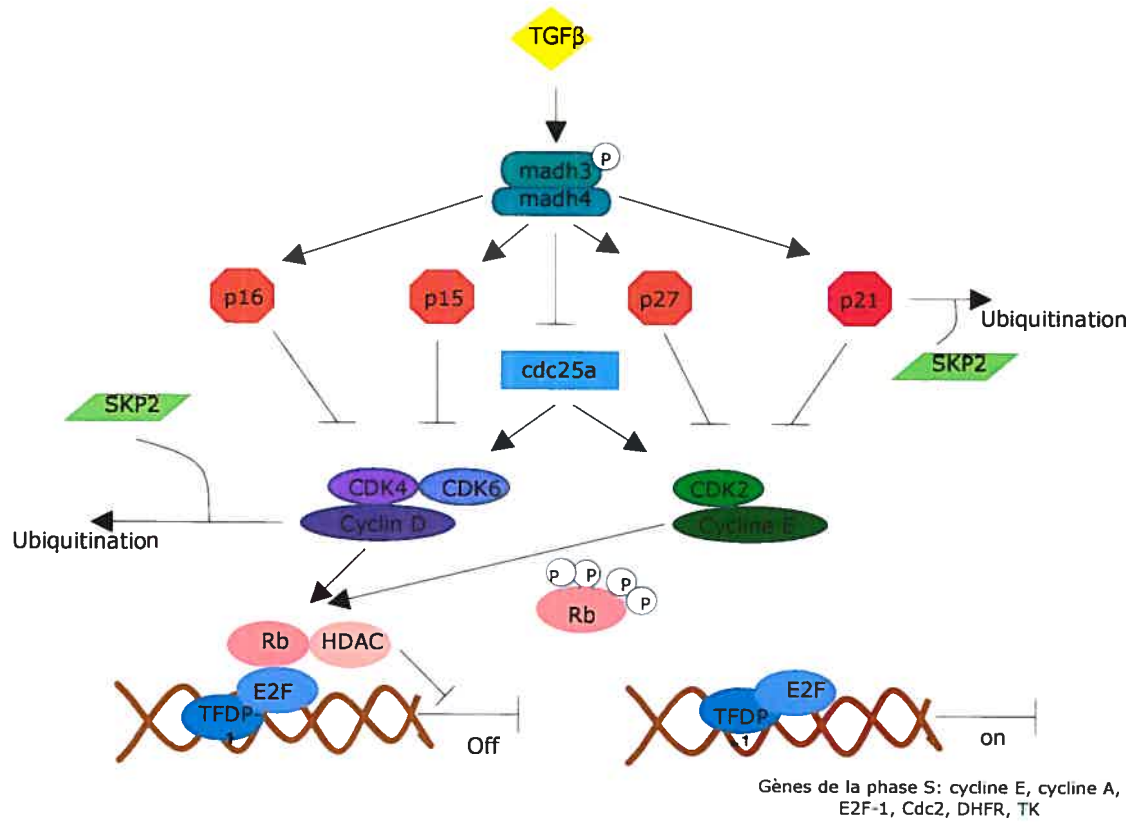


Figure 1-3: Le point de contrôle G1/S. (Voir texte pour détails)

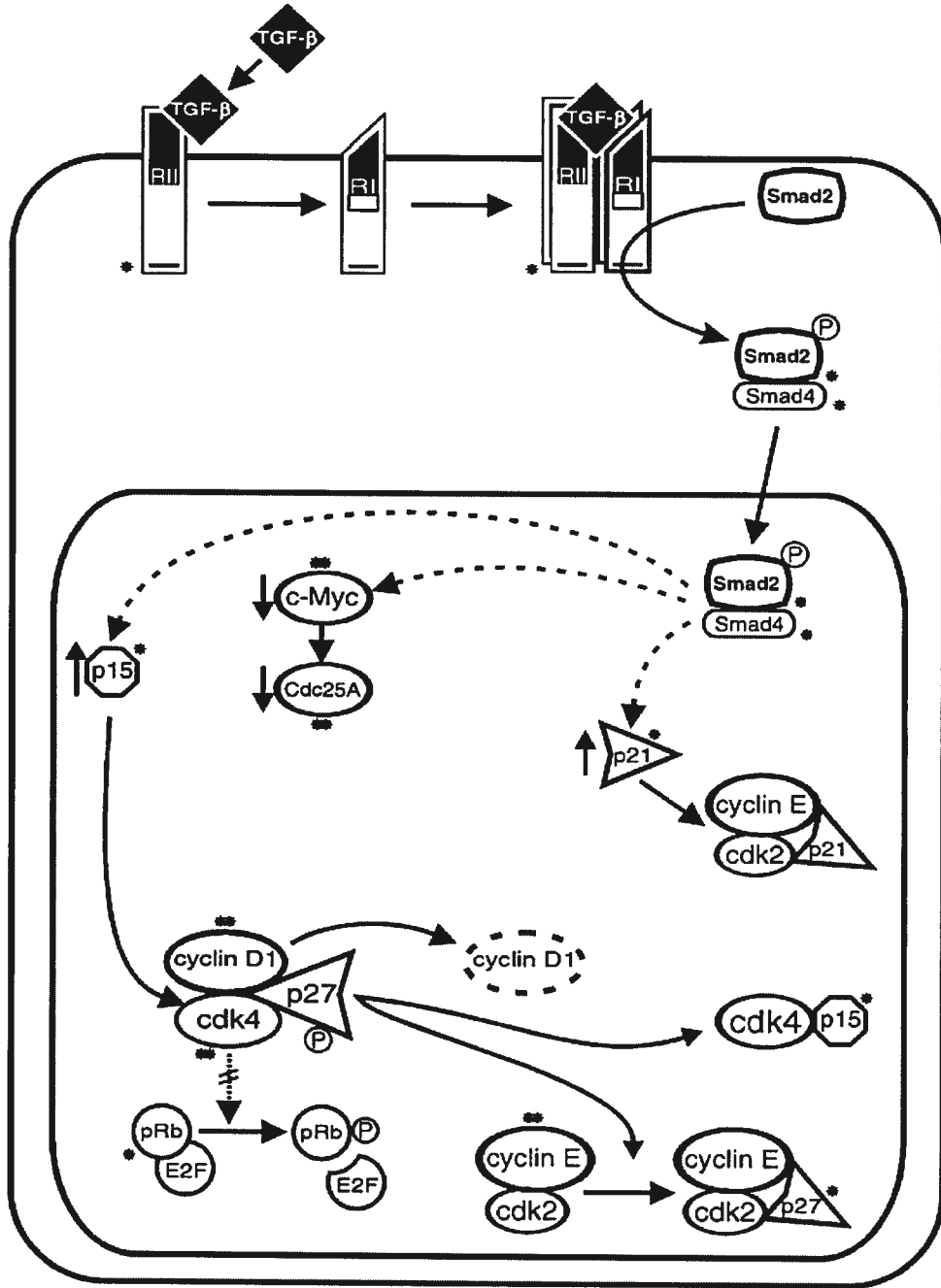


Figure 1-4: Mécanismes d'arrêt cellulaire par la voie de TGF β et la dérégulation de cette voie dans le cancer. En phase G1, TGF β 1 se lie au récepteur de type II (T β RII) qui possède une activité kinase constitutivement activée. T β RII est alors recruté au niveau du complexe TGF β 1/T β RII et phosphoryle le T β RI. La forte augmentation de l'activité kinase du récepteur I permet la phosphorylation de SMAD2. L'activation de SMAD2 permet la formation du complexe SMAD2-SMAD4 constitutivement phosphorylé qui migre vers le noyau où il régule la transcription de divers gènes. Cette cascade d'événements mène à la répression de c-myc et CDC25A, et à l'induction d'inhibiteur de kinase cycline-dépendante. Ceci va inhiber l'activation des cdks et par conséquent, inhiber la phosphorylation de RB et mener à l'arrêt cellulaire. *Composants de la voie de TGF β qui sont mutés ou fonctionnellement inactivés dans les cancers humains. **Molécules qui lorsque activés ou surexprimés peuvent contribuer à la résistance à l'arrêt via TGF β . Ligne pointillée; peut être un effet indirect[16].

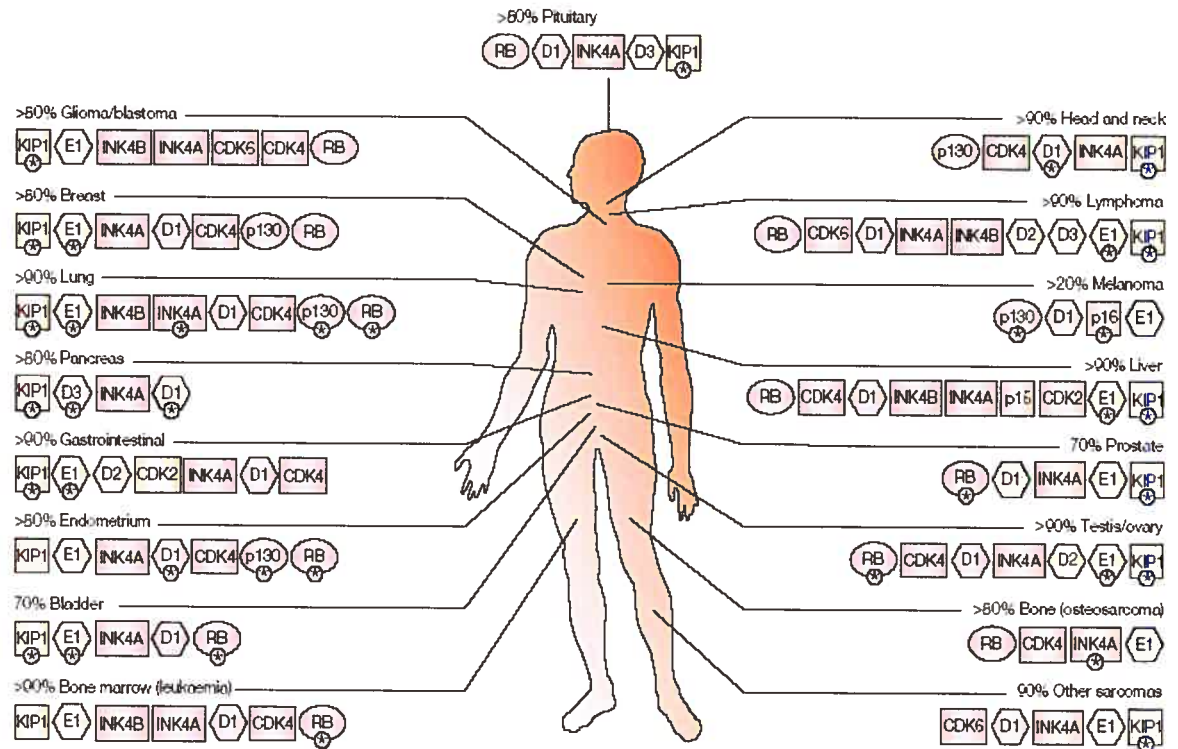


Figure 1-5: Mutations dans les régulateurs de la transition G1/S associées aux cancers humains[17]. Seulement les altérations fréquentes à plus de 10% dans les tumeurs primaires ont été considérées. Les chiffres indiquent le pourcentage de tumeur avec des altérations pour n'importe quel régulateur du cycle cellulaire énuméré. Les endroits dans lesquels la cause génétique ou épigénétique a été identifiée sont en rose. Les altérations pour lesquelles aucune explication mécanistique n'a été proposée sont en jaune. Les altérations qui sont pertinentes pour le pronostic du cancer sont indiqués par une astérisque.

l'expression anormale de CDK4, cycline D et p16 a été observée dans les cerveaux des patients atteints de la maladie d'Alzheimer[18, 19].

1.1.2.2.2 Point de contrôle G2/M

Le point de contrôle G2/M s'assure que la cellule n'entre pas en mitose tant que la réplication n'est pas totalement terminée ou tant que les lésions détectées dans l'ADN qui s'est répliqué ne sont pas réparées (Figure 1-6). CDK1 (CDC2) et la cycline B jouent un rôle clé dans cette transition. Suivant leur association, ils sont régulés par différents événements de phosphorylation, ce qui résultent en l'activation ou l'inactivation de l'activité kinase. Durant la phase G2, CDK1 est maintenu dans un état inactif par les kinases WEE1 et MYT1 en phosphorylant les résidus thréonine 14 (Thr14) et tyrosine 15 (Tyr15)[20, 21]. Lorsque les cellules approchent la phase M, la phosphatase CDC25C est activée par PLK (polo-like kinase)[22], ce qui permet la déphosphorylation de Thr14 et Tyr15 de CDK1 et son activation[23]. L'activation du complexe cycline B-CDK1 permet l'entrée en phase M en phosphorylant des protéines clés permettant la réorganisation de l'architecture de la cellule[24]. Parmi ces nombreux substrats on trouve CDC25C, ce qui permet son autorégulation, ainsi que les histone H1 et H3, les lamines et les protéines associées aux microtubules (MAPs). L'arrêt en G2 peut être activé par divers signaux (cassures des brins d'ADN, arrêt de la progression des fourches de réplication lors de la synthèse de l'ADN, etc.). Les dommages à l'ADN activent les kinases DNA-PK/ATM/ATR qui provoquent deux cascades en parallèle. La première qui inhibe rapidement la progression vers la mitose, active CHK1 qui phosphoryle CDC25C sur la sérine216[25]. Sous la forme phosphorylée, CDC25C forme un complexe avec des protéines de la famille 14-3-3 qui la séquestrent dans le cytoplasme, loin de son substrat nucléaire CDK1[26]. Dans la seconde cascade, plus lente que la première, p53 est phosphorylé, ce qui le dissocie de son partenaire MDM2 et en conséquence il ne sera pas dégradé. Ensuite, p53 active plusieurs protéines comme 14-3-3 σ , qui est maintenant considérée comme une CKI, puisqu'elle interagit avec CDK1, CDK2 et CDK4 et peut les inhiber[27]. GADD45, qui lie et dissocie le complexe

cyclineB-CDK1[28] et p21, semble inhiber l'activation de CDK1 par CAK puisqu'il diminue la phosphorylation du résidu Thr161 de CDK1[29].

Des désordres génétiques peuvent résulter de la perte ou d'altérations de la transition G2/M. Par exemple, une déficience dans ATM résulte en l'ataxie-telangiectasie, une maladie autosomale récessive caractérisée par une neurodégénération, une immunodéficience, radiosensibilité et prédisposition aux cancers (60 à 180 fois, dont 70% sont des lymphomes et des leucémies)[30]. Des mutations dans CHK2 ou p53, protéines qui agissent en aval de ATM, résultent en une maladie nommée syndrome Li Fraumeni, qui prédispose les patients dramatiquement aux cancers[31]. Un autre substrat d'ATM, BRCA1, qui est aussi impliqué dans l'arrêt du cycle cellulaire et la réparation de l'ADN, ainsi que BRCA2, sont mutés dans environ la moitié des cas de cancers du sein familial[32-34]. Aussi, la perte de fonction de 14-3-3 σ a été associée avec la transformation cellulaire. Son niveau d'expression est diminué dans les lignées cellulaires transformées et les spécimens de tumeur primaire[27, 35]. De plus, l'inactivation de 14-3-3 σ due à des changements épigénétiques tel que la méthylation a été associée avec le cancer du sein, le cancer gastrique et dans les carcinomes hépatocellulaires[35].

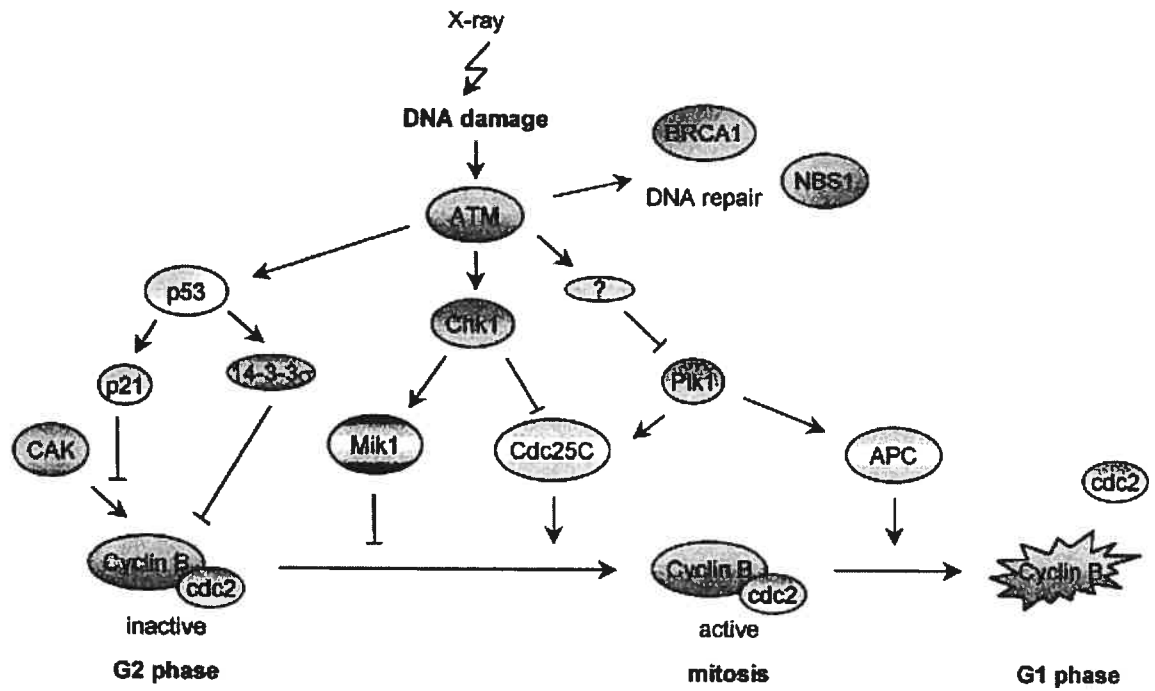


Figure 1-6: Le point de contrôle G2/M[36]. (voir texte pour détails)

1.2 Variabilité interindividuelle, cycle cellulaire et maladies complexes

L'étude de la variabilité du génome humain nous a permis d'améliorer fortement notre compréhension de l'histoire de l'humanité, de l'évolution du patrimoine génétique humain, de même que notre compréhension de base de l'évolution génomique. Cette variation génétique contribue également, en partie, à la santé et aux traits propres à chaque individu. De plus, lorsque cette variabilité interindividuelle est associée à des polymorphismes fonctionnels dans des gènes ou qui affectent l'expression, celle-ci peut parfois augmenter la susceptibilité individuelle à développer différentes maladies complexes.

Des variations au niveau des phases de transition pourraient contribuer à certaines maladies complexes tel que le cancer. Cette variabilité pourrait, entre autre être causée par des polymorphismes fonctionnels dans les gènes codant pour des composants de ces points de contrôle. Par exemple, il a été démontré qu'une substitution commune A>G au nucléotide 870 situé dans un site donneur d'épissage conservé dans l'exon 4 de CCND1 peut moduler l'épissage du ARNm[37]. En effet, l'allèle contenant le 870A encode de préférence un transcrit altéré qui résulte en une protéine avec une plus longue demi-vie. Ceci mène donc à une concentration plus importante de CCND1. De plus, cet allèle est associé avec un risque plus élevé de développer le cancer colorectal[38].

1.3 Nature d'un polymorphisme

1.3.1 Type

Un polymorphisme est défini comme une variation de la séquence d'ADN qui existe entre les individus. Les variations génétiques qui ont une fréquence supérieure à 1% dans la population sont considérées comme des polymorphismes.

Dans le génome, il existe différents types de polymorphismes dont les polymorphismes de longueur et les polymorphismes simples. Parmi les polymorphismes de longueur on retrouve les insertions et les délétions de plusieurs nucléotides et les séquences répétées en tandem (VNTR : Variable Number of Tandem Repeats). Les VNTRs sont de taille variable, constituées de répétitions en tandem d'un motif de taille également variable. Selon la taille du motif et de la répétition, on distingue les satellites[39], les minisatellites[40] et les microsatellites[41]. Les ADN satellites sont composés de séquences répétées courtes de 2-20pb ou de séquences plus complexes contenant quelques centaines de nucléotides qui peuvent être répétés jusqu'à plus de 1000 fois. Ils peuvent être retrouvés partout dans le génome mais la plupart sont situés dans les centromères. Ces séquences ne sont généralement pas transcrites. Les minisatellites sont extrêmement polymorphes et consistent de répétitions d'environ 10 nucléotides. Les répétitions peuvent s'étendre sur une région de 20kb. Ils sont généralement GC-riche et chez les humains, 90% sont situés dans la région sub-télomérique[42]. Les microsatellites comportent des séries de répétitions d'un motif court (1 à 4 pb) et sont répartis sur l'ensemble du génome. Le nombre de répétitions ne dépasse pas 25 et la taille de la séquence n'excède généralement pas 150 paires de bases. Malgré leur petite taille, ils sont fréquents dans le génome. Par exemple, les dinucléotides CA représentent environ 0,5% du génome. Le nombre de répétition dans un locus particulier varie de façon polymorphique entre les individus et est hérité de façon stable. Les répétitions de trinucléotides en tandem constituent un cas particulier, dans la mesure où un comportement anormal de microsatellites a été associé avec différentes maladies humaines. En effet, lorsque les répétitions dépassent une longueur critique celle-ci

deviennent instables au cours de la mitose et de la méiose et sont souvent transmises avec modifications à la descendance. Par exemple, des extensions ou des délétions peuvent apparaître, en fonction de la longueur de la répétition et du sexe du parent. Chez les patients atteints de la maladie de l'X fragile, le motif CGG, normalement répété de 6 à 54 fois, est présent de 200 à plus de 1000 fois[43]. Ces expansions affectent la méthylation de l'ADN et la structure de la chromatine se qui conduit à l'inactivation des gènes adjacents.

Les SNPs (single nucleotide polymorphism) sont les polymorphismes les plus simples et le plus communs. Ils sont retrouvés à un taux de 1 à chaque 300 pb ce qui représente environ 10^6 SNPs communs chez l'humain. Il existe différents types de SNPs : les insertions et les délétions de une ou quelques nucléotides et les substitutions qui consistent en des transitions (entre deux purine ou deux pyrimidine) ou des transversions (entre une purine et une pyrimidine) (Figure 1-7).

1.3.2 Impact d'un polymorphisme

Les SNPs sont distribués partout à travers le génome et sont considérés comme potentiellement responsable des différences phénotypiques entre les individus. Selon leur localisation dans l'ADN, les SNPs peuvent avoir différents effets (Figure 1-8). Un SNP situé au niveau de la séquence génomique non-codante à l'intérieur d'un gène (gSNP) n'a pas d'effet. Cependant un SNP qui se trouve au niveau de la région codante (cSNP), est susceptible d'avoir un effet direct sur la protéine et avoir un impact fonctionnel. En effet, une substitution de base peut mener à un changement d'acide aminé et peut avoir différentes conséquences sur la protéine. Elle peut altérer le repliement de la protéine ou affecter un site important pour des modifications post-traductionnelles comme un site de phosphorylation ou de glycolysation. Ces substitutions peuvent aussi causer l'apparition d'un codon stop prématuré menant à la synthèse d'une protéine incomplète. Compte tenu de la dégénérescence du code génétique, la majorité des substitutions n'entraînent pas de changement d'acide aminé. Des bases peuvent également être délétées ou ajoutées, conduisant ainsi à une

modification du cadre de lecture ou peut aussi mener à un codon stop. Un changement de forme ou de taille de la protéine peut donc changer sa fonction biologique soit en éliminant son activité, la réduisant ou en l'augmentant.

Les SNPs situés dans les introns (iSNP) peuvent affecter la production du transcrit primaire en modifiant l'épissage du gène, donc un intron pourrait être conservé ou un exon pourrait être délété. La stabilité de l'ARNm peut aussi être affectée lorsqu'un iSNP est situé dans la région non traduite en 3'. Finalement, les SNPs situés au niveau de la région promotrice sont nommés pSNPs. Ceux-ci peuvent modifier la transcription d'un gène en modifiant les sites de liaison des facteurs de transcription.

1.3.3 Détection d'un polymorphisme

La détection systématique de polymorphismes est complexe à cause du grand nombre de gènes, de la taille à explorer et de la diversité des mutations mais aussi à cause des coûts reliés et de la nécessité de travailler à haut débit. La méthode la plus sensible pour la détection de SNPs est le séquençage direct. Cependant, le séquençage n'est pas rapide à moins d'avoir à disposition des séquenceurs automatiques couplés à des outils d'analyse bioinformatique performant. Par conséquent, beaucoup de laboratoires utilisent une méthode de criblage alternative et par séquençage détermine la nature du changement. Une de ces techniques simples est le SSCP (single-stranded conformation polymorphism), qui exploite la variation de la mobilité électrophorétique des molécules d'ADN simple-brin, suite à la présence d'une mutation et au changement de conformation qui en résulte. Cependant, lorsqu'il s'agit de chercher des mutations inconnues dans des fragments d'une longueur supérieure à 250 pb, le pouvoir de détection de cette méthode diminue rapidement avec la taille. La modification du SSCP en F-SSCP a permis l'amélioration de la technique. Elle est plus adaptée à la détection haut débit, plus facile d'utilisation et permet de détecter les polymorphismes avec une efficacité d'environ 93%[44, 45].

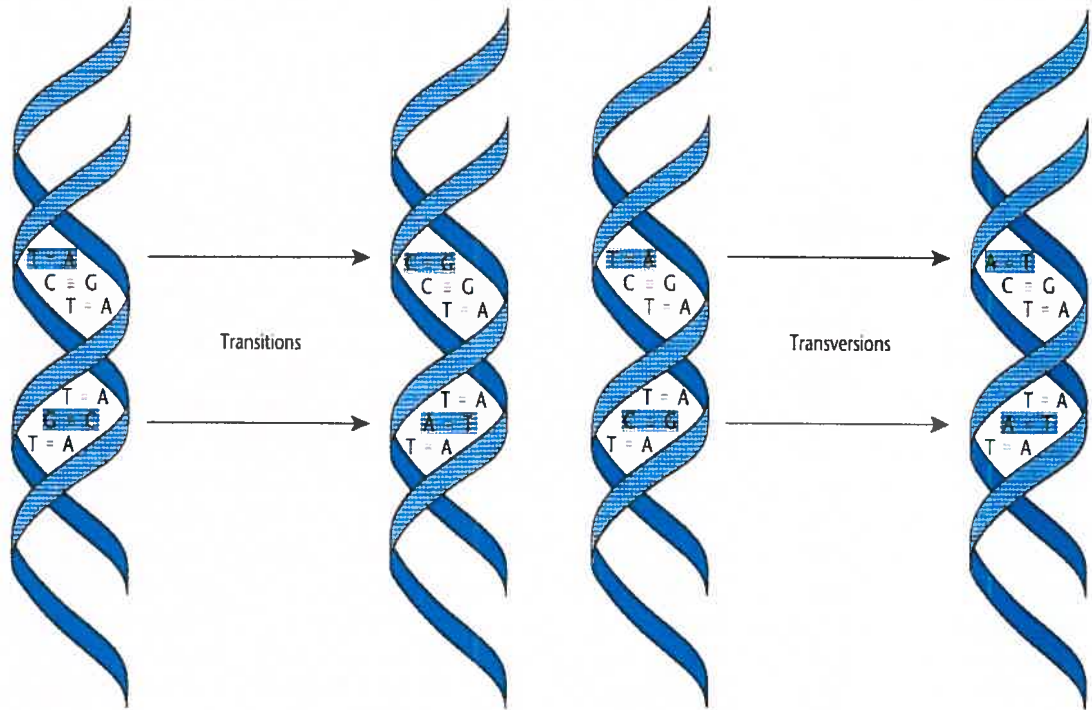


Figure 1-7 : Type de substitutions: transitions et transversions[44].

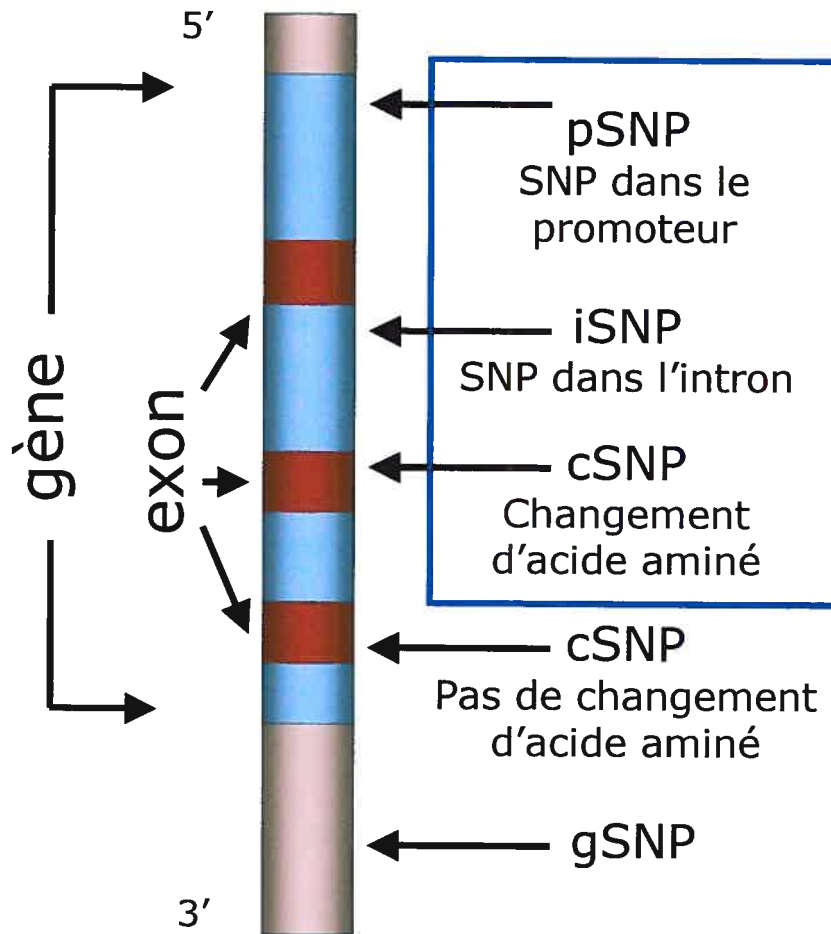


Figure 1-8 : Localisation des SNPs. pSNP, SNP dans le promoteur; iSNP, SNP dans la région intronique; cSNP, SNP dans la région codante; gSNP, SNP dans la région génomique. Les SNPs qui sont dans un encadré peuvent mener à un changement fonctionnel.

Récemment, la technique de dHPLC (denaturing high performance liquid chromatography) permettant la détection de polymorphisme a été développée[44, 45, 47, 48]. Le dHPLC permet la détection rapide de polymorphisme en identifiant les ADN hétérozygotes. La technique est simple, suite à une dénaturation, le produit PCR est lentement refroidi. Lorsqu'il s'agit d'un échantillon hétérozygote, il y aura formation d'homoduplexes et d'hétéroduplexes (Figure 1-9). Par la suite, les duplexes pourront être séparés de façon différentiel sur une colonne hydrophobe en conditions partiellement dénaturante. Puisque les hétéroduplexes commencent à se dénaturer plus rapidement à cause de leur mésappariement, ils seront élués en premier de la colonne suivie des homoduplexes. Le fragment PCR doit être partiellement dénaturé, entre 50 et 99,9%, afin de détecter un polymorphisme. Puisque la position des polymorphismes est inconnue, on sélectionne différentes températures afin d'analyser un fragment en totalité (Figure 1-10). Cette technique possède un pouvoir de détection élevé qui permet de travailler à haut débit et qui a un taux d'efficacité autour de 99% pour la détection hétérozygote[47].

1.3.4 Annotation d'un polymorphisme

Jusqu'à récemment, toute l'attention était portée sur les cSNP à cause de la possibilité de prédire leur effet sur la structure et fonction de la protéine. Les analyses du génome humain suggère que le nombre de gènes (environ 35 000) est en dessous de ce qui était attendu et sont loin d'être suffisant pour expliquer la complexité de l'organisme[49]. Ce serait plutôt les variations dans la régulation des gènes et dans l'épissage des gènes qui joueraient un rôle important. Ceci expliquerait comment une même protéine aurait diverses fonctions dans différents tissus. Par conséquent, il a été proposé que seulement une fraction des cSNP expliquerait les causes génétiques reliées à la susceptibilité individuelle de développer des maladies complexes. Ce serait plutôt les variants situés dans les séquences régulatrices comme au niveau du promoteur qui détiendraient ce rôle. Cependant, peu est connu sur les éléments régulateurs dans le génome humain. Donc, il est très difficile de distinguer les polymorphismes fonctionnels de ceux sans effets simplement en étudiant le contexte de la séquence. Présentement, de plus en plus d'efforts sont mis afin d'identifier et annoter ces variants d'ADN dans la région régulatrice.

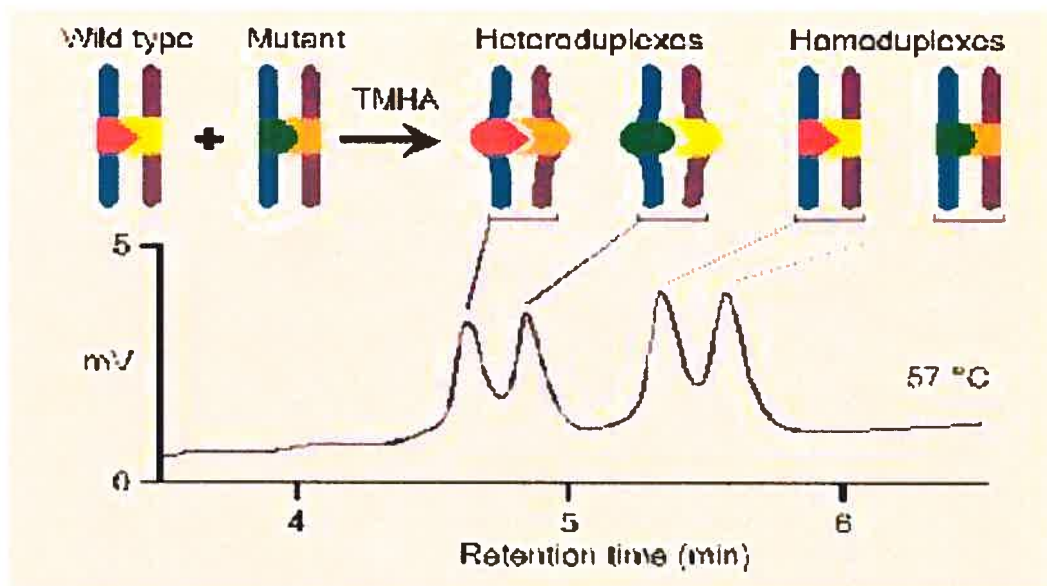


Figure 1-9 : Résolution des hétéroduplexes. (voir texte pour détails)
(www.transgenomic.com)

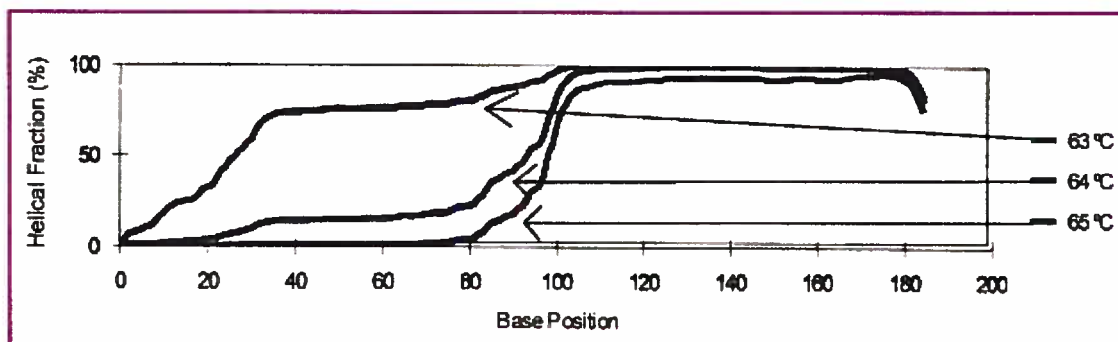


Figure 1-10 : Profil de dénaturation à 3 températures d'un fragment de 185 pb tel qu'évalué par WAVEMAKER. La recherche de polymorphismes inconnus sur la totalité du fragment nécessitera une analyse à 2 températures. Les polymorphismes situés dans la portion 5' vont être détectés à une température plus basse que ceux situés dans la région 3'.

1.3.4.1 Approche *in silico*

Un des défis dans l'évaluation des pSNPs, contrairement au cSNPs, est le développement d'algorithme robuste et de base de données pour reconnaître les séquences régulatrices fonctionnelles dans les régions non-codantes. Présentement, différentes approches existent afin de prédire des sites de liaison pour des facteurs de transcription et de trouver des modules régulateurs. Entre autre, le programme MatInspector utilise TRANSFAC, une base de donnée sur les facteurs de transcription et leurs sites de liaison, afin de comparer les similitudes entre une séquence d'ADN et des sites de liaison connus expérimentalement de facteurs de transcription[50]. Une autre approche intéressante est la génomique comparative entre les espèces afin d'identifier les régions régulatrices et les sites de facteurs de transcription, aussi appelés « phylogenetic footprinting »[51]. Cette approche se base sur le principe que l'évolution fait en sorte que les séquences importantes fonctionnellement seront beaucoup plus conservées que le reste. Plusieurs études utilisent la souris comme modèle d'analyse génomique comparative. C'est un modèle populaire à cause de la disponibilité de l'informatique génétique et de la possibilité de manipuler son génome, en plus d'être un outil efficace permettant de découvrir de nouveaux éléments régulateurs[52].

1.3.4.2 Approche *in vitro*

L'analyse de SNPs fonctionnels putatifs dans les régions promotrices est principalement faite par des techniques étudiant les interactions ADN-protéines, comme les empreintes à la DNaseI (ou DNaseI footprinting) et les gel de retardation (EMSA), et sur des techniques mesurant l'expression des gènes tel que les gènes rapporteurs[53].

La méthode de retard sur gel peut être utilisée dans l'étude de séquence d'ADN spécifique liant des protéines comme les facteurs de transcription. Cette technique consiste à incuber un fragment d'ADN marqué radioactivement avec une protéine purifiée ou des extraits nucléaires et de faire migrer sur gel non-dénaturant (Figure 1-11). Les complexe ADN-protéines migreront plus lentement (d'où le retard) que l'ADN

libre. Les expériences de retard sur gel nous permettent donc de tester et comparer la présence de sites de liaisons sur différents fragments d'ADN contenant ou non le SNP d'intérêt, de même que d'estimer l'affinité spécifique d'un facteur pour un site particulier. De plus, avec les supershifts, il est possible de vérifier la spécificité de l'interaction ADN-protéine par l'incubation préalable du complexe avec un anticorps qui se fixe à la protéine. L'empreinte à la DNase I utilise le fait qu'une protéine déposée sur l'ADN peut le protéger de l'activité de dégradation d'une nucléase comme la DNaseI. La dégradation partielle du fragment d'ADN protégé par la protéine générera une échelle de fragments de tailles différentes. En comparant l'échelle donnée par un ADN nu à celle donnée par un ADN protégé, on remarque une empreinte correspondant aux sites rendus inaccessibles à l'enzyme par la protéine (figure 1-12). Cette technique peut donc servir à l'évaluation d'une région contenant un ou plusieurs SNPs et permettrait de porter une plus grande attention au SNPs localisés près ou dans les sites de liaison.

Les études utilisant un gène rapporteur permettent d'évaluer la différence d'expression d'un promoteur comparativement à un autre dans différentes lignées cellulaires. Cette approche implique la construction d'un plasmide dans lequel est inséré la région régulatrice d'un gène d'intérêt contenant le polymorphisme couplé à un gène rapporteur comme la luciférase. Différentes études d'expression ont permis d'identifier des polymorphismes fonctionnels dans les régions promotrices[54, 55] et récemment, des études fonctionnelles à haut débit ont été développées[56, 57].

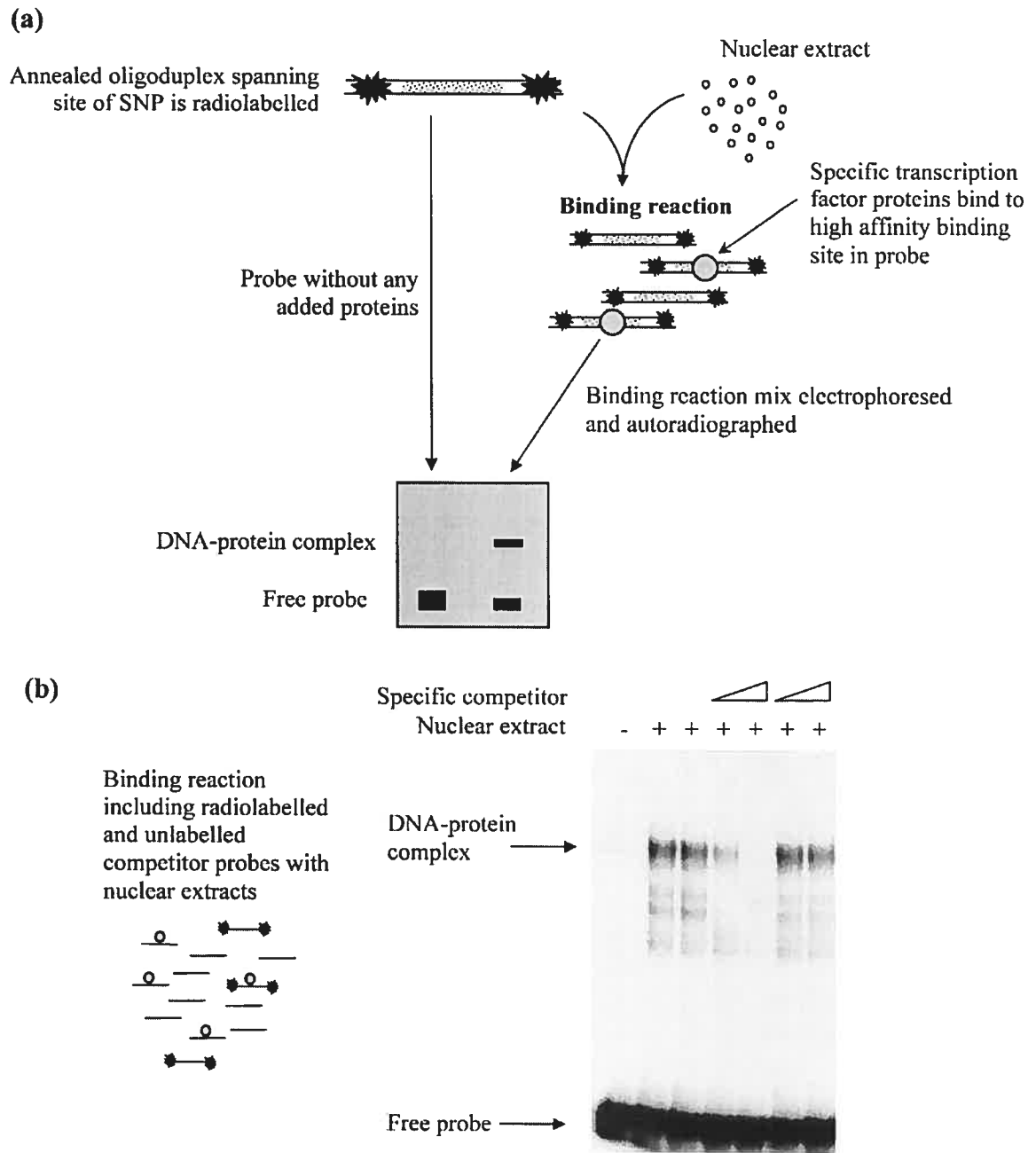


Figure 1-11 : Principe de EMSA [53]. (voir texte pour détails)

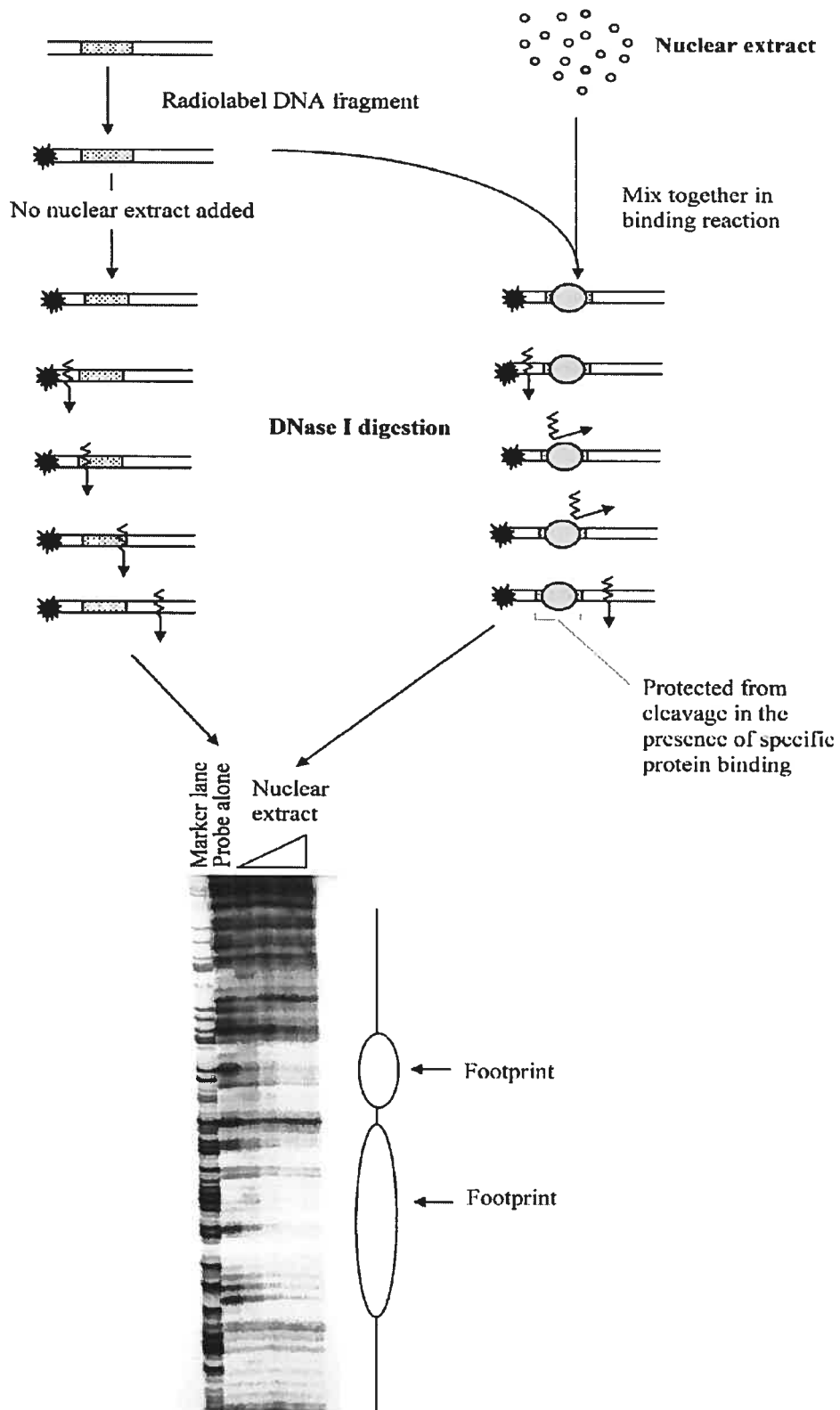


Figure 1-12 : Principe de l’empreinte à la DNase I [53]. (voir texte pour détails)

1.4 Régulation de l'expression génique

Dans un gène typique, le cœur du promoteur de classe II est situé immédiatement en amont du gène. Celui-ci lie le complexe de préinitiation qui comprend l'ARN polymérase II (Pol II) et des facteurs accessoires (« la machinerie générale de transcription »), ce qui permet à la Pol II d'initier la transcription au bon endroit. Il existe 2 classes de séquences régulatrices, les séquences proximales (ou promoteur régulateur) situées en amont du cœur du promoteur et les séquences distales, localisées à de grande distance du promoteur de base (figure 1-13). Ces dernières peuvent activer (ce sont des enhancers) ou réprimer (ce sont des silencers) la transcription de base. Le cœur du promoteur est généralement inactif. La transcription du gène est donc contrôlée par des facteurs de transcription qui peuvent être recrutés par les promoteurs régulateurs et les enhancers. Il y a activation suite au recrutement de la machinerie générale au cœur du promoteur via l'interaction entre les activateurs liés au promoteur. Certains activateurs sont exprimés de façon ubiquitaire et d'autres sont exprimés spécifiquement dans certains types cellulaires pour une fonction particulière. L'activation d'un gène nécessite aussi une modification de la chromatine, un processus complexe impliquant différents types d'enzymes, entre autre les enzymes de remodelage ATP-dépendante et les histones acétylases ainsi que d'autres facteurs qui facilitent l'assemblage du complexe de préinitiation. Or, une fois l'ADN décondensé au niveau de la région régulatrice, les sites de liaison seront accessibles aux facteurs de transcription[58].

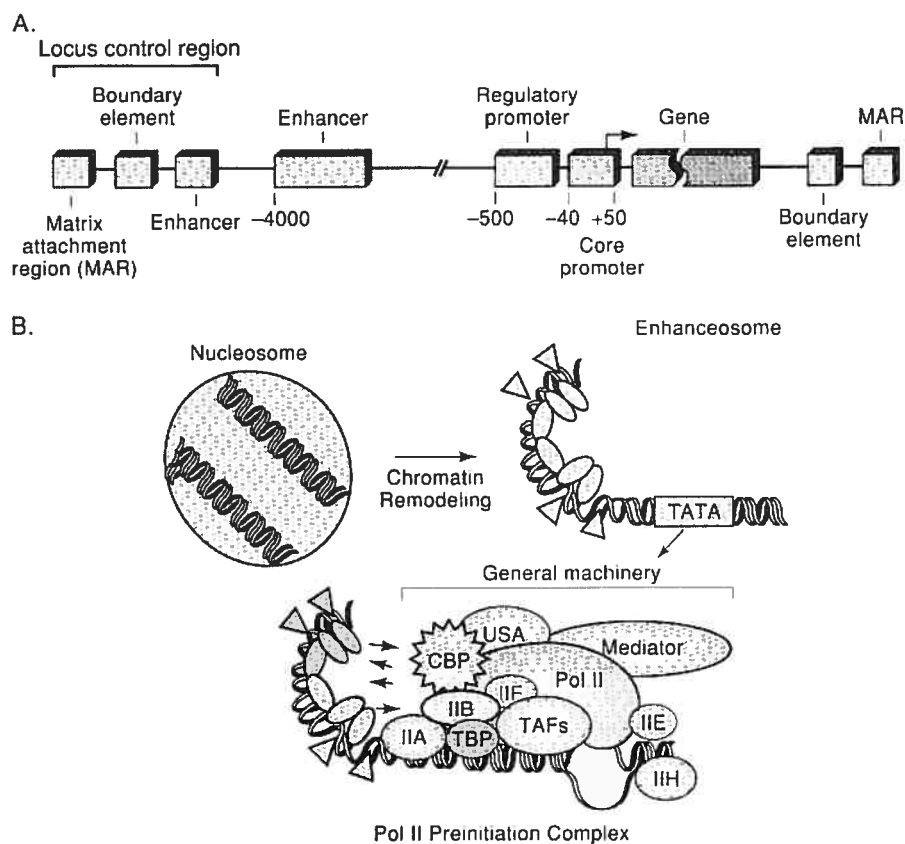


Figure 1-13 : A. Modèle typique d'un gène avec les composantes jouant un rôle dans l'activation et l'inactivation de ce gène. B. Activation d'un gène et assemblage du complexe de pré-initiation de la Pol II[58].

1.4.1 Architecture générale d'un promoteur de classe II

Typiquement, le cœur d'un promoteur de classe II chez les eucaryotes se situe à environ -40 et +50 relativement au site d'initiation de la transcription[59]. Il permet de (1) lier et contrôler le complexe de pré-initiation; (2) positionner le site d'initiation et contrôler la direction de la transcription; (3) répondre aux activateurs et répresseurs proximaux et distaux; et peut être composé des éléments suivants (Figure 1-14):

1. Le motif TATA. La boîte TATA, dont le consensus est TATAAA, est reconnue par le facteur de transcription général TFIID via l'une des sous-unités qui le compose, la protéine TBP (TATA-binding protein)[60]. La liaison de TFIID permet ensuite la formation du complexe pré-initiateur. Le point d'initiation de la transcription est généralement situé environ à 25-30 pb en aval de la boîte TATA. La boîte TATA est suffisante pour activer la transcription quand un activateur se lie à un élément régulateur qui se trouve à proximité.
2. L'élément initiateur. L'initiateur (séquence consensus PyPyA+1NT/APyPy), composé d'une séquence peu conservée, contenant le site d'initiation, possède les mêmes fonctions que la boîte TATA en permettant la formation du complexe de pré-initiation. Cette fois-ci, c'est TAF_{II}, l'une des diverses protéines associées à TBP au sein de TFIID, qui reconnaît l'initiateur[61].
3. L'élément BRE (TFIIB recognition element). L'élément BRE, est une séquence localisée à la position -32 à -38, juste en amont de la boîte TATA. La séquence consensus (G/C)(G/C)(G/A)CGCC est capable de fixer le facteur de transcription TFIIB[62].
4. L'élément DPE (downstream promoter element). En l'absence de boîte TATA, certains promoteurs possèderaient une séquence faiblement conservée, appelée DPE. DPE (séquence consensus RG(A/T)CGTG), située à environ 30 nucléotides en aval du site d'initiation, serait reconnue par TFIID[63].

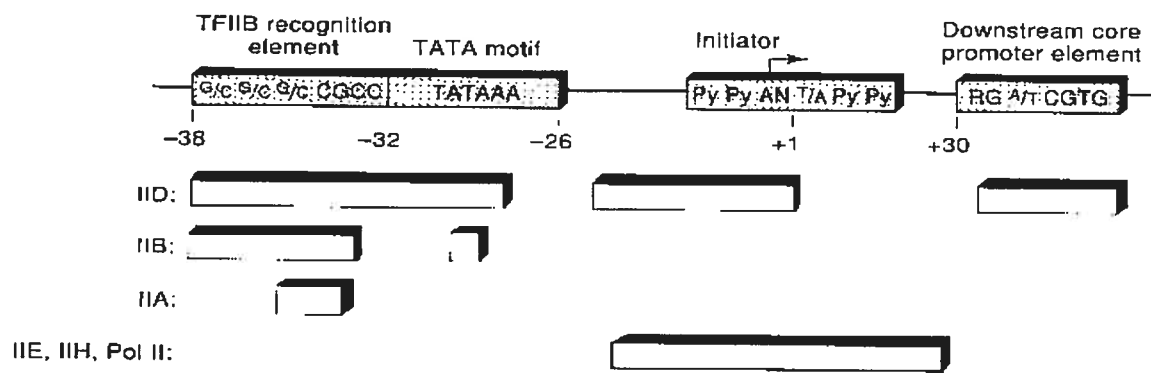


Figure 1-14 : Éléments de séquence composant le cœur de promoteur typique[58]. (voir texte pour détails)

1.4.2 La génétique régulatrice

Un polymorphisme au niveau des régions promotrices peut mener à l'apparition ou la disparition de sites de liaison des facteurs de transcription causant ainsi une perturbation au niveau de l'expression du gène. Récemment, une analyse fonctionnelle d'un polymorphisme dans le gène codant pour CYP2A6 (cytochrome P450 2A6) impliqué dans l'oxydation de la nicotine dans le foie, a été publiée[64]. Dans cette étude, l'impact fonctionnel du SNP situé en 5' a été évalué à l'aide d'études utilisant un gène rapporteur et d'EMSA. La présence du polymorphisme a montré une diminution de l'expression dans les cellules HepG2 et pourrait être causée par une baisse d'affinité d'un ou plusieurs facteurs de transcription dont l'identité est inconnue. Dans une autre étude, plusieurs polymorphismes ont été identifiés dans le gène codant pour ECE-1 (endothelin-converting enzyme) [65]. Un des SNPs localisé en 5' fait apparaître un site consensus putatif pour les protéines E2F et GATA. Les études de transfection montrent que le polymorphisme augmente l'activité du promoteur et qu'il présente une plus grande affinité pour E2F-2.

1.5 Hypothèse

L'étiologie des maladies complexes, telles que les maladies cardiovasculaires et inflammatoires ainsi que les cancers, peut en partie s'expliquer par des polymorphismes génétiques fonctionnels retracés dans la population. Jusqu'à présent, l'emphase a été mise sur la détection de polymorphismes simples dans la région codante (cSNPs) de gènes candidats. Nous proposons que les sites de liaison des facteurs de transcription dans la région régulatrice pourraient aussi être modifiés par les SNPs. Ces SNPs pourraient influencer les processus biologiques en modifiant le niveau de transcription des gènes correspondant. Les promoteurs sont impliqués dans l'initiation de la transcription et sont parmi les éléments cis-régulateurs importants qui régulent l'expression des gènes et pourraient contenir des polymorphismes (pSNPs) fonctionnels. Contrairement aux éléments régulateurs, la localisation des promoteurs est plutôt fixe relativement à leur gène respectif. Dans ce projet, nous avons arbitrairement défini la région promotrice comme un segment de 2kb situé en amont du premier exon. Nous avons sélectionné cette région 2kb parce qu'un article de revue regroupant l'étude de 151 variants fonctionnels retrouvés dans 107 gènes différents a montré que 88% (133) de ces polymorphismes se retrouvent à l'intérieur d'une région de 2kb située en amont du site d'initiateur de la transcription, 5 dans la région 3' UTR, 7 dans les introns et 6 à une distance plus grande que 2kb. Le contrôle du cycle cellulaire est essentiel au maintien de l'homéostasie cellulaire et au maintien de l'intégrité du matériel génétique. Nous proposons que des pSNPs fonctionnels dans des gènes candidats génèrent une variabilité interindividuelle dans les mécanismes contrôlant le cycle cellulaire, d'où un risque variable de développer certaines maladies, tel que le cancer.

1.6 Objectifs du projet

Les objectifs spécifiques du projet sont :

1. Identifier des polymorphismes dans les régions promotrices (pSNPs) de 17 gènes impliqués dans le point de contrôle G1/S du cycle cellulaire
2. Évaluer l'impact fonctionnel des pSNPs, à l'aide d'analyses *in silico*.
3. Valider l'impact fonctionnel des pSNPs à l'aide d'expériences *in vitro* :
 - a. Vérification des prédictions à l'aide d'expériences de gel de retardement (EMSA)
 - b. Évaluer l'impact des pSNPs sur la transcription de gènes rapporteurs dans différentes lignées cellulaires.

Chapitre 2

-Article 1-

Detection and characterization of DNA variants in the promoter regions of hundreds of human diseases candidate genes

Article en préparation

Contribution des auteurs :

D. Sinnett : a écrit l'article et supervision du laboratoire.

M.C. Théberge : a contribué aux expériences en laboratoire et a participé à l'analyse des données.

H. Bélanger : a contribué aux expériences en laboratoire et a participé à l'analyse des données.

S. Langlois : a contribué aux expériences en laboratoire et a participé à l'analyse des données.

S. Drouin : participation à la validation des données.

P. Beaulieu : a contribué aux analyses *in silico*.

D. Gelh : gestion et maintien de la base de données génomique.

C. Zotti : a effectué le séquençage.

J.-F. Lefebvre : a contribué aux analyses statistiques.

D. Labuda : participation à l'analyse des données.

**Detection and characterization of DNA variants in the promoter regions of
hundreds of human diseases candidate genes**

^{1,2}Daniel Sinnett, ¹Marie-Christine Théberge, ¹Hélène Bélanger, ¹Sylvie Langlois,
¹Simon Drouin, ¹Patrick Beaulieu, ¹Dominik Gehl, ³Corine Zotti, ¹Jean-François
Lefebvre, ¹Damian Labuda and the Genome-Quebec Regulatory Genetic
consortium*

¹Division of Hematology-Oncology, Research Center, Sainte-Justine Hospital, Montreal,
Canada; ²Department of Biochemistry and Department of Pediatrics, University of
Montreal, Montreal, Canada; ³McGill University and Genome Quebec Innovation
Centre

Running title: Identification of pSNP in human genes

Corresponding author: Daniel Sinnett, Ph.D., Division of Hematology-Oncology,
Research Center, Sainte-Justine Hospital, 3175 chemin de la Côte-Sainte-Catherine,
Montreal, H3T 1C5, Canada. Phone: (514)345-4931 #2990. Fax: (514)345-4731. ■
■

Key words: regulatory genetics, promoter polymorphisms, complex diseases, genomics.

Abstract

Understanding genetic variation might reveal the cause of individual susceptibility to a variety of complex diseases such as asthma, diabetes and cancer. Current efforts to identify functional DNA variants have essentially been oriented towards single nucleotide polymorphisms (SNPs) found in coding regions of candidate genes since they have direct impact on the structure and function of the affected proteins. Abnormal expression of finely regulated genes could also lead to disequilibria in different metabolic pathways and/or biological processes. Thus investigation of SNPs in the promoter regions (pSNPs) of genes should improve our knowledge on complex diseases aetiology. Unfortunately, little is known about the nature and the prevalence of pSNPs. The aims of this study were to identify common pSNPs in approximately 200 genes and to predict *in silico* their putative functional impact. The targeted promoter regions, arbitrary defined as a 2kb genomic segment upstream of the transcription initiation site, were screened by dHPLC for the presence of SNPs in a worldwide panel of 40 individuals. As a result 1838 pSNPs were detected, including 75% that modify (either gain or loss) the putative binding sites of known transcription factors. We also examined the distribution of these pSNPs among features such as conserved regions, repeats, dinucleotides as well as gene ontology terms. This report supports the functional relevance of several of the pSNPs investigated and suggests a putative impact on disease susceptibility.

Introduction

Understanding gene regulation is highly relevant to human health and genetic diseases. Since protein levels regulate many biological pathways, abnormalities in gene regulation could influence biological processes and thus modify disease risk. Therefore functional polymorphisms in promoter regions of relevant genes might explain some of the observed inter-individual phenotype variability such as protein levels and might be associated with common diseases as recently reported [1-7]. In other words identifying DNA variants that influence gene expression is an important issue [8].

Until now, the systematic search for functional polymorphisms has predominantly been targeting coding region of DNA (cSNPs), mainly because cSNPs include missense polymorphisms that may have a direct effect on protein structure and function. It has been estimated that only 20% of cSNPs altered the protein [9]. Polymorphism in promoter sequences (pSNPs) has been studied less intensively, despite the suggestion that most relevant mutations reside in such regulating regions [10, 11]. This is due to the lack of accurate way to predict the impact and the relative abundance of pSNPs that are associated with human disease. It has been proposed that there are as many promoter variants as there are cSNPs, assuming that the total number of bases involved in the transcriptional control of human genes is of a similar magnitude as all the protein coding bases.

We and others have shown that in approximately 20% of human genes that both alleles are not always expressed at the same levels thus leading to allelic imbalance [12, 13]. This variability might be due to functional polymorphisms in the promoter regions of genes. In this study, we screened by dHPLC 40 ethnically diverse individuals for variants in the proximal promoter region of 197 genes. We identified 1838 pSNPs that were further analyzed using *in silico* tools to assess their putative impact on transcription factor binding sites as well as their distribution in respect to various genomic sequences features.

Material and methods

Population samples. The population panel consisted of 40 unrelated individuals including 8 Africans, 8 Europeans, 8 Asians, 8 Middle-easterners and 8 Amerindians. This collection was constituted of genomic DNA samples obtained from Coriell or isolated from peripheral blood of anonymous donors following informed consent.

SNP detection in promoter regions. The SNP detection was performed by PCR-based dHPLC analysis followed by direct sequencing. The promoter region arbitrary defined as the 2kb sequence upstream of the transcription start site (based on RefSeq mRNA) of each gene was amplified sequentially in 4 to 8 ~350 bp overlapping fragments. Promoter positions were numbered with respect to the first nucleotide of the first exon as +1, and the nucleotide immediately upstream as -1. The primer sequences and detailed information on the reaction conditions are available upon request. Briefly, PCR was done in a 50 ul reaction mixture containing 15 ng DNA, 0.5 uM each primer, 0.12 mM dNTPs, 1.8 mM MgCl₂, 1.8 U platinum Taq (Invitrogene) for 34 cycles using a touchdown approach from 62C to 55C. DHPLC analysis was performed on a Transgenomic WAVE system (Transgenomic, Inc). Briefly, PCR products were denatured for 1 min at 94C and then gradually reannealed by decreasing sample temperature to 45C over a period of 30 min to form homo- and/or hetero-duplexes. The resulting PCR species were applied to the DHPLC column and eluted with a linear acetonitrile gradient at a flow rate of 1.5 ml/min. The gradient conditions were determined using the WaveMaker software (Transgenomic, Inc). Because the dHPLC approach is based on the differential retention of homo- and heteroduplex DNA fragments by ion-pair chromatography under conditions of partial heat denaturation [42] optimal discrimination of double-stranded combinations depends on the temperature at which partial denaturation of heteroduplexes occur. In this study, samples were run at one degree above and/or below the predicted temperature to ensure resolution of most possible sequence variations. All samples were analyzed by visual inspection by at least two independent investigators and samples with clear or suggestive variant peaks were sequenced.

Sequencing. The nature and the position of each pSNP were determined by direct sequencing using the same primers as used for dHPLC. For most SNPs, it was possible to evaluate both the forward and the reverse sequences. We carried out dye terminator sequencing using the ABI Prism Big Dye Systems. Samples were run on ABI 3700 automated sequencer and analyzed using the Phred/Phrap/Consed suite of software to provide base quality scores [43]. Polymorphisms were verified by manual evaluation of the individual sequence traces by two independent researchers.

Transcription factor binding site identification: Depending on the position in the regulatory region, a regulatory SNP (pSNP) may lead either to complete loss of a TF site [44, 45] or the formation of a novel TF site [46, 47]. We used the Transfac database (transfac.gbf.de/TRANSFAC/index.html) and the program matInspector to predict binding sites of known transcription factors.

Results

Detection of promoter SNPs

A total of 197 genes encoding proteins involved in double-stranded DNA repair, cell cycle checkpoint, cell metabolism, and xenobiotic biotransformation (see annexe VII) were screened by dHPLC for the presence of SNPs in their promoter regions arbitrarily defined as a 2.0 kb region upstream of the start of the 1st exon. Although regulatory sequences might be located outside the targeted region, a recent survey showed that up to 90% of validated functional *cis*-regulatory polymorphism lie indeed within the first 2kb [14]. The sample size (n=80 chromosomes) allows greater than 98% and 56% probability to detect allele frequencies of at least 0.05 either worldwide or within a specific population, respectively. Overall the screening of approximately 380 kb/individual led to the identification of 1838 pSNPs, including only 324 (26%) that were already present in public database (HGBASE, dbSNP and/or HGMD). When we analyzed the distribution of the pSNPs in respect to the different types of base substitution we found that transitions (1077 SNPs, 59%) outnumbered transversions (600 SNPs, 33%), the remaining 9% was constituted of insertions and deletions (Table

1). The observed ratio of transition on transversion of 1.8 is similar to the one recently reported for 5'UTR regions [15].

Distribution of pSNPs according to the sequence context

The high frequency of C>T and G>A transitions support the role of 5-methylcytosine deamination in the generation of DNA variability in promoter regions. Thus we looked at the distribution of the pSNPs among the different types of dinucleotides (Fig. 1), particularly because CpGs are known to be prone to point mutation. We indeed observed the highest number of pSNP in CG dinucleotides, but we did not observe the expected tenfold greater mutation rate [16-18]. This unusual pattern could be explained at least in part by the presence of CG dinucleotide in CpG islands that would be less likely to mutate [e.g. 15, 18].

Although the observed pSNPs were equally distributed over the 2 Kb region, we found an excess of small inter-SNP distances (Fig. 2) suggesting a local clustering. Finally, when we assessed the distribution of the pSNPs in respect to different types of repeats, we found a significant overrepresentation of polymorphisms only among the LTR/ERV family of repeats (Table 2).

Nucleotide diversity

We estimated the nucleotide diversity using the θ value, in which the observed number of SNPs is adjusted for the sample size [e.g. 19, 20]. We found that SNPs occurred in regulatory regions at a frequency of 1 per 204 bp, which corresponds to $\theta = 10 \times 10^{-4}$. This value is consistent with the numbers reported by Freudenberg-Hua et al ($\theta = 4.8 \times 10^{-4}$) for non-coding regions as well as Cargill et al. ($\theta = 10.0 \times 10^{-4}$) and Wong et al. ($\theta = 11.9 \times 10^{-4}$) for synonymous substitution in coding sequences.

Predicted impact of pSNPs on transcription factor binding sites (TFBSs)

In the search for SNPs with putative functional impact, the proximal promoter region of the selected genes were screened for the presence of predicted TFBSs using the matInspector program and the Transfac database

(transfac.gbf.de/TRANSFAC/index.html). Following the mapping of SNPs in the corresponding promoter regions, 428 TFBSs (23%) were predicted to be lost, 428 (23%) to be gained and 522 (28%) to lead to a combination of losses and gains. Of note, in certain cases, SNPs affected putative overlapping TFBSs. Obviously, such consensus binding sites are highly degenerated and thus most predictions might not be biologically relevant [21]. It has been suggested that the presence of evolutionary conserved sequences in the promoter regions might indicate putative functional motifs [22]. We found 91 (5%) pSNPs in regions conserved between mice and human that fall in predicted TFBSs. Of note, such comparative analysis has predictive value only for changes leading to losses of predicted TFBSs because gains would be specific to human in our study.

Distribution of pSNPs among genes

As described above in the genes and the human samples tested we observed an average 4.9 pSNPs per 1000bp. However some of the investigated promoter regions of genes are depleted in SNP while others seem to have an excess of SNP (Fig. 3). To assess whether certain genes were more or less prone for variability because of their expected function we examined their gene ontology (GO) classifications [23]. The candidate genes participate in a wide range of molecular functions and biological processes. We classified the SNPs according to functional classes (ontology groups) and then compared the SNPs contents between groups (Fig. 4). The analysis of the substitution rates of the regulatory regions of the targeted in respect to gene ontology terms indicates that certain functional groups preferentially occur in sensitive or resistant promoter regions. In most cases the sample size in the GO categories was too small to perform meaningful comparisons but several interesting trends are observed (Table 3). For example gene products that play roles in the cellular defense response, hormone activity and oxidative stress response have significant higher θ values compared to other categories (Fig. 4, right). This is consistent with the concept of balancing selection that is a known selective force affecting genes involved in the host response to endogenous and/or exogenous factors [24]. In contrast, molecular basic functions such as electron transport, apoptosis

pathway and damaged DNA binding were more prevalent in the low θ gene (Fig. 4, left).

Discussion

Now that genomic DNA sequences of mammalian species are available, one of the main challenges is the understanding of the mechanisms underlying transcription control. However, our knowledge of regulatory elements in the human genome is far from comprehensive. Consequently, no general rules have been proposed to account for the functional consequences of regulatory mutations, particularly those located in promoters, enhancers, silencers and introns. In this study, we wanted to assess the genetic variability in the proximal promoter regions of approximately 200 genes and to provide an in-depth in silico characterization of the corresponding 1838 pSNPs in order to improve the annotation of these regulatory regions. The sequence elements affecting transcriptional activity may be found within the gene sequence or many thousands of bases up or downstream from the transcribed region, but we decided to focus our study to the proximal promoter region defined here as the 2kb interval located immediately upstream of the start of transcription. This decision was supported by studies that provided evidence this was indeed an important region to study in the search for human variation affecting gene expression [14, 25]

Although, the pSNPs were equally distributed over the 2kb interval, we observed an apparent local clustering illustrated by an excess of small inter-SNP distances. This observation could be explained by many genomic context features that can influence mutagenesis, including homonucleotide tracts, sites of potential Z-DNA formation, direct and indirect repeats, microsatellite, etc. [26-29]. For instance, it has been estimated that approximately 23% of cSNPs in human genes are C->T and G->A transitions in CpG dinucleotides [e.g. 29]. Here as well, we observed a similar trend supporting the role of 5-methylcytosine deamination in the generation of DNA variability in promoter regions, which is concordant with the slight overrepresentation of pSNPs in CG dinucleotides. However, the latter is much lower than the expected tenfold increase in the mutation

rate [16-18], which could be explained by the presence of CG dinucleotide in unmethylated CpG islands that would be less prone to mutations [e.g. 15, 18]. With the exception of a significant overrepresentation of pSNPs among the LTR/ERV family of repeats we did not find any other association with the sequence features investigated. The majority of ERV groups have colonized our genome at least 25 millions years ago [30] and are non-functional due to the accumulation of mutations. However, some of these ERVs integrated into regulatory regions have been shown to alter the expression of adjacent genes [31]. In this regard, some authors have suggested that methyl-CG deamination has evolved as a specific defense against colonization of the genome by ERVs [32].

The expectation that the genomic sequence surrounding genes is large and polymorphisms in these regions will surely be abundant: how can we effectively select the true regulatory SNPs? Promoter regions contain several predicted TFBSs [33-35] and in some cases a combination of several binding sites is required for biological function [33-37]. Here, we found that 75% (1374 out of 1838) of the pSNPs had a predicted impact on putative binding sites for known transcription factors. Thus one of the obvious challenge is to accurately predict the impact of a SNP on consensus binding sites that are highly degenerated and found everywhere without functionally interfering with normal transcription control. In order to gain more confidence in predictions, we applied a phylogenetic footprinting approach [e.g. 21, 22] that is based on the assumption that functional sites in promoters should evolve much slower than other regions that do not bear any conservative function [38]. We have found 91 pSNPs in human-mouse conserved regions. However, one should be aware that species-specific regulatory differences will not be recognized by this method. For instance, it has recently been shown that 32-40% of human functional TFBSs are not functional in rodents [39]. Furthermore, this inter-species comparative approach cannot be applied on pSNPs leading to predicted gains of TFBSs.

Chuang and Li [40] provided evidence that categories of functionally related genes reside preferentially in mutationally sensitive or resistant regions. Here, we showed that

some genes more tolerant to genetic variability were biased toward host-response (receptors, cell adhesion, cellular defense and oxidative stress response, etc), whereas the resistant ones were biased toward essential cellular processes (electron transport, protein modifications, apoptosis, etc). These studies taken together would suggest that certain genes are under selective pressure against mutations because of their function or these genes are tightly regulated and therefore alterations in regulatory functions are detrimental to the expression, whereas permissive groups might reflect genes that need new mutations to face new stimuli. Akey et al. [41] suggested that these patterns of genetic variation might represent signatures of natural selection. The validation of the functional impact of the pSNPs associated with the tolerant and resistant genes might help us to develop new tools to identify “true” regulatory SNPs. Because sequence variation in the regulatory region of the selected genes might give rise to difference in the gene product’s levels, such variants would be suitable candidates for genetic susceptibility of complex traits.

Acknowledgements

This study was supported by research funds provided by Genome Quebec and Genome Canada. T.J.H. is supported by a Clinician-Scientist Award in Translational Research by the Burroughs Wellcome Fund and an Investigator Award from the Canadian Institutes of Health Research. DS holds the François-Karl Viau Chair in Pediatric Oncogenomics and is a scholar of the Fonds de la Recherche en Santé du Québec (FRSQ).

References

1. Yasuda, T., H. Takeshita, E. Nakazato, T. Nakajima, Y. Nakashima, S. Mori, K. Mogi, and K. Kishi, *The molecular basis for genetic polymorphism of human deoxyribonuclease II (DNase II): a single nucleotide substitution in the promoter region of human DNase II changes the promoter activity*. FEBS Lett, 2000. **467**(2-3): p. 231-4.
2. Li, L.C., R.M. Chui, M. Sasaki, K. Nakajima, G. Perinchery, H.C. Au, D. Nojima, P. Carroll, and R. Dahiya, *A single nucleotide polymorphism in the E-cadherin gene promoter alters transcriptional activities*. Cancer Res, 2000. **60**(4): p. 873-6.
3. Smeraldi, E., R. Zanardi, F. Benedetti, D. Di Bella, J. Perez, and M. Catalano, *Polymorphism within the promoter of the serotonin transporter gene and antidepressant efficacy of fluvoxamine*. Mol Psychiatry, 1998. **3**(6): p. 508-11.
4. Sanak, M., M. Pierzchalska, S. Bazan-Socha, and A. Szczeklik, *Enhanced expression of the leukotriene C(4) synthase due to overactive transcription of an allelic variant associated with aspirin-intolerant asthma*. Am J Respir Cell Mol Biol, 2000. **23**(3): p. 290-6.
5. Manuck, S.B., J.D. Flory, R.E. Ferrell, J.J. Mann, and M.F. Muldoon, *A regulatory polymorphism of the monoamine oxidase-A gene may be associated with variability in aggression, impulsivity, and central nervous system serotonergic responsivity*. Psychiatry Res, 2000. **95**(1): p. 9-23.
6. Pihlajamaki, J., L. Karjalainen, P. Karhapaa, I. Vauhkonen, M.R. Taskinen, S.S. Deeb, and M. Laakso, *G-250A substitution in promoter of hepatic lipase gene is associated with dyslipidemia and insulin resistance in healthy control subjects and in members of families with familial combined hyperlipidemia*. Arterioscler Thromb Vasc Biol, 2000. **20**(7): p. 1789-95.
7. Ishigami, T., K. Tamura, T. Fujita, I. Kobayashi, K. Hibi, M. Kihara, Y. Toya, H. Ochiai, and S. Umemura, *Angiotensinogen gene polymorphism near transcription start site and blood pressure: role of a T-to-C transition at intron I*. Hypertension, 1999. **34**(3): p. 430-4.
8. Hudson, T.J., *Wanted: regulatory SNPs*. Nat Genet, 2003. **33**(4): p. 439-440.

9. Sunyaev, S., V. Ramensky, I. Koch, W. Lathe, 3rd, A.S. Kondrashov, and P. Bork, *Prediction of deleterious human alleles*. Hum Mol Genet, 2001. **10**(6): p. 591-7.
10. King, M.C. and A.C. Wilson, *Evolution at two levels in humans and chimpanzees*. Science, 1975. **188**(4184): p. 107-16.
11. Pastinen, T. and T.J. Hudson, *Cis-acting regulatory variation in the human genome*. Science, 2004. **306**(5696): p. 647-650.
12. Pastinen, T., R. Sladek, S. Gurd, A. Sammak, B. Ge, P. Lepage, K. Lavergne, A. Villeneuve, T. Gaudin, H. Brandstrom, A. Beck, A. Verner, J. Kingsley, E. Harmsen, D. Labuda, K. Morgan, M.C. Vohl, A.K. Naumova, D. Sinnett, and T.J. Hudson, *A survey of genetic and epigenetic variation affecting human gene expression*. Physiol Genomics, 2004. **16**(2): p. 184-93.
13. Lo, H.S., Z. Wang, Y. Hu, H.H. Yang, S. Gere, K.H. Buetow, and M.P. Lee, *Allelic variation in gene expression is common in the human genome*. Genome Res, 2003. **13**(8): p. 1855-62.
14. Rockman, M.V. and G.A. Wray, *Abundant raw material for cis-regulatory evolution in humans*. Mol Biol Evol, 2002. **19**(11): p. 1991-2004.
15. Hellmann, I., S. Zollner, W. Enard, I. Ebersberger, B. Nickel, and S. Paabo, *Selection on human genes as revealed by comparisons to chimpanzee cDNA*. Genome Res, 2003. **13**(5): p. 831-7.
16. Labuda, D. and G. Striker, *Sequence conservation in Alu evolution*. Nucleic Acids Res, 1989. **17**(7): p. 2477-91.
17. Kondrashov, A.S., *Direct estimates of human per nucleotide mutation rates at 20 loci causing Mendelian diseases*. Hum Mutat, 2003. **21**(1): p. 12-27.
18. Tomso, D.J. and D.A. Bell, *Sequence context at human single nucleotide polymorphisms: overrepresentation of CpG dinucleotide at polymorphic sites and suppression of variation in CpG islands*. J Mol Biol, 2003. **327**(2): p. 303-8.
19. Cargill, M., D. Altshuler, J. Ireland, P. Sklar, K. Ardlie, N. Patil, N. Shaw, C.R. Lane, E.P. Lim, N. Kalyanaraman, J. Nemesh, L. Ziaugra, L. Friedland, A. Rolfe, J. Warrington, R. Lipshutz, G.Q. Daley, and E.S. Lander,

- Characterization of single-nucleotide polymorphisms in coding regions of human genes.* Nat Genet, 1999. **22**(3): p. 231-8.
20. Wong, G.K., Z. Yang, D.A. Passey, M. Kibukawa, M. Paddock, C.R. Liu, L. Bolund, and J. Yu, *A population threshold for functional polymorphisms.* Genome Res, 2003. **13**(8): p. 1873-9.
 21. Zhang, Z. and M. Gerstein, *Of mice and men: phylogenetic footprinting aids the discovery of regulatory elements.* J Biol, 2003. **2**(2): p. 11.
 22. Mottagui-Tabar, S., M.A. Faghihi, Y. Mizuno, P.G. Engstrom, B. Lenhard, W.W. Wasserman, and C. Wahlestedt, *Identification of functional SNPs in the 5-prime flanking sequences of human genes.* BMC Genomics, 2005. **6**(1): p. 18.
 23. Ashburner, M., C.A. Ball, J.A. Blake, D. Botstein, H. Butler, J.M. Cherry, A.P. Davis, K. Dolinski, S.S. Dwight, J.T. Eppig, M.A. Harris, D.P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J.C. Matese, J.E. Richardson, M. Ringwald, G.M. Rubin, and G. Sherlock, *Gene ontology: tool for the unification of biology. The Gene Ontology Consortium.* Nat Genet, 2000. **25**(1): p. 25-9.
 24. Richman, A., *Evolution of balanced genetic polymorphism.* Mol Ecol, 2000. **9**(12): p. 1953-63.
 25. Coleman, S.L., B. Hoogendoorn, C. Guy, S.K. Smith, M.C. O'Donovan, and P.R. Buckland, *A strealined approach to functional analysis of promoter polymorphisms.* Biotechniques, 2002. **33**: p. 412-418.
 26. Rogozin, I., F. Kondrashov, and G. Glazko, *Use of mutation spectra analysis software.* Hum Mutat, 2001. **17**(2): p. 83-102.
 27. Rogozin, I.B. and Y.I. Pavlov, *Theoretical analysis of mutation hotspots and their DNA sequence context specificity.* Mutat Res, 2003. **544**(1): p. 65-85.
 28. Miller, J.H., *Mutational specificity in bacteria.* Annu Rev Genet, 1983. **17**: p. 215-38.
 29. Krawczak, M., E.V. Ball, and D.N. Cooper, *Neighboring-nucleotide effects on the rates of germ-line single-base-pair substitution in human genes.* Am J Hum Genet, 1998. **63**(2): p. 474-88.

30. Herniou, E., J. Martin, K. Miller, J. Cook, M. Wilkinson, and M. Tristem, *Retroviral diversity and distribution in vertebrates*. J Virol, 1998. **72**(7): p. 5955-66.
31. Brosius, J., *RNAs from all categories generate retrosequences that may be exapted as novel genes or regulatory elements*. Gene, 1999. **238**(1): p. 115-34.
32. Yoder, J.A., C.P. Walsh, and T.H. Bestor, *Cytosine methylation and the ecology of intragenomic parasites*. Trends Genet, 1997. **13**(8): p. 335-40.
33. Kel, A., O. Kel-Margoulis, V. Babenko, and E. Wingender, *Recognition of NFATp/AP-1 composite elements within genes induced upon the activation of immune cells*. J Mol Biol, 1999. **288**(3): p. 353-76.
34. Klingenhoff, A., K. Frech, K. Quandt, and T. Werner, *Functional promoter modules can be detected by formal models independent of overall nucleotide sequence similarity*. Bioinformatics, 1999. **15**(3): p. 180-6.
35. Frech, K., J. Danescu-Mayer, and T. Werner, *A novel method to develop highly specific models for regulatory units detects a new LTR in GenBank which contains a functional promoter*. J Mol Biol, 1997. **270**(5): p. 674-87.
36. Fickett, J.W., *Coordinate positioning of MEF2 and myogenin binding sites*. Gene, 1996. **172**(1): p. GC19-32.
37. GuhaThakurta, D. and G.D. Stormo, *Identifying target sites for cooperatively binding factors*. Bioinformatics, 2001. **17**(7): p. 608-21.
38. Duret, L. and P. Bucher, *Searching for regulatory elements in human noncoding sequences*. Curr Opin Struct Biol, 1997. **7**(3): p. 399-406.
39. Dermitzakis, E.T. and A.G. Clark, *Evolution of transcription factor binding sites in Mammalian gene regulatory regions: conservation and turnover*. Mol Biol Evol, 2002. **19**(7): p. 1114-21.
40. Chuang, J.H. and H. Li, *Functional bias and spatial organization of genes in mutational hot and cold regions in the human genome*. PLoS Biol, 2004. **2**(2): p. E29.
41. Akey, J.M., G. Zhang, K. Zhang, L. Jin, and M.D. Shriver, *Interrogating a high-density SNP map for signatures of natural selection*. Genome Res, 2002. **12**(12): p. 1805-14.

42. Underhill, P.A., L. Jin, A.A. Lin, S.Q. Mehdi, T. Jenkins, D. Vollrath, R.W. Davis, L.L. Cavalli-Sforza, and P.J. Oefner, *Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography [letter] [see comments]*. *Genome Res*, 1997. 7(10): p. 996-1005.
43. Ewing, B., L. Hillier, M.C. Wendl, and P. Green, *Base-calling of automated sequencer traces using phred. I. Accuracy assessment*. *Genome Res*, 1998. 8(3): p. 175-85.
44. Vasiliev, G.V., V.M. Merkulov, V.F. Kobzev, T.I. Merkulova, M.P. Ponomarenko, and N.A. Kolchanov, *Point mutations within 663-666 bp of intron 6 of the human TDO2 gene, associated with a number of psychiatric disorders, damage the YY-1 transcription factor binding site*. *FEBS Lett*, 1999. 462(1-2): p. 85-8.
45. Boccia, L.M., D. Lillicrap, K. Newcombe, and C.R. Mueller, *Binding of the Ets factor GA-binding protein to an upstream site in the factor IX promoter is a critical event in transactivation*. *Mol Cell Biol*, 1996. 16(5): p. 1929-35.
46. Knight, J.C., I. Udalova, A.V. Hill, B.M. Greenwood, N. Peshu, K. Marsh, and D. Kwiatkowski, *A polymorphism that affects OCT-1 binding to the TNF promoter region is associated with severe malaria*. *Nat Genet*, 1999. 22(2): p. 145-50.
47. Piedrafita, F.J., R.B. Molander, G. Vansant, E.A. Orlova, M. Pfahl, and W.F. Reynolds, *An Alu element in the myeloperoxidase promoter contains a composite SP1-thyroid hormone-retinoic acid response element*. *J Biol Chem*, 1996. 271(24): p. 14412-20.

Table 1. Distribution of pSNPs according to the type of nucleotide substitution.

| | Type of nucleotide substitutions | Number of pSNPs |
|---------------|----------------------------------|-----------------|
| Transitions | G→A | 310 |
| | A→G | 225 |
| | C→T | 355 |
| | T→C | 180 |
| Transversions | C→A | 70 |
| | A→C | 55 |
| | G→T | 75 |
| | T→G | 65 |
| | T→A | 52 |
| | A→T | 55 |
| | C→G | 120 |
| | G→C | 100 |

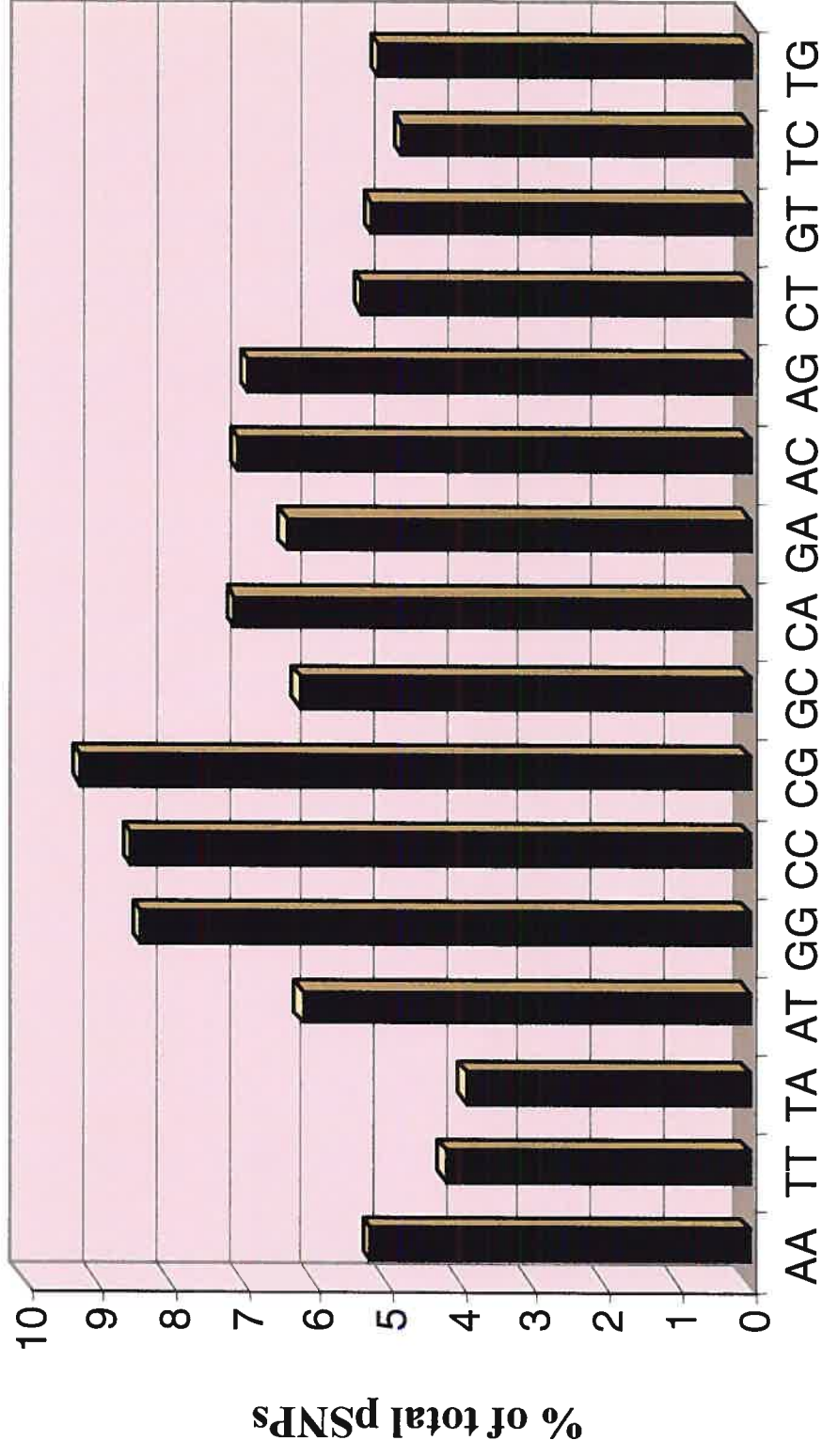
Table 2. Distribution of pSNPs among repetitive elements

| Repeat family | Number of pSNPs | SNP density (SNP/Kb)* |
|----------------|-----------------|-----------------------|
| SINE/Alu | 202 | 4.89 |
| LINE/L1 | 67 | 4.97 |
| SINE/MIR | 52 | 3.59 |
| LTR/ERV1 | 45 | 8.94** |
| LINE/L2 | 41 | 3.78 |
| DNA/MER1 | 31 | 4.57 |
| Simple repeat | 22 | 7.69 |
| LTR/MaLR | 19 | 3.87 |
| LTR/ERV1 | 15 | 18.09** |
| Low complexity | 11 | 3.41 |
| LTR/ERV1 | 11 | 10.38** |
| others | 18 | 6.62 |

Density of SNP/Kb of repeated DNA. ** Significant overrepresentation of pSNPs when compared to their presence in the sequenced regions.

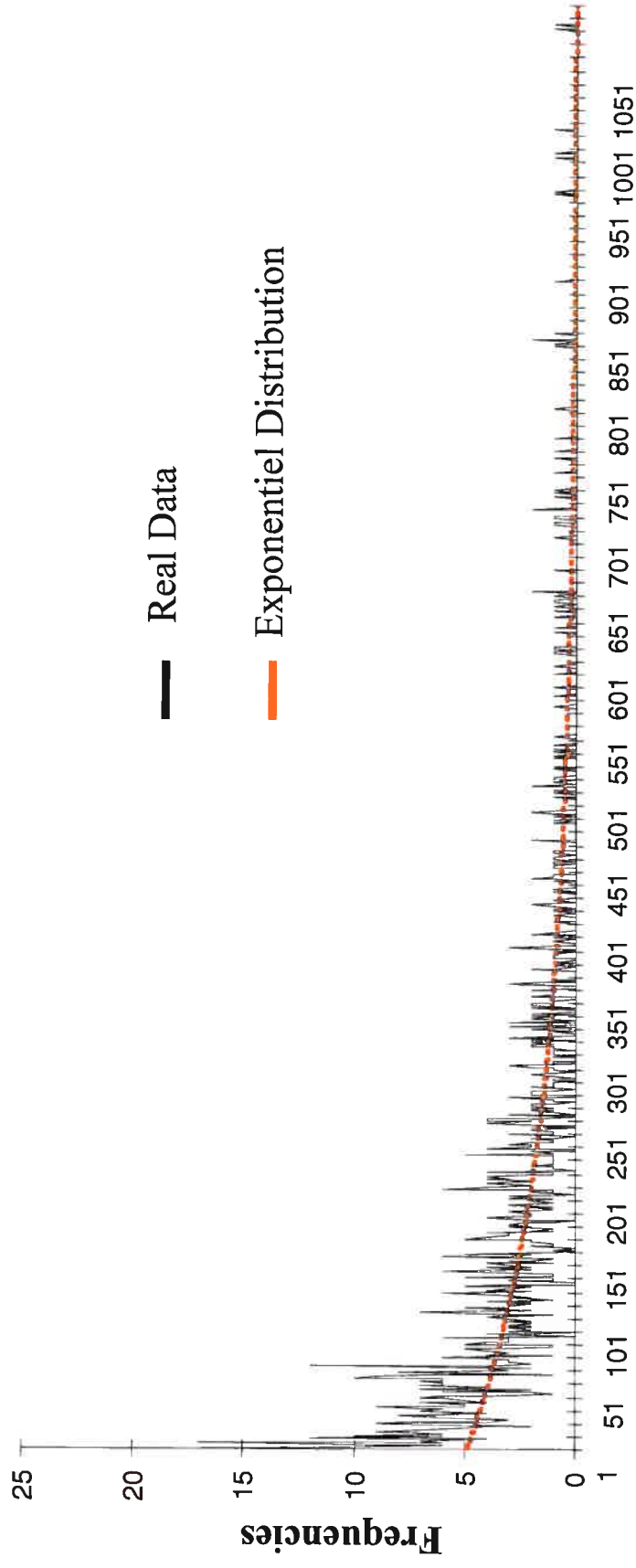
Table 3. List of Gene Ontology groups with significant differences in the promoter genetic diversity.

| Gene Ontology Group | Number of genes | SNP density (SNP/Kb)* |
|--|-----------------|-----------------------|
| Electron transport | 5 | 2.55 |
| Protein serine/threonine kinase activity | 4 | 2.8 |
| Protein amino acid phosphorylation | 10 | 3.18 |
| Anti-apoptosis | 9 | 3.21 |
| Damaged DNA binding | 11 | 3.51 |
| Cellular defense response | 4 | 8.62 |
| Hormone activity | 4 | 7.20 |
| Response to oxidative stress | 6 | 6.71 |
| Cell adhesion | 5 | 6.45 |
| Regulation of transcription from Pol II promoter | 7 | 6.11 |
| G-protein coupled receptor protein signaling pathway | 12 | 6.02 |
| Rhodopsin-like receptor activity | 11 | 5.83 |



Types of substitutions

Figure 1. Distribution of pSNPs among different dinucleotide subtypes.



Distances between pSNPs (base pairs)

Figure 2. Correlation between the prevalence of pSNPs and the physical distance. Since the mutation process follows a Poisson distribution (λ = mutation rate), the intervals between each occurrence (inter-SNPs distance) should follow an exponential distribution ($1 - e^{-\lambda x}$). Here we observe an excess of small inter-SNPs distances.

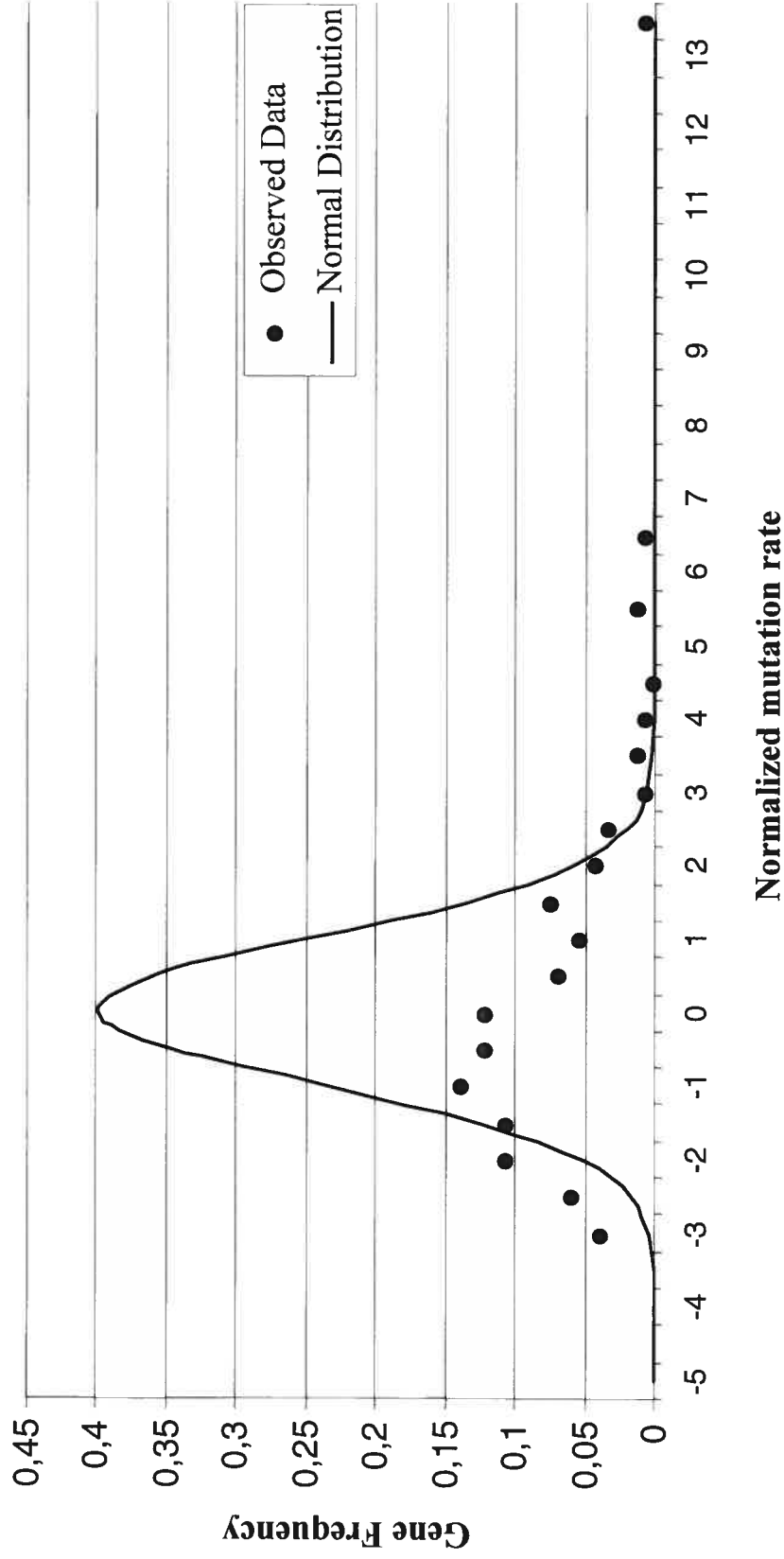


Figure 3. Distribution of normalized mutation rates in the selected genes. The mutation rates have been transformed to follow the normal distribution, but the observed data curve does not follow the expected curve. There is an excess of genes in high and low mutation rates.

Number of SNPs per Gene Group
 Mean (dot); Standard Error (box); 95% Conf Interval (Whisker)

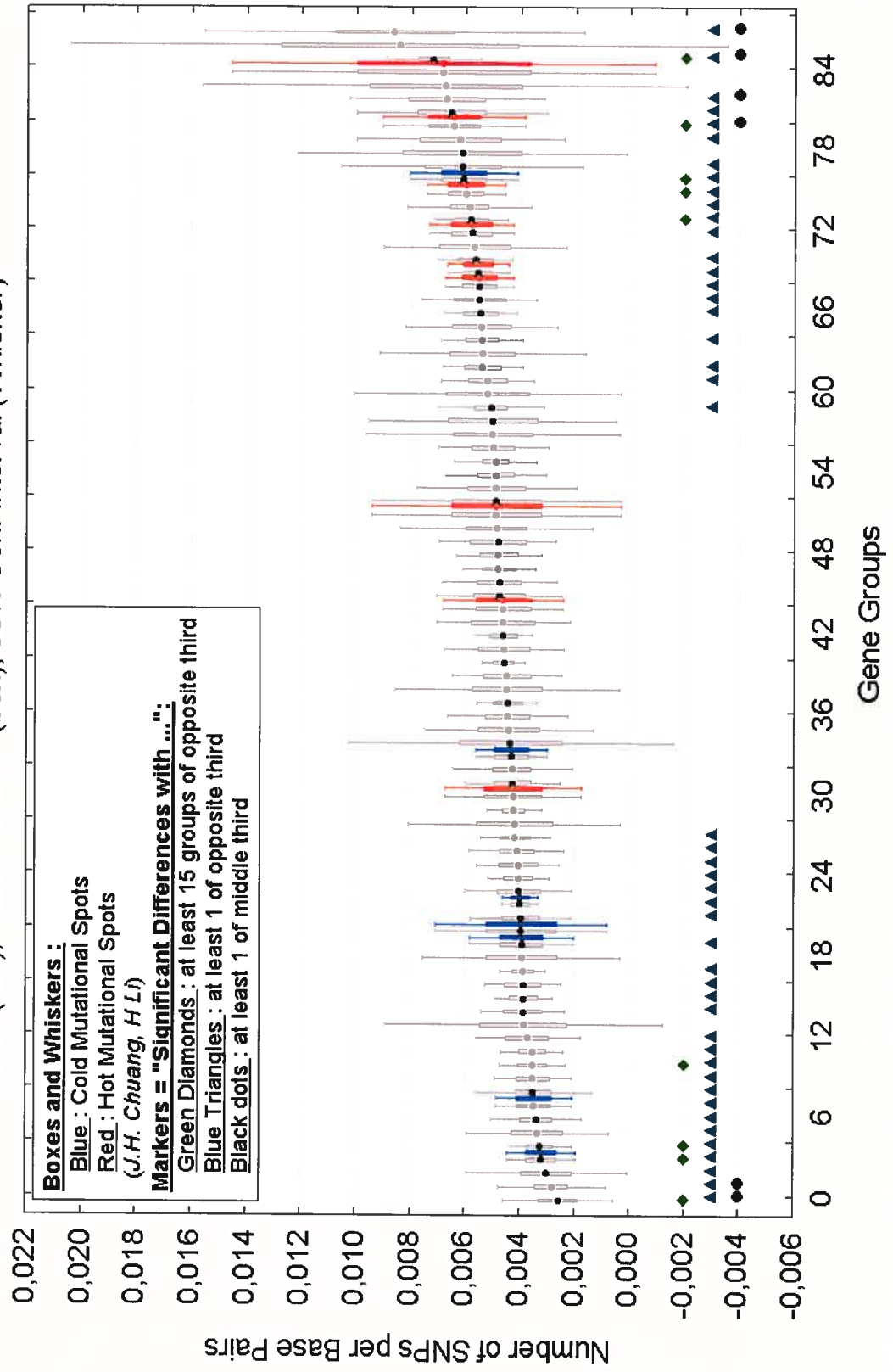


Figure 4. Correlation between the prevalence of pSNPs and the Gene Ontology groups. In this scatter plot, the gene groups associated with a lower mutation rate are depicted at the left, and those with higher mutation rate at the right. Only the categories with at least 5 genes are shown. A single gene may appear in more than one GO categories.

Chapitre 3

-Article 2-

Functional promoter SNPs in cell cycle checkpoint genes.

Article en préparation

Contribution des auteurs :

H. Bélanger : a effectué l'ensemble des expériences, a participé à l'analyse des données et a écrit le premier jet de l'article.

P. Beaulieu : a contribué aux analyses *in silico*.

D. Sinnett : supervision du laboratoire et correction de l'article.

Functional promoter SNPs in cell cycle checkpoint genes.**¹Hélène Bélanger, ¹Patrick Beaulieu and ^{1,2}Daniel Sinnett**

¹Division of Hematology-Oncology, Research Center, Sainte-Justine Hospital (3175 Côte-Ste-Catherine, Montreal, QC, Canada, H3T 1C5); ²Department of Pediatrics, University of Montreal (3175 Côte-Ste-Catherine, Montreal, QC, Canada, H3T 1C5).

Corresponding author: Daniel Sinnett, Ph.D., Division of Hematology-Oncology, Research Center, Sainte-Justine Hospital, 3175 chemin de la Côte-Sainte-Catherine, Montreal, H3T 1C5, Canada. Phone: (514)345-4931 # 2990 Fax: (514) 345-4731. ■

■ [REDACTED]

Abstract

A substantial number of genes mutated in human cancers encode factors that are involved in cell cycle components. It is therefore reasonable to consider that common variants in genes involved in G1/S cell cycle checkpoints are suitable candidates for cancer susceptibility. Since the G1/S transition is a finely regulated biological process we hypothesized that sequence variations in the promoter region of the related genes might indeed lead to abnormal expression, thus predisposing the individuals carrying these genetic variants to cancer. In this report we screened the promoter regions of 16 cell cycle checkpoint genes for DNA variants and assessed the functional impact of these promoter SNPs (pSNPs) by combining in silico analysis and in vitro functional assays. As a result we identified 127 pSNPs including 90 (71%) with predicted impact on putative binding sites of known transcription factors. Out of 11 such pSNPs, selected for electrophoresis mobility shift assays (EMSAs) because of their association with predicted gains of binding sites, 9 pSNPs showed differential allelic shifts in at least one cell line tested. Using the promoter regions cloned into a gene reporter system we found that at least 4 promoter haplotypes in CCND1, E2F1, HDAC1 and RB1 significantly influenced transcriptional activity in an allele-specific manner. Although, the biological significance of these observations still remain to be demonstrated, the expected variability of expression levels in key cell cycle components might influence individual's risk of cancer.

Introduction

The cell has developed a series of checkpoints to ensure the quality control of the cell cycle, particularly the fidelity of replication and accurate chromosomal segregation. The G1/S cell cycle stage represents a critical period for cells to commit to growth arrest or proliferation [1]. Understanding the regulation of the G1/S transition is central to the study of many diseases, particularly cancers [2, 3]. For instance, alterations leading to constitutive protein activity in positive regulators such as cyclin D1, cyclin E, CDK4, CDK6 and inactivation of repressors such as CDKN1A, CDKN1B, CDKN2A, CDKN2B, RB1 can lead to the development of tumours [4-8]. It is therefore reasonable to consider that common variants in G1/S cell cycle genes might explain some of the inter-individual variability in the risk of cancer. Until recently, the search for functional polymorphisms has predominantly been focused on single nucleotide polymorphisms in the coding region of genes (cSNPs) because of their putative direct effect on the protein structure and function. Polymorphism in non-coding DNA, particularly regulatory regions has been studied less intensively, despite the suggestion that most relevant mutations reside in regulating regions [9]. Since protein levels regulate many biological pathways, regulatory polymorphisms can indeed influence biological processes [10] and modify disease risk as suggested by several studies [11-20]. Here we propose that polymorphisms in promoter regions of G1/S cell cycle checkpoint genes could modify transcription factor binding sites (TFBSs) and thus give rise to differences in expression. To address this hypothesis we have combined *in silico* and *in vitro* approaches to detect and to validate the promoter SNPs (pSNPs) found in the regulatory region, arbitrary defined as a 2kb segment upstream exon 1, of 16 G1/S genes. We have found 127 pSNPs including 90 with predicted impact on putative TFBSs. Although we are aware that systematic effort should be made to identify and to annotate these pSNPs since our knowledge about regulatory elements is still limited [21]. We thus decided to focus on those leading to putative gains of binding sites that cannot be further investigated with phylogenetic tools. Using gene reporter and electrophoretic mobility shift assays we found that 4 such pSNPs/haplotypes associated with the genes E2F1, HDAC1, RB1 and CCND1 influenced transcriptional activity in an allele-specific manner.

Material and methods

SNP detection in promoter regions. Detection of DNA variants in a population panel consisting of 40 unrelated individuals (8 Africans, 8 Europeans, 8 Asians, 8 Middle-easterners and 8 Amerindians) was performed by PCR-based dHPLC analysis (Transgenomic WAVE Nuclei Acid Fragment Analysis System, Omaha, USA) followed by direct sequencing as described in Sinnett et al. (submitted). The promoter region arbitrary defined as the 2kb sequence upstream of the transcription start site (based on RefSeq mRNA) of each gene was amplified sequentially in 4 to 8 ~350 bp overlapping fragments. The heterozygous samples were sequenced to confirm the nature and position of the DNA variants. Promoter positions were numbered with respect to the first nucleotide of the first exon as +1, and the nucleotide immediately upstream as -1. The primer sequences and detailed information on the reaction conditions are available upon request (see annexes).

Electrophoretic mobility shift assay (EMSA). DNA-protein interactions were assessed by incubating 5'end radiolabeled double-stranded oligonucleotide probes corresponding to each allelic variants (Top strand sequences for the binding sites are listed in Table 1) in the presence of HeLa, JEG-3, and HepG2 cell nuclear extracts [22] according to the manufacturer's protocol (Promega gel shift assay system). Protein was quantified with the Bradford protein assay (BioRad). Specifically, nuclear extracts (5ug) were incubated with 35fmol radiolabeled double-stranded DNA probes and a buffer containing 50mM Tris-HCl (pH 7.5), 5mM MgCl₂, 2.5mM EDTA, 2.5mM DTT, 250mM NaCl, 0.25 ug/ul poly (deoxyinosinate-deoxycytidylate), 20% glycerol in a total volume of 10ul for 20 min at room temperature. Complexes were separated in a non-denaturing polyacrylamide gel (6%, acrylamide:bisacrylamide, 60:1) in 0.5X Tris-borate-EDTA buffer (190V at 4°C). For competition EMSA, 50-fold molar excess competitors (excess of unlabeled probe oligonucleotide or corresponding mutant oligo) were included as described in the Fig. 1

Gene reporter assay. *Constructs.* Allele/Haplotype-specific fragments corresponding to ~2.0 kb of the proximal promoter region were amplified from genomic DNA of known heterozygotes and cloned into the promoterless pGL3-basic firefly luciferase reporter vector (Promega). Constructs corresponding to the major haplotypes were sequenced to confirm the presence of the expected polymorphic sites and then purified on QIAquick PCR purification columns prior transfection.

Transfection. Approximately $3-4 \times 10^4$ cells (JEG-3, HeLa and HepG2) were plated-out and grown in 96-well plates (30mm^2) to reach 80 to 90% confluent at the time of transfection with lipofectamine according to the manufacturer's protocol (Invitrogen). These cell lines were co-transfected with 100ng of each allelic construct and 0.5 ng of SV40-driven (ratio 200:1) *renilla* luciferase pRL-SV40 plasmid to control for transfection efficiency. Similar experiments were performed with a negative control consisting of the empty promoterless pGL3-basic plasmid (Promega). Transfected cells were harvested 48h following transfection and luciferase reporter gene activity was measured with the Dual-Luciferase Reporter Assay System according to the manufacturer's instructions (Promega), in a Spectra Max 190 luminometer (Molecular Devices). The *renilla* luciferase activity of the control pRL-SV40 was used to normalize the results of the firefly luciferase activity of the allelic constructs. After background correction (subtraction of the activity in untreated cells), the results were expressed as the ratio of firefly luciferase activity divided by the pRL-CMV internal control pRL-SV40 activity, and expressed as relative luciferase (means \pm SD) of 5 replicates. Statistical significance (p-value) was determined using Student's unpaired t-test.

Results

A panel of 40 unrelated individuals from different ethnicity was screened by dHPLC to detect the presence of variants in the proximal promoter region that might alter the expression in 16 G1/S checkpoint genes: CCND1, CCNE1, CDC25A, CDK4, CDK6, CDKN1A, CDKN1B, CDKN2A, CDKN2B, E2F1, HDAC1, MADH3, MADH4, RB1, SKP2, TFDP1. The targeted promoter region was arbitrary defined as the 2Kb sequence

upstream of the transcriptional initiation site, because most of the validated regulatory SNPs were found within this interval [23]. Amplimers were designed to direct amplification of 5 to 8 overlapping PCR fragments in order to cover the entire 2kb genomic segment. Despite high GC-content, only 9% of fragments failed to amplify indicating the absence of significant bias due to PCR. We identified 127 promoter SNPs (pSNPs), including 104 (82%) that were not present in public databases. These pSNPs were equally distributed throughout the 2kb region with an average density of 4.4 pSNPs/kb. Of all DNA variants identified, 17 (13.4%) were length polymorphisms, 65 (51.2%) were transition and 45 (35.4%) were transversion, thus giving a ratio of transition on transversion of 1.4, which is similar to published figures [24].

Depending on the position in the regulatory region, a pSNP may lead either to loss of a TF site [25, 26] or the formation of a novel TFBS [27, 28]. *In silico* analysis performed with the program matInspector (transfac.gbf.de/TRANSFAC/index.html) showed that 71% (90/127) of the pSNPs found in this study had a predicted impact on transcription factor binding sites: 30 (23.6%) and 24 (18.9%) were associated with a gain or a loss of a binding site, respectively, whereas 36 (28.3%) variants were expected to lead to a combination of loss/gain of binding sites (see annexes). We decided to further analyse polymorphisms in which the major allele had no known transcription factor binding sites whereas the minor allele containing the discovered pSNP had a predicted gain of a binding site. Consequently, a panel of 11 pSNPs/haplotypes (Table 1) were validated with functional *in vitro* assays.

We assessed by EMSA the differential ability of the predicted TF to bind the polymorphic DNA sequences. In these assays, double-stranded oligonucleotide (ds-oligo) probes corresponding to the sequence surrounding the polymorphic sites (see Table 1) were radiolabelled and allowed to interact with nuclear extracts prepared from HepG2, Jeg-3 and HeLa cells. Representative data for CCND1 are shown in Fig. 1: Using ds-oligo probes spanning each form of the -568A<C pSNPs we observed a slow migrating DNA-protein complex when the allele -568C probe was incubated in presence of HeLa nuclear extracts (Fig. 1, lane 7), whereas no complexes were observed with the variant -568A (Fig. 1, lane 5). To confirm specificity of binding to the -568C cis

element, competition EMSAs were performed. The -568C related DNA-protein complex was indeed competed in presence of 50-fold excess of the corresponding unlabelled -568C ds-oligo probe (Fig. 1, lane 8). The nuclear protein bound to the -568C probe in HeLa cells thus represents a specific DNA-protein interaction. Similar results were obtained with Jeg-3 and HepG2 nuclear extracts (Table 2). The other pSNPs were processed similarly and we found that 9 of them showed differential allelic shifts in at least one cell line tested (Table 2).

To assess the functional significance of these 11 pSNPs on promoter activity, the corresponding haplotypes containing ~2.0 kb of regulatory sequences were subcloned into the promoterless pGL3 vector. Then transient transfection experiments were carried out for each of the resulting 20 allelic constructs (See Table 3). Assuming a conservative definition of promoter activity as ~ 10-fold increase activity of the “highly expressed allele” over the expression of the pGL3-basic vector [29] then all constructs showed promoter activity in at least one of the cell line tested (Table 4). To determine whether significant differences were observed between the expression levels of each allelic construct, we took into account the ratio of the highest allele over the lowest allele. Significant 1.5 fold differences ($p \leq 0.05$) in at least one cell line system combined with EMSA validation were considered as criteria for functional pSNP, at least in an in vitro environment. Histone deacetylase 1 (HDAC1) met these criteria: luciferase gene expression controlled by HDAC1 -905A>C promoter variants showed differential expression when transfected into HepG2, HeLa, and JEG-3 cells (representative data in Fig. 2). In this example, the level of luciferase driven by the -905A HDAC1 constructs (pHDAC1-905A) was 2-fold higher than the luciferase levels produced by the -905C allelic counterpart (Fig. 2, Table 4). Combined with differential DNA-protein binding detected in HeLa nuclear extract (Table 2), this support the functional significance of this pSNP/haplotype.

Similarly, we have found putative functional SNPs/haplotypes in the E2F1, RB1, and CCND1 genes (Table 4). In E2F transcription factor 1 (E2F1), the -897 T variant significantly decreased the promoter activity compared to the C allele at least in HeLa cells. Specific differential binding by the predicted cut-like homeodomain protein (gain

of binding in the T allele) was observed in HeLa and HepG2 nuclear extracts (Tables 2 and 4). For Retinoblastoma 1 (RB1), the insertion CTGATA at position -1733 (located on haplotype D) showed significant difference in promoter activity in HepG2 cells. This pSNP led to a predicted gain of binding sites for at least three transcription factors (Table 1) that was associated with differential binding between alleles for all cell lines investigated (Table 2). Finally, we observed in Cyclin D1 (CCND1) significant differences in reporter gene levels between -568A>C derived haplotypes in HeLa and Jeg-3 (Table 4), which is supported by the observation of differential binding in all 3 cell lines (Table 2). Statistically significant differences were observed in other promoter allelic constructs, but did not meet our 1.5-fold-threshold criteria. Although, these allelic differences could have an important biological relevance they will not be further discussed in this report.

Discussion

Because cell cycle is a finely regulated process that depends on combinations of cell cycle activators and inhibitors [30] it is plausible that changes in protein levels of key components might indeed have an important functional impact. Thus differential allelic expression of the corresponding genes might lead to inter-individual variability in cell cycle function. In this report we wanted to detect variants in the proximal promoter regions of 16 G1/S checkpoint genes because of their putative impact on the protein product levels. We found 127 promoter SNPs (pSNPs) for a density of 4.4 SNPs/kb, which is similar to the sequence diversity reported for other genes [31]. Since, there is no accurate way to predict the effect of these putative regulatory polymorphisms, we must rely on the combination of both *in silico* and experimental approaches to investigate their functional impact [32-35]. The presence of such functional promoter variants has been observed in 10-18% of some of the genes tested [29, 36, 37] but not all of them [38]. The scope of this report was directed toward the 11 pSNPs leading to predicted gain of binding sites because phylogenetic footprinting analysis cannot be applied in the selection process [as suggested in 39], which probably explains the lack of

validated data for this type of pSNPs. Here we provided evidence that promoter SNPs/haplotypes led to significant differential promoter activity of the genes E2F1, HDAC1, RB1 and CCND1. This is an interesting observation considering that deregulation in the G1/S machinery promotes genomic instability and can give rise to increase individual susceptibility to develop cancer and other complex diseases [e.g. 40].

E2F factors are involved in the control of multiple biological processes including cell cycle, replication, DNA synthesis and DNA repair [41]. In particular, E2F1 not only plays a role as a growth-promoting factor but also promotes apoptosis depending on the cell type and the Rb status [42]. E2F1 has been shown to be amplified in a human erythroleukemia cells [43] as well as up-regulated in thyroid carcinogenesis [44], whereas the loss of E2F1 significantly reduces the frequency of hyperplastic lesions in the thyroid glands of Rb +/- mice [45]. Given the fact that E2F1 activity is often deregulated in human tumors, one might suggest that the functional C to T substitution at position -897 C<T might indeed modify the risk of cancer among carriers. Of note, this pSNP leads to the gain of a predicted site for the binding of the cut-like homeodomain protein that has been shown to act as negative regulators of gene expression [46, 47].

HDAC1 is part of the e2F-Rb repression complex that is critical in evoking growth arrest and can also be recruited to specific genomic region by a variety of factors to mediate the repression of corresponding genes [48]. Furthermore, overexpression of HDAC1 has been shown in hormone refractory prostate cancer [49]. Therefore it is reasonable to propose that the C-905A derived allele that lead to the creation of predicted binding sites for 3 different transcription factors (C-Abl, SRY and HNF-3/fkh homolog 2) and showed significant differences in promoter activity might modify ones risk of developing certain types of cancers.

CCND1 is a regulator of cell cycle that is able to drive cells through the restriction point in the G1 phase where these cells will be committed to divide [50]. Overexpression of this gene occurs in many types of cancer including breast and colorectal cancers and it has been associated with an increase in cell proliferation [51, 52]. In this context, the

functional haplotype B carrying the pSNP $-568A < C$ could be involved in the susceptibility to cancer.

Mutations that disrupt the Rb pathway leading to inappropriate proliferation and contributes to mitotic alterations and chromosome instability [53] are common in human cancers [54]. Nonsense or frameshift mutations are the most frequent alterations followed by point mutations in intronic or exonic sequences that cause aberrant splicing [55]. Of note, two distinct point mutations in the core promoter region of RB1 gene at an ATF site and a retinoblastoma binding factor 1 (RBF-1) site caused a large reduction of the promoter activity [56]. Therefore the allelic insertion of CTGATA at position -1733 (haplotype D), associated with an increased promoter activity and leading to a predicted the gain of 3 putative transcription binding sites: ecotropic viral integration site 1 factor (EVI-1), GATA-1, and RBP-Jkappa/CBF1 might indeed play a role in cancer susceptibility. EVI-1 works as a transcriptional activator [57] although it has been shown the repress expression in some cases [58]. GATA-1 is considered a transcriptional activator of erythroid-specific transcripts but recent studies demonstrated that GATA-1 can also act as a repressor and is likely to play a major role in hematopoiesis [59-61]. Finally, the mammalian transcriptional repressor RBP-Jkappa/CBF1, has been recognized to be a transcription factor that represses mammalian gene expression [62, 63].

We also need to consider the scenario that distinct functional pSNPs/haplotypes might have a synergistic interaction because of the interconnected nature of cell cycle regulation. For instance, members of the pRB family bind to E2F transcription factors causing them to repress transcription of the genes required for cell cycle progression. This transcriptional inhibition is also done through the recruitment, via the Rb proteins, of histone deacetylases (HDAC) thus leading to chromatin compaction and decreased access to transcription factory.

In conclusion, we showed that sequence variants in the proximal promoter regions of cell cycle genes could result in abnormal transcriptional regulation. However, with the data currently available it is difficult to estimate the number of these functional pSNPs that are actually associated with human diseases.

Acknowledgements

This study was supported by research funds provided by Genome Quebec and Genome Canada. T.J.H. is supported by a Clinician-Scientist Award in Translational Research by the Burroughs Wellcome Fund and an Investigator Award from the Canadian Institutes of Health Research. DS holds the François-Karl Viau Chair in Pediatric Oncogenomics and is a scholar of the Fonds de la Recherche en Santé du Québec (FRSQ).

References

1. Sherr, C.J., *Growth factor-regulated G1 cyclins*. Stem Cells, 1994. **12**: p. 47-55.
2. Sherr, C.J., *The pezcoller lecture: cancer cell cycles revisited*. Cancer Res, 2000. **60**: p. 3689-3695.
3. Tessema, M., U. Lehmann, and H. Kreipe, *Cell Cycle and no end*. Virchows Arch, 2004. **444**: p. 313-323.
4. Steeg, P. and Q. Zhou, *Cyclins and breast cancer*. Breast Cancer Res Treat, 1998. **52**: p. 17-28.
5. Schraml, P., et al., *Cyclin E overexpression and amplification in human tumors*. J Pathol, 2003. **200**: p. 375-382.
6. Lee, M.H. and H.Y. Yang, *Negative regulators of cyclin-dependant kinases and their roles in cancers*. Cell Mol Life Sci, 2001. **58**: p. 1907-1922.
7. Wong, I.H., et al., *Aberrant p15 promoter methylation in adult and childhood acute leukemias of nearly all morphologic subtypes: potential prognostic implications*. Blood, 2000. **95**: p. 1942-1949.
8. Lloyd, R.V., et al., *p27kip1: a multifunctional cyclin-dependant kinase inhibitor with prognostic significance in human cancers*. Am J Pathol, 1999. **154**: p. 313-323.
9. King, M.C. and A.C. Wilson, *Evolution at two levels in humans and chimpanzees*. Science, 1975. **188**(4184): p. 107-16.
10. Pastinen, T., et al., *A survey of genetic and epigenetic variation affecting human gene expression*. Physiol Genomics, 2004. **16**(2): p. 184-93.
11. Yasuda, T., et al., *The molecular basis for genetic polymorphism of human deoxyribonuclease II (DNase II): a single nucleotide substitution in the promoter region of human DNase II changes the promoter activity*. FEBS Lett, 2000. **467**(2-3): p. 231-4.
12. Li, L.C., et al., *A single nucleotide polymorphism in the E-cadherin gene promoter alters transcriptional activities*. Cancer Res, 2000. **60**(4): p. 873-6.

13. Smeraldi, E., et al., *Polymorphism within the promoter of the serotonin transporter gene and antidepressant efficacy of fluvoxamine*. Mol Psychiatry, 1998. **3**(6): p. 508-11.
14. Sanak, M., et al., *Enhanced expression of the leukotriene C(4) synthase due to overactive transcription of an allelic variant associated with aspirin-intolerant asthma*. Am J Respir Cell Mol Biol, 2000. **23**(3): p. 290-6.
15. Manuck, S.B., et al., *A regulatory polymorphism of the monoamine oxidase-A gene may be associated with variability in aggression, impulsivity, and central nervous system serotonergic responsivity*. Psychiatry Res, 2000. **95**(1): p. 9-23.
16. Pihlajamaki, J., et al., *G-250A substitution in promoter of hepatic lipase gene is associated with dyslipidemia and insulin resistance in healthy control subjects and in members of families with familial combined hyperlipidemia*. Arterioscler Thromb Vasc Biol, 2000. **20**(7): p. 1789-95.
17. Ishigami, T., et al., *Angiotensinogen gene polymorphism near transcription start site and blood pressure: role of a T-to-C transition at intron I*. Hypertension, 1999. **34**(3): p. 430-4.
18. Lopez, E.R., et al., *A promoter polymorphism of the CYP27B1 gene is associated with Addison's disease, Hashimoto's thyroiditis, Grave's disease and type 1 diabetes mellitus in Germans*. Eur J Endocrinol, 2004. **151**(2): p. 193-197.
19. Ho, S.Y., et al., *Increased risk of developing hepatocellular carcinoma associated with carriage of the TNF2 allele of the -308 tumor necrosis factor-alpha promoter gene*. Cancer Causes Control, 2004. **15**(7): p. 657-663.
20. Pola, R., et al., *The -174 G/C polymorphism of the interleukin-6 gene promoter is associated with Alzheimer's disease in Italian population*. Neuroreport, 2002. **13**(13): p. 1645-1647.
21. Pastinen, T. and T.J. Hudson, *Cis-acting regulatory variation in the human genome*. Science, 2004. **306**(5696): p. 647-650.
22. Schreiber, E., et al., *Rapid detection of octamer binding proteins with 'mini-extracts', prepared from a small number of cells*. Nucleic Acids Res., 1989. **17**(15): p. 6419.

23. Rockman, M.V. and G.A. Wray, *Abundant raw material for cis-regulatory evolution in humans*. Mol Biol Evol, 2002. **19**(11): p. 1991-2004.
24. Collins, D.W. and T.H. Jukes, *Rates of transition and transversion in coding sequences since the human-rodent divergence*. Genomics, 1994. **20**: p. 286-396.
25. Vasiliev, G.V., et al., *Point mutations within 663-666 bp of intron 6 of the human TDO2 gene, associated with a number of psychiatric disorders, damage the YY-1 transcription factor binding site*. FEBS Lett, 1999. **462**(1-2): p. 85-8.
26. Boccia, L.M., et al., *Binding of the Ets factor GA-binding protein to an upstream site in the factor IX promoter is a critical event in transactivation*. Mol Cell Biol, 1996. **16**(5): p. 1929-35.
27. Knight, J.C., et al., *A polymorphism that affects OCT-1 binding to the TNF promoter region is associated with severe malaria*. Nat Genet, 1999. **22**(2): p. 145-50.
28. Piedrafita, F.J., et al., *An Alu element in the myeloperoxidase promoter contains a composite SPI-1-thyroid hormone-retinoic acid response element*. J Biol Chem, 1996. **271**(24): p. 14412-20.
29. Coleman, S.L., et al., *A streamlined approach to functional analysis of promoter polymorphisms*. Biotechniques, 2002. **33**: p. 412-418.
30. Gonze, D. and A. Goldbeter, *A model for a network of phosphorylation-dephosphorylation cycles displaying the dynamics of dominoes and clocks*. J Theor Biol, 2001. **210**(2): p. 167-186.
31. Cargill, M., et al., *Characterization of single-nucleotide polymorphisms in coding regions of human genes*. Nat Genet, 1999. **22**(3): p. 231-8.
32. Peltonen, L. and V.A. McKusick, *Genomics and medicine. Dissecting human disease in the postgenomic era*. Science, 2001. **291**(5507): p. 1224-1229.
33. Knight, J.C., *Functional implications of genetic variation in non-coding DNA for disease susceptibility and gene regulation*. Clin Sci (Lond), 2003. **104**(5): p. 493-501.
34. Knight, J.C., et al., *In vivo characterization of regulatory polymorphisms by allele-specific quantification of RNA polymerase loading*. Nat Genet, 2003. **33**(4): p. 469-475.

35. Hudson, T.J., *Wanted: regulatory SNPs*. Nat Genet, 2003. **33**(4): p. 439-440.
36. Hoogendoorn, B., et al., *Functional analysis of human promoter polymorphisms*. Hum Mol Genet, 2003. **12**: p. 2249-2254.
37. Buckland, P., et al., *A high proportion of chromosome 21 promoter polymorphisms influence transcriptional activity*. Gene Expr, 2004. **11**(5-6): p. 233-239.
38. Smith, S., et al., *Lack of functional promoter polymorphisms in genes involved in glutamate neurotransmission*. Psychiatry Genetics, 2003. **13**: p. 193-199.
39. Mottagui-Tabar, S., et al., *Identification of functional SNPs in the 5-prime flanking sequences of human genes*. BMC Genet, 2005. **6**(1): p. 18-27.
40. Park, M.T. and S.J. Lee, *Cell cycle and cancer*. J Biochem Mol Biol, 2003. **36**(1): p. 60-65.
41. Ren, B., et al., *E2F integrates cell cycle progression with DNA repair, replication, and G(2)/M checkpoints*. Genes Dev, 2002. **16**(2): p. 245-256.
42. DeGregori, J., et al., *Distinct for E2F proteins in cell growth control and apoptosis*. Proc Natl Acad Sci U S A, 1997. **94**(14): p. 7245-7250.
43. Saito, M., et al., *Amplification of the E2F1 transcription factor gene in the HEL erythroleukemia cell line*. Genomics, 1995. **25**(1): p. 130-138.
44. Onda, M., et al., *Up-regulation of transcriptional factor E2F1 in papillary and anaplastic thyroid cancers*. J Hum Genet, 2004. **49**(6): p. 312-318.
45. Yamasaki, L., et al., *Loss of E2F-1 reduces tumorigenesis and extends the lifespan of Rb1(+/-) mice*. Nat Genet, 1998. **18**(4): p. 360-364.
46. Andres, V., B. Nadal-Ginard, and V. Mahdavi, *Clox, a mammalian homeobox gene related to Drosophila cut, encodes DNA-binding regulatory proteins differentially expressed during development*. Development, 1992. **116**(2): p. 321-334.
47. Dufort, D. and A. Nepveu, *The human cut homeodomain protein represses transcription from the c-myc promoter*. Mol Cell Biol, 1994. **14**(6): p. 4251-4257.
48. Cress, W.D. and E. Seto, *Histone deacetylases, transcriptional control, and cancer*. J Cell Physiol, 2000. **184**(1): p. 1-16.
49. Halkidou, K., et al., *Up-regulation and nuclear recruitment of HDAC1 in hormone refractory prostate cancer*. Prostate, 2004. **59**(2): p. 177-189.
50. Sherr, C.J., *Cancer cell cycles*. Science, 1996. **274**(5293): p. 1672-1677.

51. Holley, S.L., et al., *Cyclin D1 polymorphism and expression in patients with squamous cell carcinoma of the head and neck*. Am J Pathol, 2001. **159**(5): p. 1917-24.
52. McKay, J.A., et al., *Cyclin D1 protein expression and gene polymorphism in colorectal cancer*. Aberdeen Colorectal Initiative. Int J Cancer, 2000. **88**(1): p. 77-81.
53. Hernando, E., et al., *Rb inactivation promotes genomic instability by uncoupling cell cycle progression from mitotic control*. Nature, 2004. **430**(7001): p. 797-802.
54. Hanahan, D. and R.A. Weinberg, *The hallmarks of cancers*. Cell, 2000. **100**: p. 57-70.
55. Lohmann, D.R. and B.L. Gallie, *Retinoblastoma: Revisiting the model prototype of inherited cancer*. Am J Med Genet, 2004. **129C**(1): p. 23-28.
56. Sakai, T., et al., *Oncogenic germ-line mutations in Sp1 and ATF sites in the human retinoblastoma gene*. Nature, 1991. **353**(6339): p. 83-86.
57. Morishita, K., et al., *EVI-1 zinc finger protein works as a transcriptional activator via binding to a consensus sequence of GACAAGATAAGATAAN1-28 CTCACTTC*. Oncogene, 1995. **10**(10): p. 1961-1967.
58. Bartholomew, C., et al., *The Evi-1 proto-oncogene encodes a transcriptional repressor activity associated with transformation*. Oncogene, 1997. **14**(5): p. 569-577.
59. Welch, J.J., et al., *Global regulation of erythroid gene expression by transcription factor GATA-1*. Blood, 2004. **104**(10): p. 3136-3147.
60. Grass, J.A., et al., *GATA-1-dependent transcriptional repression of GATA-2 via disruption of positive autoregulation and domain-wide chromatin remodeling*. Proc Natl Acad Sci U S A, 2003. **100**(15): p. 8811-8816.
61. Rylski, M., et al., *GATA-1-mediated proliferation arrest during erythroid maturation*. Mol Cell Biol, 2003. **23**(14): p. 5031-5042.
62. Dou, S., et al., *The recombination signal sequence-binding protein RBP-2N functions as a transcriptional repressor*. Mol Cell Biol, 1994. **14**: p. 3310-3319.

63. Kannabiran, C., X. Zeng, and L.D. Vales, *The mammalian transcriptional repressor RBP (CBF1) regulates interleukin-6 gene expression*. Mol Cell Biol, 1997. 17: p. 1-9.

Table 1 : List of the promoter SNPs associated with predicted gains of transcription factor binding sites and the corresponding oligonucleotides used in EMSAs.

| Gene | Selected pSNPs | Predicted gain in the minor allele* (Score) | Oligonucleotide used in EMSA |
|--------|----------------|--|---------------------------------------|
| CCND1 | -568A>C | Tumor suppressor p53 (3 half site) (>0.943) | AATGAAACTTGC(A/C)CAGGGGTTGTGTGC |
| CDC25A | -351G>A | MyT1 zinc finger transcription factor (>0.82) | GCCGTCCAGTTA(G/A)CTTCGGGCC |
| | -1440G>A | Metal transcription factor 1, MRE (>0.911) | AACAGGCCGGGAGC(G/A)GTGGCT |
| CDKN1A | -899T>G | v-Myb (>0.925) | TGGGAAAAC(T/G)GGGGCTC |
| | -1536C>T | PPAR/RXR heterodimers (>0.718) | AGGCTGTGC(C/T)GTGGCCTTCTGGGGTT |
| CDKN1B | -1608G>A | POZ/zinc finger protein (>0.842) | GGTTTCTGTCCAGG(G/A)ACATGCA |
| E2F1 | -897C>T | cut-like homeodomain protein (>0.949) | TGGTGAAAG(C/T)CCATCTCTACAAA |
| HDAC1 | -905A>C | HNF-3/Fkh Homolog 2 (FOX D3) (>0.938) | CAAAAAAAAAAAA(A/C)AAAAAAAAAACAGG |
| | | Multifunctional c-Abl src type tyrosine kinase (>0.997) | |
| | | Sex-determining region Y gene product (>0.951) | |
| | -1554C>A | c-Myb, cellular equivalent to avian myoblastosis virus oncogene v-myb (>0.983) | GTTAGGTAGGA(C/A)TGGCTGT |
| RB1 | -1733INS | TEF-1 related muscle factor (>0.854) | |
| | CTGATA | Ecotropic viral integration site 1 encoded factor (>0.842) | TTATCAGGCTCCCAGCA(-/CTGATA)GACTACTCC |
| | | GATA-binding factor 1 (>0.993) | |
| TFDP1 | -1394T>C | Mammalian RBP-Jkappa/CBF1 (>0.846) | CACACACCTGCTCTCCTCCGTG(T/C)GTCCACATAC |
| | | PAX9 | |

* Predicted impact of the pSNPs with the program matInspector. Only pSNPs leading to a predicted gain in the minor allele without having any prediction for the major allele were considered for further analyzed. Only the top strand of the double-stranded oligonucleotides centered on the predicted binding sites are shown.

Table 2: Summary of the electrophoretic mobility shift assay results for the chosen pSNP.

| Gene | Selected pSNP [†] | Validated differential DNA binding | | |
|--------|----------------------------|------------------------------------|-------|-------|
| | | HeLa | Jeg-3 | HepG2 |
| CCND1 | -568A>C | Yes | Yes | Yes |
| CDC25A | -351 G>A | No | No | No |
| | -1440G>A | Yes | No | Yes |
| CDKN1A | -899T>G | Yes | Yes | No |
| | -1536C>T | No | Yes | No |
| CDKN1B | -1608G>A | No | No | No |
| E2F1 | -897C>T | Yes | No | Yes |
| HDAC1 | -905A>C | Yes | No | No |
| RB1 | -1554C>A | No | Yes | No |
| | -1733INSCTGATA | Yes | Yes | Yes |
| TFDP1 | -1394T>C | Yes | Yes | Yes |

[†] SNP showing a gain of transcription factor binding site using *in silico* predictions (See Table 1).

Table 3: Promoter haplotypes studied in gene reporter assays

| Gene | haplotype pSNPs tags | Position relative to start site | Position relative to the Genomic DNA (GI number) |
|--------|--|---------------------------------|--|
| CCND1 | A: -568A, -766A, -1139C, -1597G, -1938T B: -568C, -766T, -1139T, -1597A, -1938C | +179 to -2375 | 218-3180 (AP001888.4) |
| CDC25A | A: -351G, -1440G, -2030C B: -351A, -1440G, -2030G C: -351G, -1440A, -2030G | +48 to -2051 | 77401-79499 (AC104190.2) |
| CDKN1A | A: -134C, -499G, -519G, -791C, -899T, -1021T, -1284C, -1466G, -1529C, -1536T B: -134G, -499C, -519C, -791C, -899G, -1021C, -1284C, -1466A, -1529T, -1536C C: -134G, -499G, -519G, -791T, -899T, -1021C, -1284T, -1466A, -1529T, -1536C | +19 to -2071 | 1138-3227 (Z85996.1) |
| CDKN1B | A: -373T, -756A, -1608G B: -373A, -756C, -1608A | +70 to -2023 | 10482-12574 (AC008115.3) |
| E2F1 | A: -897C B: -897T | -3 to -2008 | 62827-64832 (AL121906.18) |
| HDAC1 | A: -905A B: -905C | +50 to -2060 | 6104-8213 (AL109945.15) |
| RB1 | A: -768C, -1554C, -1729C, -1733- B: -768C, -1554A, -1729C, -1733- C: -768A, -1554C, -1729C, -1733- D: -768C, -1554C, -1729DEL, -1733INSCTGATA | +22 to -2022 | 2979-5022 (NT_009799.8) |
| TFDP1 | A: -1394T, -1655G, -1883G, -1887C B: -1394C, -1655A, -1883C, -1887T | +987 to -2547 | 117905-121587 (AL442125.13) |

*Only the polymorphic positions (SNP tags) in the selected promoter haplotypes are shown.

Table 4: Results of the gene reporter assays for the promoter haplotype tested.

| Gene | Haplo- type studied | SNP of interest ¹ | Ratio of highest allele activity on pGL3-basic | | | Ratio of highest allele activity on lowest allele activity | | |
|--------|---------------------------|---------------------------------|---|--------|--------|--|--------|-------|
| | | | HeLa | Jeg-3 | HepG2 | HeLa | Jeg-3 | HepG2 |
| CCND1 | A/B | -568A>C | 7.02 | 7.84 | 9.29 | 1.47* | 1.71** | 1.47 |
| CDC25A | A/B | -351 G>A | 382.61 | 125.63 | NA | 1.50 | 1.24 | NA |
| | A/C | -1440G>A | 342.16 | 102.80 | NA | 1.04 | 1.18 | NA |
| CDKN1A | A/B | -899T>G | 18.20 | 408.70 | 58.85 | 1.39** | 1.31* | 1.35 |
| | A/C | -1536C>T | 24.20 | 533.50 | 80.06 | 1.23 | 1.15 | 1.66 |
| CDKN1B | A/B | -1608G>A | 112.40 | 40.00 | 38.40 | 1.08 | 1.33* | 1.45 |
| E2F1 | A/B | -897C>T | 37.20 | 45.74 | 12.69 | 1.68* | 1.15 | 1.54 |
| HDAC1 | A/B | -905A>C | 74.00 | 65.19 | 42.24 | 1.45* | 1.79** | 2.01* |
| RB1 | A/B | -1554C>A | 616.40 | 143.29 | 32.97 | 1.19 | 1.13 | 2.12 |
| | C/D | -1733INS CTGATA | 769.07 | 245.34 | 106.45 | 1.06 | 1.10 | 1.75* |
| TFDP1 | A/B | -1394T>C | 120.18 | 474.64 | 6.14 | 1.11 | 1.15** | 1.10 |

Statistical significance for reporter gene assays that showed significant differences upon replication *P<0.05; **P<0.005; ***P<0.000. NA, levels of reporter gene activity not high enough for analysis.

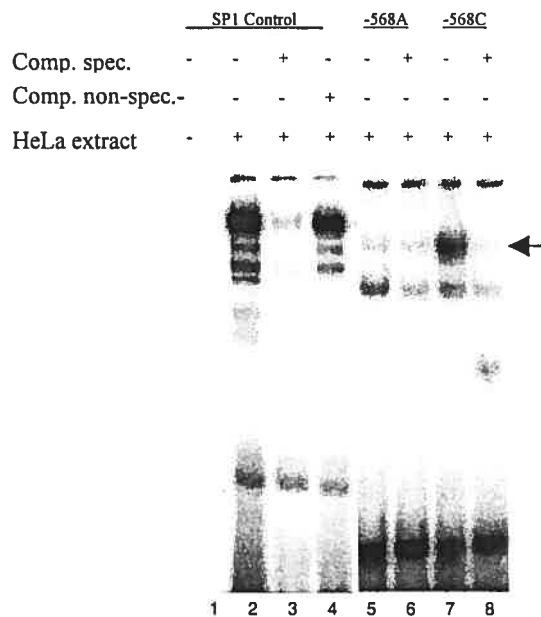


Figure 1: Representative EMSA illustrating allelic DNA-protein interactions in the promoter of *CCND1*. Labeled double-stranded oligonucleotide (ds-oligo) probe corresponding to -568A<C alleles when incubated with HeLa nuclear extracts reveals the position of the fast migrating unbound probe at the bottom of the gel and DNA-protein complexes of slower mobility (see arrow). Lane 1-4: Control ds-oligos using HeLa nuclear extracts. Lane 1, negative control; lane 2, positive control (labeled SP1 ds-oligo); lane 3, labeled SP1 ds-oligo plus unlabeled SP1 probe (competitor); lane 4, labeled SP1 ds-oligo plus unlabeled AP1 probe (noncompetitor). Lane 5-8: Differential binding of alleles -568A and -568C. Lane 5, labeled -568A ds-oligo; lane 6, labeled -568A ds-oligo plus unlabeled -568A probe (competitor). Lane 7, labeled -568C ds-oligo; lane 8, labeled -568C ds-oligo plus unlabeled -568C probe (competitor).

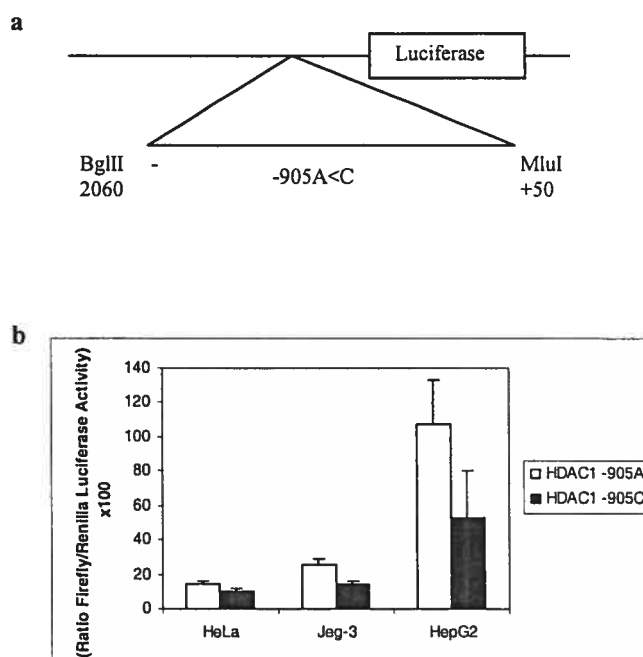


Figure 2: Assessment of the functional impact of the HDAC1 -905A<C variant on the promoter activity using a gene reporter assay. (a) 2kb regulatory haplotype inserted in the promoterless pGL3-basic vector containing the luciferase reporter gene (b) Relative luciferase levels for the allelic constructs associated either with pSNP -905A (white) or -905C (black) following transient transfection of HeLa, Jeg-3 and HepG2 cells. The C at position -905 decreases the promoter activity in all three cell lines. The luciferase activity is indicated by the ratio of firefly luciferase activity on the renilla luciferase activity multiplied by 100.

Chapitre 4

-Résultats-

4.1 Description des polymorphismes identifiés pour TGF β 1.

À l'aide de la technique dHPLC, nous avons identifié 9 polymorphismes dans la région promotrice du gène TGF β 1 dont 6 transitions, 1 transversion et 2 délétions/insertions. La densité de polymorphismes par kilobase est de 4,3. Le tableau XIV donne une description détaillée des pSNPs ainsi que l'effet de chaque pSNP sur le gain/perte de sites de liaison de facteurs de transcription. Parmi les 9 pSNPs identifiés, un seul fait apparaître un site et 5 font disparaître un ou plusieurs sites. 3 pSNPs sont situés dans la région répétitive LINE/L2. Aucun pSNP n'est retrouvé dans une région conservée entre l'homme et la souris. La représentation graphique de la région régulatrice du gène TGF β 1 est illustrée à la figure 4-2 où les pSNPs identifiés par dHPLC et confirmés par séquençage sont représentés.

4.2 Étude *in vitro* de l'impact des pSNPs du gène codant pour TGF β 1.

4.2.1 Retard sur gel

Pour TGF β 1, nous avons décidé d'étudier le polymorphisme -1550INS AGG, qui mène à une perte de liaison du facteur de transcription MAZ (myc associated zinc finger) selon les prédictions *in silico*. Pour vérifier cette prédiction, deux sondes d'ADN double-brin ont été utilisées chacune correspondant soit à la séquence de l'allèle majeur ou à l'allèle mineur. Nous avons examiné la capacité de chaque sonde à lier des facteurs nucléaires isolés à partir de cellules HeLa, Caco2, Jeg-3 et HepG2. Pour le pSNP -1550INS AGG, l'analyse des complexes ADN-protéines par électrophorèse sur gel de polyacrylamide non-dénaturant avec les extraits nucléaires de HeLa a révélé la présence de deux bandes spécifiques pour la sonde correspondant à l'allèle majeur (-1550AGG), qui sont partiellement abolies par l'ajout de la même sonde non-marquée, et l'absence de bande spécifique pour l'allèle mineur (-1550DELAGG) (Figure 4-3). Des résultats similaires sont observés avec les extraits nucléaires provenant des autres lignées cellulaires. Ces résultats confirment la prédiction *in silico*.

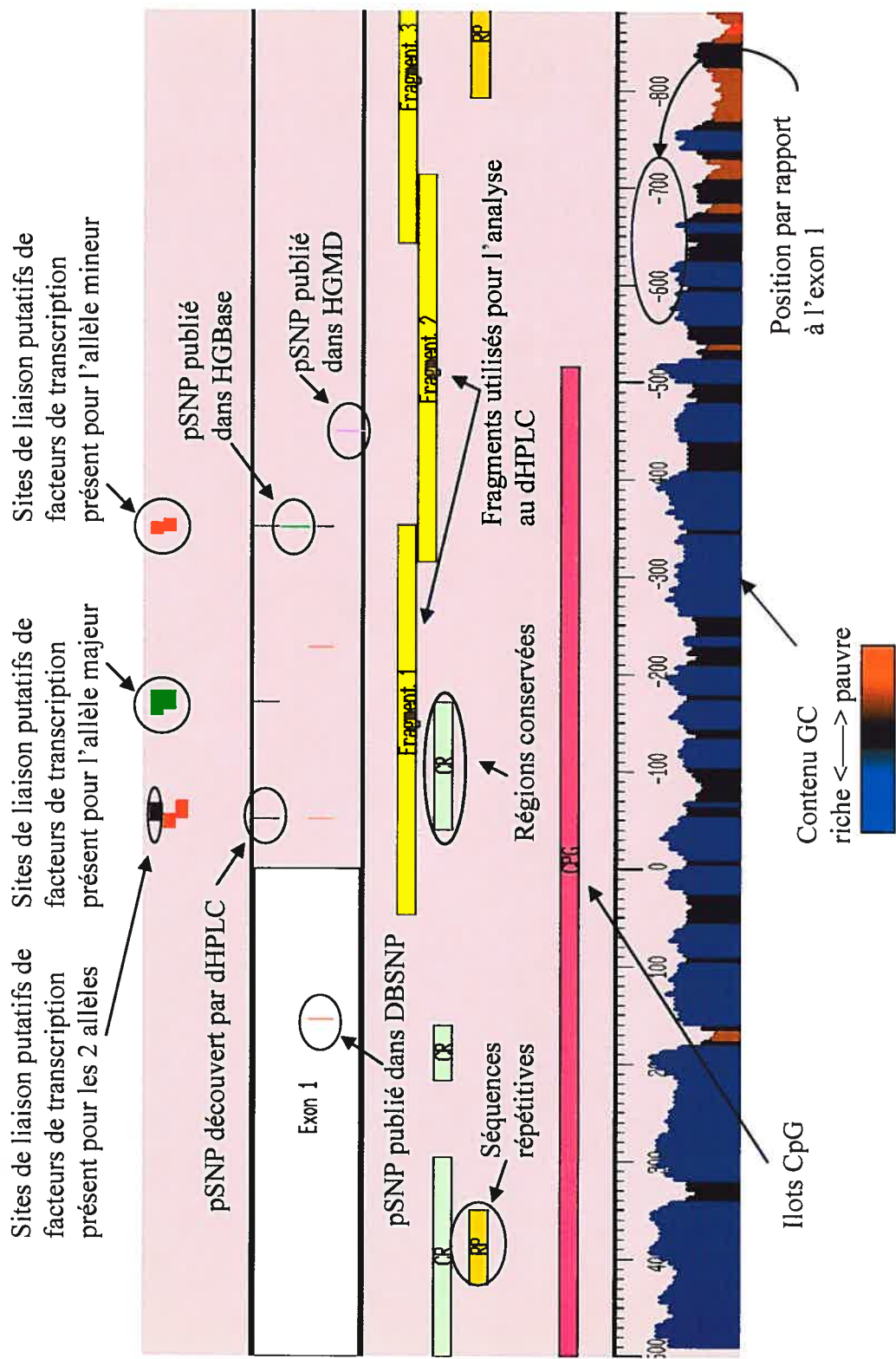


Figure 4-1. Légende pour les figures des représentations graphiques des promoteurs des gènes étudiés. DBSNP, HGMD, et HGBase sont des bases de données publiques.

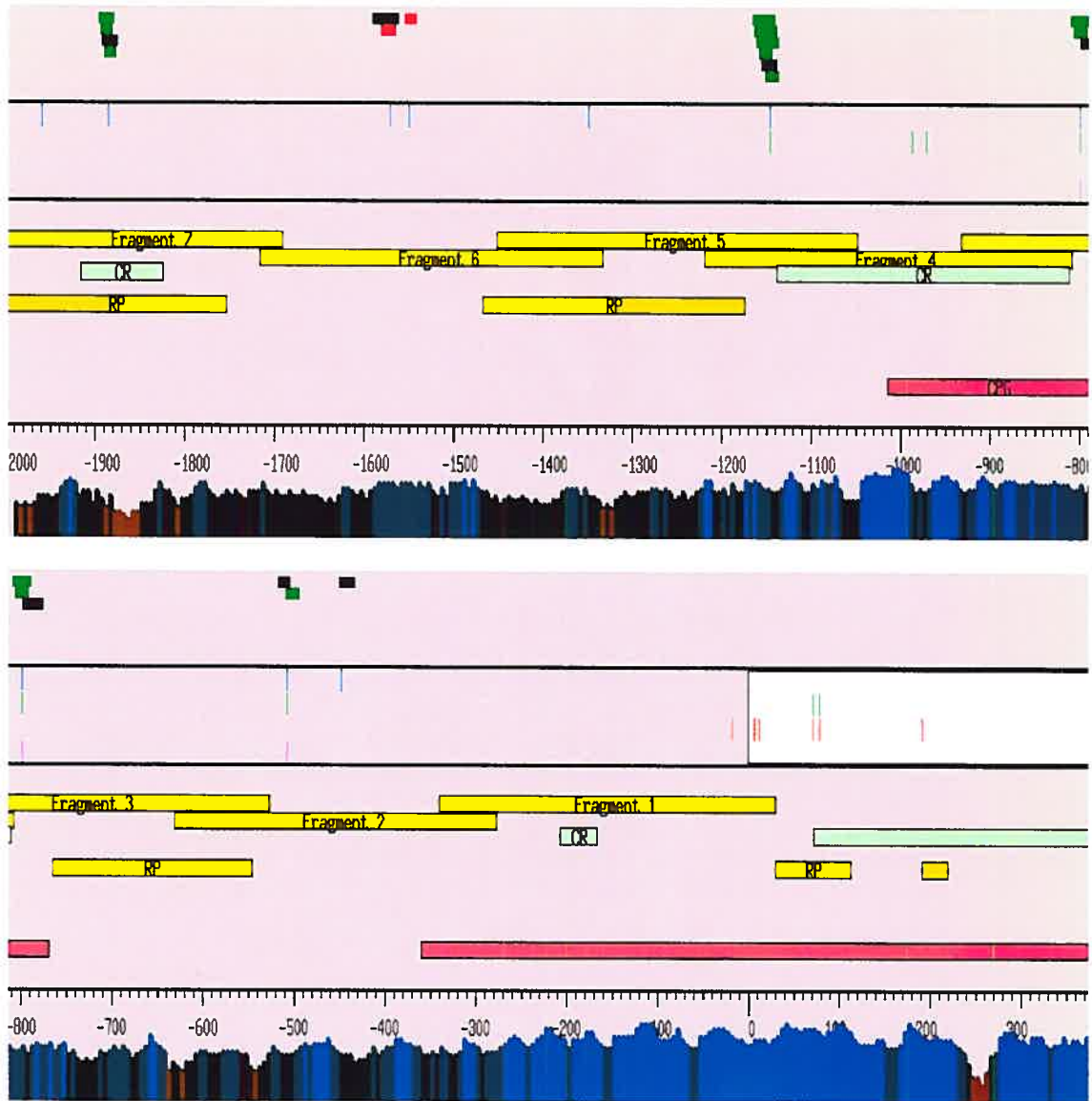


Figure 4-2: Représentation graphique du promoteur du gène TGFβ1. Voir figure 4-1 pour légende.

Tableau IV-I. Description détaillée des polymorphismes identifiés le promoteur du gène codant pour TGFβ1.

| Gène | Position | Type de polymorphisme | Sites de liaison allèle majeur | Sites de liaison allèle mineur | Sites de liaison conservés | Répétition |
|-------|-----------|-----------------------|--|---|----------------------------|-----------------|
| TGFB1 | Ex1.-1146 | C->G transversion | <u>GAGA-Box->0.82</u> <u>GAGA-Box->0.805</u> <u>GAGA-Box->0.803</u> <u>Gut-enriched Krueppel-like factor->0.912</u> <u>Gut-enriched Krueppel-like factor->0.929</u> <u>VDR/RXR Vitamin D receptor RXR heterodimer site->0.908</u> | <u>VDR/RXR Vitamin D receptor RXR heterodimer site->0.899</u> | | |
| TGFB1 | Ex1.-1349 | INS C nd | | | | LINE/L2 L2 |
| TGFB1 | Ex1.-1550 | DEL AGG nd | <u>Myc associated zinc finger protein (MAZ)->0.91</u> | | | |
| TGFB1 | Ex1.-1571 | A->G transition | <u>Myogenin / nuclear factor 1 or related factors->0.721</u> | <u>Hepatic nuclear factor 4->0.821</u> <u>Myogenin / nuclear factor 1 or related factors->0.721</u> | | |
| TGFB1 | Ex1.-1886 | G->A transition | <u>Ecotropic viral integration site 1 encoded factor->0.79</u> <u>GATA-binding factor 3->0.911</u> <u>Neurogenin 1 and 3 (ngn1/3) binding sites->0.953</u> <u>Tal-1beta/E47 heterodimer->0.957</u> | <u>Ecotropic viral integration site 1 encoded factor->0.793</u> | | LINE/L2 L2 |
| TGFB1 | Ex1.-1960 | C->T transition | | | | LINE/L2 L2 |
| TGFB1 | Ex1.-448 | C->T transition | <u>Xenopus fork head domain factor 3 (FoxA2a)->0.864</u> | <u>Xenopus fork head domain factor 3 (FoxA2a)->0.864</u> | | |
| TGFB1 | Ex1.-508 | A->G transition | <u>Gut-enriched Krueppel-like factor->0.928</u> <u>PAX 2/5/8 binding site->0.894</u> | <u>PAX 2/5/8 binding site->0.889</u> | | |
| TGFB1 | Ex1.-799 | G->A transition | <u>cAMP-responsive element binding protein->0.981</u> <u>Hypoxia induced factor-1 (HIF-1)->0.87</u> <u>Transcriptional repressor, binds to elements found predominantly in genes that participate in lipid metabolism->0.802</u> | <u>Transcriptional repressor, binds to elements found predominantly in genes that participate in lipid metabolism->0.804</u> | | |

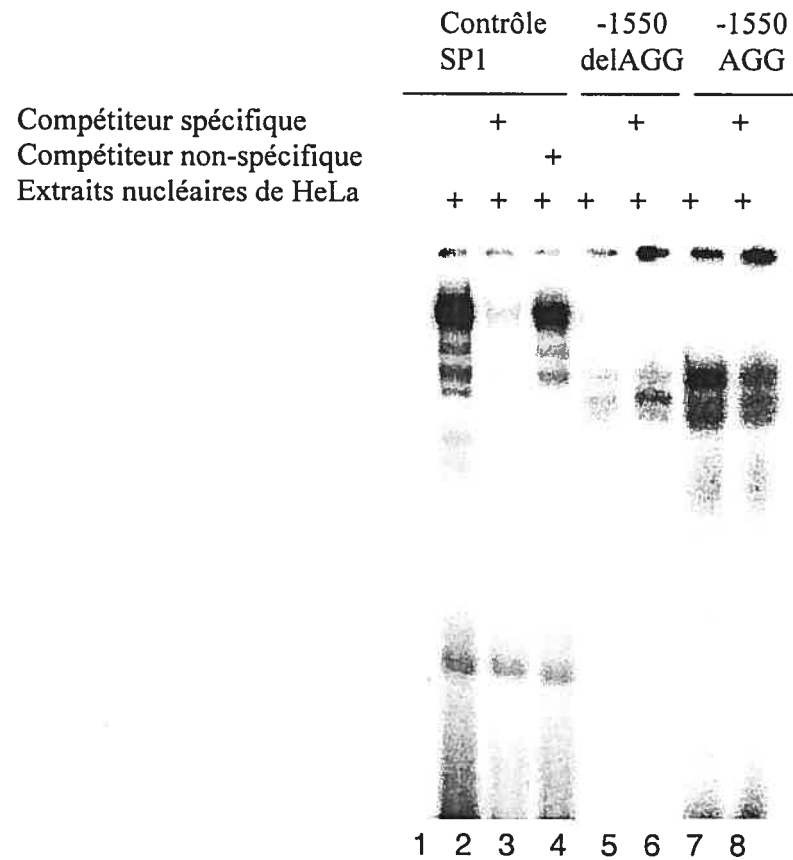


Figure 4-3: Résultats de retard sur gel pour le polymorphisme à la position -1550 du gène $TGF\beta 1$. Ligne 1, contrôle négatif; ligne 2, contrôle positif; ligne 3, contrôle avec compétiteur spécifique; ligne 4, contrôle avec compétiteur non-spécifique; ligne 5, sonde allèle mineur; ligne 6, sonde allèle mineur plus compétiteur spécifique; ligne 7, sonde allèle majeur; ligne 8, sonde allèle majeur plus compétiteur spécifique.

4.2.2 Gène rapporteur

Pour étudier le polymorphisme TGF β 1 -1550DELAGG, nous avons cloné la région 2kb à partir de l'ADN génomique d'un individu dont l'haplotype est décrit à la figure 4-4 dans le vecteur pGL3-basic qui code pour le gène de la luciférase. À l'aide de gène rapporteur, nous avons comparé l'expression de la luciférase de ces deux haplotypes de TGF β 1 (Figure 4-5). Nous avons observé une diminution d'expression significative pour l'haplotype B d'au moins 1,7 dans les trois lignées cellulaires. Cependant, dans HepG2, l'haplotype A qui avait l'activité la plus élevée, était en dessous de 10 fois l'activité du pGL3-basic.

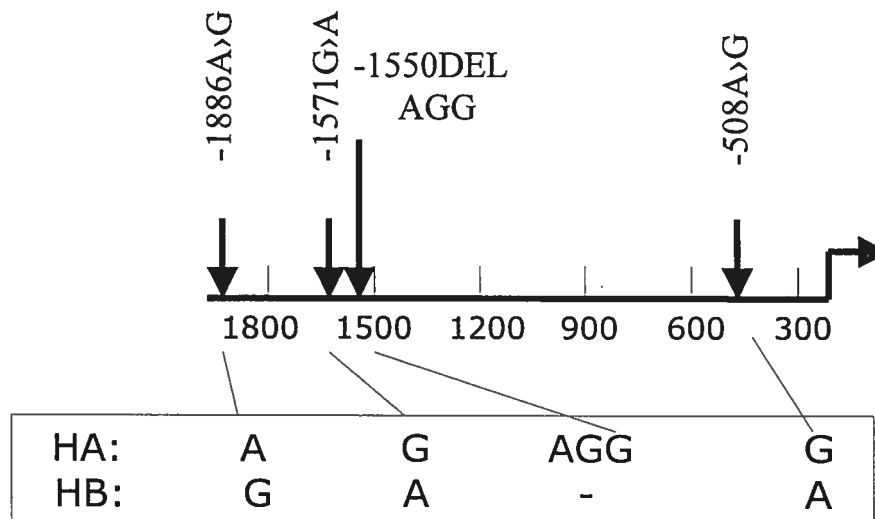


Figure 4-4: Haplotypes majeurs de TGFβ1.

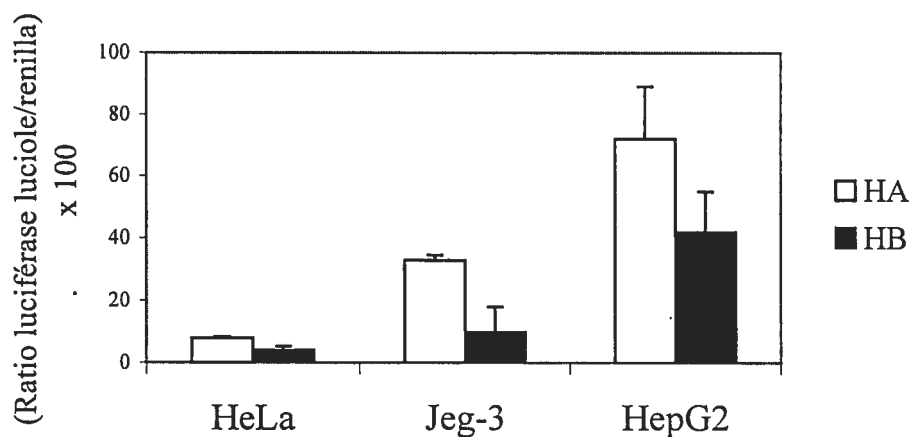


Figure 4-5: Évaluation de l'impact fonctionnel des haplotypes A et B sur l'activité du promoteur du gène TGF β 1 par expériences de gènes rapporteurs. Activité relative de la luciférase pour l'haplotype A (HA: blanc) et l'haplotype B (HB: noir) lorsque transfecté de façon transitoire dans les lignées cellulaires HeLa, Jeg-3 et HepG2. L'haplotype B diminue l'expression de la luciférase dans les trois lignées cellulaires. L'activité relative de la luciférase est indiquée par le ratio de la luciférase de la luciole sur l'activité de la luciférase de la rénilla multipliée par 100.

Chapitre 5

-Discussion-

5.1 État de la question

Suite aux études sur le génome humain qui démontrent un nombre de gènes plus faible que prévus, il a été proposé que les SNPs dans le promoteur (pSNPs) seraient responsable de la majorité des causes génétiques reliées à la susceptibilité interindividuelle de développer des maladies complexes[49]. Dans ce projet, nous avons voulu évaluer l'impact fonctionnel de ces polymorphismes dans les gènes qui contrôlent le cycle cellulaire. Les points de contrôle du cycle cellulaire sont essentiels au maintien de l'homéostasie cellulaire et au maintien de l'intégrité du matériel génétique. De plus, les anomalies dans la régulation du cycle cellulaire sont des causes importantes dans le développement du cancer[10]. Des pSNPs pourraient affecter le niveau de transcription d'un gène soit par le gain ou la perte de site de liaison de facteurs de transcription. Ceci pourrait mener à une variabilité dans les mécanismes contrôlant le cycle cellulaire et aussi modifier le risque de développer certaines maladies, tel que le cancer (Figure 5-1).

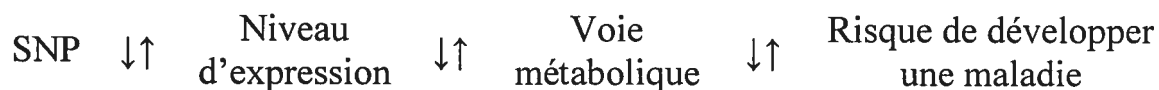


Figure 5-1: Effet d'un SNP sur le risque de développer une maladie. (voir texte pour détails)

5.2 Identification de polymorphismes

Il existe plusieurs approches permettant de détecter la présence de variations génétiques[44, 45, 66]. Nous avons choisi le dHPLC comme plate-forme de détection puisque son efficacité de détection de polymorphismes se situe entre 95% et 100%[44, 45, 47, 48, 66]. Nous avons ainsi identifié des pSNPs dans 17 de gènes impliqués dans le point de contrôle G1/S. Nous avons ainsi détecté 136 polymorphismes ce qui correspond à une densité de 4,4 pSNPs par kb, une distribution comparable aux données reportées dans la littérature. Par exemple, dans la région codante, certaines études ont observées de 2,9 à 4,1 SNPs par kb, dans les régions non-codante (introns et UTRs) de 2,8 à 6,1 et

dans la région 5' jusqu'à 5,8 SNPs par kb[67, 68]. Plusieurs évidences suggèrent que le contexte de la séquence d'ADN peut influencer la mutagenèse en agissant comme un « hotspot » pour les mutations[69-71]. Par exemple, la méthylation des dinucléotides CpG est associée avec l'apparition fréquente de mutations qui implique G→A et C→T[72]. En effet, une méthylation de la cytosine en 5-méthylcytosine augmente le risque de mutations par désamination de la 5-méthylcytosine en thymine[73]. On peut donc s'attendre à observer un taux plus élevé de polymorphismes dans la région 5' à cause de la présence accrue d'îlots CpG. Cependant, les îlots CpG ne sont pas méthylés[74, 75]. Puisque la méthylation semble être requise pour augmenter le taux de mutations, il a été proposé que les CpGs à l'intérieur des îlots ne subiraient pas de pression pour muter[76]. Nous avons observé que 64 (47%) de nos polymorphismes ce retrouvent dans les îlots CpG. Ce qui donne une fréquence de 5.4 pSNPs par kb d'îlots CpG. Cependant, cette haute fréquence peut être expliquée par trois gènes: CCND1, CDC25A et CDK4. Pour, CCND1 et CDC25A, les îlots CpG représentent 55% et 52% des régions séquencées, respectivement. Lors de l'analyse des pSNPs retrouvés dans les différentes combinaisons de dinucléotides, nous observons une surreprésentation de polymorphismes pour les dinucléotides CG ($p < 0,001$). Ensuite, nous avons observé la distribution des pSNPs en fonction de diverses caractéristiques de la séquence. 28 (21%) de nos polymorphismes sont situés dans des régions répétées. La fréquence de ces pSNPs est de 4.2 pSNPs/kb, ce qui est semblable à ce qu'on retrouve dans les autres régions. Parmi les pSNPs identifiés, il y a 71 (52,2%) transitions et 46(33,8%) transversions, et 19 (14,0%) sont des polymorphismes de longueur. Notre ratio de transition/transversion est de 1,5, ce qui est similaire à la littérature[77]. Seulement 35 (25,7%) de nos pSNPs sont présents dans les bases de données publiques (DBSNP, HGBASE, HGMD). Il est normal de ne pas retrouver les mêmes polymorphismes dans les bases de données publiques puisque nous avons utilisé un panel d'individus provenant de différents groupes ethniques, qui n'est pas nécessairement identique à celle utilisé dans les autres études.

Présentement, une problématique associée avec la caractérisation des polymorphismes dans la région régulatrice est de différencier ceux qui sont fonctionnels de ceux qui ne le

sont pas. Nous avons utilisé différentes stratégies afin d'accorder une plus grande importance aux polymorphismes qui seraient potentiellement fonctionnels. Premièrement, des analyses *in silico* : prédictions sur la perte ou la création d'un site de liaison de facteurs de transcription et la génomique comparative. Ensuite, utilisation d'approches *in vitro* : confirmation des prédictions *in silico* à l'aide de retard sur gel et évaluation de la différence d'expression par gène rapporteur.

5.3 Prédictions *in silico*

La perte ou la création d'un facteur de transcription dû à la présence d'un polymorphisme peut mener à des modifications au niveau de l'expression du gène correspondant. Les facteurs de transcription montrent un haut degré de spécificité dans leur domaine de liaison à l'ADN pour la séquence d'ADN à laquelle ils se lient. En général, la séquence consensus est d'environ 6 à 8 nucléotides et est spécifique pour une famille de facteurs de transcription[53, 78]. Il est donc possible d'analyser la séquence d'ADN correspondant aux deux formes alléliques d'un SNP et évaluer l'impact sur des sites potentiels de liaison de facteurs de transcription à l'aide d'outils bioinformatiques. À l'aide de MatInspector, a prédit que 96 (70,6%) de nos pSNPs pourraient affecter la liaison de facteurs de transcription. Parmi ces pSNPs, 32 (23,5%) font disparaître des sites de liaison de facteurs de transcription, 28 (20,6%) font apparaître des nouveaux sites et 36 (26,5%) font disparaître et apparaître des sites. L'évaluation de la prédiction est basée sur une valeur qui compare la séquence analysée avec les séquences consensus des facteurs de transcription. L'utilisation de techniques capable de valider les sites prédits tel que les retards sur gel est nécessaire afin d'éliminer les faux positifs. Aussi, l'analyse de l'impact du pSNPs est limitée par cette approche *in silico* puisque les facteurs de transcription ne sont pas tous connus. Même si cette approche permet l'identification de sites liés *in vitro* de façon efficace[79], ils sont insuffisant pour trouver des éléments fonctionnels *in vivo*[80]. En effet, les informations contenues entre l'interface du site de liaison et le facteur de transcription n'est pas suffisant afin d'établir une distinction entre les sites fonctionnels et non-fonctionnels dans l'environnement cellulaire. La spécificité d'un facteur dépend de plusieurs éléments comme l'interaction

avec d'autres protéines, l'accessibilité de l'ADN et de sa concentration. Cependant, cette analyse est une première étape nous permettant de sélectionner des pSNPs potentiellement fonctionnels parmi tous les polymorphismes identifiés.

5.4 Génomique comparative

L'analyse comparative entre les espèces est un outil intéressant puisqu'il a été proposé que les séquences conservées au cours de l'évolution seraient impliquées dans des fonctions importantes[81-83]. Nous avons utilisé la souris comme modèle pour identifier les régions conservées à cause de la disponibilité de la séquence. La comparaison entre plusieurs espèces aurait diminué le nombre de régions conservées, mais augmenté les chances de trouver des régions importantes. Parmi les 17 gènes analysés, 14 (10%) polymorphismes se retrouvent dans des régions conservées entre l'humain et la souris. Parmi ceux-ci, 12 polymorphismes ont un effet sur la liaison de facteurs de transcription. La fréquence de polymorphismes est légèrement inférieure (3,6 pSNPs/kb) dans les régions conservées que dans les régions non conservées. Ces résultats suggèrent que ces régions se trouvent sous une pression sélective pour éviter la variation génétique.

5.5 Retard sur gel

Notre attention s'est portée sur les 12 pSNPs qui, selon les prédictions *in silico*, font apparaître un nouveau site de liaison de FT où il n'y avait pas de site de liaison au préalable et 1 polymorphisme qui mène à une perte de liaison. Nous avons procédé à la validation fonctionnelle de ces pSNPs avec des expériences de retard sur gel pour vérifier les prédictions. Cette technique est efficace pour étudier les interactions entre ADN et protéines *in vitro*. Nous avons utilisé des extraits nucléaires provenant de différentes lignées cellulaires (HeLa, Caco2, Jeg-3 et HepG2) afin d'obtenir une variété de protéines exprimées dans différents tissus. En effet, nous observons dans la majorité des cas des différences entre les profils de migration utilisant les différents extraits nucléaires pour un même oligonucléotide. Nous avons ainsi démontré que 12 de ces

pSNPs mènent à une modification de la liaison à des facteurs de transcription lorsque incubés avec au moins un des extraits nucléaires. Afin de connaître la nature des protéines qui se lient à nos allèles, on pourrait faire des expériences de retard sur gel en utilisant des sondes compétitrices contenant le site consensus des différentes protéines prédites. Il serait aussi possible d'effectuer des «supershifts» avec des anticorps spécifiques pour certains facteurs de transcription prédits. La transcription chez les eucaryotes implique plusieurs facteurs de transcription et peu de polymorphismes situés dans les régions régulatrices mèneront à une variation dans le patron d'expression. Des études utilisant des gènes rapporteur sont nécessaire afin d'évaluer l'impact de ces différences observées sur l'expression du gène.

5.6 Gène Rapporteur

Nous avons effectué des études de gènes rapporteurs à l'aide de transfections transitoires dans trois lignées cellulaires (HeLa, Jeg-3 et HepG2). Pour ce faire, nous avons cloné la région de 2kb des deux formes alléliques dans le vecteur pGL3-basic (sans promoteur) afin de comparer l'effet de la présence du polymorphisme sur la transcription de la luciférase. Les lignées cellulaires ont été choisies pour représenter différents types de tissus mais aussi parce qu'elles avaient déjà été utilisées dans d'autres études de gène rapporteur[56, 84, 85] et qu'elles avaient au moins 40% de taux d'efficacité de transfection avec la lipofectamine. Parmi les 13 polymorphismes étudiés seulement 12 variants ont pu être étudiés puisque le 2kb de MADH3 n'a pas pu être amplifié. À cause de l'absence de consensus sur ce qui constitue une différence fonctionnelle importante, nous avons décidé d'utiliser différents critères. Premièrement, il doit y avoir une différence d'activité de la luciférase de 1,5 entre les deux allèles respectifs. Nous avons sélectionné 1,5 de différence pour deux raisons. Si la différence est reflétée dans des conditions *in vivo*, les homozygotes pour l'allèle ayant la plus grande activité pourraient être considérés comme s'ils avaient une copie supplémentaire de cet allèle. Cependant, il se peut que des plus petites différences mènent aussi à des changements importants. Toutefois, Hoogendoorn et al. ont démontré qu'une différence de 1,5 entre les allèles était reproductible dans des expériences indépendantes utilisant des nouveaux

clones[57]. Le deuxième critère est que la différence doit être significative à $P \leq 0,05$. Parmi les 12 variants étudiés, des différences significatives ont été détectées pour 5 polymorphismes situés dans la région régulatrice des gènes codants pour TGF β 1, E2F1, HDAC1, RB1 et CCND1 dans au moins une lignée cellulaire. Ceci représente donc 42% (5/12) de polymorphismes fonctionnellement importants. Parmi les gènes sélectionnés avec des polymorphismes (n=9), 5 gènes ont des polymorphismes fonctionnels dans leurs régions régulatrices, ce qui correspond à environ 56%. Ce nombre est plus élevé comparativement à d'autres études similaires, où on retrouve entre 20 et 30% de gènes avec des polymorphismes fonctionnels[86, 87]. Cependant, notre chiffre est biaisé puisque nous avons seulement étudié les gènes contenant des polymorphismes qui menaient à un gain ou une perte potentiel de site de liaison de facteurs de transcription.

5.7 TGF β 1

TGF β 1 (transforming growth factor beta 1) est une cytokine impliquée dans plusieurs processus biologiques tels que la régulation du cycle cellulaire, la formation de la matrice extracellulaire, la réponse immunitaire et le développement embryonnaire. TGF β 1 est un puissant inhibiteur de la prolifération dans de nombreux types cellulaires tels que les hépatocytes, les lymphocytes T et B, les kératinocytes et de nombreux types de cellules épithéliales et endothéliales[88]. La dérégulation de la transduction du signal de TGF β conduit à différentes maladies incluant les fibroses tissulaires et à divers cancers.

Le polymorphisme ciblé par notre étude est une délétion de AGG à la position -1550, qui est localisé sur l'haplotype B. Ainsi, il y avait d'autres polymorphismes qui variaient entre nos allèles. Il aurait été intéressant d'étudier l'impact de ce polymorphisme dans le contexte de divers haplotypes en utilisant la mutagenèse dirigée. Nous avons concentré nos efforts sur les haplotypes qui étaient fréquents dans la population mondiale et qui variaient pour notre polymorphisme d'intérêt. L'haplotype A est le plus fréquent et selon nos études de population est présent à environ 51%. L'haplotype B est le deuxième plus fréquent et se retrouve à environ 32% dans la population. La diminution d'expression

observée pour l'haplotype B, peut être causée par un seul variant mais est probablement le résultat de la combinaison de différents polymorphismes. Le polymorphisme à la position -508 a été décrit dans plusieurs articles[89-96]. Il a été proposé que le T à la position -508, qui équivaut à notre A à la position -508, modifie la liaison du facteur de transcription YY1 (Yin Yang-1) et qu'il augmente l'expression de TGFβ1[91]. De plus, il a été démontré qu'il était significativement associé avec une plus grande concentration de TGFβ1 dans le plasma[90]. Contrairement à ce qu'on trouve dans la littérature, nous avons observé que l'haplotype B qui contient le A à la position -508 est associé avec une diminution de l'expression. Par conséquent, la diminution doit être due à la présence des autres polymorphismes. Malgré la haute fréquence de ces variants, ils ne sont pas décrits dans aucune publication. L'absence de données sur ces trois polymorphismes peut être expliquée par le fait que la majorité des études ont mis l'emphase sur le premier kb en amont de l'exon 1. Selon nos prédictions *in silico*, la délétion de AGG mène à la perte du site de liaison du facteur de transcription MAZ (myc-associated zinc finger protein). Le facteur de transcription MAZ est ubiquitaire et indépendant du type cellulaire, et il est impliqué dans l'initiation et la terminaison de la transcription[97, 98]. Plusieurs gènes, tel que CLC-K1 et le récepteur de sérotonine 1a, sont activés par MAZ[99, 100]. Puisque nous observons des différences dans nos retards sur gel qui confirment nos prédictions informatiques, la délétion de AGG à la position -1550 pourrait être en parti responsable de la diminution d'activité observée.

A priori, une diminution de l'expression de TGFβ1 aurait comme conséquence une augmentation de la prolifération. Par conséquent, cette dérégulation du cycle cellulaire pourrait augmenter le risque de développer certaines maladies complexes tel que le cancer. Plusieurs évidences montrent que la voie de signalisation de TGFβ joue divers rôles dans la progression vers le cancer. En effet, des altérations dans cette voie contribuent de différentes façons à la progression tumorale, à l'invasion et aux métastases[101]. Premièrement, la croissance des cellules peut être inhibée par TGFβ1. Il a été démontré que la perte précoce de cette signalisation, par exemple par des mutations dans TβRII, entraîne une augmentation de la synthèse d'ADN, une croissance rapide de la tumeur et l'expansion clonale, ce qui résulte en une possibilité accrue

d'accumuler des mutations et des changements cytogénétiques[101]. Une autre étude a démontré qu'un polymorphisme dans T β RI mène à une diminution de l'activité biologique de TGF β 1 et celui-ci pourrait être associé avec un risque plus élevé de développer le cancer du colon[102]. De plus, les souris transgéniques qui ont une délétion d'une copie de TGF β 1 sont plus susceptibles à développer des tumeurs du foie et des poumons lorsqu'elles sont induites par des carcinogènes[103]. Il a aussi été démontré que l'augmentation de l'expression de TGF β 1 réduit le risque de développer des tumeurs mammaires chez des souris transgéniques[104]. Malgré, son activité suppresseur de tumeur, une augmentation de TGF β 1 peut aussi être observée dans certaines cellules tumorales[105]. TGF β 1 serait impliqué dans la progression de la tumeur à un stade plus tardif. Il serait impliqué dans l'invasion cellulaire et provoquerait des changements dans le microenvironnement de la tumeur tel que l'augmentation de l'angiogenèse, ce qui donnerait des avantages à la progression du cancer.

En résumé, nous observons une diminution significative de l'expression de TGF β 1 associée avec l'haplotype B et cette dérégulation pourrait contribuer à l'augmentation du risque de développer certains cancers.

5.8 E2F1

E2F1 joue un rôle crucial dans le contrôle du cycle cellulaire. Lorsqu'il forme un hétérodimère avec un membre de la famille TFDP, ceci permet l'activation de la transcription de gènes nécessaire à l'entrée en phase S. En plus de son rôle dans la prolifération cellulaire, il est aussi impliqué dans le processus de l'apoptose dans certains types de cellules[106]. Une dérégulation de son activité est souvent observée dans les tumeurs humaines. Cependant, peu d'études ont identifié des mutations affectant directement les membres de la famille E2F dans les tumeurs humaines[17, 107]. Récemment, une étude démontre qu'il y a une augmentation de l'expression de E2F1 dans certains cancers de la thyroïde[108]. D'autres ont montré que la perte d'E2F1 dans des souris RB1(+/-), diminuait de façon significative la fréquence des tumeurs de la thyroïde et de la glande pituitaire[109]. Puisque l'apoptose atténuée la

croissance des cancers, on pourrait s'attendre qu'une perte d'E2F1 augmenterait aussi le risque de développer des tumeurs. Cependant, des souris E2F1 (-/-) ont démontré que l'apoptose était inhibé à 80% et que la croissance de la tumeur n'était pas accélérée. Donc, les cellules tumorales agiraient différemment des cellules normales, la prolifération est moins efficace sans E2F1 ce qui viendrait contrebalancer le fait qu'il a moins d'apoptose[110].

Nous avons observé que l'allèle qui contenait le T à la position -897 diminuait l'expression de 1,7 fois comparativement à l'allèle avec le C dans les cellules HeLa. Le T à la position -897 crée une séquence consensus pour la protéine «cut-like homeodomain». Cette protéine pourrait être responsable pour les différences observées puisque nous avons noté la présence d'une liaison ADN-protéine seulement pour l'allèle mineur (-897T), tandis que l'allèle majeur (-897C) ne liait pas de protéines dans nos expériences de retard sur gel. De plus, cette protéine est reconnue comme étant un régulateur négatif de la transcription[111, 112]. Cependant, des études de «supershift» avec des anticorps seraient nécessaires afin de confirmer que c'est bien cette protéine qui se lie à l'allèle mineur.

Cette diminution observée avec l'allèle qui contient le T à la position -897 peut mener à un dérèglement au niveau de l'entrée en phase S et par conséquent diminuer la prolifération. Ceci pourrait diminuer la susceptibilité interindividuelle de développer certaines maladies complexes tel que le cancer.

5.9 RB1

Nous avons étudié deux polymorphismes qui mènent potentiellement à un gain de liaison de facteur de transcription dans la région régulatrice du gène codant pour RB1. La comparaison de l'haplotype A et B démontre que le polymorphisme à la position -1554 n'a pas d'effet sur la transcription puisqu'il est le seul polymorphisme qui varie entre ces deux haplotypes. Par contre, l'haplotype D a montré une diminution significative de l'expression dans les cellules HepG2 par rapport à l'haplotype C. Dans

nos études de retard sur gel, nous avons observé plusieurs différences entre nos allèles majeurs (sans insertion) et mineurs (avec insertion) utilisant les extraits nucléaires des quatre lignées cellulaires. Cette insertion crée une séquence consensus pour trois facteurs de transcription : EVI-1 (ecotropic viral integration site 1 encoded factor), GATA-1 et RBP (mammalian transcriptional repressor RBP-Jkappa/CBF1). EVI-1 et GATA-1 sont exprimés dans la lignée cellulaire HepG2[113]. Plusieurs évidences montrent que EVI-1 peut être impliqué dans l'activation de transcription mais aussi dans la répression[114-117]. GATA-1 est principalement vu comme un activateur, par contre de récentes études montrent qu'il peut aussi être un répresseur[118]. Par exemple, il a été démontré que la répression de GATA-2 et MYC dans les cellules érythroïdes était associée avec la liaison de GATA-1[119, 120]. RBP est reconnu comme un facteur de transcription qui réprime l'expression des gènes[121, 122]. Ces trois facteurs pourraient potentiellement être impliqués dans la diminution d'expression observée dans nos gènes rapporteurs. Le polymorphisme à la position -768A>C pourrait aussi être impliqué dans la différence d'expression observée entre les haplotypes C et D. Les prédictions *in silico* pour ce variant indiquent des différences de liaison de facteurs de transcription entre l'allèle majeur et mineur. La délétion d'un C à la position -1729 n'influence probablement pas l'expression puisqu'il n'a aucune prédiction de liaison de facteur de transcription. Lorsque nous comparons l'haplotype A et D, il y a seulement l'insertion CTGATA qui diffère entre les haplotypes et nous observons une augmentation significative de 1,9 et 1,6, dans HeLa et Jeg-3 respectivement. Cependant, les expériences de gènes rapporteurs pour ces deux haplotypes ont été réalisées dans des expériences séparées, donc certains facteurs auraient pu influencer les résultats obtenus. Il faudrait donc reprendre cette expérience afin d'en tirer une conclusion définitive.

Rb1 est un suppresseur de tumeur et joue un rôle fondamental dans le contrôle de la prolifération cellulaire exercé en G1. Des mutations affectant la voie de RB sont communes dans les cancers humains. En effet, il a été démontré que des mutations inactivant RB1 mènent à une dérégulation au niveau des facteurs de transcription E2F ce qui provoque une prolifération incontrôlée et contribue aux altérations mitotiques et à l'instabilité chromosomique. Des altérations au niveau du promoteur qui ne résulte pas

dans une perte de fonction complète de RB ont aussi été identifiées dans des cas moins sévère de rétinoblastome[123, 124].

Nous proposons que l'insertion CTGATA à la position -1733 pourrait entraîner un déséquilibre d'expression de RB1 et mener à un dérèglement des mécanismes contrôlant le cycle cellulaire et ainsi modifier la susceptibilité interindividuelle de développer certaines maladies complexes tel que le cancer.

5.10 HDAC1

Nous observons une différence significative dans nos trois lignées cellulaires pour le polymorphisme -905A>C. Le C à la position -905 crée une séquence consensus pour trois facteurs de transcription : c-Abl tyrosine kinase, SRY (sex-determining region Y), HFH-2 (hepatocyte nuclear factor 3/forkhead homolog 2). c-Abl est un proto-oncogène qui est exprimé de façon ubiquitaire. Son interaction avec RB pendant les phases G1, G2 et M permet d'abolir son activité tyrosine kinase. En plus de son activité tyrosine kinase, c-Abl peut activer la transcription de certains gènes tel que c-MYC[125]. Donc, il est improbable qu'il soit impliqué dans la diminution observée dans nos expériences de gènes rapporteurs. Plusieurs évidences montrent que SRY n'est pas seulement transcrit dans les précurseurs et cellules de Sertoli et dans les cellules germinales mâles mais aussi dans une variété de tissus. En effet, une étude montre que SRY est exprimé dans deux lignées cellulaires humaines (NT2/D1 et HepG2) à des niveaux similaires que dans les testicules[126]. HFH-2 régule principalement les gènes spécifiques au foie[127]. Cependant, dans nos expériences de retards sur gel, nous n'observons pas de liaison ADN-protéine lorsque nous utilisons les extraits nucléaires provenant de HepG2. Cette absence de liaison peut aussi être expliquée par les conditions utilisées dans nos expériences. En effet, plusieurs facteurs entrent en jeu pour la réussite d'un retard sur gel. Donc, les expériences devront être refaite en modifiant différents paramètres. Par exemple, on peut utiliser différentes concentrations de sel lors de la préparation des extraits nucléaires et dans le tampon de liaison. On pourrait aussi modifier les tampons et la température d'électrophorèse.

En phase G1, HDAC1 est recruté par RB afin de former un complexe répresseur avec le facteur de transcription E2F. Donc, une diminution de l'expression HDAC1 mènerait à une diminution de la répression et par conséquent augmenterait la prolifération. Les études utilisant des inhibiteurs HDAC suggèrent le contraire. Les inhibiteurs de HDAC active la transcription de certains gènes régulés par E2F/RB, cependant il n'y a pas d'augmentation de prolifération. Ils observent plutôt un arrêt du cycle cellulaire[128]. Cette observation semble paradoxale étant donné le rôle central du complexe RB/E2F dans le contrôle du cycle cellulaire. Le mécanisme exact n'est pas encore connu. Il a été proposé que p21 serait une cible directe ou indirecte de HDAC puisque sa transcription est fortement stimulée lors de l'addition d'inhibiteur HDAC[129]. Dans ce modèle, l'inhibition de HDAC augmente l'expression de p21, ce qui mènerait à l'activation de RB et à l'arrêt cellulaire[128]. De nombreuses évidences démontrent que HDAC1 est associé avec le cancer. Il a été proposé que HDAC1 puisse s'associer avec des oncogènes et conduire à une prolifération cellulaire incontrôlée. Par exemple, dans les leucémies aiguës, les protéines chimériques générées dans les translocations chromosomiques n'agissent plus comme des activateurs de la transcription, mais plutôt comme des répresseurs en recrutant des HDACs[128]. Une diminution de l'expression HDAC1 pourrait donc diminuer la susceptibilité interindividuelle de développer le cancer.

5.11 CCND1

Le gène CCND1 avait des différences significatives de 1,5, dans HeLa et 1,7 dans Jeg-3 entre l'haplotype A et B. Par contre, l'expression n'était jamais 10 fois celle du vecteur pGL3-basic. Le polymorphisme ciblé était le -568A>C. Tous les variants entre ces deux haplotypes pourraient être impliqués dans les différences d'expression observées. Dans nos expériences de retards sur gel, nous observons des différences pour le pSNP -568A>C qui confirment nos prédictions *in silico* avec les extraits de HeLa et qui sont opposés pour Jeg-3. Puisque ce variant mène à des changements de liaison de facteur de transcription dans nos expériences de retard sur gel, ceci suggère qu'il pourrait être

impliqué dans les différences d'expression observées. Le -568C crée une séquence consensus pour la protéine p53 suppresseur de tumeur, qui est reconnu comme étant un activateur de la transcription[130, 131]. Donc, il est improbable que cette protéine soit la cause de cette diminution d'expression. De plus, nos expériences de retard sur gel suggèrent l'implication de différentes protéines selon les lignées cellulaires.

CCND1 est un régulateur clé du cycle cellulaire car il amène les cellules au-delà du point de restriction R où elles devront se diviser[107]. Une surexpression du gène CCND1 a été associée avec une augmentation de la prolifération et a été observée dans plusieurs types de cancer comme le cancer du sein et du colon[132, 133]. La diminution d'expression observée avec l'haplotype B de CCND1 pourrait potentiellement diminuer la susceptibilité interindividuelle de développer le cancer.

5.12 CDKN1A

Pour CDKN1A, nous avons obtenu des résultats dans nos expériences de gène rapporteur qui pourraient aussi avoir un impact biologique important même s'ils ne répondent pas exactement à tous nos critères. En effet, nous avons observé une différence significative entre l'haplotype A et B de CDKN1A. Cependant, cette différence est inférieure à 1,5, c'est-à-dire de 1,4 dans les cellules HeLa et 1,3 dans les cellules Jeg-3. Ici, le polymorphisme ciblé était -899T>G. Les résultats d'EMSA pour ce pSNP confirment la prédiction informatique lorsque incubé avec les extraits nucléaires de HeLa et Jeg-3 et il y a aucune interaction entre ADN-protéine avec les extraits provenant de HepG2. Aussi, beaucoup de polymorphismes différent d'un haplotype à l'autre et pourraient aussi être impliqués dans la diminution observée. Les polymorphismes -134C>G, -1021T>C, -1466G>A, -1529C>T et -1536T>C n'ont pas d'effet sur l'expression (comparaison entre l'haplotype A et C). Nos prédictions *in silico* indiquent aucun changement dans la liaison de facteur de transcription pour -519G>C et il y a aucune liaison observée dans nos expériences de retard sur gel pour -499G>C. Ceci suggère fortement que le variant -899T>G pourrait être responsable de la diminution d'expression observée dans nos expériences de gène rapporteur.

CDKN1A est un inhibiteur de la prolifération cellulaire et est un médiateur critique de p53[134]. Une perte de fonction ou une perte d'expression de ce gène résulterait en une prolifération anormale et prédisposerait les cellules à la transformation. Aussi, une diminution d'expression du gène a été détectée dans plusieurs types de tumeurs[135]. Bien que l'haplotype B pourrait augmenter la susceptibilité interindividuelle de développer le cancer, un aussi petit effet *in vitro* est moins probable d'avoir un effet correspondant *in vivo* puisque l'homéostasie va probablement compenser pour cette petite déficience transcriptionnelle.

5.13 Limitations des études *in vitro*

Les études *in vitro* nous permettent de valider les polymorphismes qui pourraient avoir potentiellement un impact fonctionnel. Il est à noter que les changements observés au niveau de la transcription *in vitro* ne vont pas nécessairement refléter ceux observés *in vivo*. La régulation de l'expression des gènes dans leur contexte génomique naturel est très sophistiquée comparée à celle dans les gènes rapporteurs. En effet, les techniques *in vitro* étudient l'ADN dans un contexte génomique différent, en absence de sa machinerie transcriptionnelle normale et sans tous ses éléments régulateurs. Aussi, nous n'avons pas mesuré l'impact de nos SNPs sous des états dynamique différents. Par exemple, lors de dommages à l'ADN ou durant le développement. De plus, à cause des processus de régulation de la traduction et de post-traduction, il se peut que des changements de ARNm ne résultent pas en des changements au niveau de l'abondance des protéines. Par contre, la combinaison d'approches *in silico*, d'expériences de retard sur gel et de gènes rapporteurs dans plusieurs lignées cellulaires jumelée avec l'utilisation de critères conservateurs devrait augmenter nos chances de découvrir des génotypes/haplotypes ayant un impact biologique significatif.

5.14 Conclusion

En conclusion, l'analyse des promoteurs, arbitrairement définis comme étant 2kb en amont du premier exon, dans les gènes impliqués dans le contrôle du cycle cellulaire a permis d'identifier 136 pSNPs. À l'aide de prédictions *in silico*, nous avons déterminé que 70,6% de ces variants modifient potentiellement la liaison de facteurs de transcription. 13 de ces pSNPs mèneraient à l'apparition de nouveaux sites de liaison pour des facteurs de transcription. Nous avons procédé à la validation fonctionnelle de ces 13 variants avec des expériences de retard sur gel pour vérifier les prédictions. Cette analyse a permis de démontrer que 12 de ces polymorphismes mènent à une modification de la liaison à des facteurs de transcription. Les études de gènes rapporteurs ont permis de détecter des différences significatives pour 5 de ces variants situés dans la région régulatrice de gènes codant pour TGF β 1, E2F1, HDAC1, RB1 et CCND1 dans au moins une lignée cellulaire. Il reste à valider ces différences fonctionnelles dans des études *in vivo*, comme les modèles de souris. De plus, il serait intéressant d'évaluer l'impact de ces pSNPs fonctionnels dans des études d'association comprenant différentes cohortes avec des maladies complexes tel que le cancer. Finalement, cette étude a permis d'avoir une meilleure annotation de ces variations génétiques dans les régions promotrices et d'évaluer l'importance de ceux-ci dans la susceptibilité au cancer.

Références

1. Morgan, D.O., *Principles of CDK regulation*. Nature, 1995. **374**(6518): p. 131-4.
2. Fisher, R.P. and D.O. Morgan, *A novel cyclin associates with MO15/CDK7 to form the CDK-activating kinase*. Cell, 1994. **78**(4): p. 713-24.
3. Sherr, C.J. and J.M. Roberts, *Inhibitors of mammalian G1 cyclin-dependent kinases*. Genes Dev, 1995. **9**(10): p. 1149-63.
4. Magnaghi-Jaulin, L., et al., *Retinoblastoma protein represses transcription by recruiting a histone deacetylase*. Nature, 1998. **391**(6667): p. 601-5.
5. Dyson, N., *The regulation of E2F by pRB-family proteins*. Genes Dev, 1998. **12**(15): p. 2245-62.
6. Stevens, C. and N.B. La Thangue, *E2F and cell cycle control: a double-edged sword*. Arch Biochem Biophys, 2003. **412**(2): p. 157-69.
7. Johnson, D.G., et al., *Expression of transcription factor E2F1 induces quiescent cells to enter S phase*. Nature, 1993. **365**(6444): p. 349-52.
8. Diehl, J.A., et al., *Glycogen synthase kinase-3beta regulates cyclin D1 proteolysis and subcellular localization*. Genes Dev, 1998. **12**(22): p. 3499-511.
9. Sherr, C.J., *The Pezcoller lecture: cancer cell cycles revisited*. Cancer Res, 2000. **60**(14): p. 3689-95.
10. Tessema, M., U. Lehmann, and H. Kreipe, *Cell cycle and no end*. Virchows Arch, 2004. **444**(4): p. 313-23.
11. Sidorova, J.M. and L.L. Breeden, *Precocious G1/S transitions and genomic instability: the origin connection*. Mutat Res, 2003. **532**(1-2): p. 5-19.
12. Hall, M. and G. Peters, *Genetic alterations of cyclins, cyclin-dependent kinases, and Cdk inhibitors in human cancer*. Adv Cancer Res, 1996. **68**: p. 67-108.
13. Steeg, P.S. and Q. Zhou, *Cyclins and breast cancer*. Breast Cancer Res Treat, 1998. **52**(1-3): p. 17-28.
14. Schraml, P., et al., *Cyclin E overexpression and amplification in human tumours*. J Pathol, 2003. **200**(3): p. 375-82.
15. Sharpless, E. and L. Chin, *The INK4a/ARF locus and melanoma*. Oncogene, 2003. **22**(20): p. 3092-8.

16. Donovan, J. and J. Slingerland, *Transforming growth factor-beta and breast cancer: Cell cycle arrest by transforming growth factor-beta and its disruption in cancer*. Breast Cancer Res, 2000. 2(2): p. 116-24.
17. Malumbres, M. and M. Barbacid, *To cycle or not to cycle: a critical decision in cancer*. Nat Rev Cancer, 2001. 1(3): p. 222-31.
18. Busser, J., D.S. Geldmacher, and K. Herrup, *Ectopic cell cycle proteins predict the sites of neuronal cell death in Alzheimer's disease brain*. J Neurosci, 1998. 18(8): p. 2801-7.
19. McShea, A., et al., *Abnormal expression of the cell cycle regulators P16 and CDK4 in Alzheimer's disease*. Am J Pathol, 1997. 150(6): p. 1933-9.
20. Parker, L.L. and H. Piwnica-Worms, *Inactivation of the p34cdc2-cyclin B complex by the human WEE1 tyrosine kinase*. Science, 1992. 257(5078): p. 1955-7.
21. Liu, F., et al., *The human Myt1 kinase preferentially phosphorylates Cdc2 on threonine 14 and localizes to the endoplasmic reticulum and Golgi complex*. Mol Cell Biol, 1997. 17(2): p. 571-83.
22. Roshak, A.K., et al., *The human polo-like kinase, PLK, regulates cdc2/cyclin B through phosphorylation and activation of the cdc25C phosphatase*. Cell Signal, 2000. 12(6): p. 405-11.
23. Strausfeld, U., et al., *Dephosphorylation and activation of a p34cdc2/cyclin B complex in vitro by human CDC25 protein*. Nature, 1991. 351(6323): p. 242-5.
24. Nigg, E.A., *Cellular substrates of p34(cdc2) and its companion cyclin-dependent kinases*. Trends Cell Biol, 1993. 3(9): p. 296-301.
25. Sanchez, Y., et al., *Conservation of the Chk1 checkpoint pathway in mammals: linkage of DNA damage to Cdk regulation through Cdc25*. Science, 1997. 277(5331): p. 1497-501.
26. Peng, C.Y., et al., *Mitotic and G2 checkpoint control: regulation of 14-3-3 protein binding by phosphorylation of Cdc25C on serine-216*. Science, 1997. 277(5331): p. 1501-5.

27. Laronga, C., et al., *Association of the cyclin-dependent kinases and 14-3-3 sigma negatively regulates cell cycle progression*. J Biol Chem, 2000. **275**(30): p. 23106-12.
28. Jin, S., et al., *The GADD45 inhibition of Cdc2 kinase correlates with GADD45-mediated growth suppression*. J Biol Chem, 2000. **275**(22): p. 16602-8.
29. Smits, V.A., et al., *p21 inhibits Thr161 phosphorylation of Cdc2 to enforce the G2 DNA damage checkpoint*. J Biol Chem, 2000. **275**(39): p. 30638-43.
30. Shiloh, Y., *ATM and related protein kinases: safeguarding genome integrity*. Nat Rev Cancer, 2003. **3**(3): p. 155-68.
31. Bell, D.W., et al., *Heterozygous germ line hCHK2 mutations in Li-Fraumeni syndrome*. Science, 1999. **286**(5449): p. 2528-31.
32. King, M.C., J.H. Marks, and J.B. Mandell, *Breast and ovarian cancer risks due to inherited mutations in BRCA1 and BRCA2*. Science, 2003. **302**(5645): p. 643-6.
33. Nyberg, K.A., et al., *Toward maintaining the genome: DNA damage and replication checkpoints*. Annu Rev Genet, 2002. **36**: p. 617-56.
34. Venkitaraman, A.R., *Breast cancer genes and DNA repair*. Science, 1999. **286**(5442): p. 1100-2.
35. Lee, M.H. and H.Y. Yang, *Negative regulators of cyclin-dependent kinases and their roles in cancers*. Cell Mol Life Sci, 2001. **58**(12-13): p. 1907-22.
36. Smits, V.A. and R.H. Medema, *Checking out the G(2)/M transition*. Biochim Biophys Acta, 2001. **1519**(1-2): p. 1-12.
37. Betticher, D.C., et al., *Alternate splicing produces a novel cyclin D1 transcript*. Oncogene, 1995. **11**(5): p. 1005-11.
38. Le Marchand, L., et al., *Association of the cyclin D1 A870G polymorphism with advanced colorectal cancer*. Jama, 2003. **290**(21): p. 2843-8.
39. Csink, A.K. and S. Henikoff, *Something from nothing: the evolution and utility of satellite repeats*. Trends Genet, 1998. **14**(5): p. 200-4.
40. Jarman, A.P. and R.A. Wells, *Hypervariable minisatellites: recombinators or innocent bystanders?* Trends Genet, 1989. **5**(11): p. 367-71.

41. Karlin, S. and C. Burge, *Dinucleotide relative abundance extremes: a genomic signature*. Trends Genet, 1995. 11(7): p. 283-90.
42. Bois, P.R., *Hypermutable minisatellites, a human affair?* Genomics, 2003. 81(4): p. 349-55.
43. Oostra, B.A. and P. Chiurazzi, *The fragile X gene and its function*. Clin Genet, 2001. 60(6): p. 399-408.
44. Snyder, L. and W. Champness, *Molecular Genetics of Bacteria*. 1997, Washington, D.C.: ASM Press. 504.
45. Dobson-Stone, C., et al., *Comparison of fluorescent single-strand conformation polymorphism analysis and denaturing high-performance liquid chromatography for detection of EXT1 and EXT2 mutations in hereditary multiple exostoses*. Eur J Hum Genet, 2000. 8(1): p. 24-32.
46. Gross, E., et al., *A comparison of BRCA1 mutation analysis by direct sequencing, SSCP and DHPLC*. Hum Genet, 1999. 105(1-2): p. 72-8.
47. Spiegelman, J.I., M.N. Mindrinos, and P.J. Oefner, *High-accuracy DNA sequence variation screening by DHPLC*. Biotechniques, 2000. 29(5): p. 1084-90, 1092.
48. Xiao, W. and P.J. Oefner, *Denaturing high-performance liquid chromatography: A review*. Hum Mutat, 2001. 17(6): p. 439-74.
49. Peltonen, L. and V.A. McKusick, *Genomics and medicine. Dissecting human disease in the postgenomic era*. Science, 2001. 291(5507): p. 1224-9.
50. Quandt, K., et al., *MatInd and MatInspector: new fast and versatile tools for detection of consensus matches in nucleotide sequence data*. Nucleic Acids Res, 1995. 23(23): p. 4878-84.
51. Wasserman, W.W., et al., *Human-mouse genome comparisons to locate regulatory sites*. Nat Genet, 2000. 26(2): p. 225-8.
52. Hardison, R.C., J. Oeltjen, and W. Miller, *Long human-mouse sequence alignments reveal novel regulatory elements: a reason to sequence the mouse genome*. Genome Res, 1997. 7(10): p. 959-66.

53. Knight, J.C., *Functional implications of genetic variation in non-coding DNA for disease susceptibility and gene regulation*. Clin Sci (Lond), 2003. **104**(5): p. 493-501.
54. Helms, C., et al., *A putative RUNX1 binding site variant between SLC9A3R1 and NAT9 is associated with susceptibility to psoriasis*. Nat Genet, 2003. **35**(4): p. 349-56.
55. Rockman, M.V. and G.A. Wray, *Abundant raw material for cis-regulatory evolution in humans*. Mol Biol Evol, 2002. **19**(11): p. 1991-2004.
56. Coleman, S.L., et al., *Streamlined approach to functional analysis of promoter-region polymorphisms*. Biotechniques, 2002. **33**(2): p. 412, 414, 416 passim.
57. Hoogendoorn, B., et al., *Functional analysis of human promoter polymorphisms*. Hum Mol Genet, 2003. **12**(18): p. 2249-54.
58. Carey, M. and S.T. Smale, *Transcriptional regulation in eukaryotes: concepts, strategies, and techniques*. 1 ed. 2000, Cold Spring Harbor: Cold Spring Harbor Laboratory Press. 640.
59. Smale, S.T., *Core promoter architecture for eucaryotic protein-coding genes*. In *transcription: Mechanism and regulation*, ed. R.C. Conaway and J.W. Conaway. 1994, New York: Raven Press. 81.
60. Burley, S.K. and R.G. Roeder, *Biochemistry and structural biology of transcription factor IID (TFIID)*. Annu Rev Biochem, 1996. **65**: p. 769-99.
61. Smale, S.T., *Transcription initiation from TATA-less promoters within eukaryotic protein-coding genes*. Biochim Biophys Acta, 1997. **1351**(1-2): p. 73-88.
62. Lagrange, T., et al., *New core promoter element in RNA polymerase II-dependent transcription: sequence-specific DNA binding by transcription factor IIB*. Genes Dev, 1998. **12**(1): p. 34-44.
63. Burke, T.W. and J.T. Kadonaga, *The downstream core promoter element, DPE, is conserved from Drosophila to humans and is recognized by TAFII60 of Drosophila*. Genes Dev, 1997. **11**(22): p. 3020-31.
64. Pitarque, M., et al., *A nicotine C-oxidase gene (CYP2A6) polymorphism important for promoter activity*. Hum Mutat, 2004. **23**(3): p. 258-66.

65. Funke-Kaiser, H., et al., *Differential binding of transcription factor E2F-2 to the endothelin-converting enzyme-1b promoter affects blood pressure regulation*. Hum Mol Genet, 2003. **12**(4): p. 423-33.
66. Escary, J.L., et al., *Evaluation of DHPLC analysis in mutational scanning of Notch3, a gene with a high G-C content*. Hum Mutat, 2000. **16**(6): p. 518-26.
67. Schneider, J.A., et al., *DNA variability of human genes*. Mech Ageing Dev, 2003. **124**(1): p. 17-25.
68. Cargill, M., et al., *Characterization of single-nucleotide polymorphisms in coding regions of human genes*. Nat Genet, 1999. **22**(3): p. 231-8.
69. Pollock, P.M., J.V. Pearson, and N.K. Hayward, *Compilation of somatic mutations of the CDKN2 gene in human cancers: non-random distribution of base substitutions*. Genes Chromosomes Cancer, 1996. **15**(2): p. 77-88.
70. Krawczak, M., E.V. Ball, and D.N. Cooper, *Neighboring-nucleotide effects on the rates of germ-line single-base-pair substitution in human genes*. Am J Hum Genet, 1998. **63**(2): p. 474-88.
71. Rogozin, I.B., et al., *Somatic mutation hotspots correlate with DNA polymerase eta error spectrum*. Nat Immunol, 2001. **2**(6): p. 530-6.
72. Templeton, A.R., et al., *Recombinational and mutational hotspots within the human lipoprotein lipase gene*. Am J Hum Genet, 2000. **66**(1): p. 69-83.
73. Holliday, R. and G.W. Grigg, *DNA methylation and mutation*. Mutat Res, 1993. **285**(1): p. 61-7.
74. Bird, A.P., *CpG-rich islands and the function of DNA methylation*. Nature, 1986. **321**(6067): p. 209-13.
75. Cooper, D.N., M.H. Taggart, and A.P. Bird, *Unmethylated domains in vertebrate DNA*. Nucleic Acids Res, 1983. **11**(3): p. 647-58.
76. Tomso, D.J. and D.A. Bell, *Sequence context at human single nucleotide polymorphisms: overrepresentation of CpG dinucleotide at polymorphic sites and suppression of variation in CpG islands*. J Mol Biol, 2003. **327**(2): p. 303-8.
77. Collins, D.W. and T.H. Jukes, *Rates of transition and transversion in coding sequences since the human-rodent divergence*. Genomics, 1994. **20**(3): p. 386-96.

78. Jegga, A.G., et al., *Detection and visualization of compositionally similar cis-regulatory element clusters in orthologous and coordinately controlled genes.* Genome Res, 2002. **12**(9): p. 1408-17.
79. Tronche, F., et al., *Analysis of the distribution of binding sites for a tissue-specific transcription factor in the vertebrate genome.* J Mol Biol, 1997. **266**(2): p. 231-45.
80. Fickett, J.W., *Quantitative discrimination of MEF2 sites.* Mol Cell Biol, 1996. **16**(1): p. 437-41.
81. Brickner, A.G., et al., *Genomic sequence comparison of the human and mouse adenosine deaminase gene regions.* Mamm Genome, 1999. **10**(2): p. 95-101.
82. Loots, G.G., et al., *Identification of a coordinate regulator of interleukins 4, 13, and 5 by cross-species sequence comparisons.* Science, 2000. **288**(5463): p. 136-40.
83. Hardison, R., et al., *Locus control regions of mammalian beta-globin gene clusters: combining phylogenetic analyses and experimental results to gain functional insights.* Gene, 1997. **205**(1-2): p. 73-94.
84. Fishman, D., et al., *The effect of novel polymorphisms in the interleukin-6 (IL-6) gene on IL-6 transcription and plasma IL-6 levels, and an association with systemic-onset juvenile chronic arthritis.* J Clin Invest, 1998. **102**(7): p. 1369-76.
85. Su, Z., et al., *A novel allele in the promoter of the hepatic lipase is associated with increased concentration of HDL-C and decreased promoter activity.* J Lipid Res, 2002. **43**(10): p. 1595-601.
86. Hoogendoorn, B., et al., *Functional analysis of polymorphisms in the promoter regions of genes on 22q11.* Hum Mutat, 2004. **24**(1): p. 35-42.
87. Buckland, P.R., et al., *A high proportion of chromosome 21 promoter polymorphisms influence transcriptional activity.* Gene Expr, 2004. **11**(5-6): p. 233-9.
88. Massague, J., *TGF-beta signal transduction.* Annu Rev Biochem, 1998. **67**: p. 753-91.

89. de Souza, A.P., et al., *Analysis of the TGF-beta1 promoter polymorphism (C-509T) in patients with chronic periodontitis*. J Clin Periodontol, 2003. **30**(6): p. 519-23.
90. Grainger, D.J., et al., *Genetic control of the circulating concentration of transforming growth factor type beta1*. Hum Mol Genet, 1999. **8**(1): p. 93-7.
91. Hobbs, K., et al., *Interleukin-10 and transforming growth factor-beta promoter polymorphisms in allergies and asthma*. Am J Respir Crit Care Med, 1998. **158**(6): p. 1958-62.
92. Kim, Y.J., et al., *Association of transforming growth factor-beta1 gene polymorphisms with a hepatocellular carcinoma risk in patients with chronic hepatitis B virus infection*. Exp Mol Med, 2003. **35**(3): p. 196-202.
93. Luedeking, E.K., et al., *Analysis of genetic polymorphisms in the transforming growth factor-beta1 gene and the risk of Alzheimer's disease*. Hum Genet, 2000. **106**(5): p. 565-9.
94. Shu, X.O., et al., *Genetic polymorphisms in the TGF-beta 1 gene and breast cancer survival: a report from the Shanghai Breast Cancer Study*. Cancer Res, 2004. **64**(3): p. 836-9.
95. Silverman, E.S., et al., *Transforming growth factor-beta1 promoter polymorphism C-509T is associated with asthma*. Am J Respir Crit Care Med, 2004. **169**(2): p. 214-9.
96. Yang, Y.H., et al., *The association between transforming growth factor-beta gene promoter C-509T polymorphism and Chinese children with Henoch-Schonlein purpura*. Pediatr Nephrol, 2004. **19**(9): p. 972-5.
97. Bossone, S.A., et al., *MAZ, a zinc finger protein, binds to c-MYC and C2 gene sequences regulating transcriptional initiation and termination*. Proc Natl Acad Sci U S A, 1992. **89**(16): p. 7452-6.
98. Song, J., et al., *Genomic organization and expression of a human gene for Myc-associated zinc finger protein (MAZ)*. J Biol Chem, 1998. **273**(32): p. 20603-14.
99. Parks, C.L. and T. Shenk, *The serotonin 1a receptor gene contains a TATA-less promoter that responds to MAZ and Sp1*. J Biol Chem, 1996. **271**(8): p. 4417-30.

100. Uchida, S., et al., *Transcriptional regulation of the CLC-K1 promoter by myc-associated zinc finger protein and kidney-enriched Kruppel-like factor, a novel zinc finger repressor*. Mol Cell Biol, 2000. **20**(19): p. 7319-31.
101. Derynck, R., R.J. Akhurst, and A. Balmain, *TGF-beta signaling in tumor suppression and cancer progression*. Nat Genet, 2001. **29**(2): p. 117-29.
102. Pasche, B., et al., *TbetaR-I(6A) is a candidate tumor susceptibility allele*. Cancer Res, 1999. **59**(22): p. 5678-82.
103. Tang, B., et al., *Transforming growth factor-beta1 is a new form of tumor suppressor with true haploid insufficiency*. Nat Med, 1998. **4**(7): p. 802-7.
104. Pierce, D.F., Jr., et al., *Mammary tumor suppression by transforming growth factor beta 1 transgene expression*. Proc Natl Acad Sci U S A, 1995. **92**(10): p. 4254-8.
105. Dickson, R.B., et al., *Activation of growth factor secretion in tumorigenic states of breast cancer induced by 17 beta-estradiol or v-Ha-ras oncogene*. Proc Natl Acad Sci U S A, 1987. **84**(3): p. 837-41.
106. DeGregori, J., et al., *Distinct roles for E2F proteins in cell growth control and apoptosis*. Proc Natl Acad Sci U S A, 1997. **94**(14): p. 7245-50.
107. Sherr, C.J., *Cancer cell cycles*. Science, 1996. **274**(5293): p. 1672-7.
108. Onda, M., et al., *Up-regulation of transcriptional factor E2F1 in papillary and anaplastic thyroid cancers*. J Hum Genet, 2004. **49**(6): p. 312-8.
109. Yamasaki, L., et al., *Loss of E2F-1 reduces tumorigenesis and extends the lifespan of Rb1(+/-)mice*. Nat Genet, 1998. **18**(4): p. 360-4.
110. Pan, H., et al., *Key roles for E2F1 in signaling p53-dependent apoptosis and in cell division within developing tumors*. Mol Cell, 1998. **2**(3): p. 283-92.
111. Andres, V., B. Nadal-Ginard, and V. Mahdavi, *Clox, a mammalian homeobox gene related to Drosophila cut, encodes DNA-binding regulatory proteins differentially expressed during development*. Development, 1992. **116**(2): p. 321-34.
112. Dufort, D. and A. Nepveu, *The human cut homeodomain protein represses transcription from the c-myc promoter*. Mol Cell Biol, 1994. **14**(6): p. 4251-7.

113. Ogami, K., *Gene expression and transcriptional regulation of thrombopoietin*. Stem Cells, 1996. **14 Suppl 1**: p. 148-53.
114. Bartholomew, C., et al., *The Evi-1 proto-oncogene encodes a transcriptional repressor activity associated with transformation*. Oncogene, 1997. **14(5)**: p. 569-77.
115. Morishita, K., et al., *EVI-1 zinc finger protein works as a transcriptional activator via binding to a consensus sequence of GACAAGATAAGATAANI-28 CTCATCTTC*. Oncogene, 1995. **10(10)**: p. 1961-7.
116. Tanaka, T., et al., *Dual functions of the AML1/Evi-1 chimeric protein in the mechanism of leukemogenesis in t(3;21) leukemias*. Mol Cell Biol, 1995. **15(5)**: p. 2383-92.
117. Tanaka, T., et al., *Evi-1 raises AP-1 activity and stimulates c-fos promoter transactivation with dependence on the second zinc finger domain*. J Biol Chem, 1994. **269(39)**: p. 24020-6.
118. Welch, J.J., et al., *Global regulation of erythroid gene expression by transcription factor GATA-1*. Blood, 2004.
119. Grass, J.A., et al., *GATA-1-dependent transcriptional repression of GATA-2 via disruption of positive autoregulation and domain-wide chromatin remodeling*. Proc Natl Acad Sci U S A, 2003. **100(15)**: p. 8811-6.
120. Rylski, M., et al., *GATA-1-mediated proliferation arrest during erythroid maturation*. Mol Cell Biol, 2003. **23(14)**: p. 5031-42.
121. Dou, S., et al., *The recombination signal sequence-binding protein RBP-2N functions as a transcriptional repressor*. Mol Cell Biol, 1994. **14(5)**: p. 3310-9.
122. Kannabiran, C., X. Zeng, and L.D. Vales, *The mammalian transcriptional repressor RBP (CBFI) regulates interleukin-6 gene expression*. Mol Cell Biol, 1997. **17(1)**: p. 1-9.
123. Cowell, J.K., B. Bia, and A. Akoulitchev, *A novel mutation in the promotor region in a family with a mild form of retinoblastoma indicates the location of a new regulatory domain for the RB1 gene*. Oncogene, 1996. **12(2)**: p. 431-6.
124. Sakai, T., et al., *Oncogenic germ-line mutations in Sp1 and ATF sites in the human retinoblastoma gene*. Nature, 1991. **353(6339)**: p. 83-6.

125. Wong, K.K., et al., *A role for c-Abl in c-myc regulation*. *Oncogene*, 1995. **10**(4): p. 705-11.
126. Clepet, C., et al., *The human SRY transcript*. *Hum Mol Genet*, 1993. **2**(12): p. 2007-12.
127. Cereghini, S., *Liver-enriched transcription factors and hepatocyte differentiation*. *Faseb J*, 1996. **10**(2): p. 267-82.
128. Cress, W.D. and E. Seto, *Histone deacetylases, transcriptional control, and cancer*. *J Cell Physiol*, 2000. **184**(1): p. 1-16.
129. Archer, S.Y., et al., *p21(WAF1) is required for butyrate-mediated growth inhibition of human colon cancer cells*. *Proc Natl Acad Sci U S A*, 1998. **95**(12): p. 6791-6.
130. Inga, A., et al., *Differential transactivation by the p53 transcription factor is highly dependent on p53 level and promoter target sequence*. *Mol Cell Biol*, 2002. **22**(24): p. 8612-25.
131. Qian, H., et al., *Groups of p53 target genes involved in specific p53 downstream effects cluster into different classes of DNA binding sites*. *Oncogene*, 2002. **21**(51): p. 7901-11.
132. Holley, S.L., et al., *Cyclin D1 polymorphism and expression in patients with squamous cell carcinoma of the head and neck*. *Am J Pathol*, 2001. **159**(5): p. 1917-24.
133. McKay, J.A., et al., *Cyclin D1 protein expression and gene polymorphism in colorectal cancer. Aberdeen Colorectal Initiative*. *Int J Cancer*, 2000. **88**(1): p. 77-81.
134. el-Deiry, W.S., et al., *WAF1, a potential mediator of p53 tumor suppression*. *Cell*, 1993. **75**(4): p. 817-25.
135. Ng, I.O., et al., *Expression of p21/waf1 in oral squamous cell carcinomas-- correlation with p53 and mdm2 and cellular proliferation index*. *Oral Oncol*, 1999. **35**(1): p. 63-9.

Annexe I

Tableau I : Liste des amorces utilisées pour l'amplification d'ADN génomique pour l'analyse au dHPLC ainsi que les conditions PCR utilisées.

| Gène | Fragment | Nom des amorces | Séquences (5' -3') | Clamp (5' -3') | * | ** |
|--------|----------|---------------------|-----------------------------|----------------|---|----|
| CCND1 | 2 | CCND1.2F.EX1.-789 | CCGACTGGTCAAGGTAGGAA | | 1 | A |
| | | CCND1.2R.EX1.-383 | GCGACACCCCATATCCAAG | | | |
| CCND1 | 3 | CCND1.3F.EX1.-1146 | CGAACACCTATCGATTTTGC | | 1 | A |
| | | CCND1.3R.EX1.-721 | AGCATTTCTTTCATCTTGTCC | | | |
| CCND1 | 4 | CCND1.4F.EX1.-1521 | AGTAATCCTAATTCCTGGCG | | 6 | D |
| | | CCND1.4R.EX1.-1127 | GCAAAATCGATAGGTGTTCG | | | |
| CCND1 | 5 | CCND1.5F.EX1.-1776 | GAGAGGCCGGGTGATAGC | | 6 | D |
| | | CCND1.5R.EX1.-1352 | GTGTCATCAAACCACCGTACC | | | |
| CCND1 | 6 | CCND1.6F.EX1.-2069 | CITGCAATTTTGGCTCGT | GCCCCGCCG | 6 | D |
| | | CCND1.6R.EX1.-1654 | TAGCCAAGCCTCAGAGCATC | | | |
| CCNE1 | 2 | CCNE1.2F.EX1.-812 | AAGGTCTTCAGAGGCCAGG | GCCCCGCCG | 1 | A |
| | | CCNE1.2R.EX1.-404 | CGGGTGGAAATGTAAACACG | | | |
| CCNE1 | 3 | CCNE1.3F.EX1.-1189 | AAGACCCTATCCCTACCTGC | | 6 | D |
| | | CCNE1.3R.EX1.-789 | CCTTCCTGGCTCTCTGAAGAC | | | |
| CCNE1 | 4 | CCNE1.4F.EX1.-1543 | GCATTCCAGAGCCTTCTTTTAC | | 1 | A |
| | | CCNE1.4R.EX1.-1115 | CTGTCCTTCTCCCTCAAGACT | | | |
| CCNE1 | 5 | CCNE1.5F.EX1.-1875 | GTGAAACAGGATAATTGTGGC | | 1 | A |
| | | CCNE1.5R.EX1.-1502 | TCAGCTTCAGGAAGATGTGC | | | |
| CCNE1 | 6 | CCNE1.6F.EX1.-2039 | TGGTGGGTGTATACATTGCTGT | GCCCCGCCG | 1 | A |
| | | CCNE1.6R.EX1.-1737 | CCGTTAGGAATCACCAGTAACA | | | |
| CDC25A | 1 | CDC25A.1F.EX1.48 | CCGACCTACACCTCTTACCC | | 4 | C |
| | | CDC25A.1R.EX1.-353 | AGCTAACTGGACGGCAGC | | | |
| CDC25A | 2 | CDC25A.2F.EX1.-315 | GAAGCGTAGTTCCCATCGC | | 1 | A |
| | | CDC25A.2R.EX1.-713 | CTAAGAGCATCCAATAAGGGC | | | |
| CDC25A | 3 | CDC25A.3F.EX1.-642 | CAGTACAACCTGTGCTGTGG | | 1 | A |
| | | CDC25A.3R.EX1.-987 | AATCATGCACCTATGGCCC | | | |
| CDC25A | 4 | CDC25A.4F.EX1.-952 | TCCTAGACGGATTTTGAGGG | | 1 | A |
| | | CDC25A.4R.EX1.-1322 | GTGCAGAGAGGCCATAATAGC | | | |
| CDC25A | 5 | CDC25A.5F.EX1.-1301 | GGCTATTTAGGCCTCTCTGC | | 1 | A |
| | | CDC25A.5R.EX1.-1750 | GCAGTAACAGGAATGTTGAGC | | | |
| CDC25A | 6 | CDC25A.6F.EX1.-1721 | CAAAACTTGGCTCAACATTCC | GCCCCGCCG | 5 | B |
| | | CDC25A.6R.EX1.-2051 | TTTCCTTAGAAAGCACTTCCC | | | |
| CDK4 | 1 | CDK4.1F.EX1.40 | GTGACCAGCTGCCAAAGAG | | 1 | A |
| | | CDK4.1R.EX1.-329 | GGTAGTGAGACAATCCTCAGC | | | |
| CDK4 | 3 | CDK4.3F.EX1.-540 | TTTTTGGTGTCTGTCACTGG | | 1 | A |
| | | CDK4.3R.EX1.-871 | CATAATCATAGAACCCTTCTACTGG | | | |
| CDK4 | 4 | CDK4.4F.EX1.-804 | CAGGAGACAGGAATACTTGACC | | 1 | A |
| | | CDK4.4R.EX1.-1210 | GGAGCCAATAAAACAAAGTAGG | | | |
| CDK4 | 5 | CDK4.5F.EX1.-1124 | TCCAGAACACACTTGTTCCT | | 1 | A |
| | | CDK4.5R.EX1.-1475 | GGGCTACAAGTATCAGGATGG | | | |
| CDK4 | 6 | CDK4.6F.EX1.-1391 | AGATGAATATAAACCATGACTTAGAGC | | 1 | A |
| | | CDK4.6R.EX1.-1742 | CAAGCAGACAGAGAGGACG | | | |
| CDK4 | 7 | CDK4.7F.EX1.-1681 | TCCACATAAGACCTTACACAGC | | 1 | A |
| | | CDK4.7R.EX1.-2044 | TCTGAGGCCAAGTTCTACCC | | | |
| CDK6 | 1 | CDK6.1F.EX1.72 | GGGGTGGCTCAACTAGC | | 6 | D |
| | | CDK6.1R.EX1.-351 | CTTCGTAAGGAGGAGGAGG | | | |
| CDK6 | 2 | CDK6.2F.EX1.-201 | ACACGATTACATAGCCTCTGC | | 5 | B |
| | | CDK6.2R.EX1.-604 | CCGTGACATTGACGTGCC | | | |
| CDK6 | 3 | CDK6.3F.EX1.-587 | GGCACGTCAATGTACACGG | | 1 | A |
| | | CDK6.3R.EX1.-960 | AGAGAATTCGGTTTCCAACC | | | |

| | | | | | |
|--------|---|--|--|---|---|
| CDK6 | 4 | CDK6.4F.EX1.-927 CDK6.4R.EX1.-1330 | GGTAAAGGAGTCTCGGTTGG GAATGGGAAAAGGTTCCG | 5 | B |
| CDK6 | 5 | CDK6.5F.EX1.-1282 CDK6.5R.EX1.-1709 | GAGAAGGTCTCTGTCCCG AATGTAGTGAGTGCCAGAAGC | 1 | A |
| CDK6 | 6 | CDK6.6F.EX1.-1625 CDK6.6R.EX1.-2001 | AGGAATTAACAAACGGCG GTTTCTCTCCGCCAAGC | 6 | D |
| CDKN1A | 1 | CDKN1A.F.EX1.-351 CDKN1A.R.EX1.19 | AGAAAGGGGGCTCATTCTAAC AGCTGCTCACACCTCAGC | 6 | D |
| CDKN1A | 2 | CDKN1A.F.EX1.-468 CDKN1A.R.EX1.-266 | GCAGAGAGGTGCATCGTTTT GAGCCACAAATCTGGCTTTT | 1 | A |
| CDKN1A | 3 | CDKN1A.3F.EX1.-796 CDKN1A.3R.EX1.-448 | CAGGGATTCTCTGTTCAGG AAAAACGATGCACCTCTCTG | 1 | A |
| CDKN1A | 4 | CDKN1A.4F.EX1.-1077 CDKN1A.4R.EX1.-703 | TCATCTCCATAAACCATCTGC AAAAAGGGCAACCTGATCTC | 1 | A |
| CDKN1A | 5 | CDKN1A.F.EX1.-1363 CDKN1A.R.EX1.-1055 | GTCTGGCAGAGATTCCAG TTGCAGATGGTTTATGGAGATG | 1 | A |
| CDKN1A | 6 | CDKN1A.F.EX1.-1600 CDKN1A.R.EX1.-1206 | TTGGATGTATAGGAGCGAAGG TCTGTGCTGAAACATTTGC | 1 | A |
| CDKN1A | 7 | CDKN1A.F.EX1.-1796 CDKN1A.R.EX1.-1579 | TTTCATTGTGAAGCTCAGTACC ACCTTCGCTCTATACATCC | 1 | A |
| CDKN1A | 8 | CDKN1A.F.EX1.-2071 CDKN1A.R.EX1.-1704 | ATCTGATGCATGTGTGCTTG TGTCTGTCTGTTACTTTGGGG | 1 | A |
| CDKN1B | 1 | CDKN1B.1F.EX1.70 CDKN1B.1R.EX1.-277 | AGAAAAACACCCGAAAAGA CGTTTGGCTAGTTTGTGTCT | 6 | D |
| CDKN1B | 2 | CDKN1B.2F.EX1.-257 CDKN1B.2R.EX1.-612 | GACAAACAACTAGCCAAACG GCAGAACTTCTGGGTTAAGG | 6 | D |
| CDKN1B | 3 | CDKN1B.3F.EX1.-581 CDKN1B.3R.EX1.-880 | GGTTCGCTCAGCCTTAACC CTTAACTGTGCTTGGGAAGG | 4 | C |
| CDKN1B | 4 | CDKN1B.4F.EX1.-817 CDKN1B.4R.EX1.-1127 | GTTTAGTTTTGAAAACGAGGG TTCATCATCTGGAGTTTGACC | 1 | A |
| CDKN1B | 5 | CDKN1B.5F.EX1.-1062 CDKN1B.5R.EX1.-1421 | AGTGAATGCTTTCAGAGGGC TGCTTAAGAAAACTTGCTGC | 1 | A |
| CDKN1B | 6 | CDKN1B.6F.EX1.-1360 CDKN1B.6R.EX1.-1770 | GCATTTTACTGGAAACCAACC CACTTGTAAGACAGCAGGACC | 1 | A |
| CDKN1B | 7 | CDKN1B.7F.EX1.-1698 CDKN1B.7R.EX1.-2023 | GGTTCCTACTGGTTCAGGG TGAGACCTCTTGTGATCC | 1 | A |
| CDKN2A | 1 | CDKN2A.1F.EX1.68 CDKN2A.1R.EX1.-233 | TCGGAGGAGGTGCTATTAAC ATTTGGCAGTTAGGAAGGTTG | 5 | B |
| CDKN2A | 2 | CDKN2A.2F.EX1.-167 CDKN2A.2R.EX1.-534 | TGTTAAAAAGAAATCCGCC AATGTACAGTCTGAGAAACCC | 1 | A |
| CDKN2A | 3 | CDKN2A.3F.EX1.-489 CDKN2A.3R.EX1.-854 | GTGTAACCGCTGTCTGGG ACAACCCCTATAAAGCAGGG | 1 | A |
| CDKN2A | 4 | CDKN2A.4F.EX1.-701 CDKN2A.4R.EX1.-997 | GCAGAGTTGCACAGTGATCC CAAAGCATTTTCTTTCTATGCC | 1 | A |
| CDKN2A | 5 | CDKN2A.5F.EX1.-961 CDKN2A.5R.EX1.-1344 | GCAATCTGTTTTGAGGCATAG CATCCGGACTTTCTTTTATG | 1 | A |
| CDKN2A | 6 | CDKN2A.6F.EX1.-1326 CDKN2A.6R.EX1.-1712 | TAAAAGGAAAAGTCCGGATGC TTTGTAAAGTGAGGGTGGTGG | 1 | A |
| CDKN2A | 7 | CDKN2A.7F.EX1.-1671 CDKN2A.7R.EX1.-2082 | CTGAAAAGATTATGGGCACC CCCCGAGAAAATATAAATGG | 1 | A |
| CDKN2B | 1 | CDKN2B.1F.EX1.-333 CDKN2B.1R.EX1.88 | CTTGATGACAGGTGCAGAGC AATTCCGTTTTCAGCTGGG | 1 | A |
| CDKN2B | 2 | CDKN2B.2F.EX1.-631 CDKN2B.2R.EX1.-224 | CTGTACAAATATGATGAAACTGGG TTCCTGTGGAGACGTTGG | 1 | A |
| CDKN2B | 3 | CDKN2B.3F.EX1.-967 CDKN2B.3R.EX1.-538 | TCCTCACCAAAGTTTTTAGGC AGTTCGAGATTGAGAGTGGC | 1 | A |
| CDKN2B | 4 | CDKN2B.4F.EX1.-1339 | TGAAGACTAGGAAATGGGGG | 1 | A |

| | | | | | |
|--------|---|--|--|---|---|
| CDKN2B | 5 | CDKN2B.4R.EX1.-919 CDKN2B.5F.EX1.-1669 CDKN2B.5R.EX1.-1242 | GAAGAGAACC GCAAGTTATGG GCAACGTTTTCCCTACCC CAACAGAGCTGAAATCCACC | 1 | A |
| CDKN2B | 6 | CDKN2B.6F.EX1.-2043 CDKN2B.6R.EX1.-1646 | AAAATTGAAATATGGACACATGC AACTTTGGGTAGGGAAAAACG | 1 | A |
| E2F1 | 1 | E2F1.1F.EX1.-3 E2F1.1R.EX1.-492 | ACTTTTACGCGCCAAATCC GAGGCAAGTTGAGGATGGAA | 6 | D |
| E2F1 | 2 | E2F1.2F.EX1.-392 E2F1.2R.EX1.-797 | CAGCCTCAGCTTAGATTCCA GACGCCACTTCATCGTATTGT | 1 | A |
| E2F1 | 3 | E2F1.3F.EX1.-735 E2F1.3R.EX1.-1125 | CACAGCCACTGACAGGGG GAGTAAATCCTTGGGAGGGC | 1 | A |
| E2F1 | 4 | E2F1.4F.EX1.-1104 E2F1.4R.EX1.-1533 | GAGCCCTCCCAAGGATTAC GATACTGGCTTTTCTAATTCCTAA | 1 | A |
| E2F1 | 5 | E2F1.5F.EX1.-1457 E2F1.5R.EX1.-2008 | TCCCCTTTCCAGGTGATATTT CAGCAAGAGGCTTATGTGGTC | 1 | A |
| HDAC1 | 1 | HDAC1.1F.EX1.-395 HDAC1.1R.EX1.50 | AGCCCTTTTATTGGCTGAGTG CGTCCCTACCGTCAGTCG | 6 | D |
| HDAC1 | 2 | HDAC1.2F.EX1.-880 HDAC1.2R.EX1.-367 | AAAGGAAACCTGCCCTCCTAT CACAAGGTCACTCAGCCAATAA | 1 | A |
| HDAC1 | 3 | HDAC1.3F.EX1.-1307 HDAC1.3R.EX1.-822 | GCACAGAACAGCTTCACAGAA GCGTCTCTTTAGACAGGAGGAA | 1 | A |
| HDAC1 | 4 | HDAC1.4F.EX1.-1595 HDAC1.4R.EX1.-1255 | TGCACTTTTAGGCTTTTCAGC TTTTGTAGGAAGCTATTCTGTGC | 1 | A |
| HDAC1 | 5 | HDAC1.5F.EX1.-2060 HDAC1.5R.EX1.-1528 | TGAGTGAGAAACACAAGTCCC AGAACCCTCCACCAGATCA | 1 | A |
| MADH3 | 1 | MADH3.1F.EX1.-62 MADH3.1R.EX1.187 | CCGAGAGTTGAGGCGAAGTT TCTTGACCAGGCTCTTGACC | 6 | D |
| MADH3 | 3 | MADH3.3F.EX1.-825 MADH3.3R.EX1.-472 | AGTGTGGACTCCGAGAGCG TCCCAGTCTGTGTTTCATGC | 6 | D |
| MADH3 | 4 | MADH3.4F.EX1.-1157 MADH3.4R.EX1.-742 | ATTGTCCAGGTTGCTACTGG GGAAGTCTGGGAAACACG | 6 | D |
| MADH3 | 5 | MADH3.5F.EX1.-1430 MADH3.5R.EX1.-1110 | CACAAACTCTCCGAAAACCTCC GGTCTCAGAAGTGGAGAGC | 1 | A |
| MADH3 | 6 | MADH3.6F.EX1.-1722 MADH3.6R.EX1.-1312 | TCACAGGGCTTTGTAACCTC CTCCTGATAGAGAAGGATCGG | 2 | B |
| MADH3 | 7 | MADH3.7F.EX1.-1999 MADH3.7R.EX1.-1679 | CTTGCTCTCTTGACGGACG TATCTGGAACCTCAAGTGG | 1 | A |
| MADH4 | 1 | MADH4.1F.EX1.-339 MADH4.1R.EX1.9 | TTTATAAACACTTACTGGACGCC TGTTGTTGCTGCCAATTGT | 6 | D |
| MADH4 | 2 | MADH4.2F.EX1.-648 MADH4.2R.EX1.-234 | AGTGCCCTACATCTCTTCC GTTTGAGGGAGTGGTCCG | 1 | A |
| MADH4 | 3 | MADH4.3F.EX1.-947 MADH4.3R.EX1.-625 | CTGAGTATGAGTGGCGAAGG TGTTGGAAGAGATGTAGGCG | 1 | A |
| MADH4 | 4 | MADH4.4F.EX1.-1269 MADH4.4R.EX1.-850 | ACCGATTTTTGAATTGTCCC TCCAAAACCAATGATTAGCC | 1 | A |
| MADH4 | 5 | MADH4.5F.EX1.-1539 MADH4.5R.EX1.-1199 | TTTGTGGTGAGATTGTTATACAGC CCACTACTAAGGTGCATGGG | 1 | A |
| MADH4 | 6 | MADH4.6F.EX1.-1797 MADH4.6R.EX1.-1485 | CTGCCATATTGTTAGCTTCCC TCTTCTATTCCGTCTGTATTAGTTACC | 1 | A |
| RB1 | 1 | RB1.1F.EX1.-339 RB1.1R.EX1.22 | AGTCTTCCCTATCAGACCCC ACGTCCCCTGAGAAAAACC | 6 | D |
| RB1 | 2 | RB1.2F.EX1.-639 RB1.2R.EX1.-273 | GTGCTTCCCGTCTTACACC GCTATACAAAGAGTCTGGTGGG | 1 | A |
| RB1 | 3 | RB1.3F.EX1.-954 RB1.3R.EX1.-581 | TCTTAGCTAATGCATGCAAGG GGGGAATGAACGGATAGTAGTG | 1 | A |
| RB1 | 4 | RB1.4F.EX1.-1328 RB1.4R.EX1.-935 | TCTCTTTCCTTCTCAGTTGACG CTTGCATGCATTAGCTAAGAAAC | 1 | A |

| | | | | | |
|-------|---|--------------------|--------------------------|------------|---|
| RB1 | 5 | RB1.5F.EX1.-1688 | CTTGCCATAACTAATCTGGGG | 1 | A |
| | | RB1.5R.EX1.-1301 | TGAACTCGTCAACTGAGAAGG | | |
| RB1 | 6 | RB1.6F.EX1.-2022 | AGTCATGATGCTTTCTGTTGC | 1 | A |
| | | RB1.6R.EX1.-1651 | TACTTCCATTACCTCACCC | | |
| SKP2 | 1 | SKP2.1F.EX1.12 | CTCTACAGCCCCGCTCTGC | 1 | A |
| | | SKP2.1R.EX1.-247 | CGAGAAAGGAAAATCCGTCTAC | | |
| SKP2 | 2 | SKP2.2F.EX1.-220 | TGGACTGTAGACGGATTTTCC | 1 | A |
| | | SKP2.2R.EX1.-630 | ACAGACCAATTCTTAGGGC | | |
| SKP2 | 3 | SKP2.3F.EX1.-581 | GAAGCTTCTCTCCGTCG | 1 | A |
| | | SKP2.3R.EX1.-985 | GCTCCCTACTTTACTGCCC | | |
| SKP2 | 4 | SKP2.4F.EX1.-962 | TTGGGGCAGTAAGAGTAGG | 1 | A |
| | | SKP2.4R.EX1.-1355 | GATCCCTTCTTTACACCAATCTC | | |
| SKP2 | 5 | SKP2.5F.EX1.-1252 | TTTCAAGCTTGGTCCTTAGC | 3 | B |
| | | SKP2.5R.EX1.-1640 | CTCTCCATTAGCTTACAGGC | CCCCCGCCCG | |
| SKP2 | 6 | SKP2.6F.EX1.-1601 | AGGGAACAAAACAGAGAGAGC | 1 | A |
| | | SKP2.6R.EX1.-2003 | AAGGCAATAGTCATGATTAGAAGG | CCCCCGCCCG | |
| TFDP1 | 4 | TFDP1.4F.EX1.-1576 | GAGTTTTCTGTCCCTGCTTT | 1 | A |
| | | TFDP1.4R.EX1.-1093 | CGGGTATGTCAGGATGTGTGT | | |
| TFDP1 | 5 | TFDP1.5F.EX1.-2112 | CCATCAGGGCTCAAATACTGA | 1 | A |
| | | TFDP1.5R.EX1.-1555 | GAAAGCAGGGACAGGAAAAC | GCCCCCGCCG | |
| TGFB1 | 1 | TGFB1.1F.EX1.-340 | CTGGTCCTCTTTCTCTGGTGAC | 6 | D |
| | | TGFB1.1R.EX1.31 | CTCGCTGTCTGGCTGCTC | | |
| TGFB1 | 2 | TGFB1.2F.EX1.-631 | TTGGGGACAGTAAATGTATGG | 1 | A |
| | | TGFB1.2R.EX1.-277 | AGCGGGTGATCCAGATGC | | |
| TGFB1 | 3 | TGFB1.3F.EX1.-932 | ACACCATCTACAGTGGGGC | 1 | A |
| | | TGFB1.3R.EX1.-527 | GCAGACACCTGTAAGAATTGC | | |
| TGFB1 | 4 | TGFB1.4F.EX1.-1219 | GGGTCTCCAGTGAGTATCAGG | 1 | A |
| | | TGFB1.4R.EX1.-808 | CAGAGTCCCTCAGCACTCC | | |
| TGFB1 | 5 | TGFB1.5F.EX1.-1451 | AGAGAGGAAATTCAGCAGAGG | 1 | A |
| | | TGFB1.5R.EX1.-1048 | CACATGGCAAAATCCATAGC | GCCCCCGCCG | |
| TGFB1 | 6 | TGFB1.6F.EX1.-1716 | CTCCAAGGGCTTTGACTACC | 1 | A |
| | | TGFB1.6R.EX1.-1333 | AATCCATGTTACTCCCCTGC | | |
| TGFB1 | 7 | TGFB1.7F.EX1.-2062 | TGGCTAGGAGCACTTCTCC | 1 | A |
| | | TGFB1.7R.EX1.-1691 | CAGTCTGGTAGTCAAAGCCC | | |

*Référer au tableau III **Référer au tableau IV

Annexe II

Tableau II : Liste des amorces utilisées pour l'amplification d'ADN génomique pour les constructions de gène rapporteur ainsi que les conditions PCR utilisées.

| Gene | Longueur du Fragment | Nom des amorces | Séquences (5' -3') | * | ** |
|--------|--------------------------------------|-----------------------------------|--|---|----|
| CCND1 | +179 to -2375 (218-3180) | CCND1.Mlu I.F CCND1.Hind III.R | AGAGACGCGTTCGTGGTGGGTTTTAAG AGAGAAGCTTACCCCTCCTCCTCAGAAA | 7 | F |
| CDC25A | 48 to -2051 (77401-79499) | CDC25A.Bgl II.F CDC25A.Mlu I.R | AGAGAGATCTCCGACCTACACCTCTTACCC AGAGACGCGTTTTCTTAGAAAGCACTTCCC | 7 | F |
| CDKN1A | 19 to -2071 (1138-3227) | CDKN1A.Mlu I.F CDKN1A.Bgl II.R | AGAGACGCGTATCTGATGCATGTGTGCTTG AGAGAGATCTAGCTGCTCACACCTCAGC | 7 | G |
| CDKN1B | 70 to -2023 (10482-12574) | CDKN1B.Bgl II.F CDKN1B.Mlu I.R | AGAGAGATCTAGAAAAACACCCCGAAAAGA AGAGACGCGTTGAGACCTTTGCTTGATCC | 7 | F |
| E2F1 | -3 to -2008 (62827-64832) | E2F1.Bgl II.F E2F1.Mlu I.R | AGAGAGATCTACTTTTACGCGCCAAATCC AGAGACGCGTCAGCAAGAGGCTTATGTGGTC | 8 | E |
| HDAC1 | 50 to -2060 (6104-8213) | HDAC1.Mlu I.F HDAC1.Bgl II.R | AGAGACGCGTTGAGTGAGAAACACAAGTCCC AGAGAGATCTCGTCCCTACCGTCAGTCG | 7 | F |
| RB1 | 22 to -2022 (2979-5022) | RB1.Mlu I.F RB1.Bgl II.R | AGAGACGCGTAGTCATGATGCTTTCTGTTGC AGAGAGATCTACGTCCCTGAGAAAAACC | 7 | F |
| TFDP1 | +987 to -2547 (117905- 121587) | TFDP1.Mlu I.F TFDP1.Bgl II.R | AGAGACGCGTATTTACGCAACCCAATCGAG AGAGAGATCTGTGCAGCATGAAGCAAACAT | 7 | F |
| TGFB1 | 31 to -2062 (75017-77109) | TGFB1.Mlu I.F TGFB1.Bgl II.R | AGAGACGCGTTGGCTAGGAGCACTCTTCC AGAGAGATCTCTCGCTGTCTGGCTGCTC | 7 | G |

*Référer au tableau III **Référer au tableau IV

Annexe III

Tableau III: Conditions de PCR

| Kit DNA polymerase* | Amorce s (µM) | MgCl ₂ (mM) | MgS _O ₄ (mM) | dNTP (mM) | Enzyme (U) | Solution GCriche ** | 10X Pfx amplification buffer*** | 10X enrichment buffer*** | Conc. ADN (ng) | Volume final (ml) |
|---------------------|---------------|------------------------|------------------------------------|-----------|-------------------------|---------------------|---------------------------------|--------------------------|----------------|-------------------|
| 1 Platinium Taq | 0,42 | 1,5 | - | 0,1 | 1,8 | | | | 15 | 60 |
| 2 Platinium Taq | 0,42 | 2,0 | - | 0,2 | 1,8 | | | | 15 | 60 |
| 3 Platinium Taq | 0,42 | 1,5 | - | 0,2 | 1,8 | | | | 15 | 60 |
| 4 Platinium Taq | 0,42 | 0,5 | - | 0,05 | 1,8 | | | | 15 | 60 |
| 5 Platinium Taq | 0,42 | 0,5 | - | 0,2 | 1,8 | | | | 15 | 60 |
| 6 FastStart Taq | 0,42 | 2,0 | - | 0,2 | 2,0 | 1X | | | 15 | 50 |
| 7 Platinium Pfx | 0,3 | - | 1,0 | 0,3 | 1,25 | | 2X | 3X | 15 | 50 |
| 8 FastStart + pfx | 0,5 | 2,0 | - | 0,2 | 2 (Faststart) 0,1 (pfx) | 1X | | | 15 | 50 |

*Utiliser selon les directives du fabricant. **Fourni dans le kit FastStart. ***Fourni avec le kit Platinium Pfx DNA polymerase.

Annexe IV

Tableau IV: Protocole pour l'amplification PCR

| | Début | Touchdown | Amplification principale | Final |
|-------------------|------------|---|--------------------------|-------------|
| Protocol A | | | | |
| Dénaturation | 95°C, 5min | 95°C, 30s | 95°C, 30s | |
| Appariement | | 62°C -0,5°C par cycle jusqu'à 55°C, 30s | 55°C, 30s | |
| Élongation | | 72°C, 30s | 72°C, 30s | 72°C, 7min |
| Cycles | 1 | 14 | 20 | 1 |
| Protocol B | | | | |
| Dénaturation | 95°C, 5min | 95°C, 30s | 95°C, 30s | |
| Appariement | | 60°C -0,5°C par cycle jusqu'à 53°C, 30s | 53°C, 30s | |
| Élongation | | 72°C, 30s | 72°C, 30s | 72°C, 7min |
| Cycles | 1 | 14 | 25 | 1 |
| Protocol C | | | | |
| Dénaturation | 95°C, 5min | 95°C, 30s | 95°C, 30s | |
| Appariement | | 64°C -0,5°C par cycle jusqu'à 57°C, 30s | 57°C, 30s | |
| Élongation | | 72°C, 30s | 72°C, 30s | 72°C, 7min |
| Cycles | 1 | 14 | 20 | 1 |
| Protocol D | | | | |
| Dénaturation | 95°C, 5min | 95°C, 30s | 95°C, 30s | |
| Appariement | | 62°C -0,5°C par cycle jusqu'à 55°C, 30s | 55°C, 30s | |
| Élongation | | 72°C, 30s | 72°C, 30s | 72°C, 7min |
| Cycles | 1 | 14 | 22 | 1 |
| Protocol E | | | | |
| Dénaturation | 95°C, 5min | | 95°C, 45s | |
| Appariement | | | 58°C, 30s | |
| Élongation | | | 72°C, 2min | 72°C, 10min |
| Cycles | 1 | | 42 | 1 |
| Protocol F | | | | |
| Dénaturation | 94°C, 5min | | 94°C, 15s | |

| | | | | |
|-------------------|------------|--|------------|----------------|
| Appariement | | | 55°C, 30s | |
| Élongation | | | 68°C, 2min | 68°C, 10min |
| Cycles | 1 | | 35 | 1 |
| Protocol G | | | | |
| Dénaturation | 95°C, 5min | 95°C, 45s | 95°C, 30s | |
| Appariement | | 62°C -0,5°C par cycle jusqu'à 55°C, 1min | 55°C, 30s | |
| Élongation | | 72°C, 2min | 72°C, 30s | 72°C, 10min |
| Cycles | 1 | 14 | 20 | 1 |

Annexe V

Tableau V: Température d'analyse au dHPLC

| Gène | Fragment | Température (°C) | | |
|--------|----------|------------------|------|------|
| CCND1 | 2 | 59,0 | 61,0 | |
| CCND1 | 3 | 55,0 | 59,5 | 61,0 |
| CCND1 | 4 | 55,0 | | |
| CCND1 | 5 | 62,5 | 64,5 | |
| CCND1 | 6 | 63,1 | 66,0 | |
| CCNE1 | 2 | 61,8 | 65,0 | 67,0 |
| CCNE1 | 3 | 60,5 | 61,5 | |
| CCNE1 | 4 | 59,5 | 61,5 | |
| CCNE1 | 5 | 58,0 | 60,0 | 61,0 |
| CCNE1 | 6 | 57,0 | | |
| CDC25A | 1 | 63,0 | | |
| CDC25A | 2 | 62,5 | 63,5 | |
| CDC25A | 3 | 56,5 | | |
| CDC25A | 4 | 60,5 | 61,5 | |
| CDC25A | 5 | 58,2 | 62,7 | |
| CDC25A | 6 | 56,5 | 58,0 | 60,5 |
| CDK4 | 1 | 65,0 | | |
| CDK4 | 3 | 57,5 | 58,5 | |
| CDK4 | 4 | 53,0 | 58,8 | |
| CDK4 | 5 | 55,1 | 57,5 | |
| CDK4 | 6 | 59,2 | | |
| CDK4 | 7 | 57,3 | | |
| CDK6 | 1 | 61,0 | 67,0 | |
| CDK6 | 2 | 61,1 | 66,5 | |
| CDK6 | 3 | 57,5 | 63,0 | 65,0 |
| CDK6 | 4 | 60,5 | 62,5 | |
| CDK6 | 5 | 63,5 | 64,5 | |

| | | | | |
|--------|---|------|------|------|
| CDK6 | 6 | 63,5 | 66,5 | |
| CDKN1A | 1 | 65,0 | 68,0 | |
| CDKN1A | 2 | 62,0 | 63,0 | |
| CDKN1A | 3 | 58,5 | 59,5 | |
| CDKN1A | 4 | 58,5 | 59,5 | |
| CDKN1A | 5 | 58,5 | 59,5 | |
| CDKN1A | 6 | 56,0 | 61,5 | |
| CDKN1A | 7 | 56,0 | 59,0 | |
| CDKN1A | 8 | 56,0 | 58,0 | |
| CDKN1B | 1 | 60,0 | 63,0 | 68,0 |
| CDKN1B | 2 | 65,0 | 67,5 | |
| CDKN1B | 3 | 56,7 | 59,5 | |
| CDKN1B | 4 | 59,5 | 60,5 | |
| CDKN1B | 5 | 58,4 | 61,0 | |
| CDKN1B | 6 | 58,5 | 60,0 | |
| CDKN1B | 7 | 59,2 | 61,0 | |
| CDKN2A | 1 | 62,0 | 66,0 | |
| CDKN2A | 2 | 58,0 | 60,0 | |
| CDKN2A | 3 | 58,2 | | |
| CDKN2A | 4 | 57,0 | | |
| CDKN2A | 5 | 62,0 | | |
| CDKN2A | 6 | 57,0 | 61,5 | |
| CDKN2A | 7 | 54,0 | 56,0 | |
| CDKN2B | 1 | 60,5 | 65,0 | |
| CDKN2B | 2 | 56,0 | 58,0 | |
| CDKN2B | 3 | 54,0 | 56,0 | |
| CDKN2B | 4 | 57,0 | 59,0 | |
| CDKN2B | 5 | 56,2 | | |
| CDKN2B | 6 | 53,5 | 55,0 | 59,0 |
| E2F1 | 1 | 61,0 | 67,0 | |

| | | | | |
|-------|---|------|------|--|
| E2F1 | 2 | 59,5 | 61,8 | |
| E2F1 | 3 | 58,5 | 62,0 | |
| E2F1 | 4 | 55,5 | 61,0 | |
| E2F1 | 5 | 56,2 | 59,1 | |
| HDAC1 | 1 | 60,2 | 67,0 | |
| HDAC1 | 2 | 55,0 | 58,0 | |
| HDAC1 | 3 | 57,0 | 61,3 | |
| HDAC1 | 4 | 57,0 | 58,5 | |
| HDAC1 | 5 | 55,0 | 60,5 | |
| MADH3 | 1 | 64,5 | 69,0 | |
| MADH3 | 3 | 68,0 | 69,5 | |
| MADH3 | 4 | 62,5 | 66,5 | |
| MADH3 | 5 | 57,0 | 61,0 | |
| MADH3 | 6 | 57,5 | 62,5 | |
| MADH3 | 7 | 62,0 | 64,0 | |
| MADH4 | 1 | 63,0 | 67,0 | |
| MADH4 | 2 | 59,0 | 61,0 | |
| MADH4 | 3 | 57,0 | 59,0 | |
| MADH4 | 4 | 57,5 | 60,0 | |
| MADH4 | 5 | 54,5 | 56,5 | |
| MADH4 | 6 | 56,0 | 60,5 | |
| RB1 | 1 | 63,5 | 67,0 | |
| RB1 | 2 | 59,2 | 63,0 | |
| RB1 | 3 | 54,5 | 57,0 | |
| RB1 | 4 | 56,0 | 57,0 | |
| RB1 | 5 | 57,0 | 58,0 | |
| RB1 | 6 | 56,5 | 58,0 | |
| SKP2 | 1 | 64,8 | 67,0 | |
| SKP2 | 2 | 64,5 | 66,0 | |
| SKP2 | 3 | 60,0 | 62,0 | |

| | | | | |
|-------|---|------|------|------|
| SKP2 | 4 | 57,0 | 58,0 | |
| SKP2 | 5 | 56,5 | | |
| SKP2 | 6 | 54,5 | 56,5 | |
| TFDP1 | 4 | 63,9 | | |
| TFDP1 | 5 | 58,3 | 60,5 | 64,0 |
| TGFB1 | 1 | 66,0 | 67,5 | |
| TGFB1 | 2 | 61,0 | 62,5 | |
| TGFB1 | 3 | 60,5 | 62,0 | 64,0 |
| TGFB1 | 4 | 62,0 | 65,0 | |
| TGFB1 | 5 | 59,3 | 61,0 | 62,0 |
| TGFB1 | 6 | 59,0 | 61,0 | |
| TGFB1 | 7 | 58,5 | 59,0 | |

Annexe VI

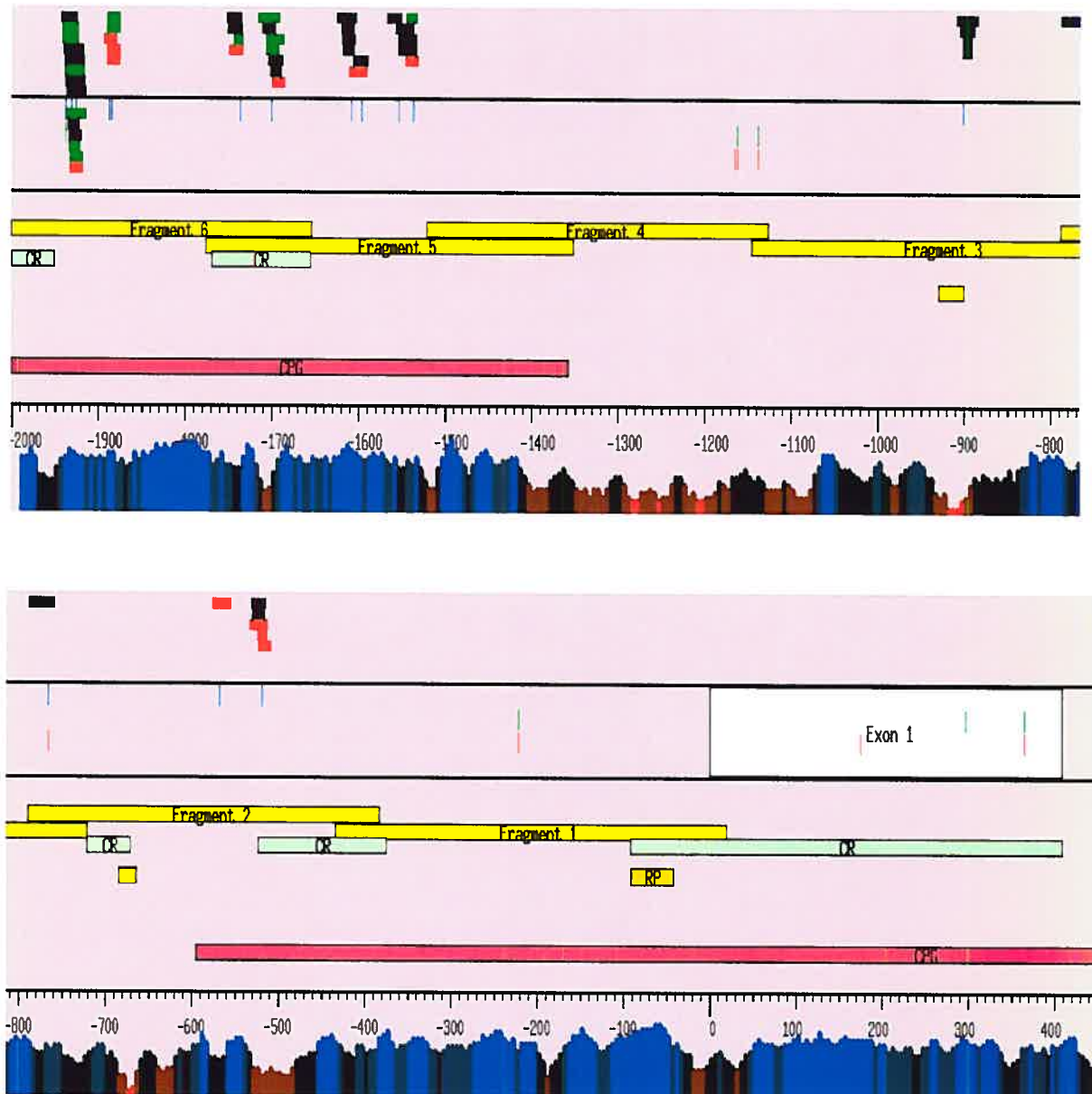


Figure 1: Représentation graphique du promoteur du gène CCND1.

Tableau VI: Description détaillée des polymorphismes identifiés dans le promoteur du gène codant pour CCND1.

| Gène | Position | Type de polymorphisme | Sites de liaison allèle majeur | Sites de liaison allèle mineur | Sites de liaison conservés | Répétition |
|-------|-----------|--------------------------|---|---|----------------------------|------------|
| CCND1 | Ex1.-1537 | DEL C nd | <u>MYC-associated zinc finger protein related transcription factor->0.905</u> <u>Pdx1 (IDX1/IPF1) pancreatic and intestinal homeodomain TF->0.757</u> | <u>Pdx1 (IDX1/IPF1) pancreatic and intestinal homeodomain TF->0.754</u> <u>Zinc finger transcription factor GLI1->0.88</u> | | |
| CCND1 | Ex1.-1554 | G->A transition | <u>Pdx1 (IDX1/IPF1) pancreatic and intestinal homeodomain TF->0.744</u> <u>Pdx1 (IDX1/IPF1) pancreatic and intestinal homeodomain TF->0.757</u> <u>POU factor Brn-2 (N-Oct 3)->0.914</u> | <u>Pdx1 (IDX1/IPF1) pancreatic and intestinal homeodomain TF->0.757</u> <u>POU factor Brn-2 (N-Oct 3)->0.914</u> | | |
| CCND1 | Ex1.-1597 | G->A transition | <u>Papilloma virus regulator E2->0.872</u> | <u>Papilloma virus regulator E2->0.875</u> <u>Tumor suppressor p53->0.66</u> | | |
| CCND1 | Ex1.-1609 | INS G nd | <u>GC box elements->0.9</u> <u>MYC-associated zinc finger protein related transcription factor->0.982</u> <u>Wilms Tumor Suppressor->0.89</u> <u>Zinc finger transcription factor ZBP-89->0.963</u> | <u>GC box elements->0.916</u> <u>MYC-associated zinc finger protein related transcription factor->0.984</u> <u>Wilms Tumor Suppressor->0.89</u> <u>Zinc finger transcription factor ZBP-89->0.977</u> | | |
| CCND1 | Ex1.-1701 | G->C transversion | <u>Activator protein 4->0.976</u> <u>Barbiturate-inducible element->0.903</u> <u>HEN1->0.828</u> <u>Neurogenin 1 and 3 (ngn1/3) binding sites->0.926</u> <u>Octamer-binding factor 1->0.843</u> <u>TCF11/MafG heterodimers, binding to subclass of AP1 sites->0.84</u> | <u>Activator protein 4->0.976</u> <u>Heterodimers of the bHLH transcription factors HAND2 (Thing2) and E12->0.816</u> <u>Neurogenin 1 and 3 (ngn1/3) binding sites->0.945</u> <u>Octamer-binding factor 1->0.845</u> | | |
| CCND1 | Ex1.-1737 | A->T transversion | <u>Muscle TATA box->0.847</u> <u>Pit1, GHF-1 pituitary specific pou domain transcription factor->0.885</u> | <u>Cut-like homeodomain protein->0.781</u> <u>Muscle TATA box->0.857</u> | | |

| | | | | | | |
|--------------|------------------|--------------------------|--|--|--|--|
| | | | <u>POU transcription factor Brn-3->0.834</u> | <u>POU transcription factor Brn-3->0.834</u> | | |
| <u>CCND1</u> | <u>Ex1.-1885</u> | G->T transversion | <u>Wilms Tumor Suppressor->0.929</u> | <u>Egr-2/Krox-20 early growth response gene product->0.792</u> | | |
| <u>CCND1</u> | <u>Ex1.-1888</u> | A->C transversion | <u>Wilms Tumor Suppressor->0.929</u> | <u>early growth response gene 3 product->0.935</u> <u>Hypoxia inducible factor, bHLH / PAS protein family->0.931</u> | | |
| <u>CCND1</u> | <u>Ex1.-1926</u> | INS G nd | <u>Muscle Initiator Sequence->0.883</u> <u>MYC-associated zinc finger protein related transcription factor->0.893</u> <u>Sterol regulatory element binding protein->0.838</u> <u>Zinc finger transcription factor ZBP-89->0.933</u> <u>Zinc finger transcription factor ZBP-89->0.949</u> | <u>Muscle Initiator Sequence->0.884</u> <u>MYC-associated zinc finger protein related transcription factor->0.933</u> <u>Sterol regulatory element binding protein->0.838</u> <u>Zinc finger transcription factor ZBP-89->0.956</u> <u>Zinc finger transcription factor ZBP-89->0.949</u> | | |
| <u>CCND1</u> | <u>Ex1.-1931</u> | G->C transversion | <u>Muscle Initiator Sequence->0.883</u> <u>MYC-associated zinc finger protein related transcription factor->0.893</u> <u>Sterol regulatory element binding protein->0.838</u> <u>Zinc finger transcription factor ZBP-89->0.933</u> <u>Zinc finger transcription factor ZBP-89->0.949</u> | <u>Nerve growth factor-induced protein C->0.859</u> | | |
| <u>CCND1</u> | <u>Ex1.-1938</u> | C->T transition | <u>Muscle Initiator Sequence->0.883</u> <u>Zinc finger transcription factor ZBP-89->0.933</u> <u>Zinc finger transcription factor ZBP-89->0.949</u> | <u>Zinc finger transcription factor ZBP-89->0.933</u> <u>Zinc finger transcription factor ZBP-89->0.949</u> | | |
| <u>CCND1</u> | <u>Ex1.-519</u> | C->T transition | <u>Homeobox transcription factor Gsh-1->0.903</u> <u>Homeodomain binding site in LIM/Homeodomain factor LHX3->0.816</u> | <u>CCAAT/enhancer binding protein beta->0.965</u> <u>Homeobox transcription factor Gsh-1->0.946</u> <u>Homeodomain binding site in LIM/Homeodomain factor LHX3->0.822</u> <u>NMP4 (nuclear matrix protein 4) / CIZ (Cas-interacting zinc finger protein)->0.975</u> | | |

| | | | | | | |
|--------------|-----------------|--------------------------|---|---|--|---------------------------------|
| | | | | <u>Pancreatic and intestinal lim-homeodomain factor->0.827</u> | | |
| <u>CCND1</u> | <u>Ex1.-568</u> | A->C transversion | | <u>Tumor suppressor p53 (3 half site)->0.943</u> | | |
| <u>CCND1</u> | <u>Ex1.-766</u> | A->T transversion | <u>Myogenin / nuclear factor 1 or related factors->0.715</u> | <u>Myogenin / nuclear factor 1 or related factors->0.715</u> | | |
| <u>CCND1</u> | <u>Ex1.-902</u> | DEL A nd | <u>Activator protein 1->1</u> <u>Activator protein 1->0.969</u> <u>Bach2 bound TRE->0.969</u> <u>RAR-related orphan receptor alpha2->0.822</u> | <u>Activator protein 1->1</u> <u>Activator protein 1->0.969</u> <u>Bach2 bound TRE->0.969</u> <u>RAR-related orphan receptor alpha2->0.822</u> | | <u>Low_complexity AT_rich</u> |

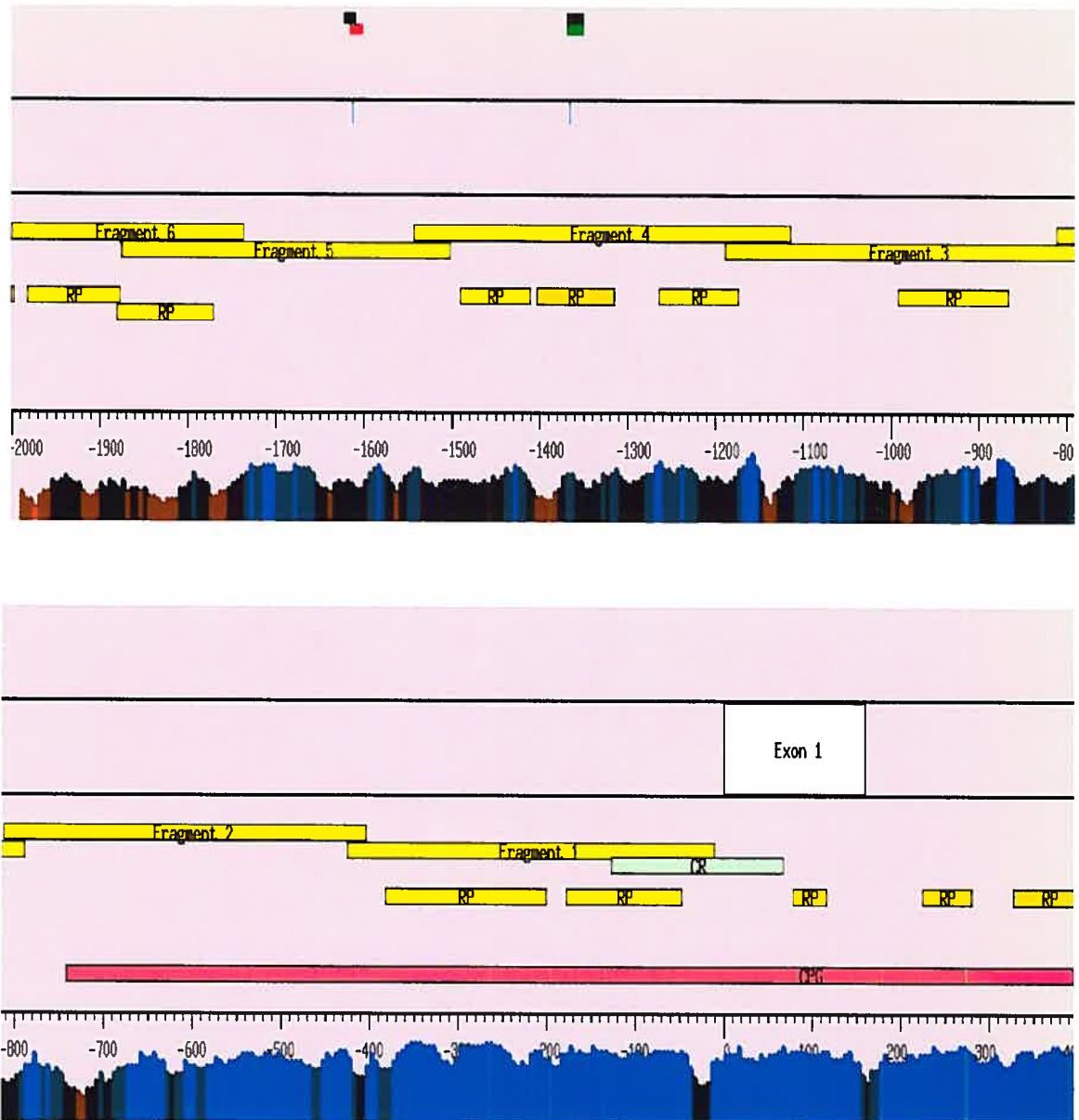


Figure 2: Représentation graphique du promoteur du gène CCNE1.

Tableau VII : Description détaillée des polymorphismes identifiés dans le promoteur du gène codant pour CCNE1.

| Gène | Position | Type de polymorphisme | Sites de liaison allèle majeur | Sites de liaison allèle mineur | Sites de liaison conservés | Répétition |
|-------|-----------|-----------------------|---|--|----------------------------|--------------|
| CCNE1 | Ex1.-1366 | G->A transition | <u>Nuclear factor 1->0.965</u> <u>Nuclear factor 1 (CTF1)->0.889</u> | <u>Nuclear factor 1->0.946</u> | | LINE/L2 L2 |
| CCNE1 | Ex1.-1613 | DEL T nd | <u>Zinc finger with interaction domain->0.911</u> | <u>Thyroid transcription factor-1 (TTF1) binding site->0.92</u> <u>Zinc finger with interaction domain->0.853</u> | | |

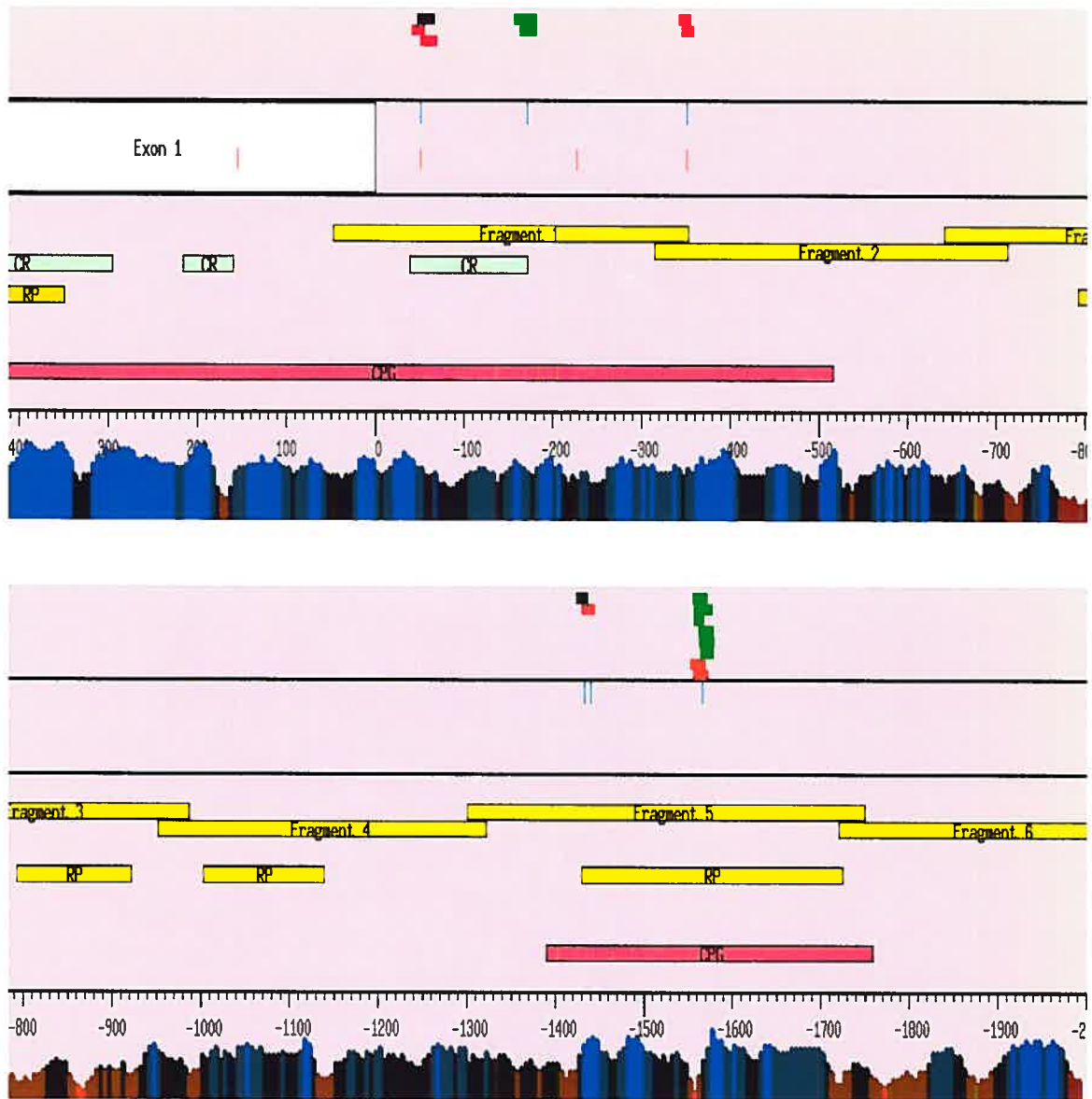


Figure 3: Représentation graphique du promoteur du gène CDC25A.

Tableau VIII : Description détaillée des polymorphismes identifiés dans le promoteur du gène codant pour CDC25A.

| Gène | Position | Type de polymorphisme | Sites de liaison allèle majeur | Sites de liaison allèle mineur | Sites de liaison conservés | Répétition |
|--------|------------------|-----------------------|---|--|----------------------------|-----------------|
| CDC25A | <u>Ex1.-1433</u> | C->T transition | <u>c-Myb, important in hematopoiesis, cellular equivalent to avian myoblastosis virus oncogene v-myb->0.944</u> | <u>c-Myb, important in hematopoiesis, cellular equivalent to avian myoblastosis virus oncogene v-myb->0.944</u> | | SINE/Alu AluY |
| CDC25A | <u>Ex1.-1440</u> | G->A transition | | <u>Metal transcription factor 1, MRE->0.911</u> | | SINE/Alu AluY |
| CDC25A | <u>Ex1.-1566</u> | C->T transition | <u>BPV bovine papilloma virus regulator E2->0.955</u> <u>BPV bovine papilloma virus regulator E2->0.884</u> <u>HMG1(Y) high-mobility-group protein 1 (Y), architectural transcription factor organizing the framework of a nuclear protein-DNA transcriptional complex->0.937</u> <u>Member of b-zip family, induced by ER damage/stress, binds to the ERSE in association with NF-Y->0.945</u> <u>Pax-4 homeodomain binding site, together with PAX-6 involved in pancreatic development->0.985</u> <u>Tumor suppressor p53->0.66</u> | <u>Cut-like homeodomain protein->0.756</u> <u>Homeobox transcription factor Gsh-1->0.899</u> | | SINE/Alu AluY |
| CDC25A | <u>Ex1.-171</u> | A->T transversion | <u>Retinoic acid receptor, member of nuclear receptors->0.972</u> <u>TCF11/MafG heterodimers, binding to subclass of AP1 sites->0.826</u> | | | |
| CDC25A | <u>Ex1.-2030</u> | T->G transversion | <u>GABP: GA binding protein->0.879</u> <u>Mammalian transcriptional repressor RBP-Jkappa/CBF1->0.955</u> | <u>GABP: GA binding protein->0.901</u> <u>Myeloid zinc finger protein MZF1->0.99</u> <u>Transcriptional repressor, binds to elements found predominantly in genes that participate in lipid metabolism->0.744</u> | | |

| | | | | | | |
|---------------|-----------------|--------------------------|--|--|--|--|
| <u>CDC25A</u> | <u>Ex1.-351</u> | G->A transition | | <u>MyT1 zinc finger transcription factor involved in primary neurogenesis->0.82</u> <u>MyT1 zinc finger transcription factor involved in primary neurogenesis->0.819</u> | | |
| <u>CDC25A</u> | <u>Ex1.-51</u> | G->C transversion | <u>Pax1 paired domain protein, expressed in the developing vertebral column of mouse embryos->0.626</u> | <u>Gut-enriched Krueppel-like factor->0.91</u> <u>Nuclear factor 1->0.941</u> <u>Pax1 paired domain protein, expressed in the developing vertebral column of mouse embryos->0.633</u> | <u>Pax1 paired domain protein, expressed in the developing vertebral column of mouse embryos</u> | |

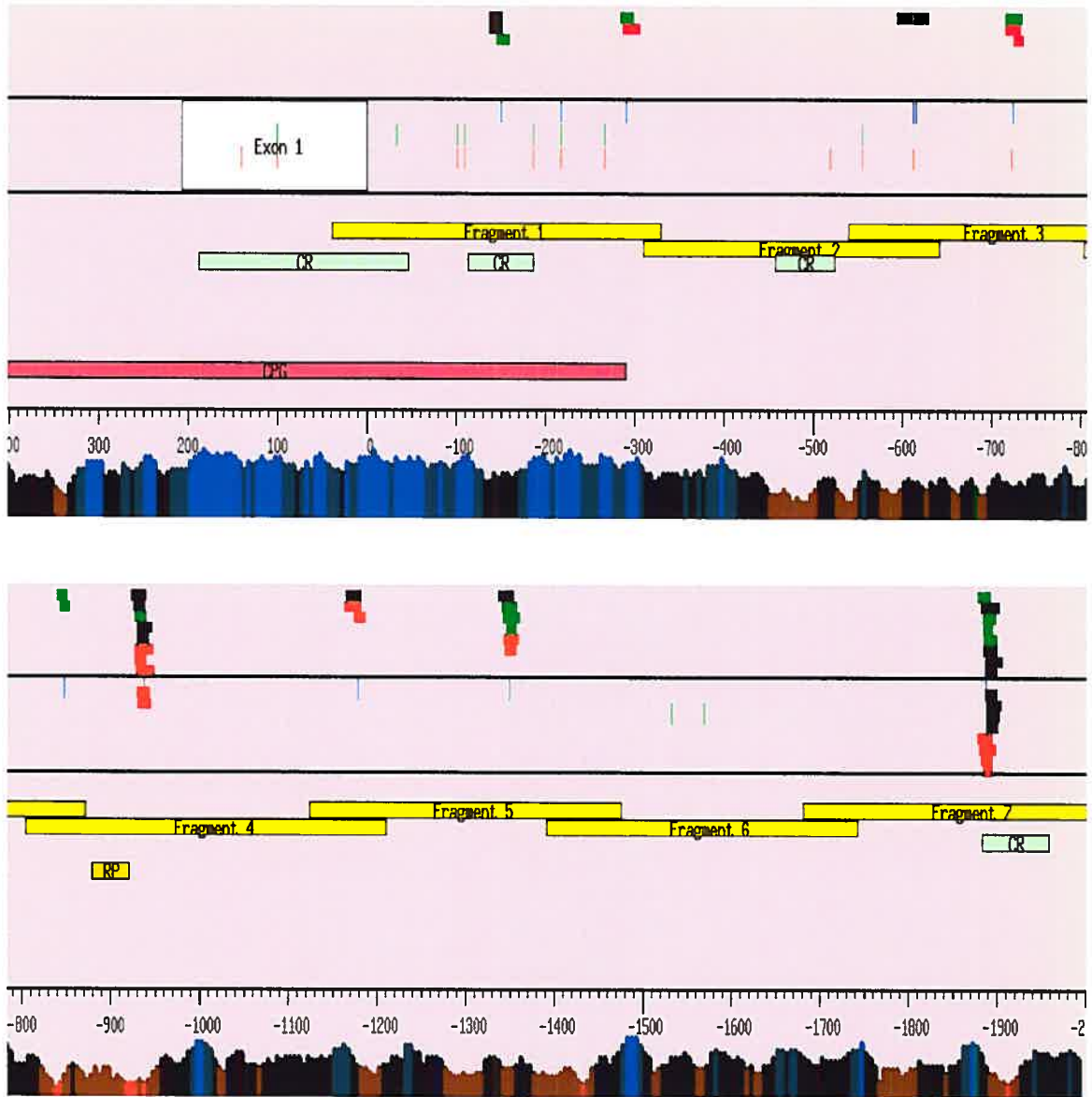


Figure 4: Représentation graphique du promoteur du gène CDK4.

Tableau IX: Description détaillée des polymorphismes identifiés dans le promoteur du gène codant pour CDK4.

| Gene | Position | Type de polymorphisme | Sites de liaison allèle majeur | Sites de liaisons allèle mineur | Sites de liaison conservés | Répétition |
|------|-----------|--------------------------|--|--|--|------------|
| CDK4 | Ex1.-1178 | A->G transition | <u>Fork head related activator-4 (FOXD1)->0.833</u> | <u>AREB6 (Atp1a1 regulatory element binding factor 6)->0.975</u> <u>Fork head related activator-4 (FOXD1)->0.845</u> <u>LBP-1c (leader-binding protein-1c), LSF (late SV40 factor), CP2, SEF (SAA3 enhancer factor)->0.875</u> | | |
| CDK4 | Ex1.-1349 | T->A transversion | <u>Ecotropic viral integration site 1 encoded factor->0.774</u> <u>Fork head related activator-4 (FOXD1)->0.788</u> <u>Interferon regulatory factor (IRF)-related protein (NF-EM5, PIP, LSIRF, ICSAT)->0.959</u> <u>Nuclear factor of activated T-cells->0.97</u> | <u>Ecotropic viral integration site 1 encoded factor->0.831</u> <u>GATA-binding factor 1->0.993</u> <u>Hepatic nuclear factor 1->0.792</u> | | |
| CDK4 | Ex1.-150 | C->G transversion | <u>Binding site of Clock/BMAL1 heterodimer, NPAS2/BMAL1 heterodimer->0.975</u> <u>Thyroid transcription factor-1 (TTF1) binding site->0.92</u> <u>Upstream stimulating factor->0.988</u> | <u>Binding site of Clock/BMAL1 heterodimer, NPAS2/BMAL1 heterodimer->0.975</u> <u>Upstream stimulating factor->0.988</u> | Binding site of Clock/BMAL1 heterodimer, NPAS2/BMAL1 heterodimer Binding site of Clock/BMAL1 heterodimer, NPAS2/BMAL1 heterodimer | |
| CDK4 | Ex1.-1887 | T->C transition | <u>Bright, B cell regulator of IgH transcription->0.928</u> <u>DLX-1, -2, and -5 binding sites->0.976</u> <u>FAST-1 SMAD interacting protein->0.826</u> <u>Gut-enriched Krueppel-like factor->0.919</u> | <u>Bright, B cell regulator of IgH transcription->0.923</u> <u>c-Myb, important in hematopoiesis, cellular equivalent to avian myoblastosis virus oncoqene v-myb->0.979</u> <u>DLX-1, -2, and -5 binding sites->0.976</u> <u>Elk-1->0.926</u> | <u>Bright, B cell regulator of IgH transcription</u> <u>Pdx1 (IDX1/IPF1) pancreatic and intestinal homeodomain TF</u> <u>Hepatic nuclear factor 1</u> <u>Octamer-binding factor 1</u> | |

| | | | | | |
|------|-----------------|-------------------|---|---|--|
| | | | <p><u>Hepatic nuclear factor 1->0.838</u></p> <p><u>Homeodomain binding site in LIM/Homeodomain factor LHX3->0.848</u></p> <p><u>Interferon regulatory factor 3 (IRF-3)->0.861</u></p> <p><u>LIM-homeodomain transcription factor->0.932</u></p> <p><u>Octamer-binding factor 1->0.859</u></p> <p><u>Pdx1 (IDX1/IPF1) pancreatic and intestinal homeodomain TF->0.742</u></p> <p><u>v-Myb->0.901</u></p> | <p><u>Hepatic nuclear factor 1->0.831</u></p> <p><u>Homeodomain binding site in LIM/Homeodomain factor LHX3->0.855</u></p> <p><u>Interferon regulatory factor 3 (IRF-3)->0.861</u></p> <p><u>LIM-homeodomain transcription factor->0.932</u></p> <p><u>Pdx1 (IDX1/IPF1) pancreatic and intestinal homeodomain TF->0.742</u></p> <p><u>Signal transducers and activators of transcription->0.909</u></p> <p><u>v-Myb, variant of AMV v-myb->0.985</u></p> | <p>DLX-1, -2, and -5 binding sites Homeodomain binding site in LIM/Homeodomain factor LHX3 v-Myb FAST-1 SMAD interacting protein</p> |
| CDK4 | <u>Ex1.-217</u> | C->T transition | | | |
| CDK4 | <u>Ex1.-290</u> | C->T transition | <u>Zinc finger transcription factor GLI1->0.896</u> | <u>Pax1 paired domain protein, expressed in the developing vertebral column of mouse embryos->0.626</u> | |
| CDK4 | <u>Ex1.-612</u> | C->A transversion | <u>Sox-5->0.986</u> | <u>Sox-5->0.986</u> | |
| CDK4 | <u>Ex1.-615</u> | C->A transversion | <u>SOX (SRY-related HMG box)->0.902</u> | <u>SOX (SRY-related HMG box)->0.91</u> | |
| CDK4 | <u>Ex1.-724</u> | DEL TC nd | <u>Interferon regulatory factor 2->0.867</u> | <u>Ecotropic viral integration site 1 encoded factor->0.771</u> <u>Nuclear factor of activated T-cells->0.971</u> | |
| CDK4 | <u>Ex1.-847</u> | C->T transition | <u>NMP4 (nuclear matrix protein 4) / CIZ (Cas-interacting zinc finger protein)->0.977</u> <u>Nuclear factor of activated T-cells->0.988</u> | | |
| CDK4 | <u>Ex1.-937</u> | A->G transition | <u>Ecotropic viral integration site 1 encoded factor->0.817</u> <u>GATA-binding factor 1->0.951</u> <u>GLI-Krüppel-related transcription factor, regulator of adenovirus E4 promoter->0.824</u> <u>Homeobox protein MEIS1 binding site->0.849</u> | <u>Activating transcription factor 6, member of b-zip family, induced by ER stress->0.982</u> <u>Basic helix-loop-helix protein known as Dec1, Stra13 or Sharp2->0.856</u> <u>Ecotropic viral integration site 1 encoded factor->0.802</u> <u>GATA binding factor->0.942</u> | |

| | | | | |
|--|--|---|--|--|
| | | <u>POU transcription factor Brn-3->0.822</u> | <u>GLI-Krüppel-related transcription factor, regulator of adenovirus E4 promoter->0.892</u> <u>Homeobox protein MEIS1 binding site->0.849</u> <u>POU transcription factor Brn-3->0.824</u> <u>v-Jun->0.854</u> <u>X-box-binding protein 1- >0.935</u> | |
|--|--|---|--|--|

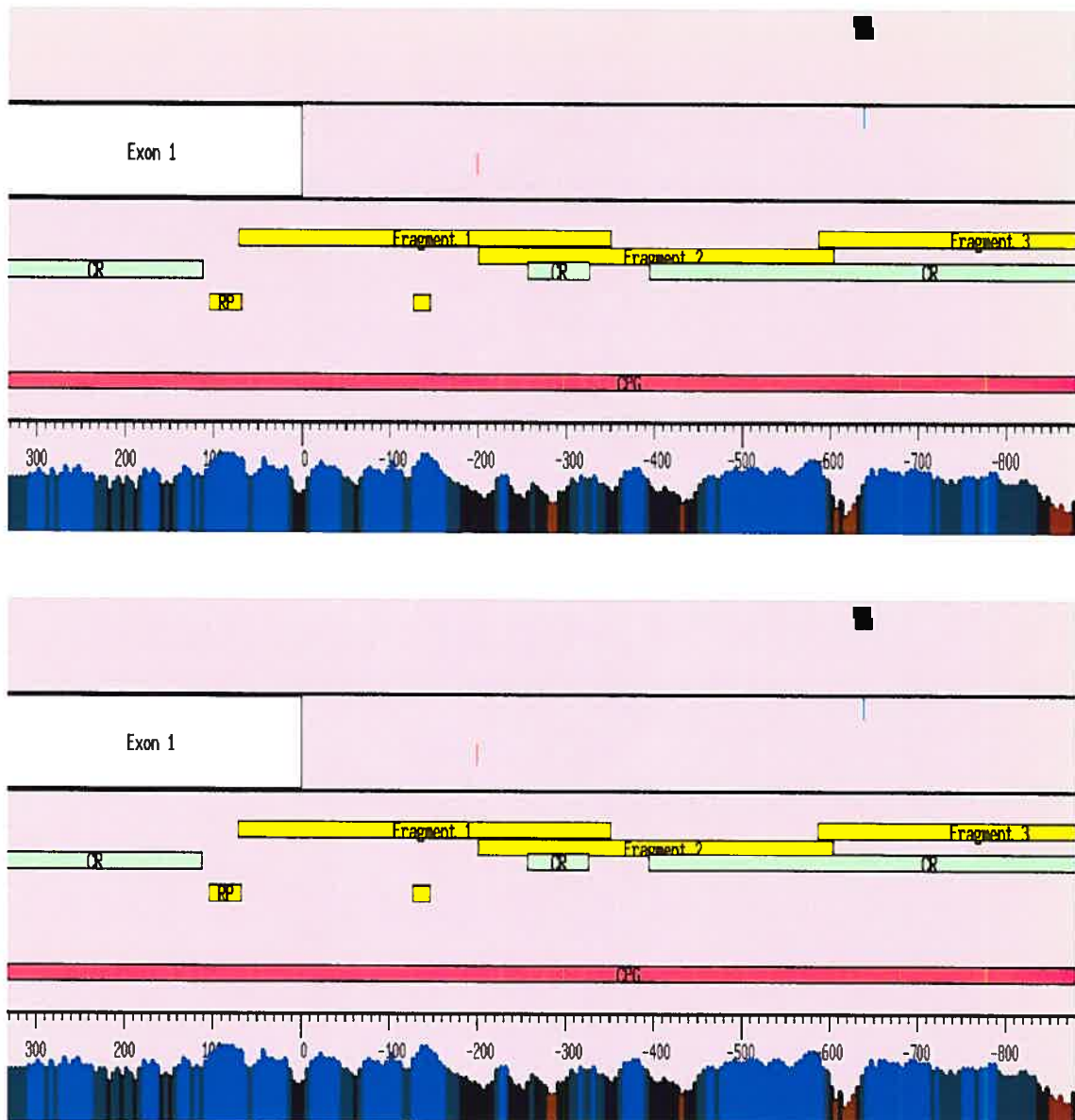


Figure 5: Représentation graphique du promoteur du gène CDK6.

Tableau X. Description détaillée des polymorphismes identifiés dans le promoteur du gène codant pour CDK6.

| Gène | Position | Type de polymorphisme | Site de liaison allèle majeur | Site de liaison allèle mineur | Sites de liaison conservés | Répétition |
|------|------------------|---------------------------|---|--|---|------------|
| CDK6 | <u>Ex1.-1124</u> | T->A transversion | <u>Interferon regulatory factor 2->0.866</u> <u>Pax-3 paired domain protein, expressed in embryogenesis, mutations correlate to Waardenburg Syndrome->0.831</u> <u>Tax/CREB complex->0.741</u> | <u>HMGI(Y) high-mobility-group protein 1 (Y), architectural transcription factor organizing the framework of a nuclear protein-DNA transcriptional complex->0.934</u> <u>Pax-3 paired domain protein, expressed in embryogenesis, mutations correlate to Waardenburg Syndrome->0.809</u> <u>Tax/CREB complex->0.741</u> | | |
| CDK6 | <u>Ex1.-1270</u> | C->T transition | <u>PAX5 paired domain protein->0.808</u> | | | |
| CDK6 | <u>Ex1.-1386</u> | A->G transition | <u>Ribonucleoprotein associated zinc finger protein MOK-2 (human)->1</u> | <u>Ribonucleoprotein associated zinc finger protein MOK-2 (human)->0.984</u> | | |
| CDK6 | <u>Ex1.-1450</u> | A->T transversion | <u>Ribonucleoprotein associated zinc finger protein MOK-2 (mouse)->0.744</u> | | | |
| CDK6 | <u>Ex1.-1789</u> | A->T transversion | <u>AREB6 (Atp1a1 regulatory element binding factor 6)->0.941</u> <u>Member of b-zip family, induced by ER damage/stress, binds to the ERSE in association with NF-Y->0.942</u> <u>MIBP-1 / RFX1 complex->0.767</u> | <u>Member of b-zip family, induced by ER damage/stress, binds to the ERSE in association with NF-Y->0.942</u> <u>X-box binding protein RFX1->0.912</u> | <u>AREB6 (Atp1a1 regulatory element binding factor 6)</u> <u>Member of b-zip family, induced by ER damage/stress, binds to the ERSE in association with NF-Y</u> | |
| CDK6 | <u>Ex1.-1902</u> | DEL CCCCCGCA nd | <u>Basic krueppel-like factor (KLF3)->0.95</u> <u>GC box elements->0.893</u> <u>Nerve growth factor-induced protein C->0.881</u> | <u>NF-kappaB->0.937</u> <u>NF-kappaB->0.937</u> <u>NUDR (nuclear DEAF-1 related transcriptional regulator protein)->0.781</u> | <u>Zinc finger transcription factor ZBP-89</u> <u>GC box elements</u> <u>Nerve growth factor-induced protein C</u> | |

| | | | | | |
|-------------|------------------|--------------------------|--|---|---|
| | | | <p><u>NF-kappaB->0.937</u></p> <p><u>NF-kappaB->0.937</u></p> <p><u>NUDR (nuclear DEAF-1 related transcriptional regulator protein)->0.779</u></p> <p><u>Olfactory neuron-specific factor->0.852</u></p> <p><u>Zinc finger transcription factor ZBP-89->0.962</u></p> | | Zinc finger transcription factor ZBP-89 |
| <u>CDK6</u> | <u>Ex1.-1958</u> | C->G transversion | <p><u>Myeloid zinc finger protein MZF1->0.99</u></p> <p><u>PAX6 paired domain and homeodomain are required for binding to this site->0.915</u></p> <p><u>Sterol regulatory element binding protein->0.849</u></p> <p><u>Transcriptional repressor, binds to elements found predominantly in genes that participate in lipid metabolism->0.863</u></p> <p><u>Wilms Tumor Suppressor->0.898</u></p> <p><u>Zinc finger transcription factor ZBP-89->0.965</u></p> | <p><u>PAX6 paired domain and homeodomain are required for binding to this site->0.904</u></p> <p><u>Transcriptional repressor, binds to elements found predominantly in genes that participate in lipid metabolism->0.86</u></p> <p><u>VDR/RXR Vitamin D receptor RXR heterodimer site->0.889</u></p> <p><u>Wilms Tumor Suppressor->0.936</u></p> | |
| <u>CDK6</u> | <u>Ex1.-638</u> | C->T transition | <p><u>MIBP-1 / RFX1 complex->0.81</u></p> <p><u>X-box binding protein RFX1->0.948</u></p> | <p><u>MIBP-1 / RFX1 complex->0.81</u></p> <p><u>X-box binding protein RFX1->0.944</u></p> | |

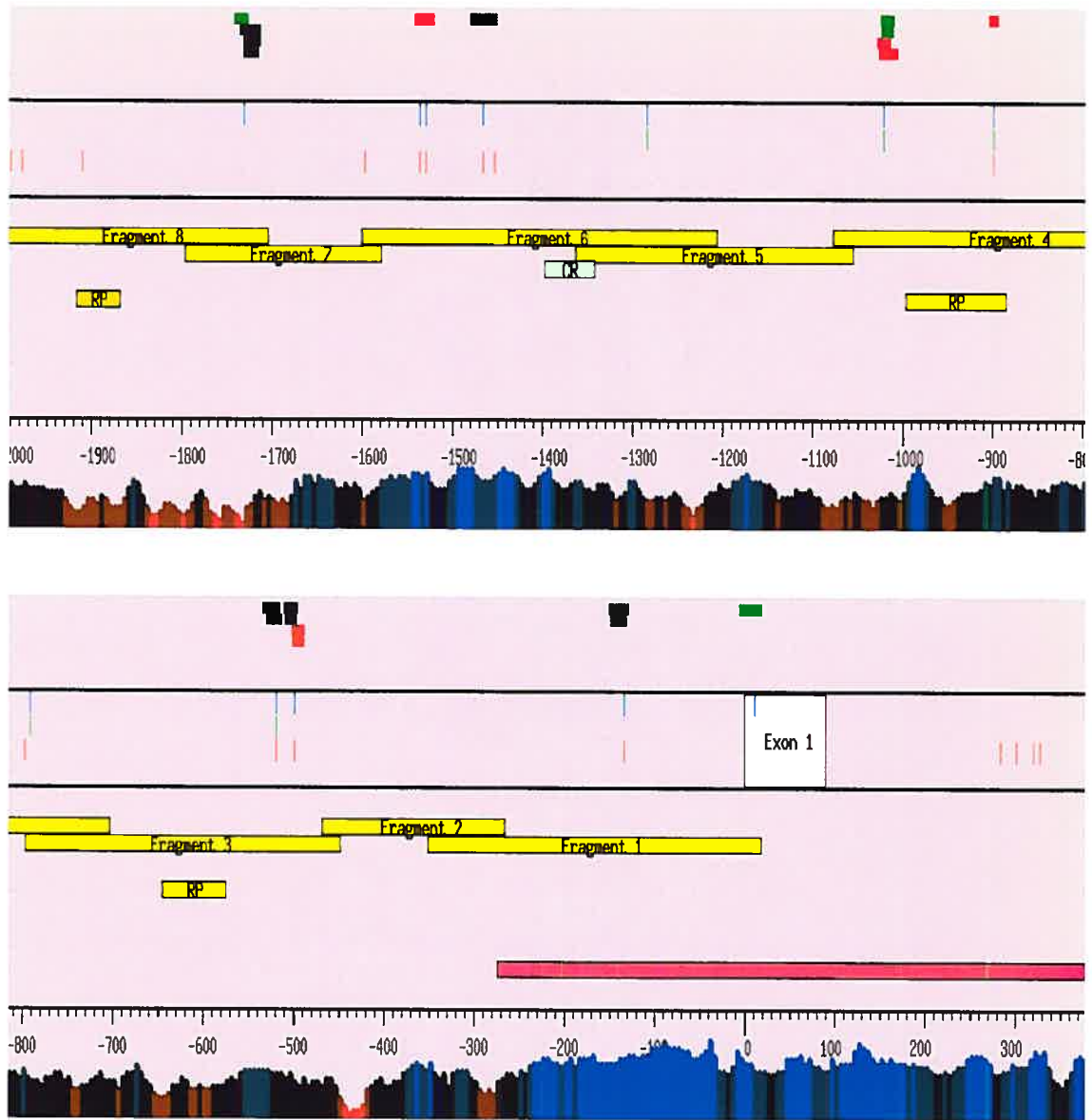


Figure 6: Représentation graphique du promoteur du gène CDKN1A.

Tableau XI : Description détaillée des polymorphismes identifiés dans le promoteur du gène codant pour CDKN1A.

| Gène | Position | Type de polymorphisme | Sites de liaison allèle majeur | Sites de liaison allèle mineur | Sites de liaison conservés | Répétition |
|--------|------------------|--------------------------|--|---|---|------------|
| CDKN1A | <u>Ex1.-1021</u> | C->T transition | <u>E2F, involved in cell cycle regulation, interacts with Rb p107 protein->0.849</u> <u>Zinc finger with interaction domain->0.856</u> | <u>AML1/CBFA2 Runt domain binding site->0.954</u> <u>Hepatic leukemia factor->0.847</u> | | |
| CDKN1A | <u>Ex1.-1284</u> | T->C transition | | | | |
| CDKN1A | <u>Ex1.-134</u> | G->C transversion | <u>Activator protein 4->0.994</u> <u>HEN1->0.819</u> | <u>Activator protein 4->0.994</u> <u>HEN1->0.858</u> <u>HEN1->0.847</u> | | |
| CDKN1A | <u>Ex1.-1466</u> | A->G transition | <u>PAX5 paired domain protein->0.851</u> | <u>PAX5 paired domain protein->0.822</u> | | |
| CDKN1A | <u>Ex1.-1529</u> | T->C transition | | | | |
| CDKN1A | <u>Ex1.-1536</u> | C->T transition | | <u>PPAR/RXR heterodimers->0.718</u> | | |
| CDKN1A | <u>Ex1.-1731</u> | DEL A nd | <u>E2F, involved in cell cycle regulation, interacts with Rb p107 protein->0.805</u> <u>Intestine specific homeodomain factor CDX-1->0.948</u> <u>Myocyte enhancer factor->0.911</u> <u>Special AT-rich sequence-binding protein 1, predominantly expressed in thymocytes, binds to matrix attachment regions (MARs)->0.971</u> | <u>Intestine specific homeodomain factor CDX-1->0.948</u> <u>Myocyte enhancer factor->0.911</u> <u>Special AT-rich sequence-binding protein 1, predominantly expressed in thymocytes, binds to matrix attachment regions (MARs)->0.971</u> | <u>E2F, involved in cell cycle regulation, interacts with Rb p107 protein</u> | |
| CDKN1A | <u>Ex1.-499</u> | G->C transversion | <u>Ikaros 2, potential regulator of lymphocyte differentiation->0.983</u> <u>Mammalian transcriptional repressor RBP-Jkappa/CBF1->0.955</u> | <u>AREB6 (Atp1a1 regulatory element binding factor 6)->0.983</u> <u>Ikaros 2, potential regulator of lymphocyte differentiation->0.983</u> | | |

| | | | | | | |
|---------------|-----------------|----------------------|---|---|---|--------------------|
| | | | | <u>Mammalian transcriptional repressor RBP-Jkappa/CBF1->0.962</u> | | |
| | | | | <u>Myoblast determination gene product->0.92</u> | | |
| <u>CDKN1A</u> | <u>Ex1.-519</u> | G->C transversion | <u>RAR-related orphan receptor alpha2->0.862</u> <u>Retinoic acid receptor, member of nuclear receptors->0.981</u> | <u>RAR-related orphan receptor alpha2->0.862</u> <u>Retinoic acid receptor, member of nuclear receptors->0.981</u> | | |
| <u>CDKN1A</u> | <u>Ex1.-791</u> | T->C transition | | | | |
| <u>CDKN1A</u> | <u>Ex1.-899</u> | T->G transversion | | <u>v-Myb, variant of AMV v-myb->0.925</u> | | SINE/MIR MIRm |
| <u>CDKN1A</u> | <u>Ex1.12</u> | G->A transition | <u>T-Box factor 5 site (TBX5), mutations related to Holt-Oram syndrome->0.998</u> | | T-Box factor 5 site (TBX5), mutations related to Holt-Oram syndrome | |

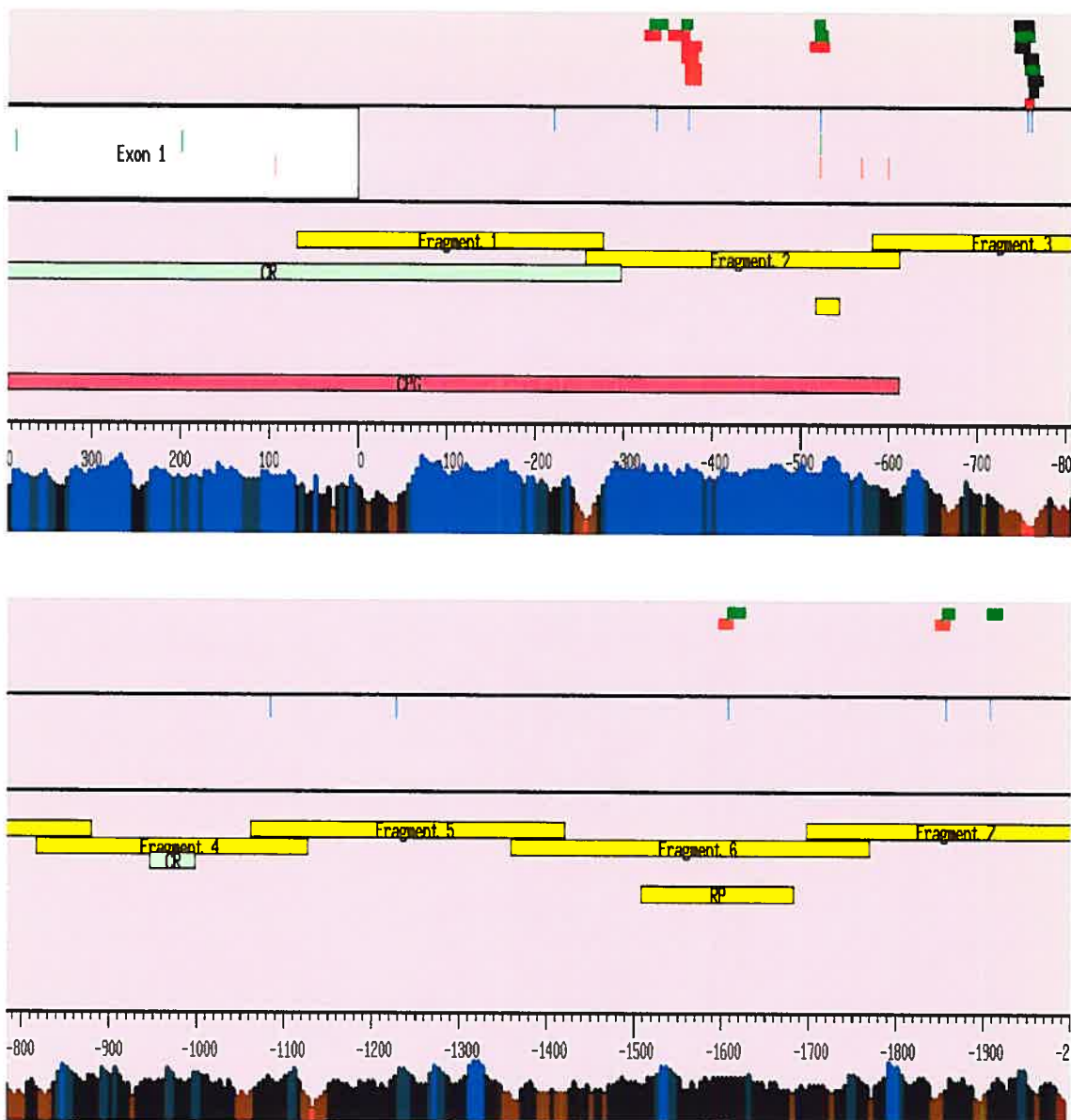


Figure 7: Représentation graphique du promoteur du gène CDKN1B.

Tableau XII : Description détaillée des polymorphismes identifiés dans le promoteur du gène codant pour CDKN1B.

| Gène | Position | Type de polymorphisme | Sites de liaison allèle majeur | Sites de liaison allèle mineur | Sites de liaison conservés | Répétition |
|--------|------------------|------------------------|---|---|-----------------------------|--------------------------|
| CDKN1B | <u>Ex1.-1084</u> | DEL AAG nd | | | | |
| CDKN1B | <u>Ex1.-1085</u> | DEL TTC nd | | | | |
| CDKN1B | <u>Ex1.-1228</u> | C->T transition | | | | |
| CDKN1B | <u>Ex1.-1608</u> | G->A transition | <u>Tumor suppressor p53 (5 half site)->0.91</u> | <u>POZ/zinc finger protein, transcriptional repressor, translocations observed in diffuse large cell lymphoma->0.842</u> | | SINE/MIR MIR3 |
| CDKN1B | <u>Ex1.-1857</u> | C->T transition | <u>Myoblast determining factor->0.983</u> | <u>Ecotropic viral integration site 1 encoded factor->0.835</u> | Myoblast determining factor | |
| CDKN1B | <u>Ex1.-1908</u> | DEL T nd | <u>Binding site for a Pbx1/Meis1 heterodimer->0.796</u> | | | |
| CDKN1B | <u>Ex1.-221</u> | T->C transition | | | | |
| CDKN1B | <u>Ex1.-337</u> | C->A transversion | <u>Neural-restrictive-silencer-element->0.677</u> | <u>Nuclear factor 1->0.941</u> | | |
| CDKN1B | <u>Ex1.-373</u> | G->T transversion | <u>Myc associated zinc finger protein (MAZ)->0.939</u> | <u>GAGA-Box->0.781</u> <u>Olfactory neuron-specific factor->0.823</u> <u>Pax1 paired domain protein, expressed in the developing vertebral column of mouse embryos->0.633</u> <u>Signal transducer and activator of transcription 1->0.815</u> <u>STAT6: signal transducer and activator of transcription 6->0.842</u> | | |
| CDKN1B | <u>Ex1.-522</u> | G->A transition | <u>MYC-associated zinc finger protein related transcription factor->0.917</u> <u>Ras-responsive element binding protein 1->0.796</u> | <u>Zinc finger transcription factor ZBP-89->0.936</u> | | Low_complexity GC_rich |
| CDKN1B | <u>Ex1.-756</u> | A->C | <u>Cone-rod homeobox-containing transcription factor / otx-like homeobox gene->0.956</u> | <u>Cone-rod homeobox-containing transcription factor / otx-like homeobox gene->0.956</u> | | |

| | | | | | | |
|---------------|-----------------|-----------------|---|---|--|--|
| | | transversion | <u>HNF-3/Fkh Homolog 2 (FOXD3)->0.938</u> <u>Myocyte enhancer factor->0.915</u> | <u>Multifunctional c-Abl src type tyrosine kinase->0.997</u> | | |
| <u>CDKN1B</u> | <u>Ex1.-761</u> | INS A nd | <u>HNF-3/Fkh Homolog 2 (FOXD3)->0.938</u> <u>Multifunctional c-Abl src type tyrosine kinase->0.997</u> <u>Myocyte enhancer factor->0.915</u> <u>Sex-determining region Y gene product->0.951</u> | <u>HNF-3/Fkh Homolog 2 (FOXD3)->0.938</u> <u>Multifunctional c-Abl src type tyrosine kinase->0.997</u> <u>Myocyte enhancer factor->0.915</u> <u>Sex-determining region Y gene product->0.951</u> | | |

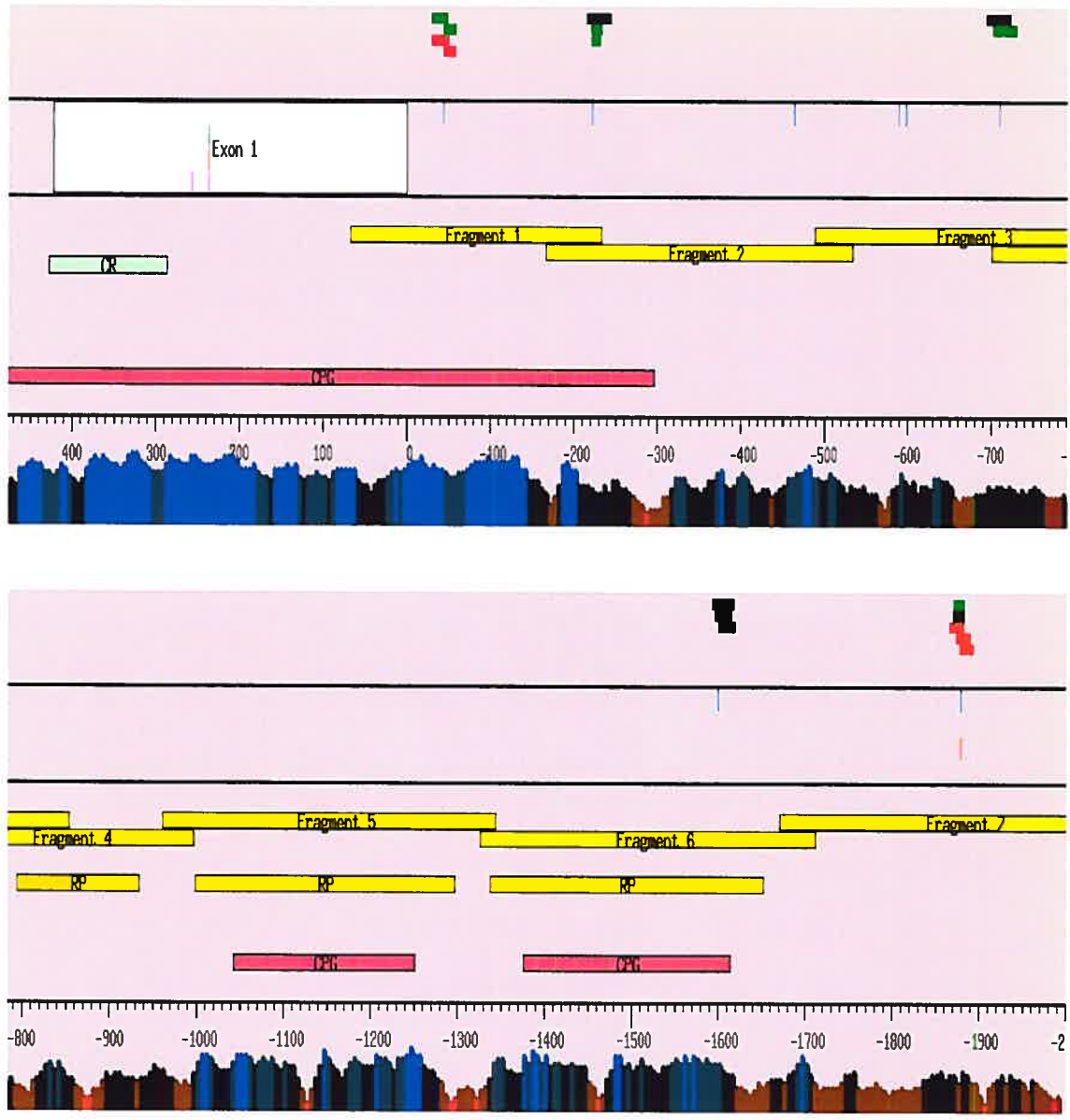


Figure 8: Représentation graphique du promoteur du gène CDKN2A.

Tableau XIII : Description détaillée des polymorphismes identifiés dans le promoteur du gène codant pour CDKN2A.

| Gène | Position | Type de polymorphisme | Sites de liaison allèle majeur | Sites de liaison allèle mineur | Sites de liaison conservés | Répétition |
|---------------|----------------------|--------------------------|--|---|----------------------------|------------------|
| <u>CDKN2A</u> | <u>Ex1.-1600</u> | A->G transition | <u>Interferon regulatory factor 1->0.861</u> <u>Mammalian C-type LTR CCAAT box->0.747</u> <u>Pax1 paired domain protein, expressed in the developing vertebral column of mouse embryos->0.622</u> | <u>Interferon regulatory factor 1->0.861</u> <u>Mammalian C-type LTR CCAAT box->0.736</u> <u>Pax1 paired domain protein, expressed in the developing vertebral column of mouse embryos->0.622</u> | | SINE/Alu AluSp |
| <u>CDKN2A</u> | <u>Ex1.-1879</u> | C->T transition | <u>GATA-binding factor 1->0.991</u> <u>Octamer-binding factor 1, POU-specific domain->0.865</u> | <u>CUT-homeodomain transcription factor Onecut-2->0.852</u> <u>Ecotropic viral integration site 1 encoded factor->0.906</u> <u>GATA-binding factor 1->0.994</u> <u>Xenopus homeodomain factor Xvent-2; early BMP signaling response->0.847</u> | | |
| <u>CDKN2A</u> | <u>Ex1.-222</u> | T->A transversion | <u>c-Myb, important in hematopoiesis, cellular equivalent to avian myoblastosis virus oncogene v-myb->0.969</u> <u>Myogenin / nuclear factor 1 or related factors->0.739</u> <u>v-Myb->0.908</u> | <u>Myogenin / nuclear factor 1 or related factors->0.739</u> | | |
| <u>CDKN2A</u> | <u>Ex1.-44</u> | A->T transversion | <u>CP2->0.945</u> <u>early growth response gene 3 product->0.774</u> | <u>Egr-2/Krox-20 early growth response gene product->0.814</u> <u>Ribonucleoprotein associated zinc finger protein MOK-2 (mouse)->0.772</u> | | |
| <u>CDKN2A</u> | <u>Ex1.-464</u> | C->T transition | | | | |
| <u>CDKN2A</u> | <u>Ex1.-589</u> — | DEL G nd | | | | |
| <u>CDKN2A</u> | <u>Ex1.-598</u> | G->A transition | | | | |
| <u>CDKN2A</u> | <u>Ex1.-710</u> | C->A transversion | <u>B-cell-specific activating protein->0.803</u> <u>PAX5 paired domain protein->0.848</u> | <u>B-cell-specific activating protein->0.799</u> | | |

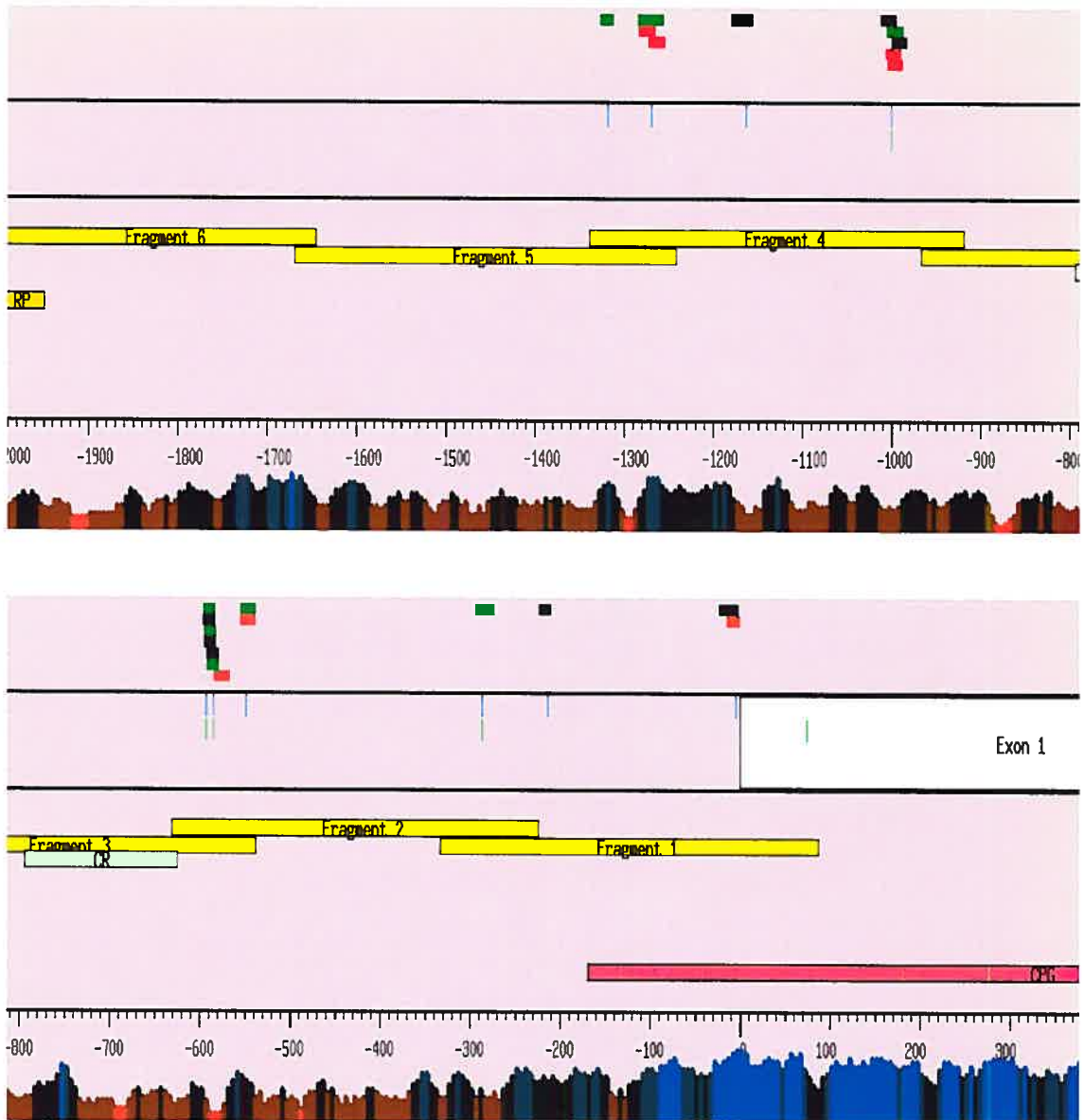


Figure 9: Représentation graphique du promoteur du gène CDKN2B.

Tableau XIV : Description détaillée des polymorphismes identifiés dans le promoteur du gène codant pour CDKN2B.

| Gène | Position | Type de polymorphisme | Sites de liaison allèle majeur | Sites de liaison allèle mineur | Sites de liaison conservés | Répétition |
|--------|-----------|------------------------|---|--|----------------------------|------------|
| CDKN2B | Ex1.-1001 | G->A transition | <u>Interferon-stimulated response element->0.813</u> <u>Xenopus fork head domain factor 3 (FoxA2a)->0.826</u> <u>Xenopus fork head domain factor 3 (FoxA2a)->0.982</u> | <u>Sex-determining region Y gene product->0.951</u> <u>Sox-5->0.983</u> <u>Xenopus fork head domain factor 3 (FoxA2a)->0.878</u> <u>Xenopus fork head domain factor 3 (FoxA2a)->0.982</u> | | |
| CDKN2B | Ex1.-1164 | G->A transition | <u>Myocyte enhancer factor->0.922</u> | <u>Myocyte enhancer factor->0.935</u> | | |
| CDKN2B | Ex1.-1270 | C->T transition | <u>B-cell-specific activating protein->0.794</u> | <u>Interferon regulatory factor 2->0.819</u> <u>Pax1 paired domain protein, expressed in the developing vertebral column of mouse embryos->0.639</u> | | |
| CDKN2B | Ex1.-1319 | A->G transition | <u>Ras-responsive element binding protein 1->0.816</u> | | | |
| CDKN2B | Ex1.-214 | INS T nd | <u>c-Myb, important in hematopoiesis, cellular equivalent to avian myoblastosis virus oncogene v-myb->0.972</u> | <u>c-Myb, important in hematopoiesis, cellular equivalent to avian myoblastosis virus oncogene v-myb->0.972</u> | | |
| CDKN2B | Ex1.-287 | C->G transversion | <u>Neural-restrictive-silencer-element->0.698</u> | | | |
| CDKN2B | Ex1.-5 | C->G transversion | <u>cAMP-responsive element binding protein->0.89</u> | <u>cAMP-responsive element binding protein->0.89</u> <u>early growth response gene 3 product->0.791</u> | | |
| CDKN2B | Ex1.-549 | C->T transition | <u>Binding site for a Pbx1/Meis1 heterodimer->0.773</u> | <u>Cone-rod homeobox-containing transcription factor / otx-like homeobox gene->0.944</u> | | |
| CDKN2B | Ex1.-577 | DEL T nd | | | | |
| CDKN2B | Ex1.-585 | DEL T nd | <u>MyT1 zinc finger transcription factor involved in primary neurogenesis->0.798</u> <u>MyT1 zinc finger transcription factor involved in primary neurogenesis->0.798</u> | <u>Ecotropic viral integration site 1 encoded factor->0.791</u> <u>MyT1 zinc finger transcription factor involved in primary neurogenesis->0.798</u> | | |

| | | | | | | |
|---------------|-----------------|-------------------------------|--|---|--|--|
| | | | <u>Prostate-specific homeodomain protein NKX3.1->0.84</u> | <u>MyT1 zinc finger transcription factor involved in primary neurogenesis->0.798</u> | | |
| | | | | <u>Prostate-specific homeodomain protein NKX3.1->0.84</u> | | |
| <u>CDKN2B</u> | <u>Ex1.-593</u> | A->(T, C) transversion | <u>MyT1 zinc finger transcription factor involved in primary neurogenesis->0.798</u> <u>MyT1 zinc finger transcription factor involved in primary neurogenesis->0.798</u> <u>Prostate-specific homeodomain protein NKX3.1->0.84</u> | | | |

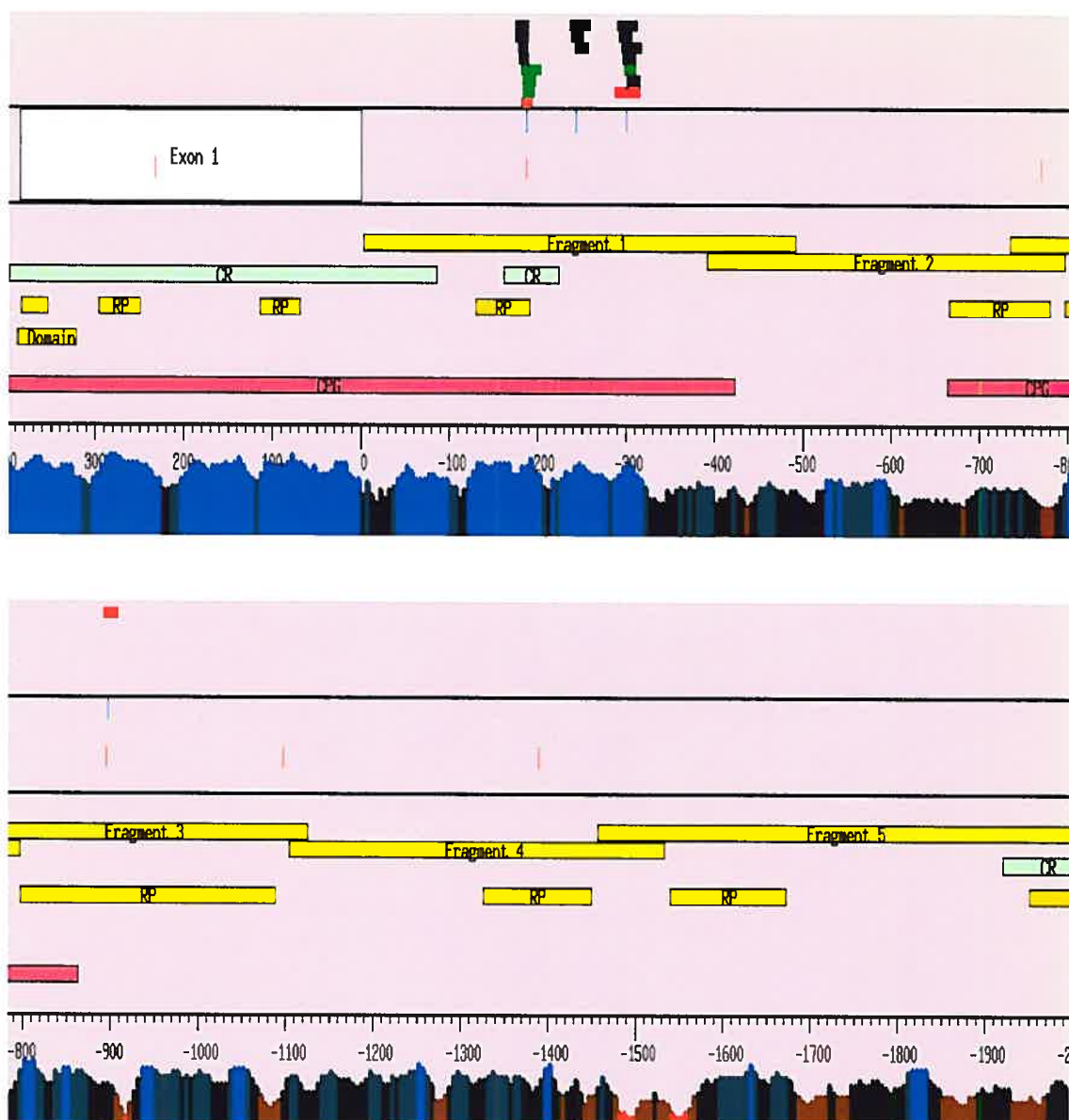


Figure 10: Représentation graphique du promoteur du gène E2F1.

Tableau XV : Description détaillée des polymorphismes identifiés dans le promoteur du gène codant pour E2F1.

| Gène | Position | Type de polymorphisme | Sites de liaison allèle majeur | Sites de liaison allèle mineur | Sites de liaison conservés | Répétition |
|-------------|------------------|-----------------------|---|--|---|--------------------------|
| <u>E2F1</u> | <u>Ex1.-1728</u> | G->A transition | <u>Albumin D-box binding protein->0.866</u> <u>Cart-1 (cartilage homeoprotein 1)->0.882</u> <u>Cone-rod homeobox-containing transcription factor / otx-like homeobox gene->0.947</u> | <u>Cart-1 (cartilage homeoprotein 1)->0.882</u> <u>CUT-homeodomain transcription factor</u> <u>Onecut-2->0.883</u> | | |
| <u>E2F1</u> | <u>Ex1.-187</u> | C->T transition | <u>Core promoter-binding protein (CPBP) with 3 Krueppel-type zinc fingers->0.912</u> <u>Drosophila hairy and enhancer of split homologue 1 (HES-1)->0.925</u> <u>Metal transcription factor 1, MRE->0.934</u> <u>MYC-associated zinc finger protein related transcription factor->0.915</u> <u>Stimulating protein 1 SP1, ubiquitous zinc finger transcription factor->0.981</u> <u>Zinc finger / POZ domain transcription factor->0.968</u> <u>Zinc finger / POZ domain transcription factor->0.965</u> | <u>Drosophila hairy and enhancer of split homologue 1 (HES-1)->0.947</u> <u>Metal transcription factor 1, MRE->0.898</u> <u>Myc associated zinc finger protein (MAZ)->0.929</u> <u>Zinc finger / POZ domain transcription factor->0.963</u> | | Simple_repeat (CGGGG)n |
| <u>E2F1</u> | <u>Ex1.-227</u> | G->C transversion | <u>BPV bovine papilloma virus regulator E2->0.924</u> <u>BPV bovine papilloma virus regulator E2->0.935</u> <u>Cell cycle-dependent element, CDF-1 binding site (CDE/CHR tandem elements regulate cell cycle dependent repression)->0.923</u> <u>Tax/CREB complex->0.723</u> | <u>BPV bovine papilloma virus regulator E2->0.917</u> <u>BPV bovine papilloma virus regulator E2->0.926</u> <u>Cell cycle-dependent element, CDF-1 binding site (CDE/CHR tandem elements regulate cell cycle dependent repression)->0.923</u> | | |
| <u>E2F1</u> | <u>Ex1.-243</u> | T->G transversion | <u>Activator protein 2->0.901</u> <u>Stimulating protein 1 SP1, ubiquitous zinc finger transcription factor->0.966</u> <u>Zinc finger transcription factor ZBP-89->0.958</u> | <u>Activator protein 2->0.905</u> <u>Stimulating protein 1 SP1, ubiquitous zinc finger transcription factor->0.93</u> <u>Stimulating protein 1 SP1, ubiquitous zinc finger transcription factor->0.97</u> <u>Zinc finger transcription factor ZBP-89->0.968</u> | <u>Stimulating protein 1 SP1, ubiquitous zinc finger transcription factor</u> | |
| <u>E2F1</u> | <u>Ex1.-278</u> | C->G | <u>Basonuclin, cooperates with USF1 in rDNA PolI transcription)->0.895</u> | <u>Basonuclin, cooperates with USF1 in rDNA PolI transcription)->0.891</u> | | |

| | | | | | | |
|-------------|-----------------|------------------------|--|---|---|--|
| | | transversion | <p><u>Core promoter-binding protein (CPBP) with 3 Krueppel-type zinc fingers->0.876</u></p> <p><u>Elk-1->0.937</u></p> <p><u>PAX5 paired domain protein->0.837</u></p> | <p><u>Basonuclin, cooperates with USF1 in rDNA Poll transcription->0.869</u></p> <p><u>Core promoter-binding protein (CPBP) with 3 Krueppel-type zinc fingers->0.876</u></p> <p><u>PAX5 paired domain protein->0.837</u></p> <p><u>X-box binding protein RFX1->0.912</u></p> | | |
| <u>E2F1</u> | <u>Ex1.-300</u> | G->A transition | <p><u>Core promoter-binding protein (CPBP) with 3 Krueppel-type zinc fingers->0.894</u></p> <p><u>Core promoter-binding protein (CPBP) with 3 Krueppel-type zinc fingers->0.958</u></p> <p><u>early growth response gene 3 product->0.814</u></p> <p><u>MYC-associated zinc finger protein related transcription factor->0.911</u></p> <p><u>Stimulating protein 1 SP1, ubiquitous zinc finger transcription factor->0.977</u></p> <p><u>Stimulating protein 1 SP1, ubiquitous zinc finger transcription factor->0.983</u></p> | <p><u>B-cell-specific activating protein->0.763</u></p> <p><u>Core promoter-binding protein (CPBP) with 3 Krueppel-type zinc fingers->0.949</u></p> <p><u>early growth response gene 3 product->0.819</u></p> <p><u>Stimulating protein 1 SP1, ubiquitous zinc finger transcription factor->0.983</u></p> | <p>Stimulating protein 1 SP1, ubiquitous zinc finger transcription factor</p> <p>Core promoter-binding protein (CPBP) with 3 Krueppel-type zinc fingers</p> <p>Stimulating protein 1 SP1, ubiquitous zinc finger transcription factor</p> <p>early growth response gene 3 product</p> <p>Core promoter-binding protein (CPBP) with 3 Krueppel-type zinc fingers</p> | |
| <u>E2F1</u> | <u>Ex1.-318</u> | C->T transition | <p><u>Barbiturate-inducible element->0.884</u></p> <p><u>Monomers of the nur subfamily of nuclear receptors (nur77, nurr1, nor-1)->0.982</u></p> | <p><u>Monomers of the nur subfamily of nuclear receptors (nur77, nurr1, nor-1)->0.982</u></p> | | |
| <u>E2F1</u> | <u>Ex1.-532</u> | C->T transition | <p><u>GATA-binding factor 1->0.969</u></p> | <p><u>GATA-binding factor 1->0.969</u></p> | <p>GATA-binding factor 1</p> | |

| | | | | | |
|-------------|-----------------|--------------------------|--|--|------------------|
| <u>E2F1</u> | <u>Ex1.-612</u> | A->G transition | | <u>Nuclear factor 1->0.949</u> <u>Serum response factor->0.856</u> <u>Serum responsive factor->0.801</u> | |
| <u>E2F1</u> | <u>Ex1.-616</u> | A->C transversion | | <u>E2F, involved in cell cycle regulation, interacts with Rb p107 protein->0.769</u> <u>Ras-responsive element binding protein 1->0.799</u> | |
| <u>E2F1</u> | <u>Ex1.-752</u> | G->A transition | | | LINE/L2 L2 |
| <u>E2F1</u> | <u>Ex1.-897</u> | C->T transition | | <u>cut-like homeodomain protein->0.949</u> | SINE/Alu AluJb |

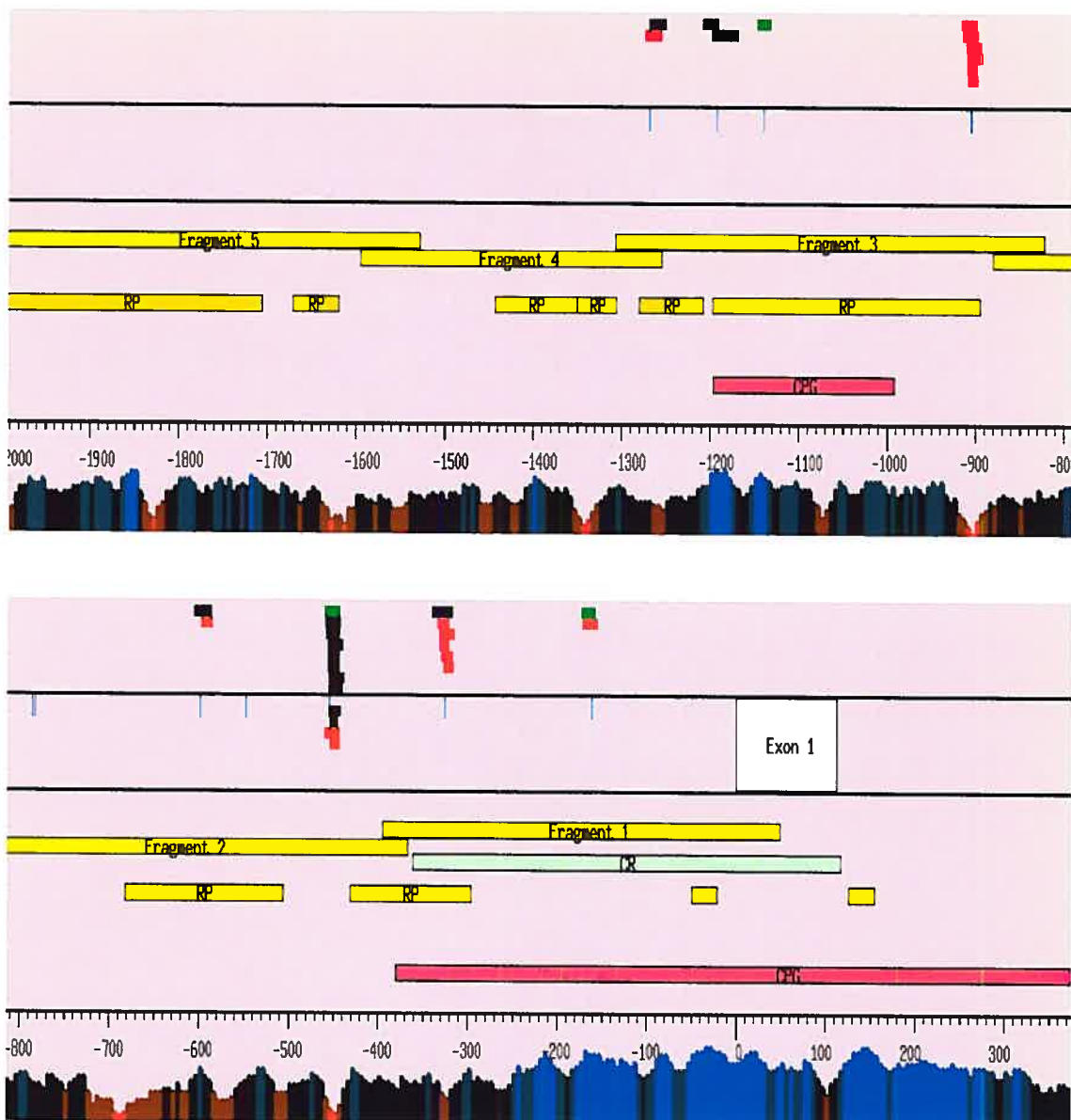


Figure 11: Représentation graphique du promoteur du gène HDAC1.

Tableau XVI : Description détaillée des polymorphismes identifiés dans le promoteur du gène codant pour HDAC1.

| Gène | Position | Type de polymorphisme | Sites de liaison allèle majeur | Sites de liaison allèle mineur | Sites de liaison conservés | Répétition |
|-------|------------------|-----------------------|--|--|---|-----------------------|
| HDAC1 | <u>Ex1.-1140</u> | G->A transition | <u>Stimulating protein 1 SP1, ubiquitous zinc finger transcription factor->0.892</u> | | | SINE/Alu AluSp |
| HDAC1 | <u>Ex1.-1193</u> | G->A transition | <u>PAX5 paired domain protein->0.822</u> <u>VDR/RXR Vitamin D receptor RXR heterodimer site->0.859</u> | <u>PAX5 paired domain protein->0.806</u> <u>VDR/RXR Vitamin D receptor RXR heterodimer site->0.859</u> | | SINE/Alu AluSp |
| HDAC1 | <u>Ex1.-1269</u> | T->C transition | <u>Signal transducers and activators of transcription->0.892</u> | <u>Progesterone receptor binding site->0.861</u> <u>Signal transducers and activators of transcription->0.892</u> | | DNA/MER1_type MER20 |
| HDAC1 | <u>Ex1.-1546</u> | G->T transversion | <u>Distal-less 3 homeodomain transcription factor->0.958</u> <u>Hepatic nuclear factor 1->0.79</u> <u>Homeo domain factor Pbx-1->0.796</u> <u>POU transcription factor Brn-3->0.955</u> <u>Transcriptional repressor CDP->0.898</u> | <u>Distal-less 3 homeodomain transcription factor->0.958</u> <u>Ecotropic viral integration site 1 encoded factor->0.9</u> <u>GATA-binding factor 1->0.996</u> <u>Hepatic nuclear factor 1->0.798</u> <u>Hepatic nuclear factor 1->0.776</u> <u>Homeo domain factor Pbx-1->0.789</u> <u>Liver enriched Cut - Homeodomain transcription factor HNF6 (ONECUT)->0.865</u> <u>POU transcription factor Brn-3->0.92</u> <u>POU transcription factor Brn-3->0.858</u> <u>Transcriptional repressor CDP->0.928</u> | | |
| HDAC1 | <u>Ex1.-162</u> | C->T transition | <u>Drosophila hairy and enhancer of split homologue 1 (HES-1)->0.92</u> | <u>VDR/RXR Vitamin D receptor RXR heterodimer site->0.905</u> | | |
| HDAC1 | <u>Ex1.-184</u> | T->C transition | <u>Alpha (1)-fetoprotein transcription factor (FTF), liver receptor homologue-1 (LHR-1)->0.961</u> <u>Monomers of the nur subfamily of nuclear receptors (nur77, nurr1, nor-1)->0.895</u> <u>Promyelocytic leukemia zink finger (TF with nine Krueppel-like zink fingers)->0.88</u> | <u>Alpha (1)-fetoprotein transcription factor (FTF), liver receptor homologue-1 (LHR-1)->0.955</u> <u>Monomers of the nur subfamily of nuclear receptors (nur77, nurr1, nor-1)->0.895</u> <u>Nuclear factor 1->0.969</u> <u>Nuclear factor 1 (CTF1)->0.826</u> <u>PAX5 paired domain protein->0.803</u> | <u>Promyelocytic leukemia zink finger (TF with nine Krueppel-like zink fingers)</u> | |

| | | | | | |
|-------|----------|--------------------------|---|--|-----------------|
| HDAC1 | Ex1.-326 | A->T transversion | <u>Zebrafish PAX2 paired domain protein->0.83</u> | <u>c-Myb, important in hematopoiesis, cellular equivalent to avian myoblastosis virus oncogene v-myb->0.96</u> <u>Homeobox protein MEIS1 binding site->0.819</u> <u>Member of the vertebrate HOX - cluster of homeobox factors->0.872</u> <u>PAR-type chicken vitellogenin promoter-binding protein->0.86</u> <u>v-Myb->0.979</u> <u>Zebrafish PAX2 paired domain protein->0.81</u> | SINE/MIR MIR3 |
| HDAC1 | Ex1.-455 | T->C transition | <u>Binding site for S8 type homeodomains->0.995</u> <u>Bright, B cell regulator of IgH transcription->0.975</u> <u>DLX-1, -2, and -5 binding sites->0.988</u> <u>DLX-1, -2, and -5 binding sites->0.988</u> <u>Hepatic nuclear factor 1->0.823</u> <u>Homeobox protein engrailed (en-1)->0.787</u> <u>Homeobox transcription factor Gsh-1->0.883</u> <u>Homeodomain binding site in LIM/Homeodomain factor LHX3->0.874</u> <u>LIM-homeodomain transcription factor->0.995</u> <u>POU transcription factor Brn-3->0.831</u> | <u>Avian C-type LTR TATA box->0.813</u> <u>Binding site for S8 type homeodomains->0.995</u> <u>Bright, B cell regulator of IgH transcription->0.945</u> <u>DLX-1, -2, and -5 binding sites->0.988</u> <u>POU transcription factor Brn-3->0.831</u> <u>Hepatic nuclear factor 1->0.818</u> <u>Homeobox transcription factor Gsh-1->0.933</u> <u>Homeodomain binding site in LIM/Homeodomain factor LHX3->0.861</u> <u>LIM-homeodomain transcription factor->1</u> <u>Pax-4 homeodomain binding site, together with PAX-6 involved in pancreatic development->0.974</u> | |
| HDAC1 | Ex1.-507 | T->G transversion | <u>Ecotropic viral integration site 1 encoded factor->0.782</u> <u>Interferon regulatory factor 3 (IRF-3)->0.953</u> <u>Myelin transcription factor 1-like, neuronal C2HC zinc finger factor 1->0.93</u> | <u>Interferon regulatory factor 7 (IRF-7)->0.862</u> <u>Myelin transcription factor 1-like, neuronal C2HC zinc finger factor 1->0.93</u> | SINE/MIR MIRb |
| HDAC1 | Ex1.-548 | G->A transition | | | SINE/MIR MIRb |
| HDAC1 | Ex1.-599 | C->T transition | <u>Nuclear factor 1->0.945</u> | <u>Myelin transcription factor 1-like, neuronal C2HC zinc finger factor 1->0.921</u> <u>Nuclear factor 1->0.947</u> | SINE/MIR MIRb |

| | | | | | | |
|--------------|-----------------|--------------------------|--|--|--|-----------------------------|
| <u>HDAC1</u> | <u>Ex1.-783</u> | A->G transition | | | | |
| <u>HDAC1</u> | <u>Ex1.-786</u> | T->C transition | | | | |
| <u>HDAC1</u> | <u>Ex1.-843</u> | T->A transversion | <u>c-Ets-1 binding site- >0.932</u> <u>Cone-rod homeobox- containing transcription factor / otx-like homeobox gene->0.948</u> | <u>Cone-rod homeobox- containing transcription factor / otx-like homeobox gene->0.948</u> <u>GATA-binding factor 1- >0.963</u> | | |
| <u>HDAC1</u> | <u>Ex1.-905</u> | A->C transversion | | <u>HNF-3/Fkh Homolog 2 (FOXD3)->0.938</u> <u>Multifunctional c-Abl src type tyrosine kinase- >0.997</u> <u>Sex-determining region Y gene product->0.951</u> | | <u>SINE/Alu AluSp</u> |
| <u>HDAC1</u> | <u>Ex1.-906</u> | A->C transversion | | <u>HNF-3/Fkh Homolog 2 (FOXD3)->0.938</u> <u>Multifunctional c-Abl src type tyrosine kinase- >0.997</u> <u>Sex-determining region Y gene product->0.951</u> | | <u>SINE/Alu AluSp</u> |

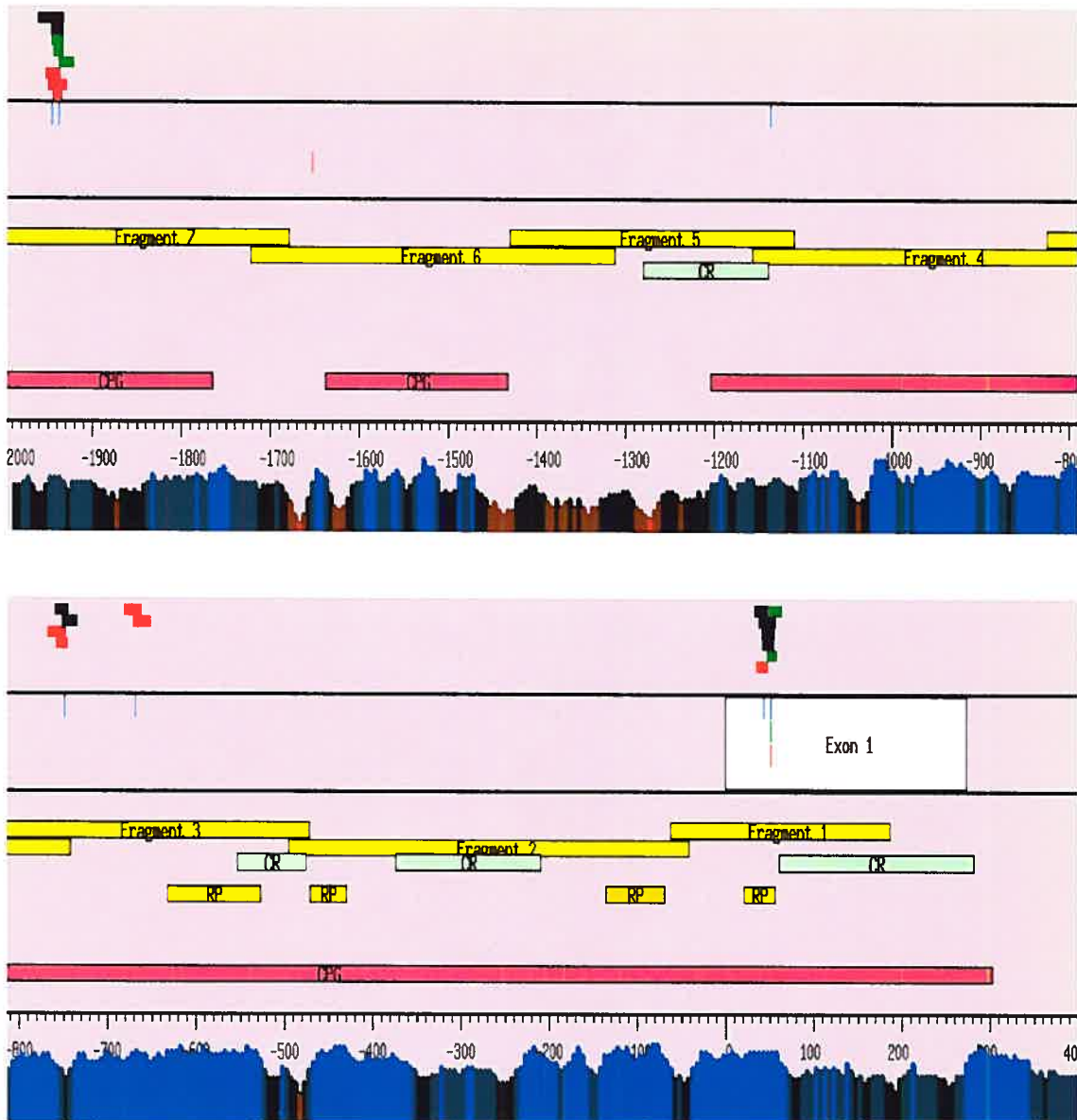


Figure 12: Représentation graphique du promoteur du gène MADH3.

Tableau XVII : Description détaillée des polymorphismes identifiés dans le promoteur du gène codant pour MADH3.

| Gène | Position | Type de polymorphisme | Sites de liaison allèle majeur | Sites de liaison allèle mineur | Sites de liaison conservés | Répétition |
|-------|------------------|------------------------|---|---|----------------------------|--------------------------|
| MADH3 | <u>Ex1.-1137</u> | T->G transversion | | | | |
| MADH3 | <u>Ex1.-1938</u> | T->C transition | <u>Cell cycle-dependent element, CDF-1 binding site (CDE/CHR tandem elements regulate cell cycle dependent repression)->0.909</u> <u>Elk-1->0.822</u> <u>v-Myb, AMV v-myb->0.886</u> | <u>cAMP-responsive element binding protein->0.95</u> <u>Winged helix protein, involved in hair keratinization and thymus epithelium differentiation->0.952</u> | | |
| MADH3 | <u>Ex1.-1946</u> | C->A transversion | <u>B-cell-specific activating protein->0.802</u> <u>Cell cycle-dependent element, CDF-1 binding site (CDE/CHR tandem elements regulate cell cycle dependent repression)->0.909</u> | <u>B-cell-specific activating protein->0.799</u> <u>Cell cycle-dependent element, CDF-1 binding site (CDE/CHR tandem elements regulate cell cycle dependent repression)->0.885</u> <u>Nuclear respiratory factor 2->0.947</u> | | |
| MADH3 | <u>Ex1.-669</u> | C->T transition | | <u>CP2->0.902</u> <u>LBP-1c (leader-binding protein-1c), LSF (late SV40 factor), CP2, SEF (SAA3 enhancer factor)->0.883</u> | | |
| MADH3 | <u>Ex1.-749</u> | G->A transition | <u>c-Ets-1 binding site->0.957</u> <u>Mammalian transcriptional repressor RBP-Jkappa/CBF1->0.944</u> | <u>c-Ets-1 binding site->0.954</u> <u>Ikaros 2, potential regulator of lymphocyte differentiation->0.983</u> <u>Mammalian transcriptional repressor RBP-Jkappa/CBF1->0.946</u> <u>Neural-restrictive-silencer-element->0.675</u> | | |
| MADH3 | <u>Ex1.44</u> | G->A transition | <u>Myc associated zinc finger protein (MAZ)->0.939</u> <u>Nerve growth factor-induced protein C->0.852</u> | <u>Myc associated zinc finger protein (MAZ)->0.909</u> <u>Nerve growth factor-induced protein C->0.805</u> <u>Pax-3 paired domain protein, expressed in embryogenesis, mutations correlate to Waardenburg Syndrome->0.773</u> | | Low_complexity GC_rich |

| | | | | | |
|--------------|---------------|--------------------|--|--|---------------------------------|
| <u>MADH3</u> | <u>Ex1.52</u> | C->T transition | <u>CP2->0.92</u> <u>Nerve growth factor-</u> <u>induced protein C->0.852</u> <u>Wilms Tumor Suppressor-</u> <u>>0.929</u> <u>Zinc finger / POZ domain</u> <u>transcription factor->0.966</u> | <u>CP2->0.92</u> <u>Wilms Tumor Suppressor-</u> <u>>0.89</u> <u>Wilms Tumor Suppressor-</u> <u>>0.929</u> | Low_compl exity GC_rich |
|--------------|---------------|--------------------|--|--|---------------------------------|

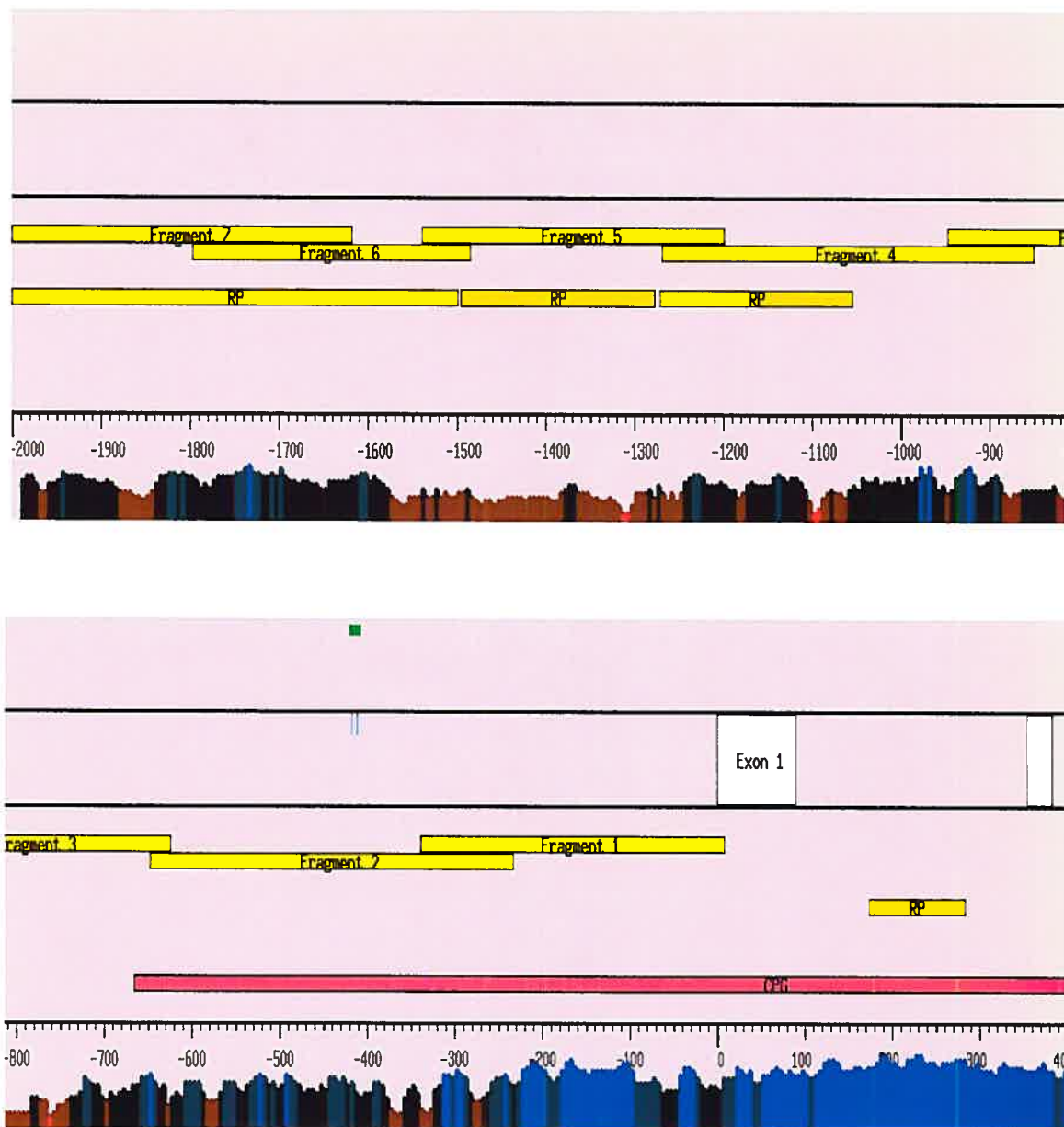


Figure 13: Représentation graphique du promoteur du gène MADH4.

Tableau XVIII : Description détaillée des polymorphismes identifiés dans le promoteur du gène codant pour MADH4.

| Gène | Position | Type de polymorphisme | Sites de liaison allèle majeur | Sites de liaison allèle mineur | Sites de liaison conservés | Répétition |
|--------------|-----------------|-----------------------|--|--------------------------------|----------------------------|------------|
| <u>MADH4</u> | <u>Ex1.-412</u> | G->T transversion | | | | |
| <u>MADH4</u> | <u>Ex1.-418</u> | A->C transversion | <u>c-Myb, important in hematopoiesis, cellular equivalent to avian myoblastosis virus oncogene v-myb->0.983</u> | | | |

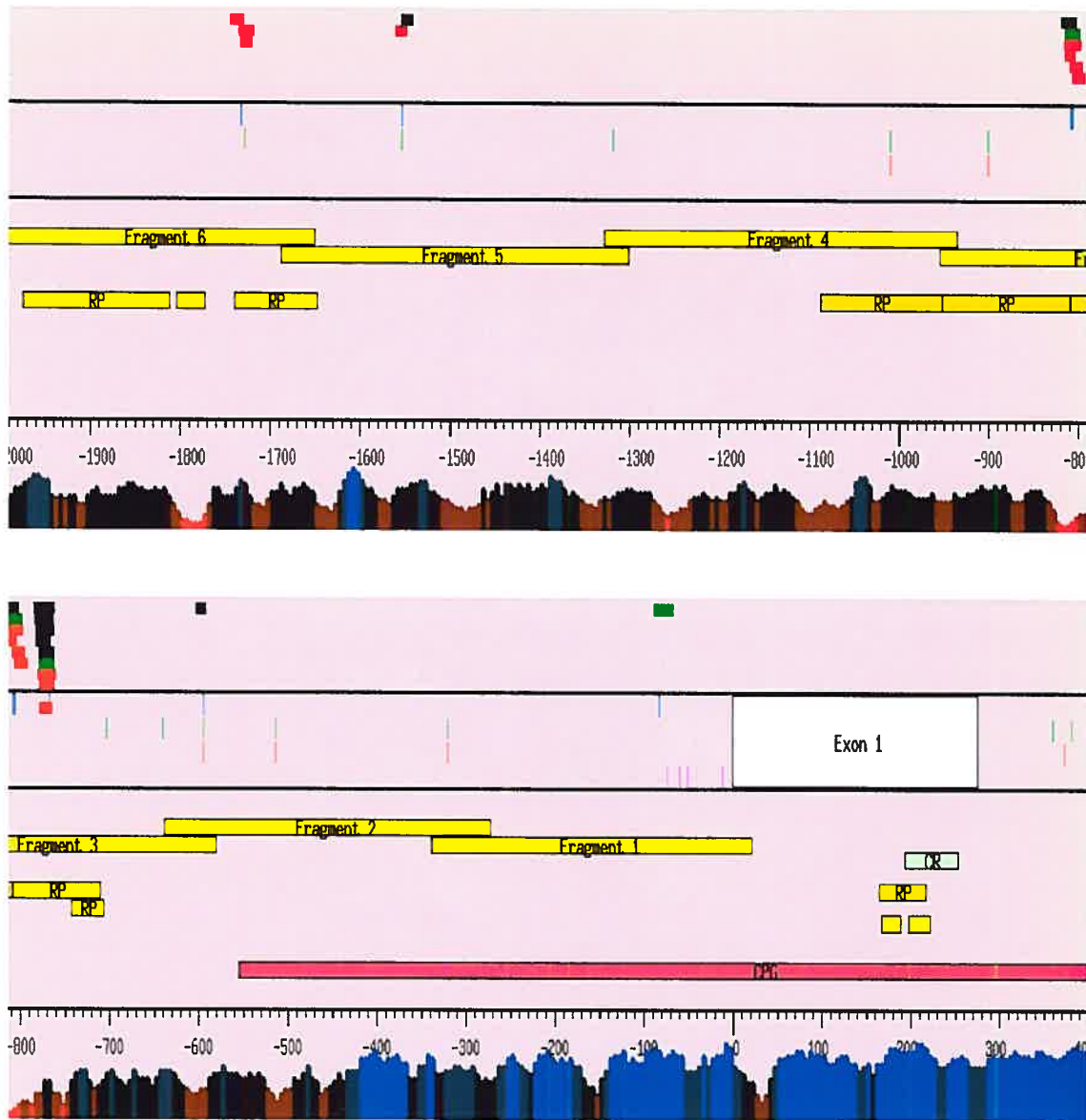


Figure 14: Représentation graphique du promoteur du gène RB1.

Tableau XIX : Description détaillée des polymorphismes identifiés dans le promoteur du gène codant pour RB1.

| Gène | Position | Type de polymorphisme | Sites de liaison allèle majeur | Sites de liaison allèle mineur | Sites de liaison conservés | Répétition |
|------|---------------|--------------------------|--|--|----------------------------|------------------------|
| RB1 | Ex1.- 1554 | C->A transversion | <u>c-Myb, important in hematopoiesis, cellular equivalent to avian myoblastosis virus oncogene v-myb->0.983</u> | <u>c-Myb, important in hematopoiesis, cellular equivalent to avian myoblastosis virus oncogene v-myb->0.983</u> <u>TEF-1 related muscle factor->0.854</u> | | |
| RB1 | Ex1.- 1729 | DEL C nd | | | | LTR/MaLR MLT1L |
| RB1 | Ex1.- 1733 | INS CTGATA nd | | <u>Ecotropic viral integration site 1 encoded factor->0.842</u> <u>GATA-binding factor 1->0.993</u> <u>Mammalian transcriptional repressor RBP-Jkappa/CBF1->0.846</u> | | LTR/MaLR MLT1L |
| RB1 | Ex1.-595 | T->G transversion | <u>v-Myb, variant of AMV v-myb->0.948</u> | <u>v-Myb, variant of AMV v-myb->0.943</u> | | |
| RB1 | Ex1.-768 | C->A transversion | <u>Avian C-type LTR TATA box->0.795</u> <u>E4BP4, bZIP domain, transcriptional repressor->0.808</u> <u>E4BP4, bZIP domain, transcriptional repressor->0.85</u> <u>Ecotropic viral integration site 1 encoded factor->0.855</u> <u>Homeodomain transcription factor Otx2 (homolog of Drosophila orthodenticle)->0.99</u> <u>Myocyte enhancer factor->0.815</u> | <u>Avian C-type LTR TATA box->0.795</u> <u>Cone-rod homeobox-containing transcription factor / otx-like homeobox gene->0.951</u> <u>E4BP4, bZIP domain, transcriptional repressor->0.808</u> <u>E4BP4, bZIP domain, transcriptional repressor->0.85</u> <u>Ecotropic viral integration site 1 encoded factor->0.855</u> <u>Homeodomain binding site in LIM/Homeodomain factor LHX3->0.814</u> <u>Myocyte enhancer factor->0.808</u> <u>Pdx1 (IDX1/IPF1) pancreatic and intestinal homeodomain TF->0.754</u> | | SINE/MIR MIRb |
| RB1 | Ex1.-807 | T->C transition | <u>Ecotropic viral integration site 1 encoded factor->0.851</u> | <u>Gut-enriched Krueppel-like factor->0.976</u> <u>Interferon regulatory factor 3 (IRF-3)->0.85</u> | | SINE/MIR MIRb |

| | | | | | | |
|------------|-----------------|---|---|--|--|--------------------|
| <u>RB1</u> | <u>Ex1.-809</u> | T->C transition | <u>Ecotropic viral integration site 1 encoded factor- >0.851</u> | <u>Ecotropic viral integration site 1 encoded factor- >0.834</u> <u>Gut-enriched Krueppel-like factor- >0.976</u> <u>Nuclear factor of activated T-cells- >0.98</u> | | SINE/MIR MIRb |
| <u>RB1</u> | <u>Ex1.-83</u> | DEL CTCGCGGACGT GACGCCGCGGCG nd | <u>Neural-restrictive-silencer-element- >0.674</u> | | | |

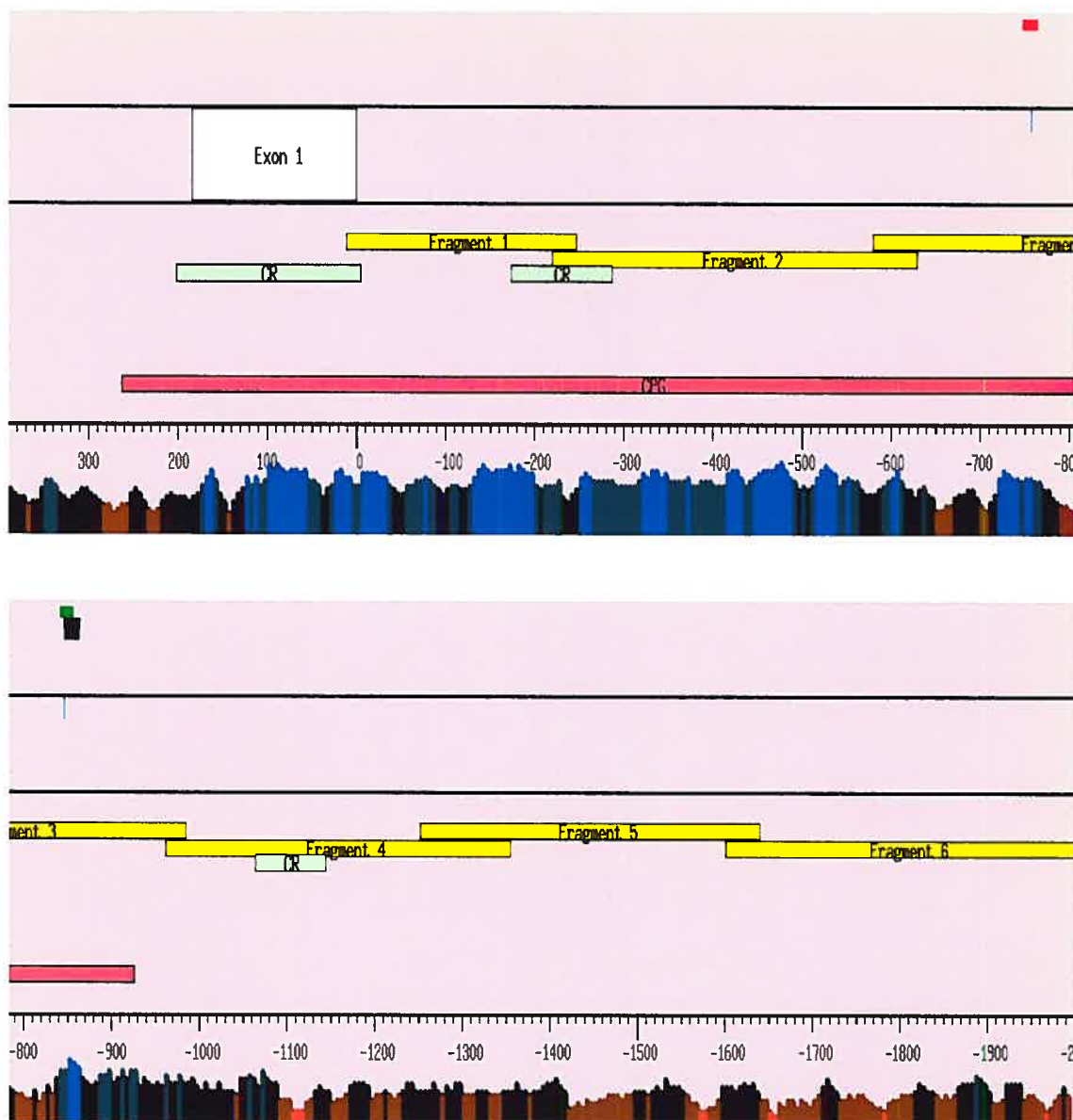


Figure 15: Représentation graphique du promoteur du gène SKP2.

Tableau XX : Description détaillée des polymorphismes identifiés dans le promoteur du gène codant pour SPK2.

| Gène | Position | Type de polymorphisme | Sites de liaison allèle majeur | Sites de liaison allèle mineur | Sites de liaison conservés | Répétition |
|-------------|-----------------|-----------------------|---|---|----------------------------|------------|
| <u>SKP2</u> | <u>Ex1.-758</u> | G->C transversion | | <u>Hepatic nuclear factor 1->0.848</u> | | |
| <u>SKP2</u> | <u>Ex1.-846</u> | A->G transition | <u>E2F, involved in cell cycle regulation, interacts with Rb p107 protein->0.77</u> <u>Glucocorticoid receptor, C2C2 zinc finger protein binds glucocorticoid dependent to GREs->0.863</u> <u>Sox-5->0.982</u> | <u>Glucocorticoid receptor, C2C2 zinc finger protein binds glucocorticoid dependent to GREs->0.866</u> <u>Sox-5->0.982</u> | | |

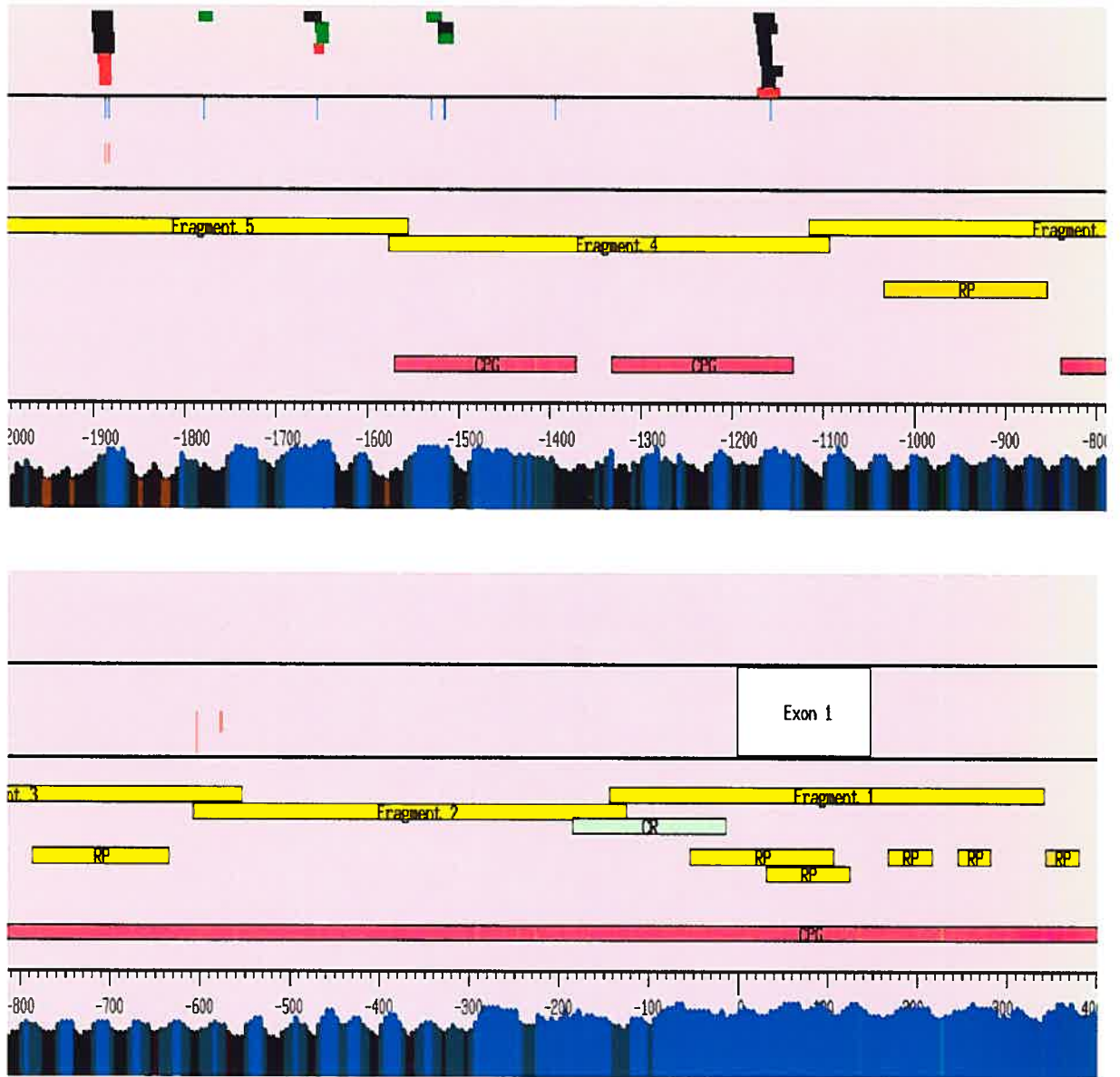


Figure 16: Représentation graphique du promoteur du gène TFDP1.

Tableau XXI : Description détaillée des polymorphismes identifiés dans le promoteur du gène codant pour TFDP1.

| Gène | Position | Type de polymorphisme | Sites de liaison allèle majeur | Sites de liaison allèle mineur | Sites de liaison conservés | Répétition |
|--------------|------------------|-----------------------|---|---|--|------------|
| <u>TFDP1</u> | <u>Ex1.-1157</u> | INS C nd | <u>GAGA-Box->0.788</u> <u>GC box elements->0.897</u> <u>MYC-associated zinc finger protein related transcription factor->0.966</u> <u>Transcriptional repressor, binds to elements found predominantly in genes that participate in lipid metabolism->0.771</u> <u>Wilms Tumor Suppressor->0.945</u> <u>Zinc finger transcription factor ZBP-89->0.959</u> <u>Zinc finger transcription factor ZBP-89->0.974</u> | <u>GC box elements->0.897</u> <u>MYC-associated zinc finger protein related transcription factor->0.968</u> <u>Stimulating protein 1 SP1, ubiquitous zinc finger transcription factor->0.893</u> <u>Transcriptional repressor, binds to elements found predominantly in genes that participate in lipid metabolism->0.771</u> <u>Wilms Tumor Suppressor->0.945</u> <u>Zinc finger transcription factor ZBP-89->0.959</u> <u>Zinc finger transcription factor ZBP-89->0.981</u> | <u>Wilms Tumor Suppressor</u> <u>Wilms Tumor Suppressor</u> | |
| <u>TFDP1</u> | <u>Ex1.-1158</u> | INS C nd | <u>GC box elements->0.901</u> <u>MYC-associated zinc finger protein related transcription factor->0.936</u> <u>Transcriptional repressor, binds to elements found predominantly in genes that participate in lipid metabolism->0.762</u> <u>Wilms Tumor Suppressor->0.893</u> <u>Wilms Tumor Suppressor->0.921</u> <u>Zinc finger transcription factor ZBP-89->0.959</u> <u>Zinc finger transcription factor ZBP-89->0.961</u> | <u>GAGA-Box->0.788</u> <u>GC box elements->0.897</u> <u>MYC-associated zinc finger protein related transcription factor->0.966</u> <u>Transcriptional repressor, binds to elements found predominantly in genes that participate in lipid metabolism->0.771</u> <u>Wilms Tumor Suppressor->0.893</u> <u>Wilms Tumor Suppressor->0.893</u> <u>Wilms Tumor Suppressor->0.945</u> <u>Zinc finger transcription factor ZBP-89->0.959</u> <u>Zinc finger transcription factor ZBP-89->0.974</u> | <u>Wilms Tumor Suppressor</u> <u>Wilms Tumor Suppressor</u> | |
| <u>TFDP1</u> | <u>Ex1.-1394</u> | T->C transition | | <u>PAX9->0.974</u> | | |
| <u>TFDP1</u> | <u>Ex1.-1515</u> | G->T transversion | <u>VDR/RXR Vitamin D receptor RXR heterodimer site->0.869</u> | | | |

| | | | | | | |
|-------|-----------|--------------------------|---|--|--|--|
| TFDP1 | Ex1.-1516 | C->T transition | VDR/RXR Vitamin D receptor RXR heterodimer site->0.869 | VDR/RXR Vitamin D receptor RXR heterodimer site->0.873 | | |
| TFDP1 | Ex1.-1530 | C->T transition | c-Ets-1(p54)->0.956 | | | |
| TFDP1 | Ex1.-1655 | G->A transition | GC box elements->0.883 MYC-associated zinc finger protein related transcription factor->0.891 PAX6 paired domain and homeodomain are required for binding to this site- >0.931 | PAX6 paired domain and homeodomain are required for binding to this site- >0.939 v-Myb->0.997 | | |
| TFDP1 | Ex1.-1779 | A->G transition | Thyroid transcription factor-1 (TTF1) binding site->0.922 | | | |
| TFDP1 | Ex1.-1883 | G->C transversion | Core promoter-binding protein (CPBP) with 3 Krueppel-type zinc fingers- >0.913 Zinc finger transcription factor ZBP-89->0.947 | Basic krueppel-like factor (KLF3)->0.95 Core promoter-binding protein (CPBP) with 3 Krueppel-type zinc fingers- >0.913 Zinc finger transcription factor ZBP-89->0.947 | | |
| TFDP1 | Ex1.-1887 | C->T transition | Core promoter-binding protein (CPBP) with 3 Krueppel-type zinc fingers- >0.913 Zinc finger transcription factor ZBP-89->0.947 | Basic krueppel-like factor (KLF3)->0.982 Basic transcription element (BTE) binding protein, BTEB3, FKLF-2->0.94 Core promoter-binding protein (CPBP) with 3 Krueppel-type zinc fingers- >0.886 Zinc finger transcription factor ZBP-89->0.94 | | |

Annexe VII

Table XXII: 197 genes screened by dHPLC for the presence of pSNPs and their predicted impact on TFBSs.

| Gene | Locuslink | Total bp screened | # pSNPs | pSNPs with predicted impact on TFBSs |
|---------|-----------|-------------------|---------|--------------------------------------|
| ABL1 | 25 | 1526 | 0 | 0 |
| ACAT1 | 38 | 1732 | 5 | 3 |
| ACPI | 52 | 2029 | 5 | 4 |
| ACTR1B | 10120 | 2070 | 3 | 3 |
| ADA | 100 | 1731 | 6 | 5 |
| ADAM17 | 6868 | 1726 | 8 | 4 |
| ADPRTL1 | 143 | 2155 | 16 | 14 |
| ADRA2B | 151 | 1869 | 5 | 3 |
| ADRB2 | 154 | 1974 | 15 | 11 |
| AGT | 183 | 2128 | 16 | 10 |
| AIRE | 326 | 2041 | 11 | 8 |
| APEX | 328 | 2160 | 10 | 6 |
| APM1 | 9370 | 2164 | 13 | 12 |
| ATF5 | 22809 | 2069 | 10 | 7 |
| ATM | 472 | 1873 | 6 | 3 |
| B2M | 567 | 1963 | 6 | 6 |
| BAT2 | 7916 | 351 | 1 | 1 |
| BCL11A | 53335 | 1969 | 1 | 1 |
| BDKRB1 | 623 | 2096 | 15 | 10 |
| BF | 629 | 2071 | 3 | 3 |
| BLM | 641 | 1968 | 8 | 4 |
| BRCA1 | 672 | 1599 | 7 | 2 |
| BRCA2 | 675 | 2343 | 9 | 7 |
| BTK | 695 | 2045 | 8 | 8 |
| BTN3A2 | 11118 | 2112 | 40 | 31 |
| C2 | -717 | 1757 | 7 | 6 |
| C5 | 727 | 2186 | 8 | 7 |
| CART | 9607 | 1962 | 20 | 18 |
| CAT | 847 | 2172 | 9 | 6 |
| CAV1 | 857 | 898 | 4 | 3 |
| CCKAR | 886 | 2483 | 12 | 9 |
| CCND1 | 595 | 1686 | 15 | 10 |
| CCND2 | 894 | 1135 | 0 | 0 |
| CCND3 | 896 | 1559 | 1 | 0 |
| CCNE1 | 898 | 924 | 2 | 2 |
| CCR2 | 1231 | 2004 | 9 | 4 |
| CCR6 | 1235 | 2124 | 15 | 13 |
| CD36L1 | 949 | 1678 | 11 | 10 |
| CD86 | 942 | 2037 | 11 | 7 |

| | | | | |
|---------|-------|------|----|----|
| CD8A | 925 | 2021 | 11 | 8 |
| CDC25A | 993 | 1784 | 7 | 6 |
| CDK4 | 1019 | 1872 | 11 | 8 |
| CDK6 | 1021 | 2072 | 8 | 6 |
| CDKN1A | 1026 | 1717 | 12 | 6 |
| CDKN1B | 1027 | 2092 | 11 | 7 |
| CDKN2A | 1029 | 2149 | 8 | 4 |
| CDKN2B | 1030 | 2130 | 10 | 8 |
| CETP | 1071 | 2051 | 19 | 13 |
| CHUK | 1147 | 1995 | 13 | 10 |
| CPE | 1363 | 1732 | 6 | 2 |
| CX3CR1 | 1524 | 2067 | 17 | 10 |
| CYP3A4 | 1576 | 1520 | 4 | 3 |
| CYP3A5 | 1577 | 867 | 2 | 2 |
| DCLRE1C | 64421 | 2141 | 8 | 7 |
| DHFR | 1719 | 2135 | 19 | 15 |
| E2F1 | 1869 | 2005 | 4 | 3 |
| EPHX1 | 2052 | 2128 | 15 | 11 |
| ERCC1 | 2067 | 1682 | 6 | 3 |
| F7 | 2155 | 2077 | 23 | 22 |
| FANCA | 2175 | 2227 | 18 | 15 |
| FANCC | 2176 | 1941 | 6 | 5 |
| FCER1A | 2205 | 2113 | 8 | 7 |
| FCER1G | 2207 | 2067 | 6 | 5 |
| FCGR2B | 2213 | 1982 | 15 | 11 |
| FEN1 | 2237 | 2072 | 5 | 3 |
| FGB | 2244 | 2031 | 15 | 12 |
| FIZZ3 | 56729 | 2338 | 20 | 12 |
| FLT1 | 2321 | 2096 | 8 | 3 |
| FPGS | 2356 | 1778 | 10 | 7 |
| FY | 2532 | 2155 | 10 | 7 |
| GADD45A | 1647 | 2118 | 4 | 4 |
| GART | 2618 | 1167 | 2 | 2 |
| GCK | 2645 | 2093 | 10 | 5 |
| GCLC | 2729 | 2068 | 7 | 5 |
| GHRHR | 2692 | 1211 | 11 | 6 |
| GPI | 2821 | 1768 | 3 | 3 |
| GPX2 | 2877 | 2157 | 17 | 15 |
| GPX3 | 2878 | 2051 | 14 | 8 |
| GPX4 | 2879 | 2234 | 29 | 22 |
| GRL | 2908 | 1488 | 7 | 6 |
| GSS | 2937 | 2011 | 9 | 7 |
| GSTM1 | 2944 | 2082 | 53 | 41 |
| GSTM3 | 2947 | 2031 | 6 | 5 |
| GSTM4 | 2948 | 1860 | 10 | 9 |
| GSTP1 | 2950 | 2140 | 13 | 10 |
| GSTT1 | 2952 | 2115 | 5 | 5 |

| | | | | |
|---------|-------|------|----|----|
| GYPA | 2993 | 1513 | 8 | 7 |
| HDAC1 | 3065 | 2109 | 12 | 8 |
| HSD11B1 | 3290 | 2281 | 6 | 6 |
| HTR2A | 3356 | 2133 | 18 | 9 |
| HUS1 | 3364 | 2015 | 9 | 6 |
| ICAM1 | 3383 | 1912 | 4 | 3 |
| IL10 | 3586 | 2086 | 7 | 7 |
| IL18R1 | 8809 | 1981 | 13 | 7 |
| IL19 | 29949 | 2039 | 6 | 5 |
| IL1A | 3552 | 2088 | 3 | 2 |
| IL1R1 | 3554 | 2155 | 11 | 9 |
| IL1RN | 3557 | 2096 | 13 | 11 |
| IL2 | 3558 | 2162 | 5 | 5 |
| IL6 | 3569 | 2054 | 14 | 13 |
| IL8RA | 3577 | 1918 | 6 | 4 |
| IL9R | 3581 | 2035 | 10 | 9 |
| IRS1 | 3667 | 2086 | 2 | 2 |
| ISL1 | 3670 | 1786 | 10 | 8 |
| ITGA2 | 3673 | 2044 | 6 | 3 |
| KL | 9365 | 1901 | 14 | 12 |
| KLK1 | 3816 | 2091 | 10 | 7 |
| KU70 | 2547 | 1968 | 17 | 14 |
| KU80 | 7520 | 1886 | 16 | 12 |
| LDLR | 3949 | 2059 | 8 | 6 |
| LEP | 3952 | 2086 | 14 | 8 |
| LIG4 | 3981 | 1314 | 6 | 5 |
| LILRB4 | 11006 | 2042 | 27 | 19 |
| LIPE | 3991 | 2410 | 5 | 4 |
| LMNA | 4000 | 2112 | 5 | 4 |
| LTA | 4049 | 2115 | 16 | 12 |
| LY64 | 4064 | 1955 | 7 | 3 |
| MADH3 | 4088 | 1526 | 7 | 6 |
| MADH4 | 4089 | 1805 | 2 | 1 |
| MBD4 | 8930 | 2207 | 9 | 5 |
| MEFV | 4210 | 2014 | 7 | 5 |
| MET | 4233 | 1082 | 5 | 4 |
| MICA | 4276 | 2244 | 33 | 28 |
| MIF | 4282 | 2097 | 22 | 17 |
| MKKS | 8195 | 2114 | 10 | 9 |
| MMP26 | 56547 | 1710 | 8 | 5 |
| MRE11A | 4361 | 2281 | 22 | 18 |
| MSH2 | 4436 | 1550 | 9 | 8 |
| MSH3 | 4437 | 1062 | 2 | 1 |
| MSH5 | 4439 | 1572 | 3 | 1 |
| NBS1 | 4683 | 1565 | 9 | 8 |
| NFATC4 | 4776 | 1992 | 7 | 7 |
| NFKB1 | 4790 | 1443 | 3 | 3 |

| | | | | |
|---------|-------|------|----|----|
| NFKB2 | 4791 | 2042 | 4 | 3 |
| NFKBIA | 4792 | 2025 | 10 | 8 |
| NOS3 | 4846 | 1781 | 0 | 0 |
| NPY | 4852 | 2094 | 13 | 9 |
| NPY5R | 4889 | 2055 | 12 | 9 |
| PAI1 | 5054 | 2161 | 13 | 8 |
| PCNA | 5111 | 1884 | 8 | 6 |
| PLAU | 5328 | 1592 | 6 | 3 |
| PLAUR | 5329 | 2025 | 6 | 5 |
| PON2 | 5445 | 1830 | 4 | 2 |
| PRKDC | 5591 | 1602 | 5 | 2 |
| PSMB9 | 5698 | 2173 | 4 | 4 |
| RAD23A | 5886 | 2018 | 1 | 1 |
| RAD50 | 10111 | 1496 | 6 | 6 |
| RAD51 | 5888 | 2077 | 10 | 7 |
| RAD54L | 8438 | 1797 | 1 | 0 |
| RB1 | 5925 | 2043 | 7 | 6 |
| RECQL | 5965 | 2008 | 7 | 4 |
| REV3L | 5980 | 1975 | 3 | 2 |
| RFC1 | 6573 | 2039 | 10 | 9 |
| RFX5 | 5993 | 1751 | 3 | 3 |
| RFXAP | 5994 | 1986 | 5 | 3 |
| SCYA2 | 6347 | 2192 | 8 | 8 |
| SCYA7 | 6354 | 1960 | 11 | 7 |
| SDF1 | 6387 | 1958 | 18 | 15 |
| SELPLG | 6404 | 2126 | 11 | 11 |
| SKP2 | 6502 | 2014 | 2 | 2 |
| SLC11A1 | 6556 | 659 | 0 | 0 |
| SOD1 | 6647 | 2057 | 13 | 11 |
| SOD2 | 6648 | 2023 | 9 | 7 |
| SOD3 | 6649 | 2134 | 9 | 7 |
| SSA1 | 6737 | 1709 | 13 | 11 |
| STAT1 | 6772 | 1698 | 13 | 9 |
| STAT6 | 6778 | 949 | 5 | 4 |
| SUR1 | 6833 | 1814 | 12 | 8 |
| TAP1 | 6890 | 1813 | 16 | 9 |
| TBX3 | 6926 | 2005 | 11 | 7 |
| TCF7 | 6932 | 737 | 7 | 4 |
| TFDP1 | 7027 | 1019 | 9 | 7 |
| TFPT | 29844 | 2044 | 7 | 4 |
| TGFB1 | 7040 | 2092 | 9 | 6 |
| TGFBR1 | 7046 | 1989 | 2 | 1 |
| TGFBR2 | 7048 | 2037 | 6 | 6 |
| TLR4 | 7099 | 2128 | 6 | 4 |
| TLR5 | 7100 | 1803 | 4 | 4 |
| TLR9 | 54106 | 2127 | 6 | 2 |
| TNFA | 7124 | 2032 | 8 | 8 |

| | | | | |
|----------|-------|------|----|----|
| TNFRSF1A | 7132 | 2031 | 13 | 11 |
| TNFRSF1B | 7133 | 1439 | 8 | 5 |
| TNFRSF6 | 355 | 2004 | 7 | 6 |
| TNFSF13B | 10673 | 2180 | 11 | 9 |
| TNFSF5 | 959 | 2230 | 3 | 3 |
| TP53BP1 | 7158 | 2148 | 2 | 2 |
| TUB | 7275 | 2124 | 11 | 7 |
| UCP1 | 7350 | 2111 | 15 | 11 |
| UCP3 | 7352 | 2131 | 9 | 7 |
| UGT1A1 | 54658 | 2025 | 4 | 3 |
| VEGF | 7422 | 2030 | 13 | 10 |
| WRN | 7486 | 1589 | 2 | 2 |
| XDH | 7498 | 2172 | 8 | 8 |
| XPC | 7508 | 2138 | 6 | 6 |
| XRCC2 | 7516 | 1257 | 7 | 5 |
| XRCC3 | 7517 | 988 | 2 | 0 |
| XRCC4 | 7518 | 2112 | 12 | 6 |

Curriculum vitae

| | | | |
|-------------------------------------|--|------------------------|-------------|
| Formation Académique | 2003 à aujourd'hui | Université de Montréal | Montréal |
| | M.Sc Biologie Moléculaire | | |
| | Titre : Variabilité génétique dans les régions régulatrices de gènes impliqués dans le contrôle du cycle cellulaire. Directeur : Daniel Sinnett | | |
| | 1999-2002 | Université de Montréal | Montréal |
| | B.Sc. Sciences biologiques : orientation microbiologie et immunologie | | |
| Expériences professionnelles | Mai 2002-Déc 2002 | Hôpital Ste-Justine | Montréal |
| | Assistante de recherche | | |
| | Design d'amorce, PCR, identification de polymorphismes à l'aide de dHPLC. | | |
| | Été 2000 et été 2001 | CIRSA | Sacré-Coeur |
| | Assistante de recherche | | |
| | Travail sur le terrain : pêche électrique, identification et manipulation de poissons. Analyse en laboratoire : identification des contenus stomacaux. | | |
| Prix | Premier prix présentation par affiche par un étudiant à la maîtrise, au congrès annuel des Étudiants du Centre de Recherche de l'hôpital Ste-Justine (Juin 2003) | | |
| Publications | <u>Bélanger, H.</u> , Beaulieu, P., Sinnett, D. (2005) Functional promoter SNP in cell cycle checkpoint genes. (en préparation) | | |
| | Sinnett, D., Théberge, M.-C., <u>Bélanger, H.</u> , Langlois, S., Drouin, S., Beaulieu P., Gehl, D., Zotti, C., Lefebvre, J.-F., Labuda, D., (2005) Detection and characterization of regulatory single nucleotide polymorphisms in hundreds of human diseases candidate genes. (en préparation) | | |
| Communications | Labuda, D., Lefebvre, J.-F., Moreau, C., Labbé, C., Langlois, S., Beaulieu P., Théberge, M.-C., <u>Bélanger, H.</u> , Sinnett, D., and Genome-Quebec Regulatory Genetics Consortium. Contribution of gene conversions to the genetic diversity of DNA segments. European Human Genetics Conference, Prague, mai 2005 (affiche) | | |
| | <u>Bélanger, H.</u> , Beaulieu, P., Lacroix, S., Labuda, D. et Sinnett, D. Impact of regulatory SNP in G1/S checkpoint genes. AACR: Cell Cycle and Cancer: Pathways and Therapies, Floride, décembre 2004 (affiche) | | |
| | Healy, J., <u>Bélanger, H.</u> , Vasquez, H., Moghrabi, A., Sinnett, D. Cyclin-Dependant Kinase Inhibitors: Regulatory Polymorphisms and their Impact on the susceptibility to Childhood Leukemia. AACR: Cell Cycle and Cancer: Pathways and Therapies, Floride, décembre 2004 (affiche) | | |

Bélanger, H., Beaulieu, P., Lariviere, M., Moreau, C., Langlois S., Labuda D. et Sinnett D. Functional polymorphism in the regulatory region of genes implicated in the G1/S checkpoint. The American Society of Human Genetics, Toronto, octobre 2004 (affiche)

Beaulieu, P., Théberge, M.-C., Bélanger, H., Langlois, S., Lariviere, M., Lefebvre, J.-F., Gehl, D., Labuda, D., Sinnett, D. In-silico assessment of the impact of single nucleotide polymorphisms in the regulatory regions of human genes. The American Society of Human Genetics, Toronto, octobre 2004 (affiche)

Théberge, M.-C., Bourgoïn, S., Beaulieu P., Langlois, S., Bélanger, H., Gehl, D., Labuda, D., Sinnett, D. High-throughput Detection of Polymorphisms in Regulatory Regions of Candidate Genes Relevant to Common Diseases. The American Society of Human Genetics, Toronto, octobre 2004 (affiche)

Labbé, C., Moreau, C., Yotova, V., Gelh, D., Lefebvre, J.-F., Beaulieu P., Langlois, S., Théberge, M.-C., Bélanger, H., Zotti, C., Sinnett, D., Labuda, D. and Regulatory Genetics Consortium. Polymorphisms in Genes of Glutathione Metabolism: Promoter Regions. The American Society of Human Genetics, Toronto, octobre 2004 (affiche)

Labuda, D. Moreau, C., Langlois, S., Gelh, D., Beaulieu P., Lefebvre, J.-F., Vasquez, H., Labbé, C., Théberge, M.-C., Bélanger, H., Zotti, C., Sinnett, D., and Genome-Quebec Regulatory Genetics Consortium. Population genetic of *bona fides* regulatory regions. The American Society of Human Genetics, Toronto, octobre 2004 (affiche)

Bélanger, H., Beaulieu P., Lariviere, M., Moreau, C., Langlois S., Labuda D. et Sinnett D. Functional polymorphism in the regulatory region of genes implicated in the G1/S checkpoint. Onogenetics, Montréal, octobre 2004 (affiche)

Bélanger, H., Sinnett, D., Variabilité génétique dans les régions régulatrices de gènes impliqués dans le contrôle du cycle cellulaire. Congrès annuel des Étudiants du Centre de Recherche de l'hôpital Ste-Justine, juin 2004 (affiche)

Bélanger, H., Sinnett, D., Variabilité génétique dans les régions régulatrices de gènes impliqués dans le contrôle du cycle cellulaire. Xe Journée scientifique des programmes de biologie moléculaire, Université de Montréal, mai 2004 (affiche)

Bélanger, H., Théberge, M.-C., Langlois, S., Bourgoïn, S., Beaulieu P., Gehl, D., Zotti, C., Labuda, D., Sinnett, D. High-throughput Detection of Polymorphisms in Regulatory Regions of Candidate Genes Related to Common Genetic Diseases. American Association for Cancer Research, Washington DC, juillet 2003 (oral)

Bélanger H., Théberge, M.C., Langlois S., Bourgoïn S., Beaulieu P., Gehl D., Zotti C., Labuda D., Sinnett D. Détection de Variants d'ADN dans les Régions Régulatrices de Gènes Candidats pour des Maladies Complexes. Congrès annuel des Étudiants du Centre de Recherche de l'hôpital Ste-Justine, juin 2003 (affiche)

Qualifications supplémentaires

Cours d'entraînement sur le dHPLC: Transgenomic WAVE® System and WAVEMAKER™ Software (23 juillet 2002)

