

Université de Montréal

Caractérisation du locus 12p12.3 impliqué dans la leucémie lymphoblastique
aiguë

par
Alexandre Montpetit

Département de biochimie
Faculté de médecine

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

Janvier 2002

© Alexandre Montpetit, 2002



Université de Montréal
Faculté des études supérieures

Cette thèse intitulée:

Caractérisation du locus 12p12.3 impliqué dans la leucémie lymphoblastique
aiguë

présentée par:
Alexandre Montpetit

a été évaluée par un jury composé des personnes suivantes:

Michel Bouvier
président-rapporteur

Daniel Sinnett
directeur de recherche

Jean Vacher
membre du jury

Jacques Simard
examineur externe

Edward Bradley
représentant du doyen de la FES

Sommaire

La leucémie lymphoblastique aiguë (LLA) est le cancer pédiatrique le plus fréquent, mais les mécanismes moléculaires en cause sont peu connus. Les délétions hémizygotiques de la région chromosomique 12p12.3 sont un événement fréquent chez les patients atteints de LLA, étant observés dans 26-40 % des cas. Ceci suggère la présence d'un important gène suppresseur de tumeurs, son identification étant le but de ce projet. La région affectée commune à tous les patients est située dans un intervalle de 750 kb, incluant au moins un gène connu, *ETV6*, essentiel pour l'angiogenèse ainsi que l'établissement de l'hématopoïèse dans la moelle osseuse. Une des étapes limitantes est l'identification des gènes présents dans la région. Dans ce but, nous avons premièrement utilisé un modèle d'analyse génomique comparative, le poisson-lune *Fugu rubripes*. Ce poisson possède un contenu en gènes similaire à l'humain malgré un génome très compact. De plus, la présence de régions génomiques synténiques entre les deux espèces en fait un modèle avantageux pour l'identification de gènes. Nous avons effectivement identifié chez *Fugu* une région synténique incluant *ETV6*. Par contre, les gènes identifiés ne se retrouvaient pas dans la région 12p12.3 chez l'humain, mais étaient distribués sur tout le chromosome 12, indiquant que *Fugu rubripes* n'est pas un bon modèle pour le clonage positionnel de gènes. Il est toutefois un excellent modèle pour caractériser les gènes: nous avons ainsi identifié par analyse comparative plusieurs régions très conservées de *ETV6*, permettant de mieux délimiter les domaines fonctionnellement importants, ainsi que d'autres pouvant jouer un rôle dans la régulation de la transcription de ce gène. Suite à la complétion du projet de séquençage du génome humain, une

approche directe d'identification des gènes a pu être envisagée chez l'humain. Nous avons construit une carte de transcription détaillée en développant une approche combinant des techniques d'analyse moléculaires et informatiques. De cette façon, nous avons pu confirmer la présence de sept gènes, tous exprimés au moins dans la moelle osseuse, dont trois n'ayant jamais été caractérisés. Par la suite, nous avons cherché, chez des patients atteints de LLA, la présence de mutations dans les gènes candidats. Aucune mutation inactivante n'a été identifiée, mais nous avons observé que l'expression d'*ETV6* était abolie chez ces patients suggérant que ce gène est la cible des délétions. Comme la translocation t(12;21) générant la protéine de fusion ETV6-AML1 est toujours associée à la perte d'expression d'*ETV6*, deux modèles, non mutuellement exclusifs, peuvent alors être proposés, soit que 1- la présence de ETV6-AML1 ou que 2- l'absence d'*ETV6* est l'événement important dans le développement de 30 % des cas de LLA.

Mots-clés: leucémie, cancer, génétique, Fugu rubripes, ETV6, gène suppresseur de tumeurs, délétions, translocations, génomique.

Summary

Acute lymphoblastic leukaemia (ALL) is the most frequent paediatric cancer. The underlying molecular bases are not known, but allelic losses on chromosome 12p12.3 are very frequent, ranging from 26 to 40% of cases. These observations suggest that an important tumour suppressor gene is present, its identification being the goal of this project. The commonly deleted region to all patients is located in a 750 kb interval including at least one gene, *ETV6*, essential for the angiogenesis and hematopoiesis processes. The limiting step is the identification of the genes present within this locus. To this goal, we first used a genomic comparative model, the pufferfish *Fugu rubripes*. Its genome is extremely small, but it contains a similar gene set than human. Moreover, syntenic regions have been observed between the two species, making it an excellent model for gene identification. Although we did observe a syntenic region in *Fugu*, including *ETV6*, none of the genes identified were present on 12p12.3 in human and were rather distributed on all of chromosome 12. This suggests that *Fugu rubripes* is not a good model for gene positional cloning. However it is an excellent model for gene characterisation: by comparative analysis, we identified many conserved regions in *ETV6* that could define more precisely functionally important domains, and other that could play essential roles in the regulation of its expression. Since we now have access to the complete human genome sequence, we developed an approach combining experimental and informatic tools to identify the genes. By this method, we identified seven genes, all expressed in the bone marrow, including three that were uncharacterised. We then looked in LLA patients for the presence of mutations in those genes. No inactivating mutations

were identified, but we observed that expression of *ETV6* was always abolished suggesting that it is the target of the deletions. We also observe that the t(12;21) translocation, generating the ETV6-AML1 fusion protein, is also always associated with the deletions. Thus, two non-mutually exclusive models can be described: 1- the presence of ETV6-AML1 or 2- the absence of ETV6 is the critical event for the development and progression of these ALL cases.

Keywords: leukaemia, cancer, genetics, *Fugu rubripes*, ETV6, tumour suppressor gene, deletions, translocations, genomics.

Table des Matières

| | Page |
|---|----------|
| Sommaire | iii |
| Summary | v |
| Table des matières | vii |
| Liste des tableaux | xi |
| Liste des figures | xii |
| Liste des abréviations | xiv |
| Remerciements | xvi |
| | |
| Chapitre 1 - Introduction | 1 |
| 1.1 Cancer | 2 |
| 1.1.1 Incidence | 2 |
| 1.1.2 Définition cellulaire | 2 |
| 1.1.3 Environnement et cancer | 3 |
| 1.1.4 Génétique du cancer | 5 |
| 1.1.4.1 Mutations génétiques | 5 |
| 1.1.4.2 Nature multi-séquentielle du cancer | 7 |
| 1.1.4.3 Cancers héréditaires | 7 |
| 1.1.4.4 Modèle génétique du cancer | 12 |
| 1.2 Cancers pédiatriques | 14 |
| 1.2.1 Incidence | 14 |
| 1.3 Leucémie | 16 |
| 1.3.1 Définition et incidence | 16 |
| 1.3.2 Causes environnementales | 17 |
| 1.3.3 Causes génétiques | 18 |
| 1.3.4 Aberrations chromosomiques | 18 |
| 1.3.4.1 Translocations chromosomiques | 19 |

| | |
|---|-----------|
| 1.3.4.1.1 Bcr-Abl | 19 |
| 1.3.4.1.2 Récepteurs des cellules T et immunoglobulines | 21 |
| 1.3.4.1.3 MLL | 22 |
| 1.3.4.1.4 RAR α | 23 |
| 1.3.4.1.5 AML1 | 24 |
| 1.3.4.2 Délétions | 27 |
| 1.3.4.2.1 Analyse LOH | 29 |
| 1.4 Identification de délétions associées à la leucémie | 31 |
| 1.4.1 Chromosome 12p12 | 35 |
| 1.5 Caractérisation des gènes dans la région 12p12.3 | 38 |
| 1.5.1 ETV6 | 38 |
| 1.5.1.1 Famille ETS | 38 |
| 1.5.1.2 Fonction de ETV6 | 39 |
| 1.5.1.3 Translocations impliquant ETV6 | 41 |
| 1.5.1.3.1 Translocation t(12;21): ETV6-AML1 | 44 |
| 1.5.2 CDKN1B | 47 |
| 1.5.3. Autres gènes connus dans la région 12p12.3 | 50 |
| 1.6 Approches utilisées pour l'identification de gènes | 51 |
| 1.6.1 Sélection d'ADNc | 52 |
| 1.6.2 Capture d'exons | 52 |
| 1.6.3 îlots CpG | 54 |
| 1.6.4 Analyse comparative | 57 |
| 1.6.5 Analyse bio-informatique | 64 |
| 1.7 Hypothèses et objectifs du projet | 64 |
| | |
| Chapitre 2 -Article 1-: Analysis of the conservation of synteny between Fugu and human chromosome 12 | 65 |
| | |
| Abstract | 67 |
| Introduction | 68 |

| | |
|---|----|
| Materials and methods | 71 |
| Results | 73 |
| Isolation of clones and shotgun sequencing | 73 |
| Sequence analysis of the <i>fETV6</i> locus | 74 |
| Discussion | 78 |
| References | 81 |

Chapitre 3 - Article 2 - Comparative analysis of the *ETV6* gene in vertebrate genomes from pufferfish to human 93

| | |
|------------------------|-----|
| Abstract | 95 |
| Introduction | 96 |
| Results and discussion | 98 |
| Acknowledgments | 104 |
| References | 104 |

Chapitre 4 - Article 3 - A detailed transcriptional map of the chromosome 12p12 tumour suppressor locus 112

| | |
|---|-----|
| Abstract | 114 |
| Introduction | 115 |
| Materials and methods | 116 |
| Results | 120 |
| Sequence analysis | 120 |
| Characterization of the candidate genes | 122 |
| Alternative termination | 125 |
| CpG islands | 126 |
| Discussion | 126 |
| Acknowledgments | 130 |
| References | 130 |

| | |
|--|------|
| Chapitre 5 - Article 4 - Expression and mutational analysis of the 12p candidate tumour suppressor genes in pre-B acute lymphoblastic leukemia. | 144 |
| Abstract | 146 |
| Introduction | 147 |
| Materials and methods | 149 |
| Results | 152 |
| Discussion | 155 |
| References | 159 |
| Chapitre 6 - Discussion - | 174 |
| 6.1 État de la question | 175 |
| 6.2 Analyse comparative du locus 12p12.3 chez <i>Fugu rubripes</i> | 176 |
| 6.3 Analyse comparative de ETV6 chez <i>Fugu</i> | 182 |
| 6.4 Analyse de la séquence génomique humaine du locus 12p12.3 | 187 |
| 6.5 Étude mutationnelle des gènes candidats chez les patients | 195 |
| 6.6 Modèle proposé | 200 |
| 6.7 Perspectives | 204 |
| 6.8 Conclusion | 207 |
| References | 209 |
| Curriculum vitae | xvii |

Liste des Tableaux

| | Page |
|---|------|
| Chapitre 1 | |
| Tableau I-I Types de mutations retrouvées dans les cellules cancéreuses | 6 |
| Tableau I-II Classification cellulaire de certains oncogènes connus | 8 |
| Tableau I-III Classification cellulaire de certains suppresseurs de tumeurs connus | 11 |
| Tableau I-IV Fréquence de la délétion de la région 12p12 dans différents cancers | 34 |
| Tableau I-V Comparaison de la taille de gènes orthologues chez l'humain et chez <i>Fugu</i> | 59 |
| Chapitre 2 | |
| Table 1 Gene identified in the <i>Fugu</i> ETV6 locus | 89 |
| Chapitre 3 | |
| Table 1 Intron sizes and exon/intron boundaries of the human and <i>Fugu</i> ETV6 genes | 109 |
| Chapitre 4 | |
| Table 1 Predicted transcriptional units within the chromosome 12p12 tumour suppressor locus | 137 |
| Table 2 Gene expression analyses by RT-PCR | 138 |
| Chapitre 5 | |
| Table 1 Primers used in the mutational analysis | 165 |
| Table 2 Primers used in the RT-PCR analysis | 166 |
| Table 3 Deletion mapping of chromosome 12p12.3 in 19 cases of ALL | 167 |
| Table 4 Polymorphisms identified | 168 |
| Table 5 Expression of 12p12.3 genes in 26 pre-B ALL patients | 169 |
| Table 6 Correlation of ETV6, BCL-G _L , MKP-7 and CDKN1B mRNA expression with clinical and biological findings in 26 pre-B LLA patients | 170 |

Liste des Figures

Page

Chapitre 1

| | | |
|-------------|---|----|
| Figure 1-1 | Modèle génétique du développement du cancer colorectal | 9 |
| Figure 1-2 | Translocations observées dans les cas de LLA et de LMA pédiatriques | 20 |
| Figure 1-3 | Schéma de la hiérarchie des précurseurs hématopoïétiques | 26 |
| Figure 1-4 | Contrôle du cycle cellulaire par les CDKI | 28 |
| Figure 1-5 | Fréquence de pertes alléliques sur différents chromosomes dans les LLA de type B et T | 33 |
| Figure 1-6 | Régions minimales de délétions sur le chromosome 12p12.3 dans différentes études | 37 |
| Figure 1-7 | Structure du gène ETV6 | 43 |
| Figure 1-8 | Représentation de la stratégie employée pour séquencer le génome humain | 53 |
| Figure 1-9 | Technique de capture d'exons | 55 |
| Figure 1-10 | Représentation de deux régions synténiques entre l'humain et Fugu | 60 |

Chapitre 2

| | | |
|----------|--|----|
| Figure 1 | Physical map of the <i>ETV6</i> locus of <i>Fugu rubripes</i> . | 90 |
| Figure 2 | Highly conserved non-coding sequence within a <i>LIG-2</i> intron. | 91 |
| Figure 3 | Synteny analysis between <i>Fugu</i> and the human chromosome 12. | 92 |

Chapitre 3

| | | |
|----------|--|-----|
| Figure 1 | Comparison of the human and <i>Fugu rubripes</i> <i>ETV6</i> genomic structures. | 110 |
| Figure 2 | Alignment of the predicted <i>ETV6</i> amino acid sequence from various vertebrates. | 111 |

Chapitre 4

- Figure 1 Physical and transcriptional map of the commonly deleted region on chromosome 12p12.3. 138
- Figure 2 Genomic structure of *BCL-G*, *MKP-7*, *LOH1CR12*, *LOH2CR12* and *LOH3CR12*. 139
- Figure 3 Northern blot analysis of *MKP-7*, *LOH1CR12* and *LOH3CR12* mRNAs. 140
- Figure 4 Amino acid alignment of hypothetical LOH1CR12 proteins from distant species. 141
- Figure 5 Amino acid alignment of hypothetical protein LOH3CR12 from mouse and human. 142
- Figure 6 The partial cDNA sequence and predicted open reading frame of the *LOH2CR12* gene. 143

Chapitre 5

- Figure 1 Physical and transcriptional map of the commonly deleted region on chromosome 12p12.3. 171
- Figure 2 SSCP analysis of exon 3 of *BCL-G* in 13 ALL patients with 12p12 LOH. 172
- Figure 3 *ETV6*, *BCL-G*, *MKP-7* and *CDKN1B* mRNA expression in 20 pre-B ALL patients. 173

Chapitre 6

- Figure 6-1 Modèle décrivant les participations possibles de *ETV6*-*AML1* et de *ETV6* à la leucémogénèse 203

Liste des Abréviations

| | |
|------|---|
| ADN | acide désoxyribonucléique |
| ADNc | ADN complémentaire |
| ARN | acide ribonucléique |
| BAC | chromosome artificiel de bactérie |
| CDK | kinase cycline-dépendante |
| CDKI | inhibiteur de complexes cyclines-CDK |
| cM | centi-Morgan |
| EST | séquence exprimée |
| ETS | séquences transformantes de E26 |
| FISH | fluorescence par hybridation <i>in situ</i> |
| GPR | récepteur couplé aux protéines G |
| HLH | hélice-boucle-hélice |
| Ig | immunoglobulines des cellules B |
| kb | kilo- paires de bases |
| LG | groupe de linkage |
| LLA | leucémie lymphoblastique aiguë |
| LMA | leucémie myéloïde aiguë |
| LOH | perte d'hétérozygotie |
| Mapk | mitogen-activated protein kinases |
| Mb | mega- paires de bases |
| MKP | phosphatase de Mapk |
| MLL | leucémie de lignées multiples |
| Myr | millions d'années |
| pb | paires de bases |
| PCR | réaction de polymérisation en chaîne |
| PNT | gène <i>pointed</i> de la drosophile |
| Rb | rétinoblastome |

| | |
|------|--|
| RNAi | interférence par l'ARN double brin |
| RT | transcription réverse |
| TEL | translocation-ETS-leucémie |
| TCR | récepteur des cellules T |
| SSCP | polymorphisme de conformation de l'ADN simple brin |
| SNP | polymorphisme au niveau d'un nucléotide |
| YAC | chromosome artificiel de levure |

Remerciements

Je voudrais remercier sincèrement mon directeur de recherche, Daniel Sinnett, pour m'avoir encouragé et avoir toujours été disponible tout au long des mes études.

Je voudrais aussi remercier tous les gens formidables du pavillon Vidéotron ainsi que du Centre de Recherche que j'ai rencontrés pour leur aide, les discussions ou tout simplement pour l'ambiance agréable au laboratoire ou durant les partys! J'aimerais remercier tout particulièrement les personnes suivantes qui ont toujours été disponibles pour m'aider dans mes travaux: Josiane, Gino, Zeina, Caroline, Chantal, Manon, Maryse, Hughes, Nicolas, Vania, Richard, Dominik, Raffaella, Dominika et Stéphanie.

Je tiens aussi à remercier ma mère, mon père et Danielle pour leur support. Je sais bien que c'est la seule page de ma thèse que vous allez lire, mais j'ai toujours bien aimé discuter avec vous de mes recherches! Finalement, merci beaucoup Marie-Eve pour avoir toujours su m'encourager et me remonter le moral durant les moments plus difficiles.

Je garderai toujours un bon souvenir de mes années à l'Hôpital Ste-Justine!

Chapitre 1

-Introduction-

1.1 Cancer

1.1.1 Incidence

Dans les pays industrialisés, le cancer est, avec les maladies coronariennes, une des maladies les plus répandues et les plus coûteuses, autant au niveau des coûts médicaux qu'elle entraîne qu'au niveau du nombre de vies perdues. On estime qu'en 2001, 134 100 nouveaux cas de cancer seront diagnostiqués et que 65 300 personnes mourront du cancer au Canada seulement (Statistiques canadiennes sur le cancer, 2001). Ce qui est plus inquiétant est que le taux d'incidence du cancer est toujours en progression, et de façon plus marquée chez les femmes. Certains progrès sont toutefois encourageants: le taux de mortalité pour tous les cancers depuis 1988, excepté le cancer du poumon chez les femmes, est en nette baisse. De plus, le taux de mortalité dû au cancer chez les enfants a diminué de près de 50 % depuis 40 ans.

1.1.2 Définition cellulaire

Le cancer est en fait une expansion clonale de cellules ayant perdu toutes formes de contrôles de la croissance cellulaire. En culture, les cellules cancéreuses ne se comportent pas comme des cellules normales et montrent plusieurs des caractéristiques qui définissent le phénotype de cellules transformées (Buick et Tannock, 1992):

1. Indépendance vis-à-vis le substrat: les cellules normales, à l'exception de la plupart des cellules du système hématopoïétique, doivent être ancrées à une matrice (substrat) afin de pouvoir croître et se diviser. Les cellules transformées peuvent croître en suspension dans un milieu liquide.

2. Inhibition de contact: les cellules normales ne se divisent plus lorsqu'elles entrent en contact avec d'autres cellules. Ainsi, elles forment une unique couche cellulaire lorsque mises en culture. Les cellules transformées ne sont pas inhibées au contact et peuvent croître l'une par-dessus l'autre formant des colonies ou foci.

3. Indépendance vis-à-vis le sérum: les cellules normales ont besoin d'un apport continu de facteurs de croissance afin d'être maintenues en culture. Les cellules transformées n'ont plus besoin de ce type de signal pour continuer de croître et de se diviser.

4. Formation de tumeurs dans un modèle animal: les cellules transformées ont la capacité de produire des tumeurs lorsque injectées dans une souris de laboratoire immuno-supprimée, ce qui n'est pas le cas des cellules normales.

De plus, la plupart des cellules cancéreuses sont immortelles et peuvent donc potentiellement se diviser à l'infini. Cette propriété d'immortalité n'est partagée qu'avec les cellules souches. Les cellules dites normales du corps humain ont une durée de vie limitée et ne peuvent effectuer que quelques dizaines de divisions lorsque mises en culture avant d'entrer dans un état de sénescence et éventuellement de mourir (Hayflick and Moorhead, 1961).

1.1.3 Environnement et cancer

Différents produits présents dans l'environnement ont été identifiés comme pouvant causer le cancer. En 1761, le Dr. John Hill publie une étude indiquant que l'inhalation de "snuff", une poudre fine de tabac, cause le cancer des voies nasales. Cette

étude est réputée comme étant la première identifiant un facteur externe causant un cancer. Quelques années plus tard, Percival Pott rapporte que les ramoneurs de Londres courent un plus grand risque de contracter un cancer du scrotum et propose que la suie en est la cause. Il faudra attendre plus d'un siècle avant que d'autres études épidémiologiques mettant en cause différents produits chimiques soient publiées et que les autorités publiques agissent afin de protéger les citoyens (revu par Bishop, 1995).

Aujourd'hui, plusieurs facteurs de risques environnementaux sont identifiés: les produits chimiques (tels la fumée de cigarette, les pesticides, le benzène, etc.), l'alcool, les médicaments (incluant des agents anticancéreux comme les inhibiteurs de la topoisomérase), la diète alimentaire, les radiations physiques (soleil, rayons-X, etc.), les hormones ainsi que les virus (revu par Harras *et al.*, 1996). Ces facteurs ont tous un point en commun: ils ont un effet direct ou indirect sur le matériel génétique de la cellule. Les mutations génétiques sont donc la cause principale du cancer. Plusieurs gènes, codant pour des protéines de fonction diverses, sont affectés dans les différents cancers. L'identification de ces gènes est donc une étape fondamentale afin de comprendre le développement et la progression des cancers. La participation des chromosomes dans le processus de la tumorigenèse a été démontré depuis près d'un siècle (revu par Bishop, 1995). Par contre, ce n'est qu'au début des années 60 que des chercheurs ont identifié une aberration chromosomique associée spécifiquement à un type de cancer, soit la découverte du chromosome de Philadelphie, chez les patients atteints de la leucémie lymphoblastique chronique, impliquant une translocation des chromosomes 9 et 22 (Novell et Hungerford, 1960).

1.1.4 Génétique du cancer

1.1.4.1 Mutations génétiques

Les gènes impliqués dans l'oncogenèse peuvent être mutés de différentes façons. Certaines mutations présentes sont héréditaires, mais celles-ci ne constitueraient que 5 à 10 % des cas de cancer (Knudson, 1993). La plupart des cancers impliquent l'apparition de mutations spontanées ou induites par l'action de carcinogènes externes (chimiques, physiques ou viraux) sur le matériel génétique. Ces mutations peuvent prendre différentes formes menant à l'expression de protéines mutantes, ou la surexpression ou l'inhibition de l'expression de certaines protéines. Elles sont décrites dans le Tableau I-I.

Les premiers gènes responsables de l'apparition de cancers ont été isolés à partir de virus. En 1970, le gène *v-src*, du génome du virus du sarcome de Rous, a été décrit comme étant responsable de la transformation cellulaire et est ainsi devenu le premier oncogène à être identifié (Martin, 1970). Par la suite, il a été démontré que *v-src* avait un homologue cellulaire normal, *c-src*. Donc, *v-src*, ainsi que la plupart des autres oncogènes viraux identifiés, sont d'origine cellulaire et ont été introduits dans les génomes viraux par transduction (revu par Bishop *et al.*, 1995). Par conséquent, ces oncogènes cellulaires ont été nommés proto-oncogènes puisqu'ils doivent être mutés afin de transformer les cellules. Les mutations transformant les proto-oncogènes en oncogènes peuvent avoir différents effets: surexpression, expression ectopique, mutation entraînant une activité constitutive, etc.

Tableau I-I. Types de mutations retrouvées dans les cellules cancéreuses

| Type de mutation | Exemple | Référence |
|---|-------------------------|----------------------------------|
| Mutation ponctuelle | <i>Ras</i> | Bos, 1989 |
| Intégration virale | <i>Wnt-1</i> | Nusse <i>et al.</i> , 1984 |
| Amplification génique | <i>Myc</i> | Schwab, 1983 |
| Délétion | <i>Rb-1</i> | Venter <i>et al.</i> , 1991 |
| Translocation chromosomique générant un gène de fusion | t(9;22): <i>Bcr-Abl</i> | Heisterkamp <i>et al.</i> , 1985 |
| Translocation chromosomique menant à la surexpression d'un gène | t(14;18): <i>Bcl-2</i> | Tsujimoto <i>et al.</i> , 1984 |
| Changements dans la méthylation de l'ADN | <i>Vhl</i> | Baylin <i>et al.</i> , 1998 |

Une liste partielle des proto-oncogènes identifiés à ce jour est présentée dans le Tableau I-II. On remarque que les voies métaboliques empruntées par les proto-oncogènes sont très variées: facteurs de croissance, transmission des signaux intracellulaires, facteurs de transcription, inhibition de l'apoptose, etc.

1.1.4.2 Nature multi-séquentielle du cancer

Il est établi depuis plusieurs années qu'une seule mutation n'est pas suffisante pour provoquer un cancer. *In vitro*, la transfection d'au moins trois oncogènes est nécessaire afin de transformer des cellules normales humaines, contrairement à des cellules de rats ou de souris où deux seulement sont suffisants (Hahn *et al.*, 1999; voir Section 1.1.4.4). Des études statistiques sur la fréquence du cancer en fonction de l'âge suggèrent que 5 à 6 événements ou mutations sont nécessaires afin de développer un cancer chez l'humain (Peto *et al.*, 1975). Le cancer du côlon est idéal pour évaluer ce processus puisqu'il se développe très lentement et progresse selon des stades cytologiquement distincts, de bénin jusqu'à métastatique. Vogelstein et Kinzler (1993) ont proposé un modèle où la combinaison de six événements génétiques séquentiels est reliée à la progression typique du cancer du côlon (Figure 1-1).

1.1.4.3 Cancers héréditaires

Certains cancers héréditaires, comme le cancer du sein impliquant des mutations dans les gènes *BRCA1* ou *BRCA2*, n'apparaissent qu'à l'âge adulte et sont ainsi beaucoup plus difficile à étudier (revu par Nathanson *et al.*, 2001). D'autres, comme le

Tableau I-II. Classification cellulaire de certains oncogènes connus

| Proto-oncogène | Fonction de la protéine | Localisation cellulaire |
|--|--|------------------------------------|
| <i>Protéines sécrétées</i> | | |
| PDGFβ | Facteur de croissance | Extracellulaire |
| Wnt1 | Facteur de croissance / développement du système nerveux central | Extracellulaire |
| <i>Récepteurs mitogènes</i> | | |
| ErbB | Récepteur tyrosine kinase d'EGF | Transmembranaire |
| Fms | Récepteur tyrosine kinase de CSF-1 | Transmembranaire |
| <i>Transducteurs de signaux intracellulaires</i> | | |
| Abl | Protéine tyrosine kinase | Cytoplasmique/nucléaire |
| Ras | Protéine liant le GTP | Associée à la membrane |
| <i>Inhibiteurs de l'apoptose</i> | | |
| Bcl-2 | Inhibiteur de l'apoptose en inhibant Bax | Membrane mitochondriale/Cytoplasme |
| <i>Facteurs de transcription</i> | | |
| Myc | Facteur de transcription à domaine HLH | Nucléaire |
| ErbA | Récepteur de l'hormone thyroïdienne | Nucléaire |
| <i>Cycle cellulaire</i> | | |
| Cycline D1 | Associée aux CDK pour la progression du cycle cellulaire | Cytoplasmique |

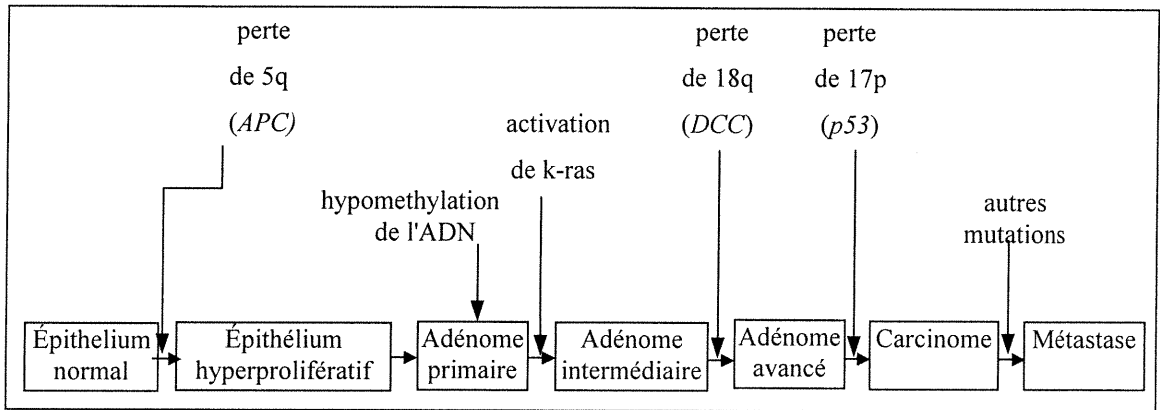


Figure 1-1. Modèle génétique du développement du cancer colorectal. Adapté de Vogelstein et Kinzler (1993).

rétinoblastome, évoluent beaucoup plus rapidement et apparaissent dès les premières années de la vie. Dans le cas de la forme héréditaire du rétinoblastome, une copie du gène défectueux, *Rb-1*, est hérité des parents et est présente dans toutes les cellules. Environ 80 % des porteurs développent une tumeur dans les deux yeux avant l'âge de 5 ans, 15 % dans un seul œil tandis que 5 % sont des porteurs asymptomatiques. Ces porteurs ont aussi un plus grand risque de développer éventuellement des tumeurs dans plusieurs autres organes. À l'aide de ce modèle, Knudson (1971) a émis l'hypothèse que la transformation d'une cellule normale en une cellule cancéreuse nécessitait au moins deux mutations, soit l'hypothèse des "two-hits".

Les analyses subséquentes ont d'ailleurs démontré que le deuxième allèle de *Rb-1* était muté dans les cas de rétinoblastome (revu dans Squire et Phillips, 1995). Étant donné que ces mutations résultent en des inactivations complètes de *Rb-1*, et non une activation aberrante comme dans le cas des oncogènes, le terme anti-oncogène ou gène suppresseur de tumeurs a été utilisé pour décrire ce type de gène impliqué dans le cancer. Une liste partielle des gènes suppresseurs de tumeurs connus est présentée dans le Tableau I-III. On peut remarquer que les gènes suppresseurs de tumeurs sont impliqués dans de nombreuses voies métaboliques, plus particulièrement dans le contrôle du cycle cellulaire.

Le facteur de transcription p53 est un des plus importants membres de ce groupe puisqu'il est altéré dans près de 50 % des cancers. Il contrôle plusieurs voies métaboliques et est responsable en grande partie de la décision de la cellule d'entrer en

Tableau I-III. Classification cellulaire de certains suppresseurs de tumeurs connus

| Gène suppresseur de tumeurs | Fonction de la protéine | Localisation cellulaire |
|--------------------------------------|--|--------------------------------|
| <i>Cycle cellulaire</i> | | |
| pRb | Inhibe la transcription (par E2F) des composants du cycle cellulaire | Nucléaire |
| p16 | Inhibiteur de complexes cycline / CDK | Nucléaire |
| <i>Apoptose</i> | | |
| Bax | Favorise l'apoptose en inhibant Bcl-2 | Mitochondrie / cytoplasme |
| <i>Réparation de l'ADN</i> | | |
| hMLH1 | Réparation de l'ADN suite aux erreurs de réplication | Nucléaire |
| BRCA1 | Réparation de l'ADN, se complexe à Rad51 | |
| <i>Toutes ces catégories</i> | | |
| p53 | Facteur de transcription / active plusieurs gènes suite aux dommages subis par l'ADN | Nucléaire |
| <i>Signalisation intracellulaire</i> | | |
| PTEN | Tyrosine-sérine/thréonine phosphatase | Cytoplasmique |
| APC | Inactive β -caténine (voie des protéines WNT) | Cytoplasmique |

adapté de Haber et Fearon (1998).

mode réparation d'ADN et d'arrêter la croissance cellulaire ou en apoptose lorsque les dommages subis à l'ADN sont trop importants. Son absence peut entraîner l'amplification ou la perte de matériel génétique et augmenter la fréquence totale des mutations génétiques (revu par Cox et Lane, 1995). Les gènes impliqués dans le contrôle de l'apoptose, tel *Bax*, sont également fréquemment touchés étant donné qu'ils déclenchent une réaction en chaîne aboutissant à la mort cellulaire lorsque des dommages irréparable ont été commis à la cellule et à son matériel génétique. Finalement, les gènes impliqués dans la réparation de l'ADN, comme *hMLH1*, muté dans certains cas de cancer du côlon et de l'estomac, forment une cible idéale pour une cellule cancéreuse puisque lorsque cette voie est dérégulée, le taux de mutations augmente de façon importante accélérant le processus oncogénique.

1.1.4.4 Modèle génétique du cancer

L'introduction d'oncogènes *in vivo* par transgénèse laisse suggérer que plusieurs événements sont essentiels à la formation de tumeurs. L'insertion d'une forme altérée de l'oncogène *myc* ou de l'oncogène *ras* induit la formation de tumeurs chez seulement 20-50 % des souris après six mois. L'introduction des deux oncogènes induit la formation de tumeurs dans 100 % des cas (Sinn *et al.*, 1987). Cependant, ces tumeurs apparaissent au hasard indiquant que d'autres événements somatiques sont nécessaires. Une analyse plus poussée a démontré dans certaines de ces tumeurs une perte du chromosome 15, indiquant possiblement l'inactivation de gènes suppresseurs de tumeurs, mais qu'aucune de ces modifications n'était responsable de l'immortalisation des cellules (Oshimura *et al.*, 1985). Toutes les observations faites sur le cancer pointent vers le modèle suivant:

activation d'oncogènes, inactivation de gènes suppresseurs de tumeurs et immortalisation cellulaire (Barrett *et al.*, 1988). L'activation de la télomérase a été démontrée récemment comme étant responsable de cette immortalisation et elle est exprimée dans plus de 90 % des cancers. Elle reforme les extrémités des chromosomes, les télomères, qui normalement rétrécissent à chaque division, empêchant ainsi le signal déclenchant la sénescence et la mort cellulaire. La télomérase n'est normalement exprimée que dans les cellules souches y compris les cellules germinales (revu par Bachand et Autexier, 1999). L'introduction de l'oncogène Ras, du facteur viral large-T qui inactive les suppresseurs de tumeurs p53 et Rb-1, et de la télomérase entraîne la transformation cellulaire de fibroblastes normaux: c'est la première démonstration de la création d'une cellule cancéreuse à partir d'éléments génétiques définis chez l'humain (Hahn *et al.*, 1999).

Ce modèle du cancer est évidemment très simplifié. Par exemple, il y a maintenant des évidences qu'il existe des gènes suppresseurs de tumeurs non-classiques, où l'inactivation d'un seul allèle est suffisante pour la carcinogenèse (haploinsuffisance). Si la surexpression d'un proto-oncogènes peut entraîner l'apparition d'un cancer en modifiant l'équilibre dans la signalisation cellulaire, il est logique de penser que la diminution de l'expression d'un gène suppresseur de tumeurs pourrait avoir le même effet. Par exemple, les souris hétérozygotes pour la présence du gène *PTEN*, encodant une protéine à activité tyrosine et sérine/thréonine phosphatase, développent spontanément des tumeurs des gonades, de la glande thyroïde et de la prostate (DiCristofano *et al.*, 1998). Aussi, il y a des gènes qui semblent pouvoir jouer un rôle d'oncogène ou d'anti-oncogène selon les circonstances. Par exemple, des formes mutées

de p53 peuvent agir comme dominants positifs en inhibant les complexes multimériques de ce facteur de transcription (revu par Deppert *et al.*, 2000). Le type cellulaire a aussi une influence importante sur le développement d'un cancer. Les cellules moins différenciées et qui se divisent rapidement, comme les fibroblastes ou certaines cellules du système hématopoïétiques, peuvent être immortalisées beaucoup plus facilement en culture que des cellules épithéliales, indiquant que les niveaux de contrôle sont différents d'un type cellulaire à l'autre (Kiyono *et al.*, 1998). Finalement, la classification des causes du cancer en éléments environnementaux et éléments génétiques n'est pas aussi bien définie dans la majorité des cas et il faut plutôt évaluer ces causes en termes d'interactions gène-environnement. Certaines prédispositions, soit des polymorphismes, ou variations génétiques, retrouvés dans une grande partie de la population normale, réduisant ou augmentant l'activité de certaines enzymes peuvent modifier significativement le risque de contracter un cancer. Par exemple, certains polymorphismes de CYP1A1, un enzyme impliqué dans le métabolisme des carcinogènes, ont un plus grand risque de développer certains cancers, dont le cancer du poumon. Ce risque est encore plus grand lorsque ces individus sont des fumeurs (revu par Bartsch *et al.*, 2000).

1.2 Cancers Pédiatriques

1.2.1 Incidence

Les cancers pédiatriques diffèrent en plusieurs points des cancers dits adultes, particulièrement au niveau du type de cancers observés, la durée de leur développement et leur incidence dans la population. Les cancers pédiatriques ne représentent que 1 % de

l'ensemble des cancers diagnostiqués (Statistiques canadiennes sur le cancer, 2001). Par contre, ils constituent la première cause de mortalité chez les enfants de 0 à 19 ans malgré une amélioration de plus de 50 % du taux de survie depuis 50 ans. Les cancers les plus fréquents sont les cancers du système hématopoiétique (leucémies) représentant 26 % des cas, les cancers du système nerveux (astrocytome, neuroblastome, etc.), représentant 22 % des cas, et les cancers du système lymphatique (lymphomes), soit 16 % des cas. Les autres types de cancers pédiatriques observés (ostéosarcome, tumeur de Wilm, rhabdomyosarcome, cancer des gonades, rétinoblastome, etc.) sont beaucoup moins fréquents et représentent moins de 5 % des cas chacun.

Plusieurs différences fondamentales existent entre le cancer pédiatrique et celui de l'adulte. Les tissus fœtaux prolifèrent beaucoup plus rapidement que les tissus adultes et sont aussi beaucoup moins différenciés, ce qui augmente les chances qu'une cellule échappe aux contrôles et que le cancer se développe rapidement. D'ailleurs, le nombre de mutations génétiques retrouvé dans les cancers pédiatriques serait plus bas que dans les tumeurs adultes (Knudson, 1993). Aussi, les principaux cancers retrouvés chez l'adulte (poumon, sein, prostate et côlon) ne sont pas observés chez l'enfant. La majorité de ces cancers sont d'origine épithéliale, donc très exposés aux carcinogènes de l'environnement, tandis que la majorité des cancers pédiatriques sont d'origine mésenchymale, donc relativement mieux protégés de l'action des carcinogènes externes (revu par Greaves, 1999).

1.3 Leucémie

1.3.1 Définition et incidence

La leucémie est le résultat de l'accumulation excessive de cellules immatures du système hématopoïétique dans la moelle osseuse et le sang. Dans les cas de leucémie aiguë, la moelle osseuse produit tellement de ces cellules qu'elle ne peut plus produire suffisamment de globules rouges, de leucocytes et de plaquettes pour les besoins du corps humain, résultant en une anémie et une susceptibilité aux infections sévères. Les leucémies constituent le cancer pédiatrique le plus fréquemment diagnostiqué (26 % des cas) et aussi celui entraînant le plus de décès (31 % des décès) (Statistiques canadiennes sur le cancer, 2001). La leucémie lymphoblastique aiguë (LLA) constitue plus de 75 % des cas de leucémies contre 15-20 % pour la leucémie d'origine myéloïde, ce qui est exactement l'inverse de ce qui est observé à l'âge adulte. Près de 85 % des cas de LLA sont le résultat de l'expansion clonale d'une cellule précurseur d'un lymphocyte B (LLA pré-B). L'expansion d'un précurseur de cellules T représente environ 15 % des cas (LLA pré-T) tandis que la leucémie à cellules B matures ne constitue qu'environ 1 % des cas. Le taux de survie sur cinq ans pour les enfants atteints de LLA était seulement d'environ 5 % au début des années 60. Aujourd'hui, surtout à cause des nouveaux protocoles de chimiothérapie plus agressifs et plus diversifiés, le taux de survie peut atteindre 75 %. Cependant, les cas de leucémie secondaire suite à une rechute sont beaucoup plus difficiles à traiter. Dans ces cas, la leucémie est le résultat de l'apparition d'une cellule cancéreuse résistante aux traitements de chimiothérapie et le taux de survie ne dépasse pas les 20 %. Les traitements actuels sont par contre non-spécifiques et entraînent ainsi plusieurs effets secondaires non-désirés. Le Gleevec (STI-571) est un nouveau

traitement utilisé chez les patients atteints de la leucémie myeloïde chronique, une forme rare de leucémie rencontrée principalement chez les adultes. Ce composé est un inhibiteur de la protéine de fusion Bcr-Abl, ayant une activité tyrosine kinase constitutive, retrouvée chez les patients ayant la translocation t(9;22) ou chromosome de Philadelphie (revu par Schindler *et al.*, 2000). Plus de 90 % des patients répondent bien au traitement initial et aucun effet secondaire majeur n'est observé. Cet exemple illustre bien la nécessité de comprendre les causes moléculaires du cancer, soit l'identification des oncogènes et des gènes suppresseurs de tumeurs, afin de développer des médicaments plus ciblés, qui seront ainsi moins toxiques et plus efficaces.

1.3.2 Causes environnementales

Les causes de la leucémie sont très peu connues. Certaines études ont mis en évidence l'effet de l'exposition carcinogènes chimiques comme le benzène, la peinture, les pesticides et certains médicaments anticancéreux sur le taux de cancers pédiatriques (revu par McBride, 1998). Dans ce dernier cas, on observe que les survivants d'un cancer pédiatriques courent six fois plus de risques de développer un cancer à l'âge adulte, fort probablement à cause de la chimiothérapie et des radiations utilisés pour les guérir (Neglia *et al.*, 2001). Les études portant par sur l'effet des radiations, par exemple des lignes électriques à haute tension ou des émissions provenant des appareils électroménagers, sont plus controversées. Par exemple, plusieurs études ont observé que les enfants exposés constamment à des champs magnétiques résidentiels de plus de 0,3 μ Tesla étaient 2 fois plus à risque de développer une leucémie (revu par Ahlbom *et al.*, 2000; Schuz *et al.*, 2001), tandis que d'autres n'observent aucune association statistique

(McBride *et al.*, 1999; Kleinerman *et al.*, 2000). Les animaux de laboratoires exposés à des radiations similaires ne développent cependant pas de leucémies ou de lymphomes (revu par Boorman *et al.*, 2000).

1.3.3 Causes génétiques

Les causes génétiques de la leucémie sont peu connues. On estime que seulement 5 % des cas de leucémies ont une composante héréditaire. Certains syndromes associés avec des défauts dans la réparation de l'ADN ont été reliés avec un taux élevé de leucémies. Aussi, les enfants atteints du syndrome de Down ont de 20 à 30 fois plus de chances de développer une leucémie (revu par Greaves, 1999). Un bon modèle pour évaluer la contribution héréditaire sur l'apparition de la leucémie est l'étude des jumeaux. Le taux de concordance de l'apparition d'une leucémie aiguë chez des jumeaux monozygotiques (versus des jumeaux dizygotiques) est de près de 50 % lorsqu'elle apparaît avant l'âge d'un an et pourrait approcher 100 % pour ceux possédant un placenta unique. Cette concordance n'est seulement d'environ 5 % lorsque la leucémie apparaît entre 1 à 15 ans et de moins de 1 % après (Buckley *et al.*, 1996). De plus, chez certains jumeaux, un décalage dans l'apparition de la leucémie a été observé, pouvant atteindre dix ans, mais impliquant toujours un clone d'origine commune. Ceci indique clairement que plus d'une mutation est nécessaire au développement de ces leucémies. La majorité des leucémies seraient donc amorcées par des mutations somatiques *in utero* ou du moins très tôt durant le développement.

1.3.4 Aberrations chromosomiques

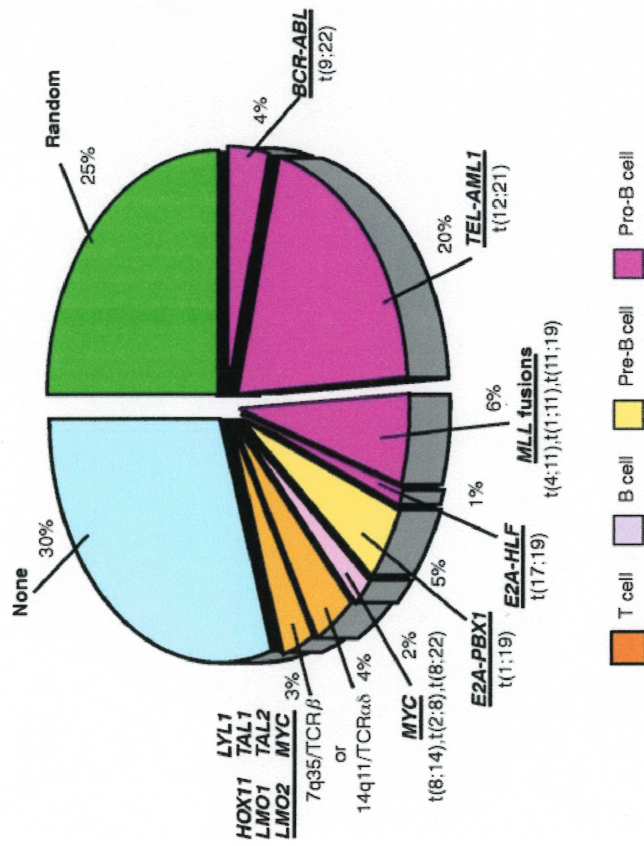
Une des caractéristiques principales de la leucémie est la fréquence élevée d'aberrations chromosomiques observées par analyse cytogénétique. Plus de 75 % des leucémies sont caractérisées par la présence de translocations ou d'inversions chromosomiques récurrentes. Aussi, dans plus de 50 % des cas, les cellules leucémiques ont un caryotype anormal, ayant moins ou plus de 46 chromosomes (Pui *et al.*, 1990). Ces observations contrastent avec les autres types de cancers où le taux d'aberrations chromosomiques est beaucoup plus faible, la majorité impliquant plutôt l'amplification de grandes régions chromosomiques (revu par Rubnitz et Look, 1998a).

1.3.4.1 Translocations chromosomiques

1.3.4.1.1 Bcr-Abl

Plus de 200 différentes translocations récurrentes différentes ont été identifiées à ce jour depuis la découverte du chromosome de Philadelphie en 1960, incluant la caractérisation de plus de 50 au niveau moléculaire ainsi que la démonstration de leur activité oncogénique dans certains cas (revu par Look, 1997; Figure 1-2). La translocation t(9;22), ou chromosome de Philadelphie, générant la protéine de fusion Bcr-Abl, est présente chez plus de 95 % des cas de leucémie myéloïde chronique et aussi 5-10 % des cas de leucémie lymphoblastique aiguë (revu par Rowley, 1998). Abl est une protéine tyrosine kinase qui est normalement activée par des dommages importants subis à l'ADN et qui amorce une cascade aboutissant à l'apoptose (revu par Wang, 2000). Son inactivation par recombinaison homologue entraîne principalement des défauts dans la maturation des cellules B et dans le processus d'ossification (Lu *et al.*, 2000). La protéine de fusion a une activité tyrosine kinase beaucoup plus importante que Abl seule

LLA



LMA

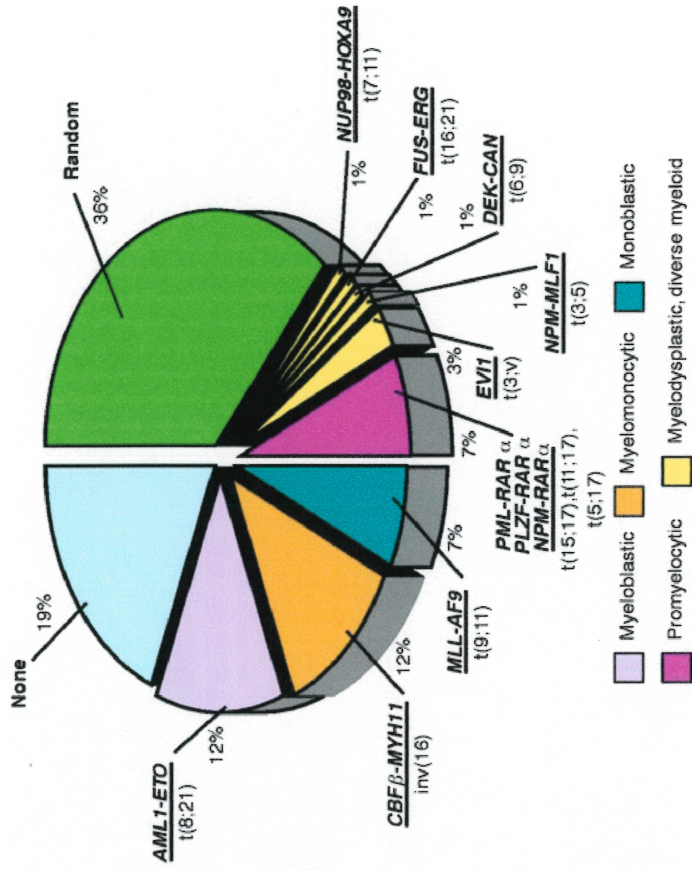


Figure 1-2. Translocations observées dans les cas de LLA et de LMA pédiatriques. Les gènes de fusion produits, sont indiqués s'ils ont été identifiés. Tiré de Look, 1997.

et se retrouve principalement dans le cytoplasme contrairement à Abl qui est essentiellement nucléaire. Son introduction dans des souris induit un phénotype comparable aux leucémies observées chez l'humain (Gishizky *et al.*, 1993; Castellanos *et al.*, 1997). Bcr-Abl est capable d'autophosphorylation et peut activer plusieurs voies de signalisation comme JAK/STAT, Ras/Mapk, Jnk, Myc, CyclinD et Bcl-2, ainsi que les protéines Crkl et Cbl ayant toutes des activités oncogéniques (revu par Maru *et al.*, 2001).

1.3.4.1.2 Récepteurs des cellules T et immunoglobulines

La majorité des translocations observées dans les leucémies lymphoblastiques à cellules T ou B matures cause la juxtaposition de facteurs de transcription, comme Tal1, Tal2, Myc, Lyl1, Lmo1, Lmo2 ou Hox11, aux chaînes α ou β des récepteurs des cellules T ou de la chaîne lourde du gènes des immunoglobulines des cellules B, entraînant leur surexpression. Dans ces cas, on croit que la recombinaison VdJ, responsable du réarrangement des récepteurs/immunoglobulines, est en cause puisqu'elle induit des cassures bicaténares dans l'ADN, facilitant la translocation. Ces types de leucémies sont très similaires cliniquement et génétiquement aux lymphomes impliquant les cellules B ou T. D'ailleurs, les leucémies de type B mature se traitent beaucoup mieux à l'aide des traitements normalement administrés aux patients atteints du lymphome de Burkitt, où 90 % des cas possèdent le réarrangement impliquant *myc* (revu par Rubnitz et Look, 1998b).

Les gènes *Tal1/2*, *Lmo1/2* et *Lyl-1*, surexprimés dans les leucémies de type T, font tous partie d'un complexe inhibiteur de E2A. E2A est un facteur de transcription essentiel dans les premières étapes de la différenciation des précurseurs des cellules B et T. L'inhibition de E2A, par recombinaison homologue chez la souris, entraîne un arrêt dans la différenciation des cellules T. Le mécanisme exact n'est pas connu, mais un des gènes activés par E2A est CD4, un récepteur des cellules T (revu par Hanrahan *et al.*, 2001). La translocation t(1;19) générant la protéine E2A-PBX1 est assez commune dans la leucémie lymphoblastique aiguë (Figure 1-2). La juxtaposition de PBX1, un facteur de transcription à homéodomaine, qui n'est normalement pas exprimé dans ce type de cellules, entraîne la transformation des cellules NIH-3T3 *in vitro* ainsi que l'induction de lymphomes chez la souris (Kamps *et al.*, 1993; Dederer *et al.*, 1993). Les protéines à homéodomaine, incluant les protéines HOX, ont un rôle extrêmement important dans la différenciation cellulaire et dans le développement de tous les systèmes du corps humain. Il a été démontré que la dérégulation de la voie de signalisation des protéines HOX entraîne des leucémies chez des souris transgéniques (Perkins *et al.*, 1990).

1.3.4.1.3 MLL

Un autre gène associé étroitement aux gènes HOX est *MLL*, qui est un homologue du gène *trithorax* (*trx*) de la drosophile et possède également un homéodomaine. MLL contrôle l'expression de plusieurs gènes HOX dans les précurseurs hématopoïétiques (Broeker *et al.*, 1996; Hanson *et al.*, 1999). C'est un gène de plus de 100 kb sur le chromosome 11q23 et qui est impliqué dans plus de 40 translocations différentes observées dans 5-10 % des cas de leucémies (Figure 1-2). Cependant, ces

translocations sont retrouvées chez plus de 80 % des patients leucémiques ayant moins d'un an, indiquant une évolution extrêmement très rapide de ces cancers (revu par Rowley, 1999). Les partenaires de fusion identifiés jusqu'à présent sont de fonctions très variées et une des hypothèses avancées est que c'est l'inactivation de *MLL* par ces translocations qui est l'événement important. Par contre, l'expression chez la souris de *MLL-AF9*, un gène créé par la translocation t(9;11) entraîne le développement d'une forme de LMA, qui n'est pas observée lorsque seule la forme tronquée de *MLL* est plutôt utilisée (Corral *et al.*, 1996). Par contre, le peu de connaissances sur les fonctions de *MLL* et de ses partenaires de fusion ne permet pas de tirer de conclusions.

1.3.4.1.4 RAR α

Les translocations t(15;17) et t(11;17) générant les protéines de fusion PML-RAR α et PLZF-RAR α , respectivement, sont spécifiques pour la leucémie promyélocytaire, une variété de leucémie myéloïde. RAR α est un récepteur de l'acide rétinoïque tandis que PML est une protéine impliquée dans l'organisation de la matrice nucléaire et qui montre une activité pro-apoptotique. Son inactivation par recombinaison homologue chez la souris entraîne une augmentation de l'incidence de leucémies, indiquant que c'est un gène suppresseur de tumeurs (Rego *et al.*, 2001). Sans ligand, RAR α inhibe la transcription, notamment de *PML*, par la liaison d'un complexe de répression, composé de N-CoR, mSin3 et de HDAC, une histone déacétylase. Le traitement de ces patients avec l'acide rétinoïque entraîne leur rémission complète en relâchant les complexes de répression (Weis *et al.*, 1994). Les patients ayant la fusion RAR α -PLZF répondent beaucoup moins bien au traitement puisque PLZF recrute

également le complexe de répression (Grignani *et al.*, 1998). Un traitement combiné avec des inhibiteurs d'histone déacétylases, soit le butyrate ou la trichostatine A, permet de traiter adéquatement ce type de leucémie (Warrell *et al.*, 1998).

1.3.4.1.5 AML1

AML1 et son partenaire, CBF β , sont les cibles les plus fréquentes des translocations, étant impliquées dans plus de 30 % dans les leucémies de l'enfant. AML1 est un membre de la famille des facteurs de transcription ayant une homologie au gène *RUNT* de la drosophile et liant sur l'ADN la séquence TGT/cGGT. AML1 doit former un complexe avec la protéine CBF β pour être active. Ce complexe est essentiel pour le développement et la différenciation de plusieurs précurseurs au cours de l'hématopoïèse, via l'activation de la transcription de plusieurs gènes importants, dont *IL-3*, *GM-CSF*, *CSF-1*, l'*élastase* et la *myéloperoxidase* (Nuchprayoon *et al.*, 1994; revu par Lorsbach *et al.*, 2001). Cependant, il faut noter que AML1 peut aussi s'associer directement avec mSin3A et TLE, un homologue humain de Groucho, pour réprimer la transcription dans certains cas (Levanon *et al.*, 1998; Imai *et al.*, 1998). Les translocations t(12;21), t(8;21) et t(21;3) générant les protéines de fusion ETV6-AML1, ETO-AML1 et AML1-EVI1, respectivement, inhiberaient la transactivation par le complexe AML1-CBF β d'une façon similaire aux translocations impliquant PML. En effet, ETV6 et ETO sont tous deux des répresseurs de la transcription recrutant le complexe de répression N-CoR, mSin3 et HDAC (Lutterbach *et al.*, 1998; Chakrabarti et Nucifora, 1999), tandis que EVI1 recrute le complexe de répression CtBP (Palmer *et al.*, 2001). Le cas de ETV6-AML1 sera discuté plus en détails à la Section 1.5.1.3.1. La protéine de fusion AML1-

ETO bloque la différenciation de cellules granulocytaires en culture, mais induit également leur mort par apoptose, indiquant que d'autres altérations cruciales sont nécessaires à l'induction de la leucémie (Burel *et al.*, 2001). Il a été montré également que ETO et EVI1 induisaient l'activation, probablement indirectement, d'oncogènes tels que Jun. Dans le cas de l'inversion inv(16)(p13;q22), retrouvée dans 12 % des cas de LMA, *CBF β* est fusionné à *MYH11*, un gène codant pour la chaîne lourde de la myosine. Ceci séquestrerait alors AML1 dans un complexe inactif localisé dans le cytoplasme (Kanno *et al.*, 1998). Aussi, la protéine de fusion E2A-HLF retrouvée chez environ 1 % des patients atteints de LLA, active par un mécanisme encore inconnu certains homologues de Groucho inhibant ainsi l'expression de *AML1*, indiquant que *AML1* est un gène crucial pour le bon fonctionnement de l'hématopoïèse (Dang *et al.*, 2001).

Chaque type de leucémie correspond à un précurseur d'une lignée hématopoïétique particulière, dont la différenciation s'est arrêtée (voir la Figure 1-3). D'ailleurs, on croit que l'arrêt de la différenciation est une étape cruciale pour le développement des leucémies. Par exemple, l'activation du récepteur CD44 de cellules leucémiques de type myéloïde, une molécule d'adhésion, par son ligand, l'acide hyaluronique, ou par un anticorps spécifique active ces cellules vers un état de différenciation terminale, inhibant ainsi toute prolifération et favorisant leur apoptose (Charrad *et al.*, 1999). On peut constater que les gènes impliqués dans les translocations jouent des rôles essentiels dans le développement et la différenciation de ces cellules. La

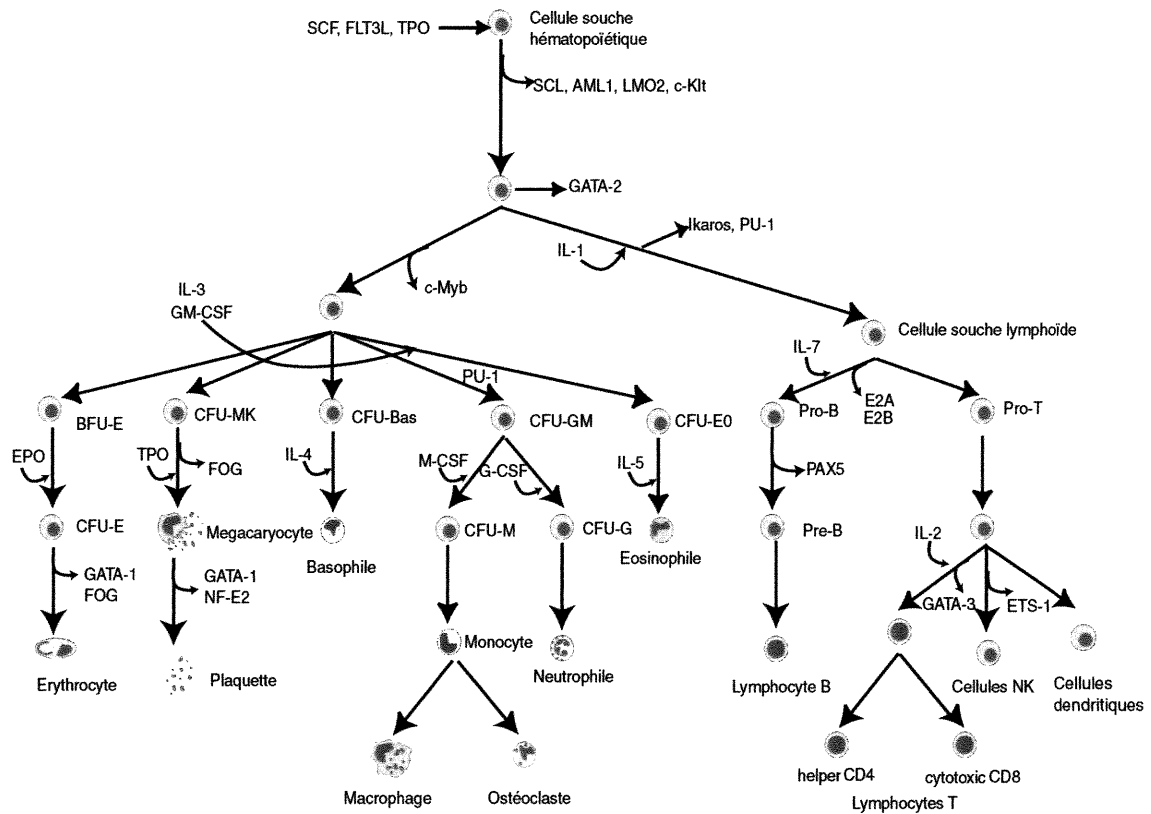


Figure 1-3. Schéma de la hiérarchie des précurseurs hématopoïétiques. Les différents progéniteurs des lignées myéloïdes et lymphoïdes sont indiqués. Les facteurs de croissance importants pour le développement des lignées sont indiqués à gauche des voies de différenciation. Les principaux facteurs de transcription impliqués dans la différenciation des lignées sont indiqués du côté droit. Modifié de Cortes et Labastie (2000).

plupart des protéines ciblées par ces translocations sont des facteurs de transcription ou des protéines kinases qui peuvent influencer l'activation, via la transcription ou la phosphorylation, de plusieurs voies de signalisation cellulaire simultanément. Ces translocations récurrentes sont retrouvées chez environ 45 % des patients atteints de LLA suggérant cependant la présence d'autres altérations génétiques.

1.3.4.2 Délétions

Les aberrations chromosomiques chez la leucémie, telles les translocations menant à l'activation d'oncogènes, ont été abondamment étudiées depuis les 40 dernières années. Cependant, les altérations plus subtiles, comme les délétions ou les mutations ponctuelles pouvant indiquer l'inactivation de gènes suppresseurs de tumeurs ont été très peu caractérisées. Les suppresseurs de tumeurs classiques, tels que *Rb-1* et *p53* sont rarement mutés dans les leucémies (Chen *et al.*, 1990; Wada *et al.*, 1993). Cependant, *p53* est muté dans près de 25 % des cas de rechute d'une LLA de type T, indiquant que son inactivation contribue au développement de leucémies résistantes (Hsiao *et al.*, 1994). Par le fait même, très peu de gènes suppresseurs de tumeurs impliqués dans le développement des leucémies ont été identifiés. Il existe quelques rares exceptions dans la littérature: $p16^{\text{INK4A}}/CDKN2$ est un régulateur négatif de la progression du cycle cellulaire qui a été initialement identifié comme suppresseur de tumeurs impliqué dans les mélanomes héréditaires (Serrano *et al.*, 1993; Hussussian *et al.*, 1994). $p16^{\text{INKA}}$ est un membre de la famille des inhibiteurs des complexes cyclines-CDK (kinases cyclines-dépendantes) qui empêche, en inactivant CDK4, l'entrée de la cellule en phase S du

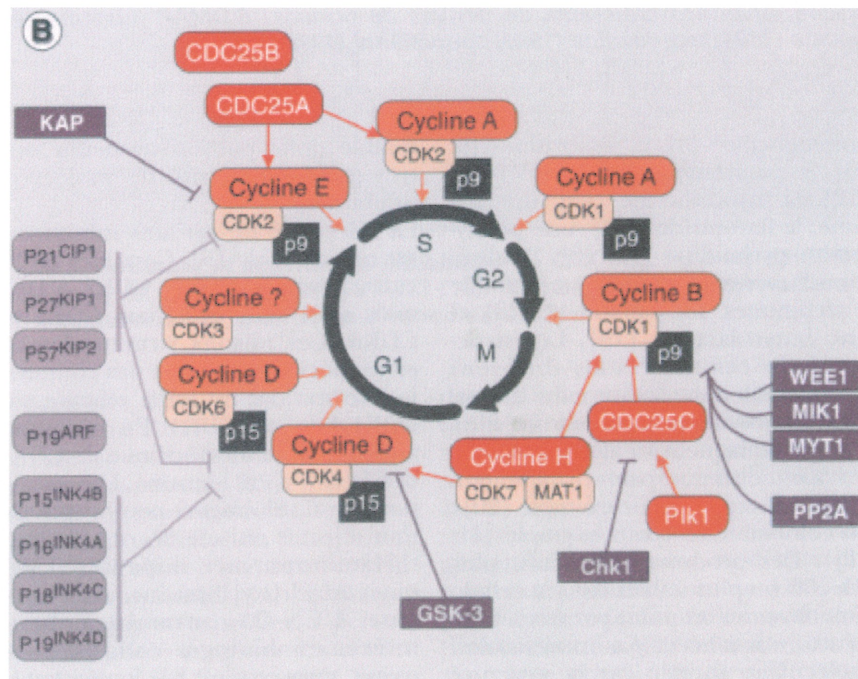


Figure 1-4. Contrôle du cycle cellulaire par les CDKI. Les complexes CDK/cyclines (orangé) contrôlent les transitions d'une phase à l'autre du cycle cellulaire en phosphorylant entre autres pRB durant la transition G1 à S. Ces complexes sont régulés positivement par des kinases ou des phosphatases (rouge) ou négativement par la liaison des CDKI (violet clair) ou l'action catalytique d'autres protéines (violet foncé). Tiré de Borgne et Meijer, 1999.

cycle cellulaire avant que certains événements, tels que la réparation de l'ADN, ne soient complétés (Figure 1-4). Ce gène, ainsi que $p15^{INK4B}$ situé tout près de $p16^{INKA}$ sur le chromosome 9, sont inactivés, principalement par délétion homozygote, dans près de 90 % des cas de LLA de type T (Hebert *et al.*, 1994). $p16^{INKA}$ est aussi inactivé, principalement par la méthylation de son promoteur, dans une importante proportion d'autres cancers, incluant le cancer des poumons et le carcinome du foie (Kamb *et al.*, 1994; Merlo *et al.*, 1995). D'autres exemples dans la littérature de gènes suppresseurs de tumeurs identifiés dans les leucémies incluent $p73$, un homologue de $p53$ (Corn *et al.*, 1999) ainsi que $FHIT$, un gène codant pour une protéine de fonction inconnue et dont l'inactivation a été observée dans de nombreux types de cancers différents (Iwai *et al.*, 1998).

1.3.4.2.1 Analyse LOH

L'analyse cytogénétique, par FISH par exemple, est très efficace pour observer de grandes altérations chromosomiques. Les observations de délétions cytogénétiques fréquentes sur le chromosome 13q, spécifiques aux cas de rétinoblastome, et la présence de mutations dans le second allèle du gène $Rb-1$ ont suggéré que ces délétions ciblaient des gènes suppresseurs de tumeurs (revu dans Gallie et Phillips, 1984). Par contre, elle ne permet pas de les localiser précisément sur un chromosome ni d'observer de très petites délétions (<2 Mb). Les analyses moléculaires sont beaucoup plus efficaces pour détecter les altérations plus subtiles. L'analyse de pertes d'hétérozygotie (loss-of-heterozygosity; LOH) est actuellement utilisée afin de détecter la présence d'un locus suppresseur de tumeurs dans des cellules cancéreuses. Cette analyse est basée sur la

présence de marqueurs polymorphiques distribués aléatoirement dans le génome humain. Ces marqueurs peuvent être de simples variations de nucléotides (SNP) ou correspondre à des microsatellites (séquences répétées de un à six nucléotides) de longueur variable dans la population. Un LOH correspond donc, expérimentalement, à la perte d'un des deux allèles polymorphiques provenant des parents et est une altération très commune dans les cellules cancéreuses. Quelques régions sont plus affectées que d'autres ou restreintes à certains types de cancers, indiquant que ces délétions ne sont pas aléatoires.

L'analyse de ces microsatellites polymorphiques, par la technique de SSCP (single-stranded conformational polymorphism) par exemple, permet d'observer de très petites délétions et ainsi cibler précisément la recherche d'un gène suppresseur de tumeur. Quatre mécanismes ont été proposés pour expliquer ce phénomène: 1- Perte d'un chromosome; 2- perte d'un chromosome suivi de la duplication du chromosome homologue; 3- délétion d'une portion d'un chromosome; 4- recombinaison mitotique et 5- conversion génique (Gupta *et al.*, 1997; Wiese *et al.*, 2001). Les causes exactes de ces pertes de matériel génétique sont inconnues. On croit cependant que l'introduction d'une cassure double brin, tout comme dans le cas des translocations est cruciale, en activant le système réparation-recombinaison. Celui-ci doit alors trouver une matrice compatible pour réparer l'ADN, ce qui peut entraîner un événement de conversion génique, une recombinaison illégitime entre deux régions hétérologues ou une perte complète d'une région d'ADN. La production de translocations balancées, soit une translocation complète entre deux chromosomes, ou non-balancées, soit une

translocation partielle entre deux chromosomes suivie d'une perte de matériel, nécessite probablement l'introduction de deux cassures double brin (Moynahan et Jasin, 1997; Wiese *et al.*, 2001). Wiese *et al.* (2001) ont transfecté une forme mutée de *p53* dans des cellules WTK1 et ont observé une très haute fréquence de recombinaison mitotique. *p53* contrôle entre autres le processus de recombinaison en interagissant avec RAD51 et RecA, responsables de l'appariement et du transfert des brins. *p53* empêcherait notamment la recombinaison entre des séquences répétées non-homologues (Gebow *et al.*, 2000).

1.4 Identification de délétions associées à la leucémie

Donc, l'analyse de la présence de LOH à l'aide de marqueurs polymorphiques distribués sur tout le génome est une approche pouvant mener à l'identification de nouveaux gènes suppresseurs de tumeurs. La première phase du projet de séquençage du génome humain consistait au développement de tels marqueurs et a donc considérablement contribué en ce sens à la recherche de gènes suppresseurs de tumeurs (Hudson *et al.*, 1995). Dans la leucémie de l'enfant, étant donné que la plupart des efforts ont été investis dans l'identification des protéines oncogéniques créées par les translocations chromosomiques, très peu de gènes suppresseurs de tumeurs ont été identifiés. Depuis 1994, notre laboratoire s'intéresse aux gènes suppresseurs de tumeurs impliqués dans la leucémie lymphoblastique aiguë. Dans une première étude, 49 marqueurs polymorphiques, distribués sur 13 bras chromosomiques et choisis initialement pour leur délétion récurrente dans la leucémie ou d'autres cancers, ont été analysés chez des patients atteints de leucémie lymphoblastique aiguë. Environ 60 % des

patients étudiés avaient perdu au moins un des marqueurs, indiquant que ce type d'altération et probablement l'inactivation de gènes suppresseurs de tumeurs est un événement très fréquent dans la leucémie. Quelques régions ont montré un taux élevé de LOH, soit sur le chromosome 9p (80 %) chez les patients atteints de LLA-T ainsi que sur 6p (8 %), 6q (13 %), 9p (16 %) et 12p (35 %) chez les patients atteints de LLA de type pré-B (Bacchichet *et al.*, 1997a) (Figure 1-5). De telles fréquences indiquent qu'il y a un avantage sélectif pour les cellules cancéreuses de perdre ces allèles et ne sont pas seulement dues qu'au hasard. La délétion sur le bras 9p est probablement reliée au locus *p16/p15* et correspond aux pourcentages observés dans ces types de leucémies (Section 1.3.4.2). La délétion du chromosome 6q est fréquemment observée dans les cas de lymphomes ou de leucémie aiguë. Récemment, une région de 3cM près de 6q21 a été définie comme étant la région délétée commune à tous les patients (Zhang *et al.*, 2000). Par contre, l'implication du chromosome 6p dans la leucémie n'a pas été étudiée davantage. Par cytogénétique, on observe des délétions hémizygotiques du bras court du chromosome 12 chez environ 5 % des patients atteints de leucémie (Raimondi *et al.*, 1986). Cependant, cette technique ne détecte pas les petites délétions qui ne peuvent être observées que par des approches moléculaires. La délétion du chromosome 12p, observée chez 34 % des patients dans notre laboratoire a été confirmée par plusieurs autres études, variant de 26 à 40 % et est donc l'altération la plus commune chez les patients atteints de LLA de type pré-B (Takeuchi *et al.*, 1996b; Cavé *et al.*, 1997; Chambon-Pautas *et al.*, 1998) (Tableau I-IV). Dans d'autres études, les chromosomes 5q31, 7p13, 7q31 et 20q12 ont été identifiés comme étant souvent délétés chez les

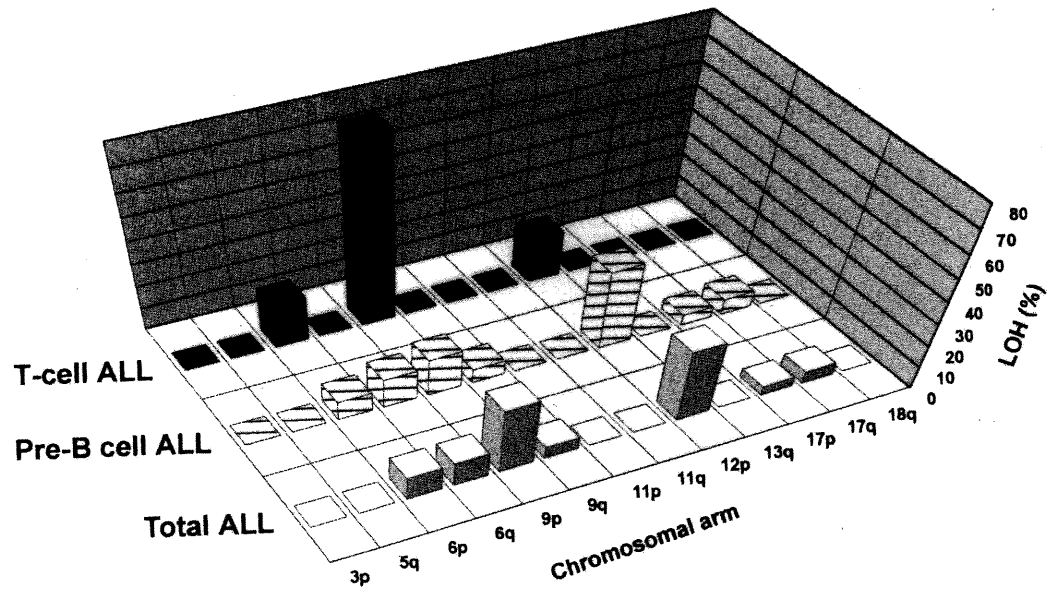


Figure 1-5. Fréquence de pertes alléliques sur différents chromosomes dans les leucémies lymphoblastiques aiguës de type B et T. Tiré de Baccichet *et al.*, 1997a.

Tableau I-IV. Fréquence de la délétion 12p12 dans différents cancers

| Cancer | # patients | Fréquence LOH | Référence |
|-----------|------------|---------------|-----------------------------|
| LLA | 81 | 15% | Stegmaier et al., 1995 |
| LLA pré-B | 183 | 26% | Cavé et al., 1996 |
| | 80 | 40% | Takeuchi et al., 1996 |
| | 24 | 34% | Baccichet et al., 1997a |
| | 63 | 31% | Chambon-Pautas et al., 1998 |
| Ovaires | 23 | 26% | Hatta et al., 1997 |
| Sein | 7 | 71% | Spirin et al., 1996 |
| Poumon | 36 | 33% | Takeuchi et al., 1996 |
| Endomètre | 12 | 10% | Aissani et al., 1999 |
| Cervix | 19 | 11% | Aissani et al., 1999 |
| Prostate | 19 | 47% | Kibel et al., 1998 |

patients atteints de LMA (Lezon-Geyda *et al.*, 2001; Koike *et al.*, 1999; MacGrogan *et al.*, 2001).

1.4.1 Chromosome 12p12

La très haute fréquence de délétions sur le bras court du chromosome 12 a amené notre laboratoire à étudier plus en détails la région affectée. Grâce à l'utilisation d'une plus grande densité de marqueurs polymorphiques, la région minimale de délétion, commune à tous les patients étudiés, soit la région critique, a été délimitée par les marqueurs *D12S89* et *D12S358*, soit une distance génétique d'environ 3 cM (Bacchichet *et al.*, 1997b). Les conclusions de plusieurs autres études convergent également vers cette région (Figure 1-6). Dans une première étude, la région critique a été située dans un intervalle génétique de 15 cM par la méthode de FISH (Kobayashi *et al.*, 1994). En utilisant l'analyse de marqueurs polymorphiques, Stegmaier *et al.* (1995) ont identifié une région critique d'environ 11 cM entre les marqueurs *D12S77* et *D12S70*. Finalement, Takeuchi *et al.* (1996), en utilisant 33 patients leucémiques possédant une délétion sur le chromosome 12, soit la plus importante étude de ce type jusqu'à présent, ont identifié cette fois deux régions critiques d'environ 4 cM, situées à l'intérieur des régions décrites précédemment. Les conclusions de leur étude sont ambiguës puisqu'il s'avère que l'ordre réel des marqueurs polymorphiques sur le chromosome 12 qu'ils ont utilisés est différent de celui qu'ils ont utilisé pour analyser leurs résultats. Après avoir refait cette analyse, nous pouvons en arriver à ces conclusions: une des régions se situe entre les marqueurs *D12S77* et *D12S98* et une autre entre les marqueurs *D12S358* et *D12S320*, suggérant la possibilité que plus d'un gène suppresseur de tumeurs soit impliqué (Figure 1-6). De

plus, la délétion du bras court du chromosome 12 est observée dans plusieurs autres cancers hématologiques, comme la LMA ou d'autres désordres myéloprolifératifs (Mitelman, 1994), ainsi que dans plusieurs types de tumeurs solides, comme le cancer de l'ovaire (Hatta *et al.*, 1997), de la prostate (Kibel *et al.*, 1998), du poumon (Takeuchi *et al.*, 1996a), du sein (Spirin *et al.*, 1996), du cervix et de l'endomètre (Aïssani *et al.*, 1999) (Tableau I-IV). Les régions de délétions ont été caractérisées avec précision, pour le cancer des ovaires, entre les marqueurs *D12S89* et *D12S364* (Hatta *et al.*, 1997), du poumon, entre les marqueurs *D12S358* et *D12S320* (Takeuchi *et al.*, 1996a) et de la prostate, entre les marqueurs *D12S391* et *D12S358* (Kibel *et al.*, 2000) (Figure 1-6).

Si l'on combine les résultats de toutes ces études, on observe qu'il y a trois régions de délétion différentes, toutes situées dans un intervalle d'environ 5 cM: une entre les marqueurs *D12S77* et *D12S98*, une entre *D12S89* et *D12S358* et une entre *D12S358* et *D12S320* (Figure 1-6). Cependant, il faut noter que dans ces études la majorité des patients ont de grandes délétions et qu'il n'y a jamais plus d'un ou deux patients ayant une délétion définissant une des deux extrémités de la région critique. Comme ces événements sont relativement fréquents dans les cellules cancéreuses, il est alors possible d'envisager que certaines de ces délétions soient dues au hasard plutôt qu'à l'inactivation d'un gène suppresseur de tumeurs, élargissant du même coup la région critique ou en créant une seconde. Alternativement, il est possible que dans chaque cancer ou même sous-type de cancer, un gène différent est la cible des délétions ou bien que plus d'un gène suppresseur de tumeurs soit impliqué.

Régions minimales de délétion

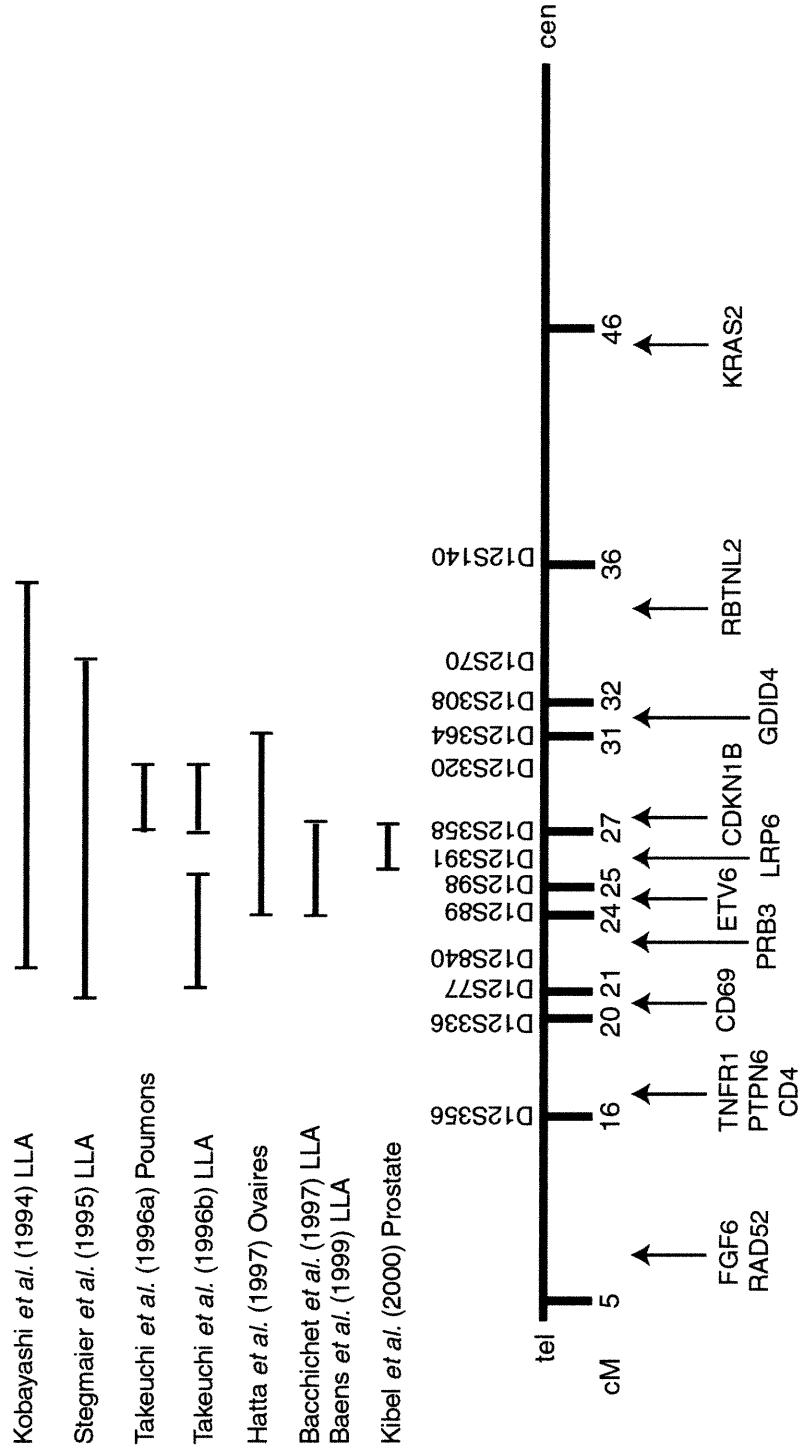


Figure 1-6 . Régions minimales de délétions observées dans différentes études sur le chromosome 12p. Les cancers étudiés sont indiqués à côté des auteurs de chacune des études. Quelques gènes et marqueurs polymorphiques connus sont indiqués sur la carte. Les distances sont indiquées en cM à partir du télomère (tel).

1.5 Caractérisation des gènes dans la région 12p12.3

Récemment des cartes physiques du locus suppresseur de tumeurs sur le chromosome 12p ont été construites, à l'aide de clones génomiques de cosmides ou de YAC, permettant de positionner les gènes connus avec plus de précision et de pouvoir évaluer les distances entre ceux-ci (Aïssani *et al.*, 1999; Baens *et al.*, 1999). Parmi les gènes connus, *ETV6* et *CDKN1B/p27^{Kip1}* ont attiré l'attention des chercheurs en raison de leur fonction.

1.5.1 ETV6

1.5.1.1 Famille ETS

Le gène *ETV6* (ETS variant 6) ou *TEL* (translocation ETS leukemia) a été identifié initialement comme un des partenaires dans la fusion *ETV6/TEL-PDGFR β* découlant de la translocation t(5;12)(q33;p13) observée dans plusieurs cas de leucémie myélomonocytaire chronique (Golub *et al.*, 1994). *ETV6* est un gène possédant huit exons distribués sur près de 250 kb et est exprimé dans la plupart des tissus (Golub *et al.*, 1994; Baens *et al.*, 1996). Il code pour une protéine ayant deux domaines très conservés: un domaine ETS de liaison à l'ADN en carboxy-terminal ainsi qu'un domaine HLH (helix-loop-helix) ou PNT (par homologie avec *pointed*, un gène de la drosophile) de dimérisation en amino-terminal présent dans un grand nombre des protéines de la famille ETS (Figure 1-7). Les protéines ETS sont des facteurs de transcription qui forment une famille composée de plus d'une trentaine de membres. Un de ceux-ci, Ets-1, a été identifié comme étant un des deux oncogènes, avec Myb, à être transduits par le virus de la leucémie du poulet E26 (Leprince *et al.*, 1983; Nunn *et al.*, 1983). Les

protéines ETS sont impliquées dans divers processus du développement. Elles contrôlent l'expression de gènes participant notamment à la prolifération, la différenciation, l'apoptose, la migration ainsi que les interactions cellulaires dans plusieurs systèmes ou organes. De plus, l'expression de la plupart des gènes ETS est contrôlée par d'autres gènes ETS et plusieurs interactions protéine-protéine entre des membres de la famille ont été documentées (revu par Sementchenko et Watson, 2000). L'expression et l'importance de ces gènes au niveau de l'hématopoïèse a été relativement bien connue. L'expression des gènes *Ets-1*, *Ets-2*, *Pu-1*, *Fli-1*, *Erg*, *Elf1*, *SpiB* ainsi que de *ETV6* a été très bien caractérisée dans chaque lignée hématopoïétique (revu dans Maroulakou et Bowe, 2000). Par exemple, *Ets-1* est exprimé seulement dans les lignées lymphocytaires. Son absence induit l'apoptose des lymphocytes T et bloque le développement des cellules NK, mais favorise la différenciation terminale des lymphocytes B (Bories *et al.*, 1995; Barton *et al.*, 1998). De plus, les gènes *Pu-1* et *Fli-1* sont surexprimés dans des cas d'érythroleucémies, les gènes *Erg*, *Fli-1*, *ETV1* et *PEA3* sont impliqués dans des translocations dans des cas de sarcome d'Ewing et de LMA et *ETV6* est transloqué dans plus de 25 % des cas de LLA (voir Section 1.5.1.3), indiquant que la plupart des protéines ETS ont un potentiel oncogénique (revu dans Laudet *et al.*, 1999).

1.5.1.2 Fonction de ETV6

L'inactivation de *ETV6* par la technique de knock-out est létale chez la souris homozygote à un stade embryonnaire précoce due à l'absence d'angiogenèse dans le sac vitellin (Wang *et al.*, 1997). Des études subséquentes ont démontré que *ETV6* était régulé par le VEGF via la voie MAPK/ERK2 lors de l'angiogenèse (Edel, 1998).

L'observation des souris chimériques a toutefois permis de conclure que les cellules homozygotes pour l'absence de *ETV6* participaient à l'hématopoïèse primitive dans le sac vitellin, ainsi que la rate et le foie fœtal, mais pas à l'hématopoïèse dans la moelle osseuse adulte, indiquant l'incapacité de ces cellules à migrer ou à répondre aux stimuli spécifiques de la moelle osseuse (Wang *et al.*, 1998). Dans les cellules normales, ETV6 est localisée presque exclusivement dans le noyau et lie la séquence consensus GGAA/T caractéristique des protéines de la famille ETS (Poirel *et al.*, 1997). Chakrabati *et al.* (2000) ont récemment montré que ETV6 était modifiée par SUMO-1, via une interaction avec l'ubiquitine UBC9 sur le domaine PNT, à un moment précis de la phase S du cycle cellulaire. Cette modification entraîne la formation de corps nucléaires, nommés TEL, actifs dans certaines fonctions spécifiques du noyau comme la transcription, l'apoptose ou la modification de la chromatine, ainsi qu'une protection contre la dégradation.

Récemment, il a été montré que ETV6 est un répresseur de la transcription (Chakrabarti *et al.*, 1999). Cette propriété a été observée chez seulement quatre autres protéines ETS, soit ERF, NET, TEL-2, un homologue de ETV6 chez l'humain, ainsi que YAN, l'orthologue de ETV6 chez la drosophile (revu par Mavrothalassitis et Ghysdael, 2000). ETV6 peut lier les protéines SMRT, mSin3A et N-CoR, une histone déacétylase, qui forment un complexe de répression de la transcription (Chakrabarti et Nucifora, 1999; Guidez *et al.*, 2000). La liaison requiert la section centrale de ETV6, plus particulièrement entre les acides aminés 171 à 285 (Lopez *et al.*, 1999). ETV6 peut aussi lier Fli-1 via son domaine PNT et inhiber son activité transcriptionnelle (Kwiatkowski *et al.*, 1998). La surexpression de ETV6 dans des fibroblastes NIH3T3 en culture induit

une diminution de leur croissance ainsi que plusieurs changements morphologiques (Van Rompaey *et al.*, 1999). De plus, sa surexpression dans des cellules transformées par Ras inhibe leur croissance dans l'agar mou ou la formation de tumeurs dans des souris immunodéficientes, indiquant que ETV6 pourrait agir comme un suppresseur de tumeurs (Fenrick *et al.*, 2000; Van Rompaey *et al.*, 2000). À l'aide d'une analyse d'expression différentielle, il a été montré que ETV6 entraîne dans des cellules en culture une augmentation de l'expression de l'entactine, la fibronectine, Smad5 Col3a1 et CD44 ainsi que la diminution de l'expression de Colla1 et de l'inhibiteur de protéases leucocytaire. La plupart de ces gènes sont essentiels pour l'angiogenèse (Van Rompaey *et al.*, 2000). Dans une autre étude, Fenrick *et al.* (2000) ont montré que ETV6 lie directement le promoteur et réprime l'expression de la *stromelysine-1*, codant pour une métalloprotéinase de la matrice, qui est actuellement le seul gène connu qui est ciblé directement par ETV6. Les métalloprotéinases de cette famille dégradent la matrice cellulaire permettant la migration cellulaire et leur activation est un événement important dans la formation de métastases (revu par John et Tuszynski, 2001).

1.5.1.3 Translocations impliquant ETV6

Malgré ces caractéristiques le catégorisant comme un gène suppresseur de tumeurs, ETV6 est depuis longtemps considéré comme un oncogène puisqu'il est impliqué dans des translocations dans 25 % des cas de LLA ainsi que dans plusieurs autres cancers hématologiques (Figure 1-2). Depuis son identification comme partenaire de fusion avec PDGFR β dans les cas de CMML (Golub *et al.*, 1994), une trentaine d'autres translocations l'impliquant ont été identifiées et 14 de ses partenaires

caractérisés au niveau moléculaire (Figure 1-7). Cinq de ces gènes codent pour des récepteurs tyrosines kinase: *PDGFR β* , *Abl*, *Jak2*, *ARG* et *NTRK3/TRKC*. Dans la plupart des cas, il a été démontré que la protéine de fusion a une activité tyrosine kinase constitutive et cause la transformation de cellules hématopoiétiques BaF3 (ETV6-PDGFR β : Carroll *et al.*, 1996; ETV6-ABL: Hannemann *et al.*, 1998; ETV6-JAK2: Lacronique *et al.*, 1997; ETV6-NTRK3: Liu *et al.*, 2000) ou même l'apparition de leucémies lorsque injectées chez la souris (ETV6-PDGFR β : Tomasson *et al.*, 1999; ETV6-JAK2: Schwaller *et al.*, 1998; ETV6-NTRK3: Liu *et al.*, 2000). Ces expériences indiquent que ces protéines de fusion possèdent un fort potentiel oncogénique activant probablement plusieurs voies de signalisation cellulaires de façon similaire à Bcr-Abl. D'ailleurs une étude a montré que les protéines Bcr-Abl et ETV6-Abl activaient les mêmes voies de signalisation, soit les protéines rasGAP, Shc, SH-PTP2, SH-PTP1, CRKL, CBL, paxillin et les protéines STAT (Okuda *et al.*, 1996). La voie des STAT est également activée par ETV6-PDGFR β et ETV6-JAK2, la phosphorylation de STAT5 étant essentiel à l'activité oncogénique de cette dernière (Wilbanks *et al.*, 2000; Lacronique *et al.*, 2000; Schwaller *et al.*, 2000). Le domaine PNT (ou HLH) d'oligomérisation, encodé par les exons 3 et 4, est essentiel à l'activité oncogénique de ces protéines de fusion (Figure 1-7). On croit que la dimérisation de la protéine entraîne son autophosphorylation et ainsi son activation constitutive. Ce mécanisme est similaire à la transformation par la protéine de fusion Bcr-Abl dans les cas de leucémie chronique où Bcr contribue un domaine d'oligomérisation à la kinase Abl (Section 1.3.4.1.1).

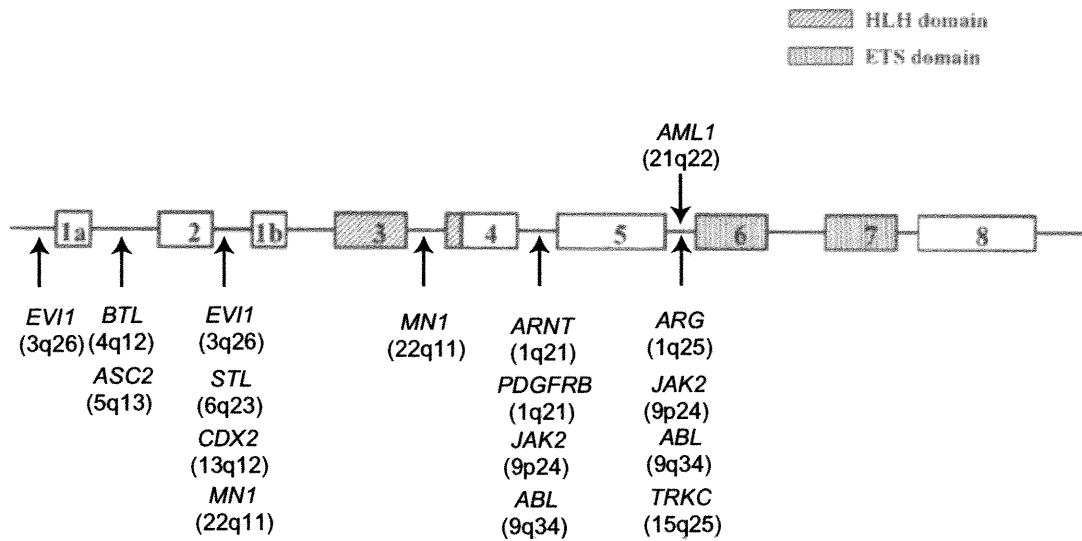


Figure 1-7. Structure du gène ETV6. Les boîtes représentent les exons. Les domaines protéiques encodés HLH et ETS sont représentés. Le domaine de répression de la transcription est encodé par l'exon 5. La localisation des points de translocations ainsi que les partenaires de fusion clonés sont indiqués. Tiré de Odero *et al.*, 2001.

ETV6 est aussi fusionné à des facteurs de transcription, ARNT, AML1/CBFA2, CDX2, MN1, PAX5, HLXB9, MDS1/EVI1, à ACS2, une synthétase d'Acyl-CoA, ou à des protéines de fonction inconnue, BTL et STL (revu dans Odero *et al.*, 2001; Beverloo *et al.*, 2001). Contrairement aux translocations avec les kinases où le transcrit exprimé utilisait le promoteur de *ETV6*, celui-ci contribue plutôt à la partie carboxy-terminale de la protéine dans les fusions impliquant MN1, PAX5, HLXB9, ACS2, BTL et STL. Dans ces cas, tous les domaines de ETV6 sont présents, mais seul le domaine ETS de liaison à l'ADN est essentiel à la transformation cellulaire par MN1-ETV6 (Bujis *et al.*, 2000). Dans les translocations impliquant EVI1, ACS2 et CDX2, ETV6 ne contribue aucun domaine connu et l'on croit que c'est l'expression aberrante du partenaire de ETV6 qui pourrait contribuer à l'oncogenèse.

1.5.1.3.1 Translocation t(12;21): ETV6-AML1

La translocation t(12;21) est observée dans 20 à 30 % des cas de leucémie lymphoblastique aiguë de l'enfant et est donc la plus fréquente (Figure 1-2). Le gène de fusion *ETV6-AML1* dérivant de cette translocation a été cloné indépendamment par deux équipes (Golub *et al.*, 1995; Romana *et al.*, 1995). Dans tous les cas, le point de cassure sur le chromosome 12 se situe dans l'intron 5 du gène *ETV6*, quoique localisés à divers endroits, tandis que sur le chromosome 21, les bris surviennent le plus souvent dans l'intron 1 d'*AML1* (Wiemels et Greaves, 1999). Le transcrit et la protéine ETV6-AML1 sont détectés chez tous les patients possédant la translocation. Par contre, le transcrit réciproque *AML1-ETV6* n'est détecté que chez un tiers de ces patients et ne génère qu'une très petite protéine (Poirel *et al.*, 1998). La protéine résultante ETV6-AML1

possède ainsi les domaines de dimérisation et de répression de ETV6 ainsi que le domaine RUNT de AML1, permettant la liaison à l'ADN, l'interaction avec CBF β et avec les répresseurs mSin3A et TLE.

La présence de la translocation t(12;21) a été associée à un excellent pronostic et à un faible taux de rechute et est un des seuls éléments, avec l'hyperdiploïdisme, qui permettent d'identifier un tel pronostic clinique chez les patients atteints de LLA (Shurtleff *et al.*, 1995; McLean *et al.*, 1996; Maloney *et al.*, 1999). Cette fréquence élevée de réarrangements chromosomiques impliquant *ETV6* laisse envisager que ce locus pourrait être fragile. L'analyse des points de cassure dans les cas de translocation t(12;21) montre que ceux-ci sont distribués sur tout l'intron 5 (14 kb) de *ETV6*, mais de façon non-aléatoire. Il y a entre autres un regroupement important dans une petite région de 384 paires de base, contenant entre autres une séquence répétée Alu (Wiemels et Greaves, 1999). Les hypothèses récentes suggèrent que les séquences répétées pourraient favoriser l'évolution du génome via des événements de recombinaison non-homologues. Une autre étude a révélé que plusieurs des points de cassure se situaient près de séquences alternantes purine/pyrimidine (Thandla *et al.*, 1999). Ces séquences sont reconnues pour favoriser la conformation Z de l'ADN et induire la présence de sites hypersensibles à la DNase I, ce qui pourrait faciliter l'attaque de l'ADN par des agents physiques ou chimiques (Boehm *et al.*, 1989). Ces études n'ont pu démontrer si la recombinaison VdJ des lymphocytes ou que la topoisomérase étaient impliquées dans la génération de ces translocations. La recombinaison VdJ génère la grande variété des récepteurs des cellules T et des anticorps des cellules B par des événements de

recombinaison somatique. La présence de séquences similaires à la séquence consensus de la VdJ peuvent provoquer des réarrangements et ainsi juxtaposer des oncogènes à proximité du promoteur des TCR ou des Ig, ce qui est observé dans plusieurs cas de leucémies (Aplan *et al.*, 1990). Des séquences consensus de la topoisomérase sont fréquemment observées près des points de cassure dans les cas de translocation impliquant *MLL*. La topoisomérase est essentielle pour la réplication de l'ADN et induit des cassures double brin dans son mécanisme d'action. Les inhibiteurs de topoisomérase II, comme les etoposides et l'anthracycline, qui sont utilisés fréquemment en chimiothérapie, sont reconnus pour induire la formation de translocations et ainsi favoriser l'apparition de formes plus agressives de cancer (revu par Felix *et al.*, 2001; Andersen *et al.*, 2001).

Jusqu'à présent, les analyses de la translocation t(12;21) ont plutôt mis en évidence des régions de micro-homologies (quelques bases de chaque côté de la translocation) ainsi que des petits gains ou pertes de matériel, qui sont caractéristiques de la réparation des bris double brin par la voie NHEJ (non-homologous end joining) (Wiemels et Greaves, 1999). D'ailleurs des cellules B immatures en culture exposés à des agents apoptogéniques, tels la privation de serum, l'acide salicylique ou l'étoposide génèrent des bris double-brin dans l'intron 5 de *ETV6* et l'intron 1 de *AML1* ainsi que la présence du gène de fusion *ETV6-AML1* (Eguchi-Ishimae *et al.*, 2001). De plus, ces transcrits sont détectés dans des échantillons sanguins chez plus de 5 % de la population normale indiquant que cet événement est extrêmement commun et que d'autres événements sont essentiels pour le développement d'une leucémie (Eguchi-Ishimae *et*

al., 2001). D'ailleurs, contrairement à la majorité des gènes de fusion impliquant *ETV6* ou *AML1*, la transfection de ETV6-AML1 dans différentes lignées hématopoiétiques ou son introduction dans des souris transgéniques n'entraînent pas d'effets notables tant au niveau de la croissance des cellules ou de l'apparition d'une leucémie chez les souris (Andreasson *et al.*, 2001). En corrélation avec ce dernier point, des analyses chez des jumeaux identiques atteints de LLA ont révélé la présence de clones leucémiques possédant la translocation t(12;21), suggérant que celle-ci se soit produite *in utero*. De plus, l'apparition de la leucémie chez certains ne s'est manifesté que plusieurs années après leur jumeau, indiquant que d'autres altérations sont nécessaires pour la progression de ce cancer (Ford *et al.*, 1998; Wiemels *et al.*, 1999).

La présence de la translocation t(12;21) a été associée à la délétion, partielle ou complète, de la région incluant *ETV6* sur le chromosome 12 dans environ 80 % des cas, indiquant possiblement que l'allèle intact de *ETV6* interfère avec l'activité de ETV6-AML1 (Cavé *et al.*, 1997). Toutefois, l'existence de plusieurs cas de LLA possédant la translocation, mais sans la présence d'une délétion ou de mutations inactivatrices dans le second allèle de *ETV6*, ou ayant une délétion hémizygote de la région, mais sans la translocation t(12;21), semble indiquer que le tableau est beaucoup plus complexe (Stegmaier *et al.*, 1996; Cavé *et al.*, 1997).

1.5.2 CDKN1B

CDKN1B (ou *p27^{Kip1}*), situé à environ 1 Mb de *ETV6* sur le chromosome 12 (Figure 1-6), encode un membre de la famille des inhibiteurs de l'activité kinase des

complexes cyclines / kinases cyclines-dépendantes (CDKI) qui comprend le groupe des INK4 ayant des motifs ankyrins: p15^{INK4B}, p16^{INK4A}/CDKN2, p18^{INKC} et p19^{INK4D} et le groupe des protéines Kip/Cip: p21^{Cip1}/Waf1, p27^{Kip1}/CDKN1B et p57^{Kip2}. Ces complexes sont directement responsables, via l'inactivation de Rb-1 par phosphorylation, de la progression du cycle cellulaire entre les phases G1, S, G2 et M (voir Figure 1-4). Le groupe des protéines INK4 inhibe plutôt les complexes composés de la cycline D et des kinases CDK4 et CDK6 actives au milieu de la phase G1. CDKN1B et les autres membres du groupe Kip/Cip ont un champ d'action plus vaste, inhibant les complexes formés des cyclines A, B, D ou E et de la plupart des kinases. CDKN1B peut même induire le clivage, par une protéase, de la cycline A (Bastians *et al.*, 1998). Contrairement à CDKN1B, les autres membres du groupe KIP/CIP peuvent lier et inhiber PCNA, un composant de l'ADN polymérase. La surexpression de tous ces CDKIs bloque la transition G1 à S et ainsi entraîne l'arrêt de la prolifération cellulaire (Hunter, 1993; Serrano *et al.*, 1995). Toutes ces protéines sont essentielles pour le contrôle de la croissance cellulaire et donc une cible de choix pour les mutations ou délétions dans les cellules cancéreuses. Les membres du groupe INK4, p15^{INK4B}, p16^{INK4A} et p18^{INKC}, sont fréquemment inactivées dans les cancers incluant les leucémies aiguës (Section 1.3.4.2). L'expression de p21^{Cip1} est directement régulée, lors d'un dommage à l'ADN, par p53 qui est la cible de mutations dans une très grande proportion de cancers (Shiohara *et al.*, 1994). Cependant très peu de mutations ont été observées chez les membres du groupe KIP/CIP malgré une analyse exhaustive dans de nombreux cancers différents, suggérant qu'il ne sont pas une cible directe (Kawamata *et al.*, 1995; Hayette *et al.*, 1997; Morosetti *et al.*, 1995). Les analyses de la région codante de

CDKN1B dans les cas de leucémie lymphoblastique aiguë ayant une délétion du chromosome 12p n'ont révélé aucune mutation inactivante (Pietenpol *et al.*, 1995; Stegmaier *et al.*, 1996). Par contre, le niveau de la protéine est très variable dans les tissus cancéreux et une réduction, ou même une perte totale de son expression ont été associées à un mauvais pronostic clinique et à une grande agressivité tumorale dans plusieurs cancers, dont le cancer du pancréas, de la prostate, du sein et du colon (revu par Lloyd *et al.*, 1999), ainsi que la leucémie myéloïde (Yokozawa *et al.*, 2000).

L'inactivation de *CDKN1B*, tout comme de *p21^{Cip1}*, *in vivo*, par recombinaison homologue, est viable, indiquant une certaine redondance dans les fonctions, mais plusieurs défauts ont été observés (Kiyokawa *et al.*, 1996; Fero *et al.*, 1996; Nakayama *et al.*, 1996). Entre autres, plusieurs organes montrent une hyperplasie indiquant que les cellules ont effectué plusieurs divisions supplémentaires avant leur différenciation terminale, ce qui est également observé chez les souris *Rb-1 -/-*. D'ailleurs plusieurs études ont indiqué un rôle pour *CDKN1B* dans la différenciation. *CDKN1B* est exprimée à de faibles niveaux dans plusieurs types de cellules précurseurs, mais à un niveau élevé lorsque les cellules atteignent leur niveau final de différenciation (revu dans Lloyd *et al.*, 1999). Aussi, les souris homozygotes pour l'inactivation de *CDKN1B* ont un plus grand risque de développer certaines tumeurs, en faisant un excellent candidat à titre de suppresseur de tumeurs. Cependant Fero *et al.* (1998) ont analysé plus en détails leurs souris et ont observé que, suite à l'irradiation aux rayons gamma ou à l'administration de produits cancérogènes, les souris *CDKN1B -/-* et les souris *CDKN1B +/-* développaient plus de tumeurs que les souris sauvages. Dans les tumeurs des souris

hétérozygotes, l'allèle intact de *CDKN1B* n'avait pas été muté indiquant que ce gène pouvait être haploinsuffisant pour la suppression de tumeurs. De telles observations avaient également été faites chez les souris *p53 +/-*, entraînant notamment des lymphomes (Jiang *et al.*, 2001). Une récente étude chez l'humain a révélé que la méthylation du promoteur de *CDKN1B* était un événement relativement fréquent (environ 10 %) dans les cas de mélanomes métastatiques (Worm *et al.*, 2000). De plus, cette méthylation n'affectait qu'un seul des deux allèles indiquant encore une fois que l'haploinsuffisance de ce gène pourrait contribuer à la carcinogenèse.

1.5.3 Autres gènes connus dans la région 12p12.3

La région du chromosome 12p12.3 entre les gènes *ETV6* et *CDKN1B* a une taille d'environ 1Mb si l'on se fie aux plus récentes cartes physiques de la région. Il existe d'autres gènes connus à l'intérieur de cet intervalle. LRP6 (lipoprotein receptor protein 6) est une protéine transmembranaire faisant partie de la famille des récepteurs aux LDL. Elle partage avec LRP5 71 % d'identité ainsi que, contrairement aux autres membres de la famille, quatre domaines extracellulaires EGF (epidermal growth factor), mais n'ayant toutefois pas de domaine cytoplasmique catalytique (Brown *et al.*, 1998). Le phénotype des souris knock-out pour ce gène suggère que d'autres fonctions lui sont associées. Ces souris meurent à la naissance ayant de sévères malformations incluant une troncation du squelette, des défauts dans le développement des membres ainsi qu'une malformation du système urogénital, ressemblant étrangement aux souris déficientes pour les gènes *Wnt-1*, *Wnt-3A* et *Wnt-7A* (Pinson *et al.*, 2000). Les protéines Wnt sont des homologues de wingless, un facteur sécrété déclenchant un signal crucial dans le développement de l'axe

postérieur de la drosophile (revu dans Peifer et Polakis, 2000). Plus d'une dizaine de membres de la famille Wnt ont été identifiés. Des mutations dans ces gènes, engendrant une augmentation de leur cascade de signalisation, ou dans des gènes impliqués dans la transmission de celle-ci, notamment *APC*, *GSK3*, *FRAT* et *β -caténine*, ont été identifiées dans les cancers humains. À l'aide de ces données, trois groupes ont réussi à démontrer que LRP6 est le co-récepteur, avec les protéines Frizzled, de plusieurs protéines Wnt chez la drosophile, la grenouille *Xenopus laevis* ainsi que la souris (Pinson *et al.*, 2000; Tamai *et al.*, 2000; Wehrli *et al.*, 2000).

Deux autres gènes ont également été identifiés dans la région: *CREBL2* et *GPR19*. *CREBL2* encode une protéine ayant un domaine bZip de liaison à l'ADN similaire à celui des protéines CRE-BP (c-AMP response element binding proteins). Celles-ci jouent un rôle important dans le contrôle du cycle cellulaire en activant la transcription des cyclines A et D1 (Hoornaert *et al.*, 1998). Finalement *GPR19* encode un membre de la famille des récepteurs des protéines G exprimé plus particulièrement dans le cerveau. Cependant sa fonction précise demeure inconnue (O'Dowd *et al.*, 1996; Montpetit et Sinnott, 1999).

1.6 Approches utilisées pour l'identification de gènes

L'approche "classique" utilisée afin de caractériser un gène consiste premièrement à le cloner, à l'aide de différentes techniques dont l'utilisation de bibliothèques d'ADNc, de le séquencer puis d'identifier sa localisation chromosomique, en cytogénétique à l'aide de l'hybridation d'une sonde marquée, soit la technique de FISH,

ou par PCR grâce à l'utilisation d'hybrides de radiation somatiques hamster-humain (Cox *et al.*, 1990). L'identification de gènes inconnus à l'intérieur d'une grande région chromosomique donnée, ou clonage positionnel, est une étape longue et limitante. Plusieurs approches complémentaires peuvent être utilisées, mais la plupart requièrent l'utilisation de clones d'ADN génomique, de type YAC (environ 1 Mb), cosmides (40kb) ou BAC (100-200 kb), dérivés de la région d'intérêt. Ceux-ci peuvent être sélectionnés et positionnés sur le chromosome par PCR, à l'aide de marqueurs polymorphiques connus et des cartes physiques existantes, et par analyse de restriction (Figure 1-8).

1.6.1 Sélection d'ADNc

La technique de sélection d'ADNc permet d'isoler les gènes d'intérêt par hybridation (Lovett *et al.*, 1991; Parimoo *et al.*, 1991). Tout d'abord, l'ADN provenant des clones génomiques est coupée en petits fragments de 500-750 pb, puis ceux-ci sont biotinylés et liés à des billes magnétiques à l'aide de la streptavidine. Ensuite, ces fragments sont incubés en présence d'une librairie d'ADNc, puis, à l'aide d'un aimant, seuls les ADNc qui se sont hybridés sont conservés. Cette méthode a été utilisée pour caractériser les gènes situés sur le chromosome 12p menant à l'identification de 76 gènes différents (Baens *et al.*, 1995). Cependant, cette approche engendre un biais pour les gènes les plus exprimés. Aussi, il y a une sélection de la taille des cDNAs utilisés pour la construction de ces librairies. Les plus gros transcrits (>2 kb) sont alors absents ou très partiels, ce qui rend leur caractérisation plus difficile.

1.6.2 Capture d'exons

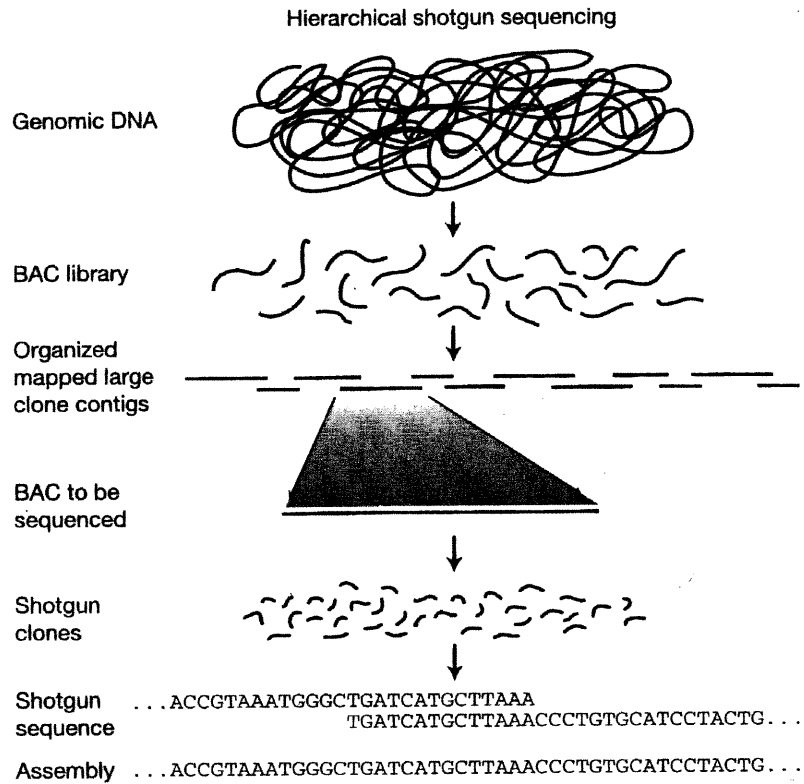


Figure 1-8. Représentation de la stratégie employée pour séquencer le génome humain. Le génome est tout d'abord fragmenté puis cloné dans un vecteur de type BAC. Ces clones sont ensuite organisés sur une carte physique selon la présence de marqueurs connus. Des BACs individuels sont ensuite sélectionnés et sous-clonés en plus petits fragments (2-4 kb) qui sont par la suite séquencés de façon aléatoire puis alignés pour reconstruire la séquence complète. Tiré de Lander *et al.*, 2001.

La technique de capture d'exons, ou "exon trapping", possède un avantage vis-à-vis la technique de sélection d'ADNc puisqu'elle est indépendante du niveau d'expression de chaque gène (Duyk *et al.*, 1990; Buckler *et al.*, 1991). Brièvement, les clones d'ADN génomique provenant de la région d'intérêt sont coupés en plus petits fragments (2-4 kb) puis sous-clonés dans un vecteur d'expression contenant des sites consensus donneur et accepteur d'épissage. Si la séquence ayant été clonée contient un exon dans la bonne orientation, celui-ci sera épissé et pourra être récupéré par RT-PCR suivi d'un clonage (Figure 1-9). Cependant, les premiers et derniers exons de chaque gène ne peuvent être isolés par cette méthode et la caractérisation d'un gène complet à partir d'un seul ou de quelques exons peut être compliquée. L'utilisation combinée des approches de sélection d'ADNc et d'amplification d'exons est couramment utilisée et permet d'identifier une grande partie des transcrits dans une région choisie.

1.6.3 Ilots CpG

Le taux de dinucléotides CpG dans le génome humain correspond à environ 20 % du taux prévu si tous les dinucléotides étaient représentés de façon équivalente. Dans le génome humain, la majorité de ces dinucléotides sont méthylés en position 5 sur la cytosine et une corrélation inverse entre le taux de méthylation dans les promoteurs et le taux d'expression des gènes a été observé (revu par Leonhardt et Cardoso, 2000). On croit que la désamination spontanée des cytosines, qui est un événement relativement commun, est à l'origine de la sous-représentation de ces dinucléotides. La désamination d'une cytosine (C) génère une uridine (U) normalement absente dans l'ADN et qui est immédiatement reconnue et clivée par l'uridine glycosylase (UDG) (Duncan et Miller,

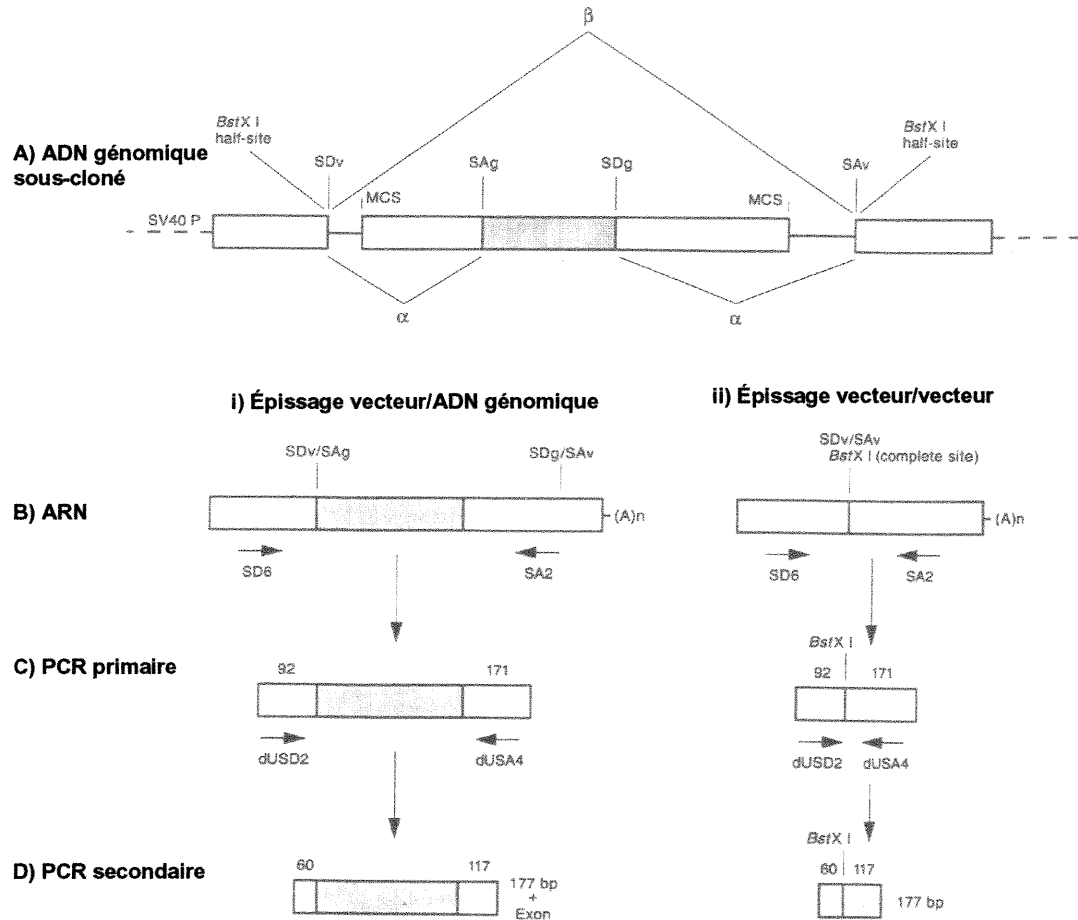


Figure 1-9. Technique de capture d'exons. A) L'ADN génomique est sous-cloné dans un vecteur contenant des sites donneurs (SDv) et accepteurs (SAv) d'épissage. B), Les produits d'épissage possibles sont représentés selon que l'ADN cloné contient (i) ou non (ii) un exon. C) L'ARN est isolé de cellules COS, réverse transcrit en ADNc puis amplifié par PCR. Le produit en ii) peut être digéré par *BstX1* pour éviter son amplification. D) Le produit de PCR primaire est réamplifié avec des amorces nichées.

1980). Cependant, la désamination d'une cytosine méthylée génère une thymine (T) qui est une base naturelle de l'ADN. Le mauvais appariement T-G pourra être reconnu spécifiquement par la thymine DNA glycosylase (TDG), mais dans quelques cas, il sera corrigé par d'autres membres du complexe de réparation de l'ADN pour un appariement T-A, introduisant ainsi une mutation et une perte d'un CpG (Hendrich *et al.*, 1999). On observe une grande concentration de dinucléotides CpG, correspondant jusqu'à 10 fois le taux prévu, dans la région située immédiatement en 5' de plus de la moitié des gènes. Ces régions ont été nommées îlots CpG (Bird *et al.*, 1985). On croit que la présence de ces îlots permettrait de contrôler la transcription dans les cellules germinales, via la méthylation des cytosines dans les promoteurs des gènes, et cette tendance aurait été conservée durant l'évolution (revu par Monk, 1995). Dans les cellules somatiques toutefois, ces îlots sont habituellement déméthylés et sont principalement associés avec des "housekeeping genes" ou gènes exprimés de façon constitutive dans tous les tissus (revu par Bird, 1986). Il existe quelques exceptions tels que les gènes situés sur le chromosome X chez la femelle et les gènes à empreinte parentale (imprinted genes) où un des deux allèles est inactivé par méthylation. Dans certaines tumeurs, on observe que l'activité globale DNA-méthyltransférase est plus élevée et que des gènes suppresseurs de tumeurs, tels *p15*, *p16* et *hMLH1*, sont inactivés par méthylation (revu par Baylin *et al.*, 1998). Il est possible d'identifier ces îlots CpG par l'utilisation d'enzymes de restriction dont la séquence de reconnaissance possède un ou plusieurs dinucléotides CpG, tels *NotI* (GCGGCCGC), *SfiI* (GGCCN₅GGCC). La présence de deux ou plusieurs de ces sites de restriction, normalement très rares dans le génome humain, indique

habituellement qu'on se situe dans un îlot CpG. Le sous-clonage de ces fragments peut ainsi permettre l'identification de nouveaux gènes (Gillespie *et al.*, 1991).

1.6.4 Analyse comparative

L'analyse comparative de régions génomiques entre des espèces relativement éloignées est un outil puissant pour l'identification de gènes. En effet, le principe de l'évolution fait en sorte que les séquences importantes fonctionnellement, donc les régions codantes des gènes, seront beaucoup plus conservées que le reste. Donc, malgré la complexité des génomes des vertébrés, il est possible de cette façon d'identifier relativement facilement l'emplacement des gènes. Aussi, l'observation que l'ordre des gènes sur les chromosomes est relativement bien conservé à travers l'évolution des espèces animales, du moins chez les vertébrés fait de l'analyse comparative un outil fort avantageux pour le clonage positionnel de gènes. Le modèle génomique vertébré le plus utilisé actuellement est la souris. Malgré le fait que la plupart des gènes identifiés dans la région chromosomique 12p12, chez l'humain sont situés dans une région synténique, i.e. région contenant les mêmes gènes, sur le chromosome 6 chez la souris, l'absence de séquences génomiques murines ordonnées dans les bases de données publiques et la complexité de son génome, comparable à l'humain, n'en fait pas un bon modèle pour le clonage positionnel de gènes (revu par Clark, 1999).

Récemment, les poissons-lune ou "pufferfishes" ont été proposés comme étant un excellent modèle d'analyse génomique comparative à cause de leur caractéristiques étonnantes (Brenner *et al.*, 1993). La taille de leur génome est environ 7,5 fois plus

petite que celle de l'humain (400 Mb vs 3000 Mb) (Hinegardner, 1968). Malgré cette petite taille, les poissons-lune sont des animaux complexes partageant plusieurs caractéristiques physiologiques, neurologiques et immunologiques avec tous les vertébrés suggérant qu'ils possèdent un nombre de gènes similaire à l'humain. L'analyse de séquences aléatoires indique que seulement 5-7 % de leur génome est composé de séquences répétées comparativement à 60 % pour l'humain. Conséquemment, seulement 2-3 % du génome humain est transcrit comparativement à 15-20 % pour les poissons-lune. De plus, la taille des introns, mais pas celle des exons, est considérablement réduite (revu dans Koop et Nadeau, 1996). Par exemple, le gène responsable de la maladie de Huntington possède 67 exons chez l'humain et chez *Fugu rubripes*, un poisson-lune. Cependant, la taille des introns varie de 47 bp à 1476 bp chez *Fugu* comparativement à 131 bp à 12 286 bp chez l'humain, résultant en une réduction totale de 7,5 fois de la taille du gène (Baxendale *et al.*, 1995). La réduction observée aurait aussi un lien avec la composition en GC de la région correspondante chez l'humain, ou isochore (Elgar, 1996) (Tableau I-V).

Plusieurs études ont démontré la présence de régions synténiques entre *Fugu* et l'humain sur des distances de plus de 1 Mb et comprenant plusieurs gènes. Par exemple, la région responsable du syndrome de WAGR et de la tumeur de Wilm comprend trois gènes connus, *PAX6*, *RCNI* et *WT1*. Ces trois gènes, ainsi que deux autres gènes de fonction inconnus, ont été identifiés dans le même ordre dans une région d'environ 100 kb chez *Fugu*, soit une réduction de 15 fois par rapport à l'humain (Figure 1-10). Cette approche a permis de confirmer l'existence de deux nouveaux gènes, mais aussi de

Tableau I-V Comparaison de la taille de gènes orthologues chez l'humain et chez Fugu

| Gène | Isochore | Taille (kb) | | | | Réduction |
|-------|----------|-------------|---------|------|---------|-----------|
| | | Humain | | Fugu | | |
| | | Gène | Introns | Gène | Introns | |
| APP | L | 285 | 282 | 8 | 5.8 | 35.5 X |
| C9 | L | 90 | 87 | 3 | 1.1 | 30 X |
| HD | H1/H2 | 172 | 162 | 23 | 13.4 | 7.5 X |
| p55 | H1/H2 | 25 | 23.5 | 5.5 | 4 | 4.5 X |
| TSC2 | H3 | 43 | 37.5 | 17.3 | 11.8 | 2.5 X |
| L1CAM | H3 | 13.2 | 9.4 | 14 | 10.2 | .95 X |

isochore L: région riche en AT

isochore H: région riche en GC

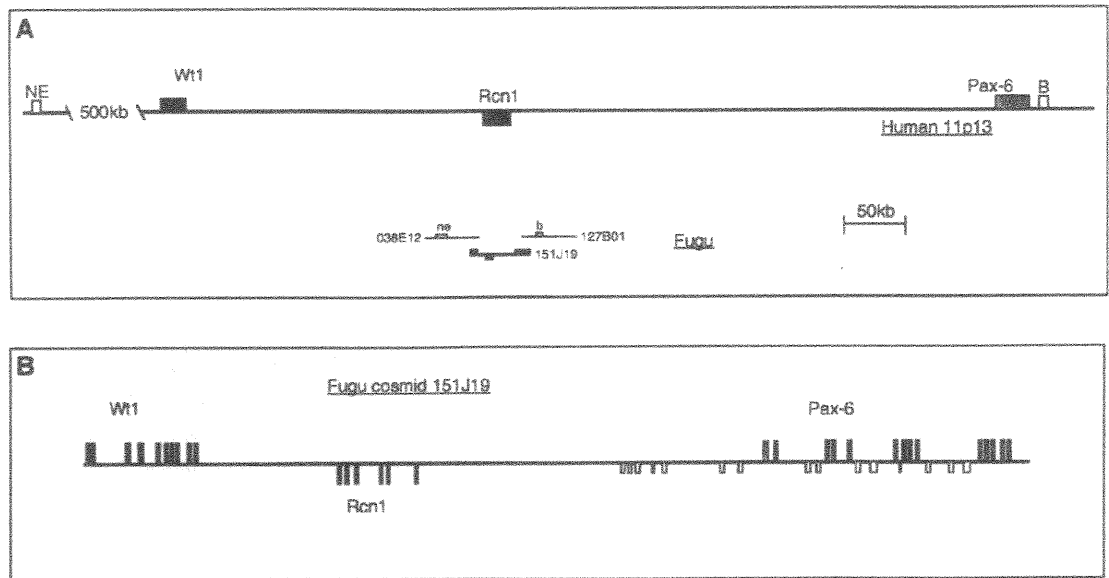


Figure 1-10. Représentation de deux régions synténiques entre l'humain et Fugu. A) Représentation de la région 11p13 chez l'humain comprenant les gènes WT1, RCN1 et PAX-6. Le cosmide 151J19 provenant de Fugu contenant les mêmes gènes y est représenté à l'échelle. B) Vue agrandie du cosmide 151J19 mesurant 45 kb. Les exons sont représentés par des boîtes noires tandis que les régions conservées non-codantes sont représentées en blanc. Tiré de Clark, 1999.

fortement suggérer l'absence de gènes dans les grandes séquences intergéniques de ce locus chez l'humain (Miles *et al.*, 1998). Aussi, Trower *et al.* (1996), ont identifié sur un même cosmide quatre gènes, incluant *cFOS*, un gène relié à la maladie d'Alzheimer, sur le chromosome 14q24, en synténie avec leur orthologues répartis chez l'humain sur plus de 600 kb, soit une réduction de près de 50 fois, faisant de *Fugu* un outil de clonage positionnel extrêmement intéressant. Cependant, d'autres gènes présents sur ce cosmide avaient leur orthologue situés sur différents chromosomes chez l'humain indiquant que plusieurs réarrangements chromosomiques étaient survenus au cours de l'évolution entre les deux espèces. De plus, Gilley *et al.* (1997), ont observé dans la région *surfeit*, comprenant les six gènes *surf*, de grands réarrangements dans l'ordre de ces gènes entre l'homme et *Fugu* ainsi que plusieurs bris de synténie. Cependant, la grande majorité des études publiées sur le sujet montre un bon niveau de synténie entre l'humain et *Fugu*.

Une récente étude estime qu'il y aurait eu entre 6000 à 12000 réarrangements pendant les 450 millions d'années d'évolution séparant le génome de *Fugu* et celui de l'humain, générant des régions synténiques moyennes d'environ 250-500 kb (McLysaght *et al.*, 2000). Ces estimations ont été faites sur un faible échantillon de gènes séquencés chez *Fugu* et à partir d'une estimation élevée du nombre total de gènes, soit 60000, ce qui surestime le nombre de réarrangements. L'estimation actuelle du nombre de gènes chez l'humain se situe autour de 30 000 (Venter *et al.*, 2001; Lander *et al.*, 2001). Cependant, ce nombre de réarrangements est, de façon relative, largement supérieur à celui estimé entre la souris et l'humain, soit de 150-200 pour 100 Myr d'évolution.

L'utilisation des poissons-lune comme modèle comparatif peut avoir aussi d'autres avantages. Étant donné la distance évolutive séparant les poissons de l'humain, seuls les éléments essentiels ont été conservés. Ceci facilite l'identification des gènes incluant les régions codantes et régulatrices. Par exemple, le gène *MLL*, fréquemment réarrangé dans plusieurs cas de leucémie (Section 1.3.4.1.3) et dont la fonction n'est pas clairement établie, possède 90,8 % d'identité au niveau de sa séquence protéique prédite avec son homologue murin. Cette grande homologie ne permet pas d'identifier des domaines importants pour la fonction de la protéine qui normalement ont un taux de mutation inférieur au reste de la protéine au cours de l'évolution. À l'autre extrémité du spectre, *MLL* et son homologue chez la drosophile, *TRX*, ne partagent que deux domaines, *ATA* et *SET*, de liaison à l'ADN et d'interaction protéine-protéine respectivement, qui sont présents chez plusieurs autres protéines ne permettant pas l'identification de domaines fonctionnels spécifiques à *MLL*. La comparaison avec *Fugu* est plus avantageuse, étant donné qu'il est beaucoup plus distant de l'homme que la souris et qu'il possède toute la même complexité physiologique que l'humain, contrairement à la drosophile. À l'aide de cette analyse comparative, plusieurs régions de très grande similarité, ayant possiblement un rôle fonctionnel important, ont été identifiées (Caldas *et al.*, 1998).

De plus, certaines régions non-codantes conservées peuvent également être observées de cette façon. Plusieurs études ont identifié des séquences non-répétées ayant une très grande identité dans les promoteurs, les régions 3' non-traduites et même les introns de gènes (Aparicio *et al.*, 1995; How *et al.*, 1996; Rowitch *et al.*, 1998; Miles *et*

al., 1998; Gellner et Brenner, 1999). Ces séquences ont probablement un rôle dans la régulation de l'expression de ces gènes. Aparicio *et al.* (1995) ont montré que ces séquences homologues, identifiées dans des régions non-codantes de *Hoxb-4*, étaient des enhancers et permettaient l'expression de gènes dans des souris transgéniques. Ces études pourraient permettre d'étudier plus facilement les régions régulatrices des gènes humains puisqu'il est souvent impossible de les cloner étant donné leur très grande taille (Venkatesh *et al.*, 1997). Cependant l'utilisation de gènes de *Fugu* peut être problématique: Sathasivam *et al.* (1997) ont montré que le gène *HD* de *Fugu*, incluant son promoteur, était bel et bien exprimé, mais n'était pas correctement épissé lorsqu'introduit dans des souris transgéniques.

La séquence complète du génome de *Fugu* sera un outil très important pour l'identification et la caractérisation des gènes humains. Jusqu'à maintenant plus de 25 Mb du génome de *Fugu* (Elgar *et al.*, 1999) et plus de 150 Mb, soit un tiers du génome de *Tetraodon nigroviridis*, un autre poisson-lune, ont été séquencés et publiés (Roest-Crollius *et al.*, 2000). Ce dernier groupe a publié un programme, EXOFISH, permettant la comparaison d'une séquence nucléique d'intérêt avec le génome de *T.nigroviridis* au niveau protéique. Par cette méthode, les auteurs ont pu identifier 66 % d'un groupe de gènes servant de contrôle situés sur le chromosome 22 et ainsi estimer le nombre total de gènes chez l'humain entre 20 000 et 30 000 (Roest-Crollius *et al.*, 2000). Les estimations utilisant la séquence complète, maintenant obtenue, du génome humain arrivent à un nombre similaire (Lander *et al.*, 2001; Venter *et al.*, 2001).

1.6.5 Analyse bio-informatique

Depuis 1999, le consortium public du projet Génome Humain publie ses données de séquence brutes, c'est-à-dire non analysées. En juin 2000, ce consortium a annoncé, conjointement avec la compagnie privée Celera, que le génome humain avait été entièrement séquencé (Lander *et al.*, 2001; Venter *et al.*, 2001). Par contre dans les faits, environ 90-95 % du génome est séquencé et seulement un peu plus du tiers des séquences sont correctement ordonnées le long des différents chromosomes. L'analyse de ces séquences est évidemment une méthode extrêmement utile et efficace pour identifier les gènes présents dans une région donnée. La comparaison avec les bases de données comme dbEST ou Unigene, qui sont une collection de séquences exprimées, permet d'identifier sur l'ADN génomique les régions transcrites. D'autres programmes informatiques, tels Hexon, Genescan, Grail, HMMGene, etc. permettent de prédire la présence d'exons en se basant sur l'existence de séquence consensus à la jonction intron-exon et/ou de cadre de lecture. Cependant, il a été montré que ces programmes, malgré une bonne sensibilité, détectaient une importante proportion de faux positifs rendant l'analyse compliquée (Bursset et Guigo, 1996).

1.7 Hypothèses et objectifs du projet

La fréquence de la délétion hémizygote du chromosome 12p12.3 observée dans 26-40 % des cas de LLA ainsi que dans plusieurs autres cancers suggère la présence d'un important gène suppresseur de tumeurs dans cette région d'environ 750 kb, incluant le gène ETV6. Notre hypothèse est qu'une mutation quelconque entraîne l'inactivation de l'allèle intact, ce qui correspond à la définition d'un gène suppresseur de tumeurs. Une

des étapes limitantes cependant est l'identification des gènes présents dans la région. Étant donné la grande complexité du génome humain, nous proposons d'utiliser le poisson-lune *Fugu rubripes* comme modèle génomique comparatif pour le clonage positionnel pour étudier cette région. Son petit génome, la présence d'un nombre similaire de gènes ainsi que de régions synténiques devraient nous permettre d'identifier plus facilement les gènes présents dans la région critique chez l'humain. Nous avons donc utilisé le gène *ETV6* comme sonde pour identifier son orthologue chez *Fugu rubripes* et ainsi identifier plus facilement, après séquençage de la région chez *Fugu*, les gènes liés à *ETV6* chez l'humain.

La présence de la séquence génomique humaine dans les bases de données à partir de la fin de 1999 nous a poussé à développer une approche combinée, faisant appel à des techniques biologiques et informatiques, afin d'identifier les gènes présents dans le locus 12p12.3. Après leur identification, nous avons caractérisé ces gènes chez un groupe de patients ayant la leucémie lymphoblastique aiguë et porteurs d'une délétion de la région 12p12.3. Nous avons vérifié l'expression ainsi que la présence de mutations dans ces gènes afin d'identifier le gène suppresseur de tumeurs ciblé par ces délétions.

Chapitre 2

-Article 1-

Analysis of the conservation of synteny between Fugu and human chromosome 12

Article en préparation

Contribution des auteurs:

A. Montpetit: a contribué aux expériences en laboratoire, a effectué toutes les analyses et a écrit l'article.

M.D. Wilson: a réalisé l'étape de séquençage shotgun.

M. Chevrette: a contribué à la cartographie chez zebrafish.

B. Koop: supervision du laboratoire (Victoria).

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

Analysis of the conservation of synteny between *Fugu* and human chromosome 12

Alexandre Montpetit^{1,2}, Michael D. Wilson³, Mario Chevrette⁴, Ben F. Koop³ and Daniel Sinnett^{*,1,2,4}

¹Division of Hematology-Oncology, Charles-Bruneau Cancer Center, Research Center, Sainte-Justine Hospital, 3175 Côte Ste-Catherine, Montreal, QC H3T 1C5, Canada;

²Department of Biochemistry and ⁴Department of Pediatrics, University of Montreal, Montreal, QC, Canada.

³Department of Biology, Centre for Environmental Health, University of Victoria, Victoria, BC V8W 2Y2, Canada

⁴Montreal General Hospital Research Institute and Department of Surgery, McGill University, Montreal, QC H3G 1A4, Canada

*Correspondence: Dr Daniel Sinnett, Charles-Bruneau Cancer Center, Sainte-Justine Hospital, 3175 Côte Ste-Catherine, Montreal, QC H3T 1C5, Canada. Tel: (514) 345-2990; Fax: (514) 345-4731; E-mail: daniel.sinnett@umontreal.ca.

Running title: Chromosome 12 synteny in *Fugu*

Abstract

The pufferfish *Fugu rubripes* with its compact genome is increasingly recognized as an important vertebrate model organism for comparative genomic studies. In particular, large regions of conserved synteny between human and *Fugu* genomes indicate its utility as a tool for positional cloning. Chromosome 12p12-13 is frequently deleted in various human hematological cancers and solid tumors, but the actual tumor suppressor gene remains unidentified. We investigated approximately 200 kb of the genomic region surrounding the *ETV6* locus in *Fugu rubripes* in order to find conserved functional features, such as genes or regulatory regions, that could give insight into the recurrence of the deletions in human. Seven genes were identified in this region, but even if the synteny with human chromosome 12 was conserved, extensive genomic rearrangements have occurred between the *Fugu* and human *ETV6* locus, thereby limiting the usefulness of *Fugu* as a model for positional disease gene cloning. However, this comparative analysis of about 200 kb of sequence led to the identification of previously uncharacterized genes in the human genome and some potentially important regulatory sequences as well. This is a good demonstration that the genome of *Fugu rubripes* is a valuable tool for the characterization of the human genome.

Introduction

Part of the short arm of chromosome 12 is frequently deleted in a wide variety of hematological malignancies (Berger *et al.*, 1986; Raimondi *et al.*, 1986; Chan *et al.*, 1992) as well as in certain solid tumours including breast, lung, ovarian, and prostate carcinomas (Kibel *et al.*, 1999; reviewed in Aïssani *et al.*, 1999). Loss of heterozygosity studies showed hemizygous deletions of chromosome 12p12 in 26 to 47 % of pre-B acute lymphoblastic leukemia (ALL) cases, making it one of the most common genetic alterations found in this disease (Stegmaier *et al.*, 1995; Takeuchi *et al.*, 1996; Bacchichet *et al.*, 1997). The frequent loss of genetic material in tumour cells is usually indicative of the inactivation of tumour suppressor genes. The construction of both high-resolution genetic and physical maps led to the delineation of the shortest commonly deleted region within a 750kb interval delimited by microsatellite markers *DI2S89* and *DI2S358* in ALL patients (Bacchichet *et al.*, 1997; Aïssani and Sinnett, 1999; Baens *et al.*, 1999).

Two genes were known to be present in this region, *ETV6* and *LRP6*. *LRP6* encodes a member of the LDL receptor family (Brown *et al.*, 1998) and was recently shown to act as the WNT co-receptor (Pinson *et al.*, 2000; Tamai *et al.*, 2000; Wehrli *et al.*, 2000). The deregulation of this signaling pathway is involved in many human cancers (reviewed in Taipale and Beachy, 2001). *ETV6*, also known as TEL, is a member of the ets-like family of DNA binding proteins, and is involved in numerous translocations in hematological malignancies, leading to the production of fusion proteins with oncogenic activities (Golub *et al.*, 1994; reviewed in Rubnitz *et al.*, 1999).

Recently, it has been shown that the normal ETV6 protein acts as a transcriptional repressor of several matrix metalloproteinases and that its expression could counteract the transformation of cultured cells by ras (Van Rompaey *et al.*, 1999; Fenrick *et al.*, 2000). However, mutational analysis failed to reveal any inactivating mutations in the second *LRP6* or *ETV6* allele in ALL patients with hemizygous deletion at this locus as expected for a classical tumour suppressor gene (Stegmaier *et al.*, 1996; Baens *et al.*, 1999). These observations led to the suggestion that an as yet unidentified tumour suppressor gene might be located between *D12S89* and *D12S358*. In order to characterize the genes present in that region, we used different approaches. In one study, we used the now almost complete sequence of the human genome and EST database searches. With this approach, we mapped five genes including *BCL-G*, *MKP-7* and three other previously uncharacterized genes with unknown function in the critical region, although others could have been missed (Montpetit *et al.*, 2002). In this present study, we wanted to use a comparative approach, by using *Fugu* as a model, to analyse the *ETV6* locus.

The pufferfish *Fugu rubripes* has a compact 400-Mb genome, which is one of the smallest of the vertebrate kingdom, and has been presented as a model vertebrate genome (Brenner *et al.*, 1993). Many of its characteristics offer real advantages for studying the human genome. The *Fugu* genome is eight times smaller than its human counterpart, but has a similar gene set, in relation to both the total gene number and the gene structures. Only 3 % of the human genome is coding, compared to nearly 20 % for *Fugu*, which greatly facilitates gene identification and cloning. The *Fugu* model has

been successful in characterizing important functional protein domains (Baxendale *et al.*, 1995; Caldas *et al.*, 1998) and even non-coding sequences, located in promoters, 3' UTR and introns. Some of them have been shown to have functional significance *in vivo*, indicating that they act as enhancers of transcription (Marshall *et al.*, 1994; Aparicio *et al.*, 1995; Kimura *et al.*, 1997; Rowitch *et al.*, 1998). Also, many reports presents evidence of gene order conservation (synteny) over relatively long stretches of sequence between human and *Fugu* indicating that *Fugu* could be used to clone potential new disease genes (reviewed in Venkatesh *et al.*, 2000). However, other reports showed considerable difference in the order of the genes in some regions between these two species and gave rise to a controversy on the potential of *Fugu* as a model organism for positional cloning (Gilley *et al.*, 1997; Gilley and Fried, 1999).

In this study, we characterized about 200 kb by shotgun sequencing around *ETV6* in *Fugu*. We identified seven genes, including three uncharacterized genes in the human genome and some potentially important regulatory sequences. However, although some synteny with chromosome 12 was conserved, extensive genomic rearrangements have occurred between the *Fugu* and human *ETV6* locus. These observations raise important questions on the usefulness of *Fugu* as a model for disease gene cloning.

Materials and methods

Isolation of BAC clones

An *ETV6* cDNA probe was used to hybridize a *Fugu* genomic BAC library from Genome Systems and cosmid library no.66 from Carola Burgtorf purchased from RZPD

(<http://www.rzpd.de>) as described in Montpetit and Sinnott (2001). Five overlapping clones were obtained, BACs 220c16, 225j23, 227h21 and 235e13 and cosmid ICRFc66A1236Q1.2. A physical map was constructed using rare-cutters (*NotI*, *SfiI* and *SrfI*) and Southern hybridization with T7, Sp6 and human *ETV6* probes. Direct sequencing of BAC 227 with T7 primer enabled us to determine specific primers for BAC walking (227T7F: AAGCTTGGATCTGGGTCCGTC and 227T7R: GTCCTTTCATTCCACCACAG). A 200bp fragment was amplified by PCR, labeled by random priming and hybridized back to the BAC and cosmid libraries and led to the isolation of one overlapping cosmid (ICRFc66G248Q1.3) and 2 BAC clones (207f2 and 231j9).

DNA sequencing

Fugu BACs 220c16 and 231j9 were isolated using NucleoBond Plasmid Maxi Kits (Clontech) and randomly sheared by nebulization. Blunt ends were assured using mung bean nuclease, klenow fragment and DNA polymerase I and fragments from 1.5 to 3 kb were size-fractionated by agarose gel electrophoresis. Fragments were ligated into *SmaI*-digested M13mp9 vector and single stranded templates were purified using Qiagen M13 plates. Random clones from each sub-library were sequenced using dye-terminator and dye-primer chemistry (Amersham and ABI) and ABI 373 and 377 automated DNA sequencers. BAC 220 was sequenced to a 3.1 fold redundancy and BAC 231 was sequenced to a 2.2 fold redundancy.

Sequence analysis

DNASStar software was used for gel trace analysis, contig assembly and to obtain global alignments of genomic data. Repeat elements were identified by using RepeatMasker2 (<http://ftp.genome.washington.edu/cgi-bin/RepeatMasker>). NIX software (<http://www.hgmp.mrc.ac.uk/NIX>) was used to identify the putative genes. It integrates more than 15 different programmes for exon prediction (FGenes, Genescan, Grail, Hexon, etc.) and Blast analysis (Swissprot, dbEST, Unigene, etc.). All other sequence analysis were performed using BLAST or CD-Search programs on the NCBI site (<http://www.ncbi.nlm.nih.gov>). Multiple sequence alignments were obtained using the ClustalW algorithm (Thompson *et al.*, 1994). The physical map of chromosome 12 was obtained from the Human Genome Sequencing Center at the Baylor College of Medicine (<http://www.hgsc.bcm.tmc.edu/>).

Zebrafish radiation hybrid screening

The zebrafish LN54 radiation hybrid (RH) panel was probed with primers zEX3F: GCATGCAGCCGGTGTCTGGA and zEX3R:GTCCTCTTTGGTGAGCAGCAG corresponding to the third exon of the zebrafish *ETV6* gene (GenBank acc. no. AF339838) on the 93 hybrids. Mapping results were obtained at ZFIN (<http://zfish.uoregon.edu/ZFIN>).

Results

Isolation of clones and shotgun sequencing

We previously isolated four BACs and one cosmid that hybridized positively

with a human *ETV6* probe and characterized this gene (Montpetit and Sinnett, 2001). We obtained a physical map by digesting the clones with rare-cutters (*SfiI*, *NotI* and *SrfI*) and Southern hybridization with T7, Sp6 and various *ETV6* probes (Figure 1). We then isolated two other BACs and one cosmid by extending the contig in the direction of the 3' end of *ETV6*. We thus obtained a *Fugu* contig covering about 200kb and containing the *Fugu* ortholog of *ETV6* (*fETV6*). Two non-overlapping, but contiguous *Fugu* BAC clones 220c16 (approximately 65 kb) 231j9 (approximately 115 kb) were analyzed using a shotgun cloning and sequencing approach. 506 sequence reads corresponding to more than 400 kb were assembled into 87 contigs using DNASTar software. About 60 kb of sequence, from 22 contigs, were obtained for BAC 220 and 95 kb of sequence, from 65 contigs, were obtained for BAC 231. Most contigs could be assembled precisely assembled using the complete *ETV6* genomic sequence (GenBank acc. no. AF340230) and shotgun reads from a cosmid clone (188P21) characterized from the *Fugu* genome project (Elgar *et al.*, 1999; see Figure 1). After this assembly, it was revealed that both BAC clones ended at a common *HindIII* site used for cloning.

Sequence analysis of the *fETV6* locus

We used the NIX program that includes several sequence analysis tools to identify protein or EST matches and possible exons. Blast analysis of the sequence revealed homology to several known or hypothetical proteins in addition of *ETV6*. Beginning from the Sp6 extremity of 220 the gene order is as follow: *Timp3*, *SynIII*, hypothetical gene similar to *CAB70657*, *ETV6*, hypothetical gene similar to *BAB01579*, *MCT2*, *Lig-2* and *KUB3*. The orientation of the *KUB3* gene could not be inferred (Figure

1 and Table 1).

Strong sequence identity to the *SynIII* (synapsin III) and *Timp3* (tissue inhibitor of metalloproteinases-3) genes was observed at the sp6 extremity of BAC 220 (Table 1). Only the first 5 exons of *SynIII* and the last 4 exons of *Timp3* are present on this BAC and could be characterized. *Timp3* is located within the fifth intron of *SynIII* on the opposite strand and in human, the physical relation between these two genes is the same. In fact, the *Drosophila* homologs for these two family are organized in the same way and gave rise to the four mammalian members of the *Timp* and *Syn* family due to a fourfold amplification of the genome during invertebrate to vertebrate transition (Pohar *et al.*, 1999; Kao *et al.*, 1999). Both genes map to chromosome 22 in human. Zebrafish and *Fugu* are two teleost fishes and should have approximately the same genomic structure. Since physical maps are available in zebrafish, we mapped the *ETV6* gene using a zebrafish radiation hybrid panel. *ETV6* mapped between marker Z1366 and Z21743 on zebrafish LG4 chromosome. The Z21743 marker is identified as *SynII*. However, close analysis revealed stronger identity (88 % vs 85 %) to *SynIII* at the protein level indicating that this marker is probably *Syn III* and that *ETV6* and *SynIII* are syntenic in zebrafish and *Fugu*, but not in human.

Immediately upstream of the synapsin gene in *Fugu* and on the same strand, we identified an homolog for hypothetical gene *FLJ22693* (Table 1). The predicted protein has similarity to the TRF1-interacting ankyrin-related ADP-ribose polymerase, but its actual function is unknown. Blast analysis revealed 3 human PAC clones that contained

hypothetical gene *FLJ22693* (AC004961, AC004849 and AC025816) and that mapped on 7q33-35. Even though the overall identity is low, the position of the splicing sites are conserved. In *Fugu*, the gene has 11 exons, compared to 12 in human, since exons 3 and 4 are fused together. At least ten copies of parts of the gene, containing often interrupted coding regions, are found within this region. Several stretches of sequences are almost identical (>99 % sequence identity) over a few kb which complicated the contig assembly process. At least two complete copies are present, but only one could encode a complete protein. However, it is to be noted that homology is low between the human and the *Fugu* genes (48 % identity at the nucleotide level and 38 % identity at the protein level) and that the actual ortholog could be different. We observed another homolog, *C12ORF6* (Gen Bank acc. no. NM_020367) that lies on chromosome 12p13 near *ETV6* on BACs AC006207 and AC027283. The overall similarity was similar, but the actual sequence of *C12ORF6* corresponded to only the last 7 exons of the *Fugu* gene. However, it could reflect the fact that the actual *C12ORF6* gene is longer. Downstream of *fETV6*, only one gene is observed in approximately 60 kb of sequence on BAC 231. This gene has strong identity (71 % and 76 %) to two uncharacterized genes, *BAB71606* and *LOC128320*, located on human chromosomes 1 and 3 respectively (Table 1). They encode proteins with unknown function.

Two genes are observed near the sp6 extremity of BAC 231. First, the *Fugu* ortholog of *SLC16A7* (solute carrier family 16 member 7) which encodes a monocarboxylic acid transporter, was identified. It has 6 exons and 5 are coding in both human and *Fugu*. Then we identified a gene with significant homology (72 % identity)

to a partial cDNA from human (GenBank acc.no. BC012380) (Table 1). It encodes a protein with high similarity (67 %) to *LIG-1* an integral membrane protein and was named *LIG-2*. A complete *LIG-2* sequence from human could be deduced from the alignment of the human *LIG-1* and *Fugu LIG-2* genes to the genomic sequence (data not shown). In both *Fugu* and human, the gene has 17 coding exons. A sequence without coding potential, but which is extremely conserved between *Fugu* and human, was observed within the *LIG-2* intron 2 (Figure 2). This sequence was only conclusively identified on chromosome 12 thus identifying *LIG-2* as the true ortholog of the *Fugu* gene. Finally, the ortholog of *KUB3* gene was observed in the data released from cosmid 188P21. All three gene, *SLC16A7*, *LIG-2* and *KUB3* are present on 12q13 BACs (GenBank acc.nos. AC079905, AC078963 and AC084033 respectively). *LIG-2* and *KUB3* are only about 200 kb apart according to the Baylor College of Medicine physical map while *SLC16A7* is at least 1 Mb from those two genes.

If we assume that *C12ORF6* is the actual ortholog of the *Fugu* gene located just downstream from *ETV6*, then synteny with the human chromosome 12 is observed with 5 *Fugu* genes (*C12ORF6*, *ETV6*, *SLC16A7*, *LIG-2* and *KUB3*). However, if we combine the results of this analysis and from another study on chromosome 12 genes from *Fugu* (Power *et al.*, 2000), we observe that while synteny is conserved, gene order is not and that multiple rearrangements and linkage breaks occurred since 430 Myr when the two species diverged (Figure 3).

Discussion

Conservation of synteny between distant species is a characteristic that many researchers identified as an important tool for the cloning of disease genes. *Fugu rubripes* was described as an excellent model for this since its genome is extremely compact, while it possesses a similar gene content with humans, and many examples of conservation of synteny with human are observed (reviewed in Venkatesh *et al.*, 2000). Chromosome 12p12-13 is frequently deleted in various human hematological cancers and solid tumors, but the actual tumor suppressor gene remains unidentified. Many genes, including *ETV6*, have been identified in this region using a combination of techniques such as cDNA selection, exon trapping and EST database screening, but many genes could have been missed (Baens *et al.*, 1999; Montpetit *et al.*, 2002). In this study, we characterized approximately 200 kb of *Fugu rubripes* genomic sequence around *ETV6* and identified eight genes. Human homologues of these genes were observed on various chromosomes, but five of them are present on chromosome 12 indicating that synteny with *ETV6* was somehow maintained throughout 430 Myr of evolution. However, by combining the results of another chromosome 12 study in *Fugu*, we have a more global picture of the actual rearrangements that took place during evolution (Figure 3). Many chromosomal inversions must have occurred between human and *Fugu* to explain the position of each gene. But more importantly, we observe that neighbouring genes in the *Fugu* genomic sequence are separated by many megabases in the human chromosome. Many human genes map in those regions, including homologues that are present on a different linkage group, in *Fugu*. The closest

homologues identified on human chromosome 12, *LIG-2* and *KUB3*, are separated by at least 200 Mb according to recent physical maps that include at least one human gene, *PLEC1*.

Many reports have observed conservation of synteny between human and *Fugu* genes (reviewed in Venkatesh *et al.*, 2000), but only a few reports could demonstrate that gene order was also conserved. That includes the *PAX6* region on chromosome 11p (Miles *et al.*, 1998) and the Alzheimer disease locus on chromosome 14 (Trower *et al.*, 1999). Some reports actually showed extensive rearrangements between human and *Fugu* genomic regions, i.e. the surfeit locus (Gilley *et al.*, 1997) and even the Hox locus (Aparicio *et al.*, 1997), but once the complete human genome is analyzed in details, these rearrangements will probably be observed in many more cases, including the *PTHLH* analysis (Power *et al.*, 2000; Figure 2) and other already published reports. A recent computer simulation showed that 4000 to 16000 chromosomal rearrangements occurred between *Fugu* and human which is a much faster rate than seen with human/mouse comparisons (180 rearrangements) even when normalized for the actual evolutionary divergence time (McLysaght *et al.*, 2000). One plausible explanation is that when the bony fish lineage underwent a complete genome duplication it has been followed by differential gene loss. As many as 80 % of these duplicated genes were selectively lost (Amores *et al.*, 1998; Meyer and Schartl, 1999; Taylor *et al.*, 2001). This mechanism should in principle maintain synteny but not absolute gene order, which is what we observe in our study. We also observe that *ETV6* and *SynIII*, present on two different chromosomes in human, are on the same linkage group in both *Fugu* and

zebrafish, indicating that the actual observed chromosomal rearrangement is specific to the teleost lineage, and thus probably related to the whole genome duplication, although not specific to the *Fugu* genome. Therefore, it excludes these fishes as model for disease gene cloning by the positional cloning approach.

During the last year we have seen a dramatic increase in the number of pufferfish sequence in the database. By shotgun sequencing of cosmid clones, 2 5Mb of *Fugu* sequence were generated by the *Fugu* Landmark Project (Elgar *et al.*, 1999; <http://Fugu.hgmp.mrc.ac.uk>). Recently, 155 Mb of genomic sequence from the closely related *Tetraodon nigroviridis*, which represents one third of its genome, was obtained (Roest Crolius *et al.*, 2000). Moreover, the authors of this study developed a program, Exofish, that can detect at least two third of the human genes, with very few false positives, using the comparison of the translated sequences from both species. It was also used to predict that the human genome contains 20 000 to 30 000 genes in agreement with other studies. Thus, in addition to its importance in understanding vertebrate evolution, the availability of a complete pufferfish genome sequence will also be of great importance for the annotation and characterization of the human genome. In this study, by characterizing a short region of 200 Mb in *Fugu*, the detailed structure of one unknown gene and two other previously uncharacterized genes in human could be defined using the *Fugu* sequence, indicating the importance of this model. Furthermore, many studies have now identified conserved non-coding sequences between *Fugu* and human. One such sequence was identified within a *LIG-2* intron and another one in an *ETV6* intron (Montpetit and Sinnett, 2001). These sequences are more conserved than

the actual coding sequence indicating their functional importance. These sequences may act as transcriptional enhancers (Marshall *et al.*, 1994; Aparicio *et al.*, 1995), although other experiences will be needed to prove it. All these characteristics show that the *Fugu* genome is an essential tool to characterize the human genome.

References

1. Aissani B, Bonan C, Baccichet A and Sinnett D. 1999. Childhood acute lymphoblastic leukemia: is there a tumor suppressor gene in chromosome 12p12.3? *Leuk Lymphoma* **34**: 231-239.
2. Aissani B, Sinnett D. 1999. Fine physical and transcript mapping of a 1.8 Mb region spanning the locus for childhood acute lymphoblastic leukemia on chromosome 12p12. *Gene* **240**: 297-305.
3. Amores A, Force A, Yan YL, Joly L, Amemiya C, Fritz A, Ho Langeland J, Prince V, Wang YL, Westerfield M, Ekker M, Postlethwait JH. 1998. Zebrafish hox clusters and vertebrate genome evolution. *Science* **282**: 1711-1714.
4. Aparicio S, Morrison A, Gould A, Gilthorpe J, Chaudhuri C, Rigby P, Krumlauf R, Brenner S. 1995. Detecting conserved regulatory elements with the model genome of the Japanese puffer fish, *Fugu rubripes*. *Proc Natl Acad Sci U S A* **92**: 1684-1688.

5. Aparicio S, Hawker K, Cottage A, Mikawa Y, Zuo L, Venkatesh B, Chen E, Krumlauf R, Brenner S. Organization of the *Fugu rubripes* Hox clusters: evidence for continuing evolution of vertebrate Hox complexes. *Nat Genet.* 1997 May;16(1):79-83.
6. Baccichet A and Sinnett D. 1997. Frequent deletion of chromosome 12p12.3 in children with acute lymphoblastic leukaemia. *Br J Haematol* **99**: 107-114.
7. Baens M, Wlodarska I, Corveleyn A, Hoornaert I, Hagemeyer A and Marynen P. 1999. A physical, transcript, and deletion map of chromosome region 12p12.3 flanked by ETV6 and CDKN1B: hypermethylation of the LRP6 CpG island in two leukemia patients with hemizygous del(12p). *Genomics* **56**: 40-50.
8. Baxendale S, Abdulla S, Elgar G, Buck D, Berks M, Micklem G, Durbin R, Bates G, Brenner S, Beck S. 1995. Comparative sequence analysis of the human and pufferfish Huntington's disease genes. *Nat Genet* **10**: 67-76.
9. Berger R, Bernheim A, Le Coniat M *et al.* 1986. Abnormalities of the short arm of chromosome 12 in acute nonlymphocytic leukemia and dysmyelopoietic syndrome. *Cancer Genet Cytogenet* **19**: 281-289.
10. Brenner S, Elgar G, Sandford R, Macrae A, Venkatesh B, Aparicio S. 1993. Characterization of the pufferfish (*Fugu*) genome as a compact model vertebrate genome. *Nature* **366**: 265-268.

11. Brown SD, Twells RC, Hey PJ *et al.* 1998. Isolation and characterization of LRP6, a novel member of the low density lipoprotein receptor gene family. *Biochem Biophys Res Commun* **248**: 879-888.
12. Caldas C, Kim MH, MacGregor A, Cain D, Aparicio S, Wiedemann LM. 1998. Isolation and characterization of a pufferfish MLL (mixed lineage leukemia)-like gene (fMll) reveals evolutionary conservation in vertebrate genes related to *Drosophila trithorax*. *Oncogene* **16**: 3233-3241.
13. Chan LC, Kwong YL, Liu HW, Lee CP, Lie KW and Chan AY. 1992. Deletion 12p in de novo acute myeloid leukemia. An association with early progenitor cell. *Cancer Genet Cytogenet* **62**: 47-49.
14. Elgar G, Clark MS, Meek S, Smith S, Warner S, Edwards YJ, Bouchireb N, Cottage A, Yeo GS, Umrania Y, Williams G, Brenner S. 1999. Generation and analysis of 25 Mb of genomic DNA from the pufferfish *Fugu rubripes* by sequence scanning. *Genome Res* **9**: 960-971.
15. Fenrick R, Wang L, Nip J *et al.* 2000. TEL, a putative tumor suppressor, modulates cell growth and cell morphology of ras-transformed cells while repressing the transcription of stromelysin-1. *Mol Cell Biol* **20**: 5828-5839.

16. Gilley J, Fried M. 1999. Extensive gene order differences within regions of conserved synteny between the *Fugu* and human genomes: implications for chromosomal evolution and the cloning of disease genes. *Hum Mol Genet* **8**: 1313-1320.
17. Gilley J, Armes N, Fried M. 1997. *Fugu* genome is not a good mammalian model. *Nature* **385**: 305-306.
18. Golub TR, Barker GF, Lovett M and Gilliland DG. 1994. Fusion of PDGF receptor beta to a novel ets-like gene, tel, in chronic myelomonocytic leukemia with t(5;12) chromosomal translocation. *Cell* **77**: 307-316.
19. Kao HT, Porton B, Hilfiker S, Stefani G, Pieribone VA, DeSalle R, Greengard P. 1999. Molecular evolution of the synapsin gene family. *J Exp Zool* **285**: 360-377.
20. Kibel AS, Freije D, Isaacs WB and Bova GS. 1999. Deletion mapping at 12p12-13 in metastatic prostate cancer. *Genes Chromosomes Cancer* **25**: 270-276.
21. Kimura C, Takeda N, Suzuki M, Oshimura M, Aizawa S, Matsuo I. 1997. Cis-acting elements conserved between mouse and pufferfish Otx2 genes govern the expression in mesencephalic neural crest cells. *Development* **124**: 3929-3941.

22. Marshall H, Studer M, Popperl H, Aparicio S, Kuroiwa A, Brenner S, Krumlauf R. 1994. A conserved retinoic acid response element required for early expression of the homeobox gene *Hoxb-1*. *Nature* **370**: 567-571.
23. McLysaght A, Enright AJ, Skrabanek L, Wolfe KH. 2000. Estimation of synteny conservation and genome compaction between pufferfish (*Fugu*) and human. *Yeast* **17**: 22-36.
24. Meyer A, Schartl M. 1999. Gene and genome duplications in vertebrates: the one-to-four (-to-eight in fish) rule and the evolution of novel gene functions. *Curr Opin Cell Biol* **11**: 699-704.
25. Miles C, Elgar G, Coles E, Kleinjan DJ, van Heyningen V, Hastie N. 1998. Complete sequencing of the *Fugu* WAGR region from WT1 to PAX6: dramatic compaction and conservation of synteny with human chromosome 11p13. *Proc Natl Acad Sci U S A* **95**: 13068-13072.
26. Montpetit A, Sinnett D. 2001. Comparative analysis of the ETV6 gene in vertebrate genomes from pufferfish to human. *Oncogene* **20**: 3437-3442
27. Montpetit A, Boily G, Sinnett D. 2002. A detailed transcriptional map of the chromosome 12p12 tumor suppressor locus. Accepted for publication in *Eur J Hum Gen.*

28. Pohar N, Godenschwege TA, Buchner E. 1999. Invertebrate tissue inhibitor of metalloproteinase: structure and nested gene organization within the synapsin locus is conserved from *Drosophila* to human. *Genomics* **15**: 293-296.
29. Pinson KI, Brennan J, Monkley S, Avery BJ and Skarnes WC. 2000. An LDL-receptor-related protein mediates Wnt signalling in mice. *Nature* **407**: 535-538.
30. Power DM, Ingleton PM, Flanagan J, Canario AV, Danks J, Elgar G, Clark MS. 2000. Genomic structure and expression of parathyroid hormone-related protein gene (PTHrP) in a teleost, *Fugu rubripes*. *Gene* **30**: 67-76.
31. Raimondi SC, Williams DL, Callihan T, Peiper S, Rivera GK and Murphy SB. 1986. Nonrandom involvement of the 12p12 breakpoint in chromosome abnormalities of childhood acute lymphoblastic leukemia. *Blood* **68**: 69-75.
32. Roest Crollius H, Jaillon O, Bernot A *et al.* 2000. Estimate of human gene number provided by genome-wide analysis using *Tetraodon nigroviridis* DNA sequence. *Nat Genet* **25**: 235-238.
33. Rowitch DH, Echelard Y, Danielian PS, Gellner K, Brenner S, McMahon AP. 1998. Identification of an evolutionarily conserved 110 base-pair cis-acting regulatory

- sequence that governs Wnt-1 expression in the murine neural plate. *Development*. **125**: 2735-2746.
34. Rubnitz JE, Pui CH, Downing JR. 1999. The role of TEL fusion genes in pediatric leukemias. *Leukemia* **13**: 6-13.
35. Stegmaier K, Pendse S, Barker GF *et al.* 1995. Frequent loss of heterozygosity at the TEL gene locus in acute lymphoblastic leukemia of childhood. *Blood* **86**: 38-44.
36. Stegmaier K, Takeuchi S, Golub TR, Bohlander SK, Bartram CR and Koeffler HP. 1996. Mutational analysis of the candidate tumor suppressor genes TEL and KIP1 in childhood acute lymphoblastic leukemia. *Cancer Res* **56**: 1413-1417.
37. Takeuchi S, Bartram CR, Miller CW *et al.* 1996. Acute lymphoblastic leukemia of childhood: identification of two distinct regions of deletion on the short arm of chromosome 12 in the region of TEL and KIP1. *Blood* **87**: 3368-3374.
38. Tamai K, Semenov M, Kato Y *et al.* 2000. LDL-receptor-related proteins in Wnt signal transduction. *Nature* **407**: 530-535.
39. Taylor JS, Van de Peer Y, Braasch I, Meyer A. 2001. Comparative genomics provides evidence for an ancient genome duplication event in fish. *Philos Trans R Soc Lond B Biol Sci* **356**: 1661-1679.

40. Trower MK, Orton SM, Purvis IJ, Sanseau P, Riley J, Christodoulou C, Burt D, See CG, Elgar G, Sherrington R, Rogaev EI, St George-Hyslop P, Brenner S, Dykes CW. 1996. Conservation of synteny between the genome of the pufferfish (*Fugu rubripes*) and the region on human chromosome 14 (14q24.3) associated with familial Alzheimer disease (AD3 locus). *Proc Natl Acad Sci U S A* **20**: 1366-1369.
41. Van Rompaey L, Dou W, Buijs A and Grosveld G. 1999. Tel, a frequent target of leukemic translocations, induces cellular aggregation and influences expression of extracellular matrix components. *Neoplasia* **1**: 526-536.
42. Venkatesh B, Gilligan P, Brenner S. 2000. *Fugu*: a compact vertebrate reference genome. *FEBS Lett* **476**: 3-7.
43. Wehrli M, Dougan ST, Caldwell Ket al. 2000. Arrow encodes an LDL-receptor-related protein essential for Wingless signalling. *Nature* **407**: 527-530.

Table 1 Gene identified in the *Fugu* ETV6 locus

| Gene | Human Homologs (Gen Bank acc.nos.) | % identity | Chromosomal location in human |
|--|--|--------------|----------------------------------|
| Timp3 | (NM_00036) | 81 % | 22q13 |
| Syn3 | (NM_003490) | 64 % | 22q13 |
| Unknown (homology to ADP- ribose phosphorylase) | FLJ22693 (NM_022750) C12ORF6 (NM_020367) | 47 % 45 % | 7q33-35 12p13 |
| ETV6 | NM_001987 | 63 % | 12p12.3 |
| Unknown (no homology to known domains/proteins) | BAB71606 (AK057890) LOC128320 (XM_060942) | 71 % 76 % | 3p11-12 chr 1 |
| SLC16A7 | (NM_004731) | 68 % | 12q13 |
| LIG-2 | clone IMAGE:3862624 (BC012380) | 72 % | 12q13 |
| KUB3 | (AF078164) | 70 % | 12q13 |

^aAll localizations were determined by Blast analysis (<http://www.ncbi.nlm.nih.gov/Blast>) on htgs and dbSTS databases or by using the physical chromosome 12p physical map at the Baylor College of Medicine site (<http://www.hgsc.bcm.tmc.edu/>).

Figure 1. Physical map of the *ETV6* locus of *Fugu rubripes*. The direction of transcription of the identified genes, when known, is indicated with an arrow. BACs and cosmids used in this study are indicated under the physical map. The approximate size of the inserts are indicated in parentheses. The position of the Sp6 and T7 extremities of the clones and the presence of *NotI*, *SfiI* and *SrfI* restriction sites are also indicated. The asterisk (*) indicates a putatively non-functional copy of the C12ORF6 gene.

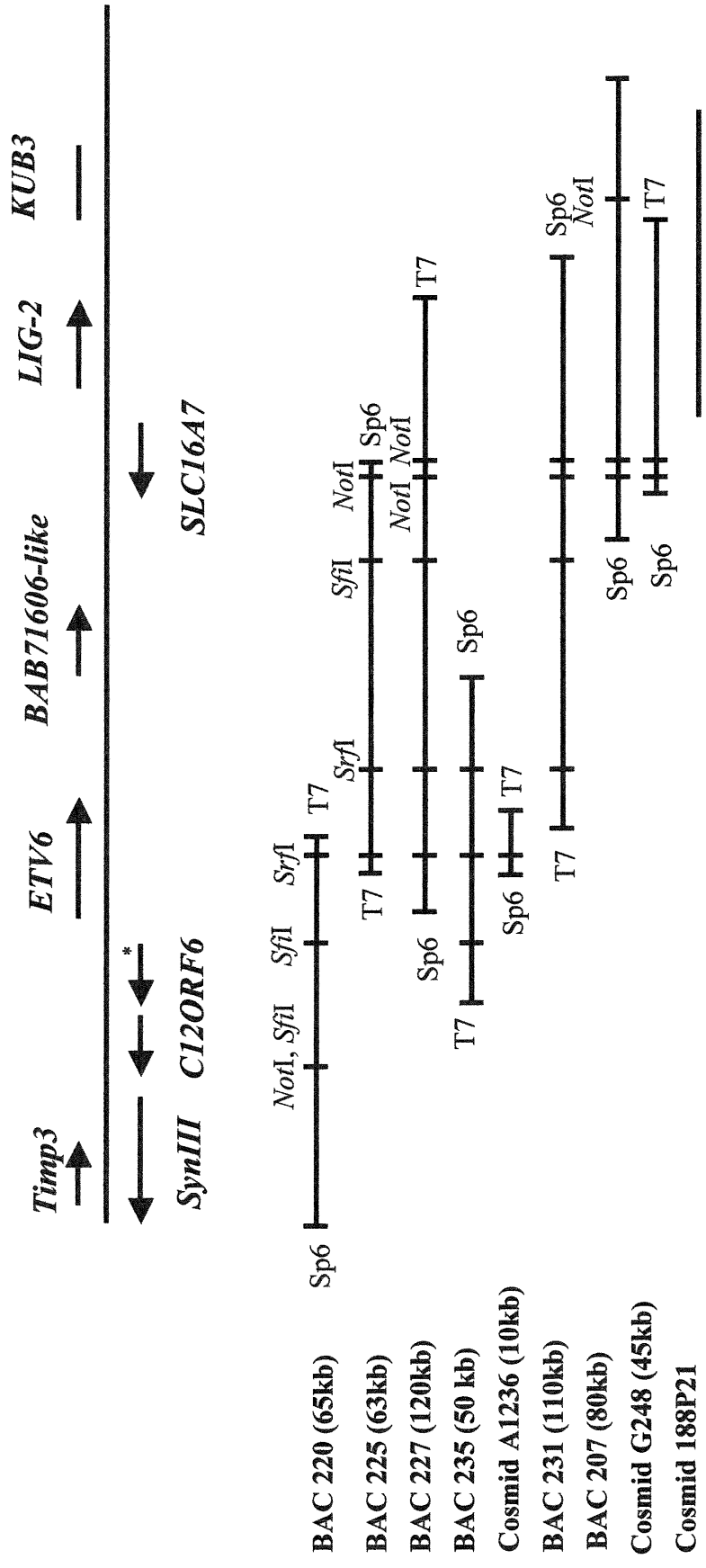


Figure 2. Highly conserved non-coding sequence within a LIG-2 intron. Vertical bars indicate identical nucleotides. Dashes were introduced to maintain the maximal alignment.

Figure 3. Synteny analysis between *Fugu* and the human chromosome 12. The physical relationship between genes present on the ETV6 locus (this study) or on the PTHLH locus (Power *et al.*, 2000) and their homologues on the human chromosome 12 is indicated with straight lines.

Chapitre 3

- Article 2 -

Comparative analysis of the *ETV6* gene in vertebrate genomes from pufferfish to human

Article publié dans *Oncogene* **20**: 3437-3442 (2001).

Contribution des auteurs:

A. Montpetit: a effectué toutes les expériences et toutes les analyses et a écrit l'article.

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

Comparative analysis of the *ETV6* gene in vertebrate genomes from pufferfish to human

Alexandre Montpetit^{1,2} and Daniel Sinnett^{1,2,3}

¹Division of Hematology-Oncology, Charles-Bruneau Cancer Center, Sainte-Justine

Hospital, Montreal, Quebec H3T 1C5, Canada; ²Department of Biochemistry and

³Department of Pediatrics, University of Montreal, Montreal, Quebec, Canada

Keywords: ETV6, transcription factor, comparative genetics, vertebrate genomes, *Fugu rubripes*

Running title: *Fugu rubripes* homologue of the ETV6 gene

Correspondence:

Daniel Sinnett,

Division of Hematology-Oncology,

Sainte-Justine Hospital, 3175 Côte Ste-Catherine,

Montreal, Quebec H3T 1C5, Canada

E-mail: sinnettd@ere.umontreal.ca

Abstract

The *ETV6* gene encodes an Ets-like transcription factor that is frequently rearranged in leukemias. While some of the functions of ETV6 have been uncovered recently, little is known about the key structural elements involved. Comparative genome analysis may provide novel insights into gene evolution and functions. In this study, we cloned and sequenced the homologue of *ETV6* from the compact genome of the pufferfish *Fugu rubripes* (*fETV6*). The genomic structure of the *fETV6* gene was investigated by sequence analysis of a contig of genomic clones. The *fETV6* gene, composed of 8 exons, spans about 15 kb and is 16 times smaller than its human counterpart mainly because of the reduced intron size. Three of the seven introns of the *fETV6* are unusually large (more than 2 kb), including the 8.2 kb intron 2. The gene codes for a protein of 465 amino acids that is highly related to its human homologue, exhibiting an overall identity of 58% (72% similarity). To investigate the functional and evolutionary aspects of ETV6, we undertook a comparative analysis of this gene from various vertebrates (human, mouse, chicken, zebrafish and *Fugu*). As expected, the PNT and ETS domains were highly conserved, with on average 81% and 95% peptide sequence identity, respectively. In addition, we found several new highly conserved regions within the central section of the protein that are likely to represent further functional or structural domains, which may be associated with the transcription repression capacity of this protein. We also found conserved putative regulatory elements in the promoter as well as in the large intron 2 of *fETV6*. The information derived from this comparative analysis will serve as the basis for more precise functional studies of *ETV6* gene regulation and function.

Introduction

ETV6 (also known as TEL) is a member of the Ets family of DNA-binding proteins. It is composed of 8 exons distributed over 240 Kb of genomic DNA (Baens *et al.*, 1996). *ETV6* was originally identified as the 5' partner of PDGFR β in a fusion gene created by a t(5;12) translocation found in a subset of chronic myelomonocytic leukemias (Golub *et al.*, 1994). Subsequently, several translocations involving *ETV6* and various partners have been documented in different hematological malignancies as well as in congenital fibrosarcoma, most of these leading to the formation of chimeric fusion products (reviewed in Rubnitz *et al.*, 1999). Many of the resulting fusion proteins have been shown to transform cells *in vitro* (Carroll *et al.*, 1996; Lacronique *et al.*, 1997; Hannemann *et al.*, 1998; Buijs *et al.*, 2000; Liu *et al.*, 2000; Wai *et al.*, 2000) or to induce hematological malignancies in mice (Tomasson *et al.*, 1999; Schwaller *et al.*, 1998; Carron *et al.*, 2000). Also, allelic loss at the *ETV6* locus on the short arm of chromosome 12 has been observed in various cancers (reviewed in Aissani *et al.*, 1999) suggesting that the human *ETV6* gene lies within a region frequently altered in cancer cells.

ETV6 is a nuclear phosphoprotein that is widely expressed in all normal tissues (Poirel *et al.*, 1997) and plays an important role in angiogenesis (Wang *et al.*, 1997) as well as in the normal development of the hematopoietic system (Wang *et al.*, 1998). Like all Ets proteins, ETV6 binds a specific DNA sequence consisting of a purine-rich GGAA/T core motif within the promoters of the target genes through the highly conserved ETS domain located at its C-terminus (e.g. Donaldson *et al.*, 1996). In

addition, ETV6 can dimerize through its conserved pointed oligomerization domain (termed PNT domain), located at its N-terminus, with itself or with other partners such as UBC9 (Chakrabarti *et al.*, 1999), Fli-1, another Ets member (Kwiatkowski *et al.*, 1998), and TEL2, a protein highly related to ETV6 (Potter *et al.*, 2000; Poirel *et al.*, 2000). Contrary to most of the Ets proteins, ETV6 acts as a transcription repressor by recruiting proteins involved in the histone-deacetylase pathway. The domain responsible for this activity is not well characterized but has been located in the large central section of the protein (Lopez *et al.*, 1999; Chakrabarti and Nucifora 1999).

Given the extensive involvement of Ets factors in normal hematopoiesis and in hematological malignancies, a complete understanding of these complex processes relies on the characterization of known Ets genes such as *ETV6*. We propose that the comparative analysis of *ETV6* in different vertebrate species will lead to the identification of key domains of the protein and regulatory elements that may perform important functions. However, the overall high degree of amino acid identity between the human and mouse proteins precludes the use of this comparison to identify small conserved blocks that could have functional significance. The teleost fish *Fugu rubripes* has been shown to be a useful model genome for comparative sequence analysis by virtue of its small size (~400Mb) and its compact gene organization (Brenner *et al.*, 1993). The greater evolutionary distance between *Fugu* and human can be exploited to determine functional domains, as these regions are likely to show a higher degree of sequence conservation. Hence, this approach has previously led to the discovery of important functional protein domains (Baxendale *et al.*, 1995; Caldas *et al.*, 1998).

Results and Discussion

To identify the *Fugu* homologue of *ETV6*, two gridded genomic *Fugu* libraries, Cosmid library 66, cloned in Lawrist 4 (distributed by RZPD, Berlin, Germany), and a BAC library cloned in pBeloBAC11 (obtained from Genome Systems), were initially screened at low stringency with a radiolabeled cDNA probe corresponding to exons 2-4 of the human *ETV6* cDNA (GenBank: NM_001987). This led to the identification of five positive clones: four BACs (220c16, 225j23, 227h21 and 235e13) and one Cosmid (ICRFc66A1236Q1.3). A strongly hybridizing 1kb *EagI* fragment was present in all five genomic clones tested. Following the subcloning and sequencing of this restriction fragment, BlastN analysis revealed strong homology to the third exon of the human and mouse *ETV6* genes, thus confirming the identification of the *Fugu rubripes* homologue (*fETV6*). Thereafter the entire sequence of the *fETV6* gene with its flanking regions was determined from overlapping genomic clones. The sequence was deposited in GenBank (Accession number AF340230).

The positions and sizes of the coding exons in *fETV6* were determined by comparison with the human and zebrafish gene and by using exon prediction programs to detect splicing donor and acceptor sites (Table 1). The *Fugu* and human *ETV6* genes showed 63% sequence identity at the nucleotide level (data not shown). They are both composed of 8 exons of similar sizes and the positions of the introns are perfectly conserved between the two species (Figure 1A). Exons 3-8 were well predicted by all exon prediction programs and clearly corresponded to the human gene structure. Only the FGene program predicted exons 1 and 2 correctly, although it predicted many other

false positives. It predicted a 27bp exon with an in-frame methionine that could be aligned with *ETV6* exon 1 from other species and a 130bp exon that had strong homology to a zebrafish *ETV6* cDNA at the nucleotide level. As predicted, the introns are much smaller in *Fugu*, ranging from 81 to 8185 bp compared to 1.3 kb to more than 100 kb for the human gene (Figure 1A and Table I). Consequently, the *fETV6* gene (~15 kb long from ATG to Stop codon) is 16 times smaller than its human counterpart (238 kb). In comparison, the *Fugu* G6PD gene is reduced 3.8 fold whereas the APP gene is compacted 30 times compared to their human counterparts (Mason *et al.*, 1995; Villard *et al.*, 1998). The degree of compaction most probably depends on the degree of AT-richness of the corresponding human chromosomal location (Villard *et al.*, 1998). However, in *fETV6*, the introns 1 (2.1kb), 2 (8.1kb) and 3 (2.3 kb) are much larger than the average size of 150bp reported for the *Fugu* genome (Elgar *et al.*, 1996). No major repeated element such as Alu-like and LINE sequences was found in these introns or in the whole *fETV6* gene. The only repetitive elements identified were dinucleotide repeats, in the 5' and 3' flanking regions as well as in intron 2, and GGA and GGGGA repeats in introns 2 and 3, respectively (data not shown). Thus, this may suggest the presence of additional exons, regulatory sequences or even other genes. Blast analysis of these intronic sequences failed to reveal homology with known proteins or the presence of the putative human alternative exon 1b. However, it revealed the presence of a conserved sequence of ~80 bp in both the human and the *Fugu* intron 2 (Figure 1B). No consensus 5' splice donor site or the presence of wobble bases could be identified in or near the homologous sequence thus ruling out the presence of an exon. Further studies will be

required to determine what role this sequence plays in the regulation of the transcription of *ETV6*, but it illustrates the importance of the *Fugu* genome as a comparative model.

We sequenced approximately 2.5 kb upstream of the first *Fugu* exon. The comparison of the promoter region of the *Fugu* and human *ETV6* genes was performed with the TFSEARCH algorithm (<http://pdap1.trc.rwcp.or.jp/research/db/TFSEARCH.html>). Of note the transcriptional start has not been determined experimentally in both genes. The *Fugu* promoter, like its human counterpart, does not seem to contain a TATA box. However, both human and *Fugu* promoters contain consensus binding sites for GATA, STAT, AML-1, MZF1 and Ets factors (data not shown) that are all known to play a role in hematopoiesis or angiogenesis. In contrast, the human promoter contains numerous Sp1 sites in the 5' upstream region but none could be found in the *Fugu* promoter, which could reflect some differences in their respective mechanism of transcription.

Further studies will be required to confirm these preliminary results, but this is consistent with several reports of conserved non-coding sequences, between *Fugu* and human genes, located in promoters, 3' UTR and even introns (Aparicio *et al.*, 1995; How *et al.*, 1996; Kimura *et al.*, 1997; Rowitch *et al.*, 1998; Miles *et al.*, 1998; Gellner and Brenner 1999). Some of them have been shown to have functional significance *in vivo* (Aparicio *et al.*, 1995; Kimura *et al.*, 1997; Rowitch *et al.*, 1998).

An open reading frame could be assembled from the 8 predicted exons encoding a protein of 465 amino acids. The predicted fETV6 protein sequence was aligned against the sequence of other vertebrates (zebrafish, chicken, mouse and human) using ClustalW and manual refinements (Figure 2). Being a teleost like *Fugu*, the zebrafish sequence

was helpful to identify the less-conserved coding regions, but the large evolutionary distance between the two fishes (in the order of a few hundred Myr) adds weight to the significance of the comparative analysis (e.g. Venkatesh *et al.*, 2000). A zebrafish *ETV6* cDNA was obtained from RZPD (WashU clone fc03a10; RZPD Clone ID MPMGp609I1922) and the complete sequence was determined in our laboratory and deposited in GenBank (Accession number AF339838). The overall amino acid sequence identity and similarity between the *Fugu* and human predicted ETV6 proteins is 58% and 72%, respectively. As expected the highest sequence identity was found in the regions corresponding to the PNT (encoded by exons 3 and 4) and ETS (encoded by exons 6 to 8) domains. The PNT and ETS domains show sequence identity of 81% and 95% respectively between human and *Fugu*, and are strongly conserved even in Yan, the *Drosophila* homologue of ETV6 (Figure 2). These domains are crucial for hetero/homo dimerization and DNA-binding through a conserved purine GGAA motif, respectively (e.g. Donaldson *et al.*, 1996; Slupsky *et al.*, 1998).

The central section of the protein, represented by the entire exon 5, is less well conserved. However, several domains termed boxes A, B, and C as well as smaller motifs were shown to be shared by all studied species (Figure 2), suggesting an important function or structural role. These conserved regions would have remained undiscovered if only mouse/human comparisons were performed. We can only speculate about the putative function of these sequences, but it was recently shown that ETV6 acts as a repressor of transcription (Lopez *et al.*, 1999) and that the central domain could mediate this function through the binding of SMRT, mSin3A and NCoR, which are well-known transcriptional repressors (Chakrabarti and Nucifora 1999). The limits of

this putative functional domain were narrowed down to residues 171 to 285 of the human ETV6 protein (Lopez *et al.*, 1999) that encompasses the conserved Boxes A, B and C as well as a putative phosphorylation site (see below).

The activity of several mammalian Ets-like proteins as well as *Drosophila* Yan and Pointed proteins is regulated by the Ras/MAPK pathway through phosphorylation (reviewed in Wasyluk *et al.*, 1998). The human ETS-1, ETS-2, and *Drosophila* Pointed proteins contain a MAPK phosphorylation consensus site (P-X(1-2)-T/S-P) near their pointed domains that are phosphorylated on a conserved threonine (Slupsky *et al.*, 1998; Wasyluk *et al.*, 1998). ETV6 has been shown to be hyperphosphorylated *in vivo* (Poirel *et al.*, 1997). The human ETV6 contains at least three putative consensus phosphorylation sites for MAPK including one located at an equivalent position, upstream of the PNT domain, to the actual MAPK phosphorylation site in ETS-1. It is worth mentioning that it is still unknown whether ETV6 is indeed a target of these protein kinases. By looking at the ETV6 alignment we see clearly that this site is not conserved in *Fugu* (dotted box in Figure 2). However there is another putative MAPK site located in the central section, involving residues 255 to 258 of the human protein, that is conserved across all species. This phosphorylation site may represent a conserved regulatory mechanism in ETV6 proteins. It is important to note that mutation of the first MAPK site did not alter the function of ETV6 indicating that this site is not important or that another site is present in the protein (Van Rompaey *et al.*, 2000). Furthermore, a construction resulting in ETV6 initiated at the second in-frame ATG, thus missing the first MAPK site, could still be phosphorylated in transfected Cos cells (Poirel *et al.*, 1997).

Analysis of the human *ETV6* transcript species suggested the presence of splicing variants: one was shown to have skipped exon 2 (Romana *et al.* 1995) and one that had an alternative exon (termed exon 1b) spliced directly to exon 3 (Baens *et al.*, 1996). Those two transcripts could only use the third in-frame ATG (position 89 of the human protein) as their translation start to generate full-length proteins, but their functional significance is still unknown. In addition, Poirel *et al.*, (1997) suggested that an *ETV6* fast-migrating species (50kDa vs. 57kDa) was generated by its translation beginning at the second in-frame ATG (pos.57 in human). As mentioned earlier, we were unable to detect a sequence with similarity to the human alternative exon 1b in the corresponding *Fugu* intron. This could reflect the lack of function and thus conservation of this sequence (it is a non-coding exon in human) or indicate a mammalian-specific (at least in human) splicing variant. However, the alignment of the *ETV6* proteins reveals that the second and the third in-frame methionine are well conserved in all species except for zebrafish, although in this case another methionine is found 13 residues from the expected second methionine (Figure 2).

In conclusion, the use of *Fugu rubripes*, which has one of the smallest genomes of the vertebrates, is an important strategy in the molecular study of gene functions. Because of the large evolutionary distance between *Fugu* and mammals and because of the almost complete absence of repeated sequences and other "junk" DNA in its genome, only essential, coding and non-coding, features are conserved. In this study, the comparative analysis of *ETV6* has demonstrated conservation of the gene product across some 400 Myr of vertebrate evolution. The identification of novel conserved domains and the delineation of smaller blocks of conserved residues within the *ETV6* predicted

proteins will provide important clues for further functional studies and target mutagenesis experiments. These analyses will be important for the understanding of angiogenesis and hematopoiesis and will give insights into the mechanism of oncogenesis by the ETV6-AML1 fusion protein, present in 25% of childhood pre-B ALL cases, in which the central section of ETV6 is always retained (reviewed in Rubnitz *et al.*, 1999).

Acknowledgements

D.S. is a scholar of the Fonds de la Recherche en Santé du Québec. A.M. has received a studentship from the Medical research Council of Canada. This work was supported by a grant from the Medical Research Council of Canada.

References

- Aissani B, Bonan C, Baccichet A and Sinnott D. (1999). *Leuk. Lymphoma*, **34**, 231-239.
- Aparicio S, Morrison A, Gould A, Gilthorpe J, Chaudhuri C, Rigby P, Krumlauf R and Brenner S. (1995). *Proc. Natl. Acad. Sci. USA*, **92**, 1684-1688.
- Baens M, Peeters P, Guo C, Aerssens J and Marynen P. (1996). *Genome Res.*, **6**, 404-413.
- Baxendale S, Abdulla S, Elgar G, Buck D, Berks M, Micklem G, Durbin R, Bates G, Brenner S and Beck S. (1995). *Nat. Genet.*, **10**, 67-76.
- Brenner S, Elgar G, Sandford R, Macrae A, Venkatesh B and Aparicio S. (1993). *Nature*, **366**, 265-268.

- Buijs A, van Rompaey L, Molijn AC, Davis JN, Vertegaal AC, Potter MD, Adams C, van Baal S, Zwarthoff EC, Roussel MF and Grosveld GC. (2000). *Mol. Cell. Biol.*, **20**, 9281-9293.
- Caldas C, Kim MH, MacGregor A, Cain D, Aparicio S and Wiedemann LM. (1998). *Oncogene*, **16**, 3233-3241.
- Carroll M, Tomasson MH, Barker GF, Golub TR and Gilliland DG. (1996). *Proc. Natl. Acad. Sci. USA*, **93**, 14845-14850.
- Carron C, Cormier F, Janin A, Lacronique V, Giovannini M, Daniel MT, Bernard O and Ghysdael J. (2000). *Blood*, **95**, 3891-3899.
- Chakrabarti SR, Sood R, Ganguly S, Bohlander S, Shen Z and Nucifora G. (1999). *Proc. Natl. Acad. Sci. USA*, **96**, 7467-7474.
- Chakrabarti SR and Nucifora G. (1999). *Biochem. Biophys. Res. Commun.*, **264**, 871-877.
- Donaldson LW, Petersen JM, Graves BJ and McIntosh LP. (1996). *EMBO J.*, **15**, 125-134.
- Elgar G, Sandford R, Aparicio S, Macrae A, Venkatesh B and Brenner S. (1996). *Trends Genet.*, **12**, 145-150.
- Gellner K and Brenner S. (1999). *Genome Res.*, **9**, 251-258.
- Golub TR, Barker GF, Lovett M and Gilliland DG. (1994). *Cell*, **77**, 307-316.
- Hannemann JR, McManus DM, Kabarowski JH and Wiedemann LM. (1998). *Br. J. Haematol.*, **102**, 475-485.
- How GF, Venkatesh B and Brenner S. (1996). *Genome Res.*, **6**, 1185-1191.

- Kimura C, Takeda N, Suzuki M, Oshimura M, Aizawa S and Matsuo I. (1997). *Development*, **124**, 3929-3941.
- Kwiatkowski BA, Bastian LS, Bauer TR Jr, Tsai S, Zielinska-Kwiatkowska AG and Hickstein DD. (1998). *J. Biol. Chem.*, **273**, 17525-17530.
- Lacronique V, Boureux A, Valle VD, Poirel H, Quang CT, Mauchauffe M, Berthou C, Lessard M, Berger R, Ghysdael J and Bernard OA. (1997). *Science*, **278**, 1309-1312.
- Liu Q, Schwaller J, Kutok J, Cain D, Aster JC, Williams IR and Gilliland DG. (2000). *EMBO J.*, **19**, 1827-1838.
- Lopez RG, Carron C, Oury C, Gardellin P, Bernard O and Ghysdael J. (1999). *J. Biol. Chem.*, **274**, 30132-30138.
- Mason PJ, Stevens DJ, Luzzatto L, Brenner S and Aparicio S. (1995). *Genomics*, **26**, 587-591.
- Miles C, Elgar G, Coles E, Kleinjan DJ, van Heyningen V and Hastie N. (1998). *Proc. Natl. Acad. Sci. USA*, **95**, 13068-13072.
- Poirel H, Oury C, Carron C, Duprez E, Laabi Y, Tsapis A, Romana SP, Mauchauffe M, Le Coniat M, Berger R, Ghysdael J and Bernard OA. (1997). *Oncogene*, **14**, 349-357.
- Poirel H, Lopez RG, Lacronique V, Della Valle V, Mauchauffe M, Berger R, Ghysdael J and Bernard OA. (2000). *Oncogene*, **19**, 4802-4806.
- Potter MD, Buijs A, Kreider B, van Rompaey L and Grosveld GC. (2000). *Blood*, **95**, 3341-3348.
- Romana SP, Mauchauffe M, Le Coniat M, Chumakov I, Le Paslier D, Berger R and Bernard OA. (1995). *Blood*, **85**, 3662-3670.

- Rowitch DH, Echelard Y, Danielian PS, Gellner K, Brenner S and McMahon AP. (1998). *Development*, **125**, 2735-2746.
- Rubnitz JE, Pui CH and Downing JR. (1999). *Leukemia*, **13**, 6-13.
- Schwaller J, Frantsve J, Aster J, Williams IR, Tomasson MH, Ross TS, Peeters P, Van Rompaey L, Van Etten RA, Ilaria R Jr, Marynen P and Gilliland DG. (1998). *EMBO J.*, **17**, 5321-5333.
- Slupsky CM, Gentile LN, Donaldson LW, Mackereth CD, Seidel JJ, Graves BJ and McIntosh LP. (1998). *Proc. Natl. Acad. Sci. USA*, **95**, 12129-12134.
- Thompson JD, Higgins DG and Gibson TJ. (1994). *Nucleic Acids Res.*, **22**, 4673-4680.
- Tomasson MH, Williams IR, Hasserjian R, Udomsakdi C, McGrath SM, Schwaller J, Druker B and Gilliland DG. (1999). *Blood*, **93**, 1707-1714.
- Van Rompaey L, Potter M, Adams C and Grosveld G. (2000). *Oncogene*, **19**, 5244-5250.
- Venkatesh B, Gilligan P and Brenner S. (2000). *FEBS Lett.*, **476**, 3-7.
- Villard L, Tassone F, Crnogorac-Jurcevic T, Clancy K and Gardiner K. (1998). *Gene*, **210**, 17-24.
- Wai DH, Knezevich SR, Lucas T, Jansen B, Kay RJ and Sorensen PH. (2000). *Oncogene*, **19**, 906-915.
- Wang LC, Kuo F, Fujiwara Y, Gilliland DG, Golub TR and Orkin SH. (1997). *EMBO J.*, **16**, 4374-4383.
- Wang LC, Swat W, Fujiwara Y, Davidson L, Visvader J, Kuo F, Alt FW, Gilliland DG, Golub TR and Orkin SH. (1998). *Genes Dev.*, **12**, 2392-2402.

Wasylyk B, Hagman J and Gutierrez-Hartmann A. (1998). *Trends Biochem. Sci.*, **23**, 213-216.

Table 1 Intron sizes and exon/intron boundaries of the human and *Fugu ETV6* genes

| # | Position in human cDNA | Human exon/intron junctions | | | Fugu exon/intron junctions | | | Reduction |
|---|------------------------|-----------------------------|------------------|--------------------|----------------------------|------------------|--------------------|-----------|
| | | 5' splice donor | intron size (bp) | 3' splice acceptor | 5' splice donor | intron size (bp) | 3' splice acceptor | |
| 1 | 1-57 | ATTAAG gtaaaa | 102301 | taacag CAGGAA | AGCAAAG gtgaca | 2158 | tcatag CAGGAG | 47X |
| 2 | 58-187 | ACCTGC gtgagt | 86573 | caacag GCTTGC | ACCTGC gtgagt | 8185 | ctgcag GCCTGC | 11X |
| 3 | 188-352 | ATTCAG gtgaga | 14107 | ttccag GTGATG | ATTCCG gtaatc | 2325 | g'gcag GCGACG | 6X |
| 4 | 353-487 | AAGAAg gtactg | 15856 | ccacag ATAAct | TTGAAG gtcagt | 592 | cctcag AAACGG | 27X |
| 5 | 488-1033 | TAGCAG gtgagt | 11456 | ctgtag ACTGTA | TCGCAG gtagga | 119 | ccccag ACTGCA | 96X |
| 6 | 1034-1176 | CATAAG gtaaaa | 1338 | aaacag AACAGA | CACAAG gtgaag | 82 | ttgcag AACAGG | 16X |
| 7 | 1177-1277 | GTTCAG gtagca | 4924 | ccaaag GTTTAT | ATTCAG gtagat | 81 | cttcag GTTCAT | 61X |
| 8 | 1278- | | | | | | | |

Position in cDNA and human exon/intron junctions are taken from Baens *et al.* (1996). Intron sizes were calculated from GenBank Accession numbers U81830, U81831, U81832, U81834, AC005989 and NT_000601 forming the *ETV6* contig. The exon/intron junctions of *fETV6* were deduced by using NIX program (<http://www.hgmp.mrc.ac.uk/>) which uses several different exon prediction programs.

Figure 1 Comparison of the human and *Fugu rubripes ETV6* genomic structures. **(a)** Filled boxes represent coding exons (numbered 1 to 8) from the human (top) and *Fugu* (bottom) *ETV6*. The non-coding human alternative exon 1A is represented as an empty box. The position of a conserved motif between human and *Fugu* genes in intron 2 is represented by an empty triangle. The sizes of each exon (in bp) and introns (in kb) are indicated above and below each map, respectively. **(b)** BlastN analysis (<http://www.ncbi.nlm.nih.gov/BLAST/>) of non-coding sequence from *ETV6* intron 2 from human (pos.13664-13744; GenBank: AC005989) and *Fugu* (pos. 2856-2935; GenBank: AF340230).

Figure 2 Alignment of the predicted ETV6 amino acid sequence from various vertebrates. Human (Swissprot: P41212), mouse (Swissprot: P97360), chicken (partial; GenBank: AAC97200), zebrafish (GenBank: AF339838) and *Fugu* (GenBank: AF340230) ETV6 sequences were aligned using the ClustalW algorithm (Thompson *et al.*, 1994) and manually edited using the SeaView program (<http://pbil.univ-lyon1.fr/software/seaview.html>). Alignments were displayed using the BOXSHADE program (http://www.ch.embnet.org/software/BOX_form.html). *Drosophila* PNT and ETS sequences from Yan protein (GenBank: A43315) are shown for comparison. Dashes indicate gaps introduced to maintain optimal alignment. The corresponding exons (numbered from 1 to 8) in the genomic sequence are indicated at the top. BOX A, B, C: regions of strong conservation in the central section (see text). MAPK: consensus phosphorylation site for MAP kinase. Dashed box indicates MAPK putative sites present in human but not conserved across all species studied (see text). Positions in the human protein are indicated on the right.

Chapitre 4

- Article 3 -

A detailed transcriptional map of the chromosome 12p12 tumour suppressor locus

Article publié dans *Eur J Hum Gen* **10**: 62-71 (2002)

Contribution des auteurs:

A. Montpetit: a effectué la majorité des expériences, toutes les analyses et a écrit l'article

G. Boily: a effectué quelques expériences de RT-PCR et de séquençage.

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

A detailed transcriptional map of the chromosome 12p12 tumour suppressor locus

Alexandre Montpetit^{1,2}, Gino Boily¹ and Daniel Sinnett^{*,1,2,3}

¹Division of Hematology-Oncology, Charles-Bruneau Cancer Center, Research Center, Sainte-Justine Hospital, 3175 Côte Ste-Catherine, Montreal, QC H3T 1C5, Canada;

²Department of Biochemistry and ³Department of Pediatrics, University of Montreal, Montreal, QC, Canada.

*Correspondence: Dr Daniel Sinnett, Charles-Bruneau Cancer Center, Sainte-Justine Hospital, 3175 Côte Ste-Catherine, Montreal, QC H3T 1C5, Canada. Tel: (514) 345-2990; Fax: (514) 345-4731; E-mail: daniel.sinnett@umontreal.ca.

Running title: Transcription map of the chromosome 12p12 locus

Abstract

Loss of heterozygosity of the short arm of chromosome 12 is a frequent event in a wide range of haematological malignancies and solid tumours. In previous studies, the shortest commonly deleted region was delimited to a 750-kb interval, defined by the markers *DI2S89* and *DI2S358*, in pre-B acute lymphoblastic leukemia patients, suggesting the presence of a tumour suppressor locus. Here we report the construction of a transcriptional map that integrates the data obtained by genomic sequence analysis, EST database search, comparative analysis and exon amplification. We identified seven putative transcriptional units as well as six pseudogenes. Four of these candidate genes were already known: *ETV6*, encoding an ets-like transcription factor, *LRP6*, a member of the LDL receptor gene family, *BCL-G*, a recently identified pro-apoptotic gene and *MKP-7*, encoding a new member of the dual-specificity phosphatase family. The products encoded by the three new genes identified in this study, *LOH1CR12*, *LOH2CR12* and *LOH3CR12*, have no clear homology to known proteins. The gene predictions were all confirmed by expression analysis using RT-PCR and Northern blot. This transcriptional map is a crucial step toward the identification of the tumour suppressor gene at 12p12.

Key Words: Tumour suppressor gene, leukemia, chromosome 12, transcriptional units, gene expression.

Introduction

The short arm of chromosome 12 is frequently rearranged in a wide variety of haematological malignancies of both myeloid and lymphoid origins.¹⁻³ Most of these 12p alterations result from unbalanced translocations and deletions that lead to a loss of genetic material.⁴ The frequent loss of genetic material in tumour cells is usually indicative of the inactivation of tumour suppressor genes, which prompted the search for a suppressor locus on chromosome 12p. Loss of heterozygosity studies showed hemizygous deletions of chromosome 12p12 in 26 to 47% of pre-B acute lymphoblastic leukemia (ALL) cases, making it one of the most common genetic alterations found in this disease.⁵⁻⁷ The construction of both high-resolution genetic and physical maps led to the delineation of the shortest commonly deleted region within a 750kb interval delimited by markers *D12S89* and *D12S358*.⁸⁻¹⁰ The existence of a putative 12p12 tumour suppressor gene is further substantiated by the observation of hemizygous deletions in a variety of haematological malignancies, as well as in certain solid tumours including breast, lung, ovarian and prostate carcinomas.^{11, 12}

This region was shown to contain two known genes, *LRP6* and *ETV6*. *LRP6* encodes a member of the LDL receptor family¹³ and was recently shown to act as the WNT co-receptor.¹⁴⁻¹⁶ However Baens *et al.*¹⁰ failed to reveal any inactivating mutations in leukemia patients. *ETV6*, also known as TEL, is a member of the ets-like family of DNA binding proteins, that was originally identified by virtue of its fusion to the platelet-derived growth factor receptor β (PDGFR β) in chronic myelomonocytic leukemia with the translocation t(5;12) (q33;p13).¹⁷ Recently, it has been shown that *ETV6* acts as a transcriptional repressor of several matrix metalloproteinases and that its

expression could counteract the transformation of cultured cells by *ras*.^{18,19} However, mutational analysis failed to detect deleterious mutations in the second *ETV6* allele of ALL patients with hemizygous deletion at this locus^{6,20} as expected for a classical tumour suppressor gene. These observations taken together led to the suggestion that an as yet unidentified suppressor locus might be located between *D12S89* and *D12S358*.¹¹

Here, we describe the application of a strategy that integrates the use of data mining tools, exon amplification experiments, gene expression studies and comparative genomic analysis to identify putative transcriptional units in the chromosome 12p12 locus. This strategy led to the identification of 7 genes, including 3 uncharacterized genes, and 6 pseudogenes, indicating the usefulness of integrating different complementary genome-based approaches to identify candidate genes in a large chromosomal interval.

Materials and methods

Genomic clones BAC clones RP11-525I13, RP11-267J23, RP11-757G14 and RP11-253I19 (GenBank accession numbers AC022222, AC007537, AC007621 and AC007619 respectively) were obtained from Incyte Genomics, Inc (St-Louis, Mo). The sequence of these BACs, of BAC RPC111-180M15 (GenBank accession number AC008115) as well as the whole *ETV6* contig (GenBank accession numbers NT_000600, AC005989 and NT_000601) were selected for analysis based on their localization within or near the shortest commonly deleted region according to the chromosome 12 physical map at the Human Genome Sequencing Center, Baylor College of Medicine (<http://www.hgsc.bcm.tmc.edu>).

Sequence analysis The genomic sequences available from the selected BAC clones were used as queries for BLAST searches and analysed using the NIX software (<http://www.hgmp.mrc.ac.uk/NIX>) that displays the results from 15 different prediction algorithms (Grail-CpG, Grail-exons, FGenes, Hexon, FEX, GenScan, etc) and BLAST analysis on many public databases (dbEST, Unigene, htgs, etc.). The program Exofish (<http://www.genoscope.cns.fr/externe/tetraodon>) was used to identify the presence of conserved regions at the protein level between the human sequences and the genome of *Tetraodon nigroviridis*. All other sequence analysis were performed using BLAST or CD-Search programs on the NCBI site (<http://www.ncbi.nlm.nih.gov>). Multiple sequence alignments were obtained using the ClustalW algorithm²¹ and manually edited using the SeaView program (<http://pbil.univ-lyon1.fr/software/seaview.html>). Alignments were displayed using the BOXSHADE program (http://www.ch.embnet.org/software/BOX_form.html).

Exon amplification Exon amplification was performed using the Exon Trapping System (GibcoBRL-Life Technologies, Burlington, Ontario, Canada) according to the manufacturer's recommendations. Briefly, BAC clones were partially digested with Sau3A. The resulting restriction fragments ranging from 2.5 to 4kb were isolated from 1% agarose gels and subcloned into BamH1-digested pSPL3. Individual subclones were pooled and transfected into 80%-confluent COS-7 cells. Total cellular RNA was extracted 24h after transfection and used as a template for RT-PCR using vector-specific

primers. The fragments containing putative trapped exons were subcloned into pAMP10 and sequenced using T7 primer.

5'- Rapid amplification of cDNA ends (RACE) Human testis Marathon-Ready cDNA (Clontech, Palo Alto, CA) was used to amplify the transcript 5' ends according to the manufacturer's instructions. Briefly, 200 ng of Marathon-ready cDNA was amplified with AP1 primer and a gene-specific primer: 11962RACE (CTGTCCATGACTCATTGCTGAACA), 11962RACE2 (AGGTGGTTAGTTCTGCCAGGCGAGT) and LOH2RACE (CCTTTAAGGTGGCATCGCCCTAAGT) and LOH2RACE2 (CTAATCCGGAAGTGAAGGTCTTGCG) by Touchdown PCR. The resulting RT-PCR products were analysed on 1% agarose gel, transferred to nylon membranes (Hybond N+, Amersham-Pharmacia Biotech, Inc, Baie d'Urfé, Québec, Canada) and hybridised with a radioactive internal oligonucleotide to assess the specificity of the bands. The sequence of the putative 5'end-derived exons was determined by dideoxy sequencing using the Thermo Sequenase Radiolabelled Terminator Cycle Sequencing kit (USB, Cleveland OH) following the subcloning of the PCR products into pGEMT-easy (Promega, Montreal, Canada).

RNA expression analysis by RT-PCR Normal human brain, testis and prostate RNAs were obtained from Clontech (Palo Alto, CA). Placenta, peripheral blood and bone marrow RNAs were isolated from healthy subjects using standard procedures. Following DNase-treatment, the RNA samples were incubated at 37°C for 60 minutes in the

presence of 15 μM of random hexamers, 0.2 mM of each dNTPs and 200 U of MMLV reverse transcriptase (GibcoBRL-Life Technologies, Burlington, Canada). The cDNAs were amplified using 0.2 mM dNTP, 0.4 μM each primer and 2 U of ELONGase polymerase (GibcoBRL-Life Technologies, Burlington, Canada) in a 60 mM Tris-SO₄ pH 9.0 buffer containing 18 mM of (NH₄)₂SO₄, 8% of DMSO and 1.5 mM of MgSO₄.

The primers amplifying the entire coding regions were 11962F (GTGGGTGTGACCTGGAAGAAAT) and 11962R (TGCACAGGACCACAGTAGACA) for the *BCL-G_L* gene; 97818F (CAAATTGTTGAGCTGCTGAAAT) and 97818R (GGGGAAGTGCAGAAGTAGGTC) for the *BCLG_M* gene; DUSPF (CTCGAGGGTGGGAAAAGAGGACTTATTG) and DUSPR (GGATCCTTTCAGATTTACAGGGAATTTTT) for the *MKP-7* gene; LOH1F (CTCGAGCCGCCGTTCTTCTGCTG) and LOH1R (GGATCCCCTTAGCTCCACACGTCCTC) for the *LOH1CR12* gene; LOH2F (GGCTACGCAAGACCTTCAGTT) and LOH2R (ACTGCTTAGTGGGCCCTTCCCAGTC) for the *LOH2CR12* gene; FLJF (CTCGAGCCTTGACCTTTGAAGACCAAAA) and FLJR (AAGCTTAGAGACACCGAGTTCCATCC) for the *LOH3CR12* gene. The cycling conditions were as follows: initial denaturation of 2 min at 94°C, 35 cycles of 94°C for 30 s, 60-64°C for 30s, and 72°C for 60 s, and a final extension at 72°C for 10 minutes. RT-PCR products were revealed by 1 or 2% agarose gel electrophoresis. Then they were excised from the gel, cloned into pGEMT-easy (Promega, Montreal, Canada) and sequenced completely.

Northern blot hybridisation Gel-purified RT-PCR products (see above) were radiolabelled by random priming and used as probes to hybridise multiple Tissue Northern Blots (Human MTN Blot I, Human MTN Blot II and Human Immune System MTN Blot II, Clontech, Palo Alto, CA). The hybridisations were carried out at 50°C in a solution containing 50% formamide, 10% dextran sulfate, 1% SDS, 1M NaCl, 50 µg/mL fish sperm DNA and 5×10^6 cpm/ml of a given probe. Following hybridisation, the filters were washed at high stringency with 0.1 X SSC, 0.1 % SDS at 55°C and autoradiographed at -80°C with X-OMAT Kodak films.

Results

Sequence analysis

The availability of BAC sequences within the chromosome 12p12 suppressor locus, as part of the human genome sequencing project, provided the molecular framework to search for transcribed sequences. Four BAC clones, RP11-525I13, RPCI11-267J23, RP11-757G14 and RP11-253I19, mapped between markers *D12S89* and *D12S358*. Only partial and unordered sequences were available for clones 525I13 and 253I19 at the time of the writing of this paper. The genomic gap between the marker *D12S89* and the BAC 525I13 (Figure 1) was filled by using the data from the complete sequence of the *ETV6* locus.²² This genomic sequence was analysed with the NIX program to identify putative transcribed sequences. We identified 6 pseudogenes based on the absence of introns and the presence of stop codons within the consensus coding sequence. In addition, we found 4 CpG islands and over 90 EST clusters, 38 of them

corresponding to Unigene entries. Many of these ESTs appear to be artefacts as suggested by their low representation (i.e. either unique or only two members) or by the presence of a polyA tail at their 3' end in the corresponding genomic sequence without any upstream polyadenylation signal. This could have allowed oligo dT priming during RT-PCR on mRNA, but also on genomic DNA or on unspliced RNA. To restrict the number of ESTs to analyse, they were classified according to the following characteristics: 1) presence of spliced sequence, 2) homology with genes/ESTs or proteins, when translated, from other species, 3) evolutionarily conserved as shown by Exofish analysis, 4) matches with trapped exon(s) and 5) presence of a polyadenylation signal at the 3' end of the cluster. In this study, only the EST clusters that fulfill at least one of the above characteristics were further investigated.

The EXOFISH algorithm compares genomic sequences to 150 Mb of *T.nigroviridis* sequence (corresponding to 33% of its genome) at the protein level. This algorithm has been successful in detecting two thirds of the known genes on human chromosome 22 with an estimated specificity of over 90%.²³ In this study, 33 hits (termed ecores) corresponded to four EST clusters (Table 1) and five pseudogenes (data not shown), but to no other regions. The exon trapping technique identified seven exons corresponding to three of the clusters found in the region analysed (Table 1). Six other trapped sequences did not reveal any match in the EST database, but the absence of a translation frame may suggest false positives rather than actual exons. Applying all these criteria, only 20 clusters were apt for further analysis (Table 1). None of the other clusters analysed contained an ORF that could encode a protein. Of the 20 clusters selected, eight contained spliced sequences and were considered the most interesting.

Except for two clusters, *BCL-G_M* and *LOH2CR12*, they all had clear orthologs from other species in the dbEST. Unigene cluster Hs.97818 corresponded clearly to an alternative splicing form of the *BCL-G* gene (see below). The physical relationship of these predicted genes within the commonly deleted region is presented in Figure 1. To be more informative, we have extended the map on the proximal side in order to include the known genes *CREBL2*, *GPR19* and *CDKN1B*. No other putative transcriptional units were present on the corresponding BAC clone RPC111-180M15 (data not shown). Two of the clusters corresponded to the *ETV6*²² and *LRP6*¹³ genes that were previously characterized in detail and mapped to this region^{13, 22} and as such will not be further discussed in this section.

By using the assembled ESTs and exon prediction programs it was possible to obtain an open reading frame (ORF) starting with a methionine for all the remaining selected clusters except for *LOH2CR12* (Figure 2). The region encompassing the putative ORFs of all the clusters could be amplified by RT-PCR in multiple tissues, confirming the gene prediction (Table 2). Furthermore, many alternatively splicing species could be observed by this analysis (see below).

Characterization of the candidate genes

The human *BCL-G* gene was originally characterized as composed of 6 exons that encode at least two known splicing variants, *BCL-G_L* (long form, 327 amino acids) and *BCL-G_S* (short form, 252 amino acids) isoforms.²⁴ The BH3 domain, one of four BCL-2 homology domains (BH domain), is found in both isoforms and confers pro-apoptotic activity. The longer variant also possesses a BH2 domain, which negatively

regulates this activity.²⁴ RT-PCR (Table 2) and Northern blot (data not shown) analyses revealed that while *BCL-G_S* is only expressed in the testes, *BCL-G_L* is expressed in many tissues including bone marrow, prostate, pancreas, colon, but predominantly in the testes, which is in agreement with Guo *et al.*²⁴ Our Blast analyses and 5'RACE experiments suggest the presence of seven additional exons (Figure 2). The first four 5' upstream exons (exons 1A to 1D) are non-coding and various combinations of these exons were observed in the dbEST. Alternative splicing involving the four additional 3' exons (exons 7 to 10) could generate a new isoform, termed *BCL-G_M* for median, whose 276 amino acid product would also lack box BH2. By RT-PCR, *BCL-G_M* was shown to be expressed only in the testes (Table 2).

MKP-7 has 7 exons (Figure 2) and codes for a new member of the dual-specificity phosphatase family that dephosphorylates MAP kinases.²⁵ This gene shows 55% sequence identity with DUSP8/hvh-5, a closely related dual-specificity phosphatase.²⁵ Northern blot hybridisation with a *MKP-7* probe revealed the expression of two mRNA species of 4.0 and 6.0 kb in all tissues tested (Figure 3). The detection of weakly hybridising 3.0 and 5.5 kb bands in some tissues, especially in the brain, could be due to cross-hybridisation with the related DUSP8/hvh-5 species, which correspond to the observed size and distribution of this mRNA.²⁶ By RT-PCR, we also observed a splicing variant, in which exon 4 is skipped and that is present in every tissue analysed by RT-PCR, although expressed at lower levels (Table 2). The predicted protein would be truncated and lack its phosphatase domain, keeping only the conserved cdc25 domain putatively involved in protein-protein interactions.²⁷

LOHICR12 has 4 exons (Figure 2) that encode a predicted 195 amino acid

protein that shows high homology with hypothetical proteins from distant species: mouse BAB25030 (95% identity), *D.melanogaster* CG11802 (35% identity and 59% similarity) and *C.elegans* F59E12.11 (27% identity and 55% similarity) (Figure 4). The function of these proteins is not known, but they have some homology to one of the spectrins repeats of the TRIO protein and to an uncharacterized region from Mekk4 protein (Figure 4). Three major mRNA species at 1.2, 1.4 and 4.5 kb were detected in all tissues tested by Northern blot (Figure 3). Bands at 1.7 and 6kb were also detected at lower levels. RT-PCR and sequencing experiments revealed the presence of a splicing variant skipping exon 2, thus explaining the 1.2 and 1.4kb species. Interestingly this variant was not observed in the SJNB-7 cell line (Table 2). Both bands at 4.5 and 6 kb could be explained by the alternative use of termination sites (see below).

The *LOH3CR12* gene is constituted of 5 exons (Figure 2). The function of this gene product is unknown, but it possesses low homology to the syndecan domain, which is characteristic of membrane heparan sulfate proteoglycans (Figure 5). Blast analysis also indicated low homology to hepatocyte growth factor activator inhibitor protein and to glycoprotein 1a (data not shown). Homologues from various species were observed, including the mouse RIKEN cDNA 9130403P13 with 62% identity and 66% similarity at the protein level (Figure 5). We have also found a transcript variant (GenBank accession number AK023622) that is characterized by the presence of the 102 bp coding intron 2 that is translated in phase when left unspliced (hatched box in Figure 2). Both transcripts were present in every tissue tested by RT-PCR, the longest being the most abundant (data not shown).

The definitive structure of the *LOH2CR12* gene is still unknown at this moment:

only two exons were identified from the human EST database. All those ESTs, as well as the most upstream sequence obtained by 5'RACE analysis, stopped in a GC-rich region corresponding to the *LOHICR12* CpG island (Figure 6). By Blast analysis, a 500-bp mouse genomic fragment (clone Ti no.10921056) revealed clear homology to 5' exons from both *LOHICR12* and *LOH2CR12*. A common ORF could be constructed for *LOH2CR12*, although the start codon was not observed and is thus presumably located in another unidentified exon (Figure 6). No homology was found with proteins from other species. RT-PCR and sequencing experiments confirmed the gene prediction and revealed very low levels of expression were observed in the bone marrow, prostate and testes. It also revealed that its 3' end corresponded to Unigene cluster Hs.67553 (Table I). The weak expression could explain the absence of signal in the Northern blot analysis (data not shown).

Alternative termination

In previous studies, *ETV6* was shown to use alternative termination sites.²² Three of the EST clusters found in our study, Hs.169081, Hs.293972, Hs.146381, clearly correspond to the 3' ends of the *ETV6* transcripts of 2.4, 4.3 and 6.2 kb, respectively, when placed in the genomic context (Table 1). *LRP6* possesses two transcripts of 6 and 11 kb.¹³ Using the same reasoning, it is clear that the 3' end of clusters Hs.23672, which includes the coding sequence of *LRP6*, and Hs.41269, containing over 90 ESTs, correspond to the 3' end of the short and long transcripts, respectively. All the other identified transcribed sequences, with the exception of *LOH2CR12*, were characterized by at least two different mRNA species by Northern hybridisation (Figure 3) suggesting

the presence of alternative splicing or alternative usage of termination sites. The latter is supported by the identification of large Unigene clusters with consensus 3' polyadenylation sites (but without coding potential), that could explain the 6 kb- *LOH3CR12*, 6 kb-*MKP-7* and 4.5- and 6 kb- *LOH1CR12* mRNAs respectively (see Table 1 and Figures 2 and 3). RT-PCR and sequencing experiments confirmed the existence of the longer transcripts for each gene (data not shown).

CpG islands

The presence of a CpG island usually indicates the 5' extremity of an housekeeping gene. The analysis of the genomic sequences with Grail/CpG predicted 4 CpG islands in the *D12S89-D12S358* interval. Three of them seem to be associated with the 5' ends of *LRP6*, *LOH3CR12* and *LOH1CR12*. *ETV6* also possesses a 5' CpG island,²² which is located outside the interval analysed (Figure 1). As predicted, all of them are expressed in every tissue analysed in contrast with *BCL-G* and *LOH2CR12* whose expression is more restricted (Table 2). The 5' end of *MKP-7* could not be identified since it lies in a gap between the BACs 253I19 and 180M15 (Figure 1), but Baens *et al.*,¹⁰ placed a CpG island in that region. The other CpG island is associated with the pseudo-prothymosin gene and probably does not reflect the presence of a transcribed gene. Indeed, it has been shown that this CpG island is methylated *in vivo*.¹⁰

Discussion

Here we report the construction of a detailed transcriptional map of the putative chromosome 12p12 tumour suppressor locus using a combination of different

approaches (comparative analysis, exon trapping and gene prediction algorithms). We identified seven distinct candidate transcriptional units, including *ETV6*, *LRP6*, *BCL-G* and *MKP-7* that were characterized previously.^{13,22,24,25} *LOH1CR1*, *LOH2CR12* and *LOH3CR12* are three novel genes characterized in this study.

ETV6 is a nuclear phosphoprotein that is widely expressed in all normal tissues²⁸ and plays an important role in angiogenesis²⁹ as well as in the normal development of the hematopoietic system.³⁰ Unlike most Ets-like proteins, *ETV6* acts as a transcriptional repressor recruiting proteins involved in the histone-deacetylase pathway.^{31,32} It can also block the transformation process induced by the overexpression of *ras* in cultured cells.¹⁹ *LRP6* is a member of the LDL-receptor family that has been recently shown to be an essential component of the Wnt pathway.¹⁴⁻¹⁶ This pathway is involved in many different cancers and developmental defects in humans.³³ *BCL-G* is a new member of the BCL-2 family that possesses proapoptotic activity.²⁴ *BCL-G* has multiple splicing variants generating three proteins of different lengths. *BCL-G_L* contains both conserved boxes BH3 and BH2, whereas *BCL-G_S* lacks box BH2. The latter is a more potent apoptosis inducer than the longer variant indicating that Box BH2 has an inhibitory role.²⁴ We found a third variant, *BCL-G_M*, that also lacks box BH2 thus predicting a function similar to *BCL-G_S*. Considering their function, *ETV6*, *LRP6* and *BCL-G* are good candidate tumour suppressor genes, but the failure to detect any inactivating mutations in these genes in leukemia patients does not support a role as classical tumour suppressor genes.^{10,20,24}

MKP-7 is a new member of the dual-specificity ser/thr and tyr phosphatases. Phosphatases play important roles in the regulation of intracellular signalling and have

been associated with either oncogenic or tumour suppressor activity.³⁴ For instance, PTEN, another dual-specificity phosphatase, is often inactivated in human cancers, particularly endometrial carcinomas and glioblastomas, and germline mutations have been reported in patients affected with either of these two tumour predisposing syndromes.³⁵ MKP-7 was recently shown to bind and inactivate p38 MAPK and JNK/SAPK involved in the transmission of cellular growth signals to the nucleus.²⁵

LOH1CR12 and LOH3CR12 have no strong sequence similarity with proteins of known function, but some speculations can be made. LOH1CR12 has homologues at least in mouse, *C.elegans* and *D.melanogaster* and Blast analysis revealed some sequence similarities at the protein level with a region encompassing one of the spectrin repeats of TRIO and other closely related proteins. TRIO is a guanine nucleotide exchange factor that regulates the actin cytoskeleton organization, cell motility and cell growth via activation of Rho GTPases.^{36,37} Spectrin repeats are usually found in proteins associated with the cytoskeleton. Although the specific function of LOH1CR12 cannot be inferred, the fact that it is conserved throughout evolution points to an important role for this protein. Homologues of protein LOH3CR12 are present in various mammals including the mouse, but no homologues have been identified in lower vertebrates. It possesses a syndecan domain and has also low homology with glycoprotein 1a and hepatocyte growth factor activator inhibitor. All of them are membrane-bound or transmembrane proteins. Syndecans are transmembrane heparan sulfate proteoglycans that play an important role in cell-matrix and cell-cell interactions.³⁸ Finally, the structure of *LOH2CR12* is still incomplete and the characterization of the remaining 5' exons will be necessary to infer a putative function for its encoded protein.

Previous expression analysis has demonstrated that *ETV6* and *LRP6* are present in a wide range of tissues.^{13,28} In this study, we showed that *LOH1CR12*, *LOH3CR12* and *MKP-7* also have a broad range of expression. Only *BCL-G* and *LOH2CR12* have a more restricted pattern of expression. *BCL-G* variants, *BCL-G_S* and *BCL-G_M* are only expressed in the testes while *BCL-G_L* is expressed predominantly in the testes and also at lower levels in other tissues including prostate, pancreas and the bone marrow, which is consistent with the work done by Guo *et al.*²⁴ The expression of *LOH2CR12* was observed only by RT-PCR analysis in certain tissues including testes, prostate and bone marrow. The fact that all the genes in the commonly deleted region from ALL patients are expressed in the bone marrow makes them interesting tumour suppressor gene candidates.

The observation that more EST clusters were identified than the number of genes in the region was puzzling at first. One explanation is the fact that most of the genes identified in our study seem to use alternative termination sites, giving rise in the database of multiple independent clusters, since most of the ESTs are primed at their 3' end. Whether the latter phenomenon is specific to the genomic context of this locus is not known, but many examples of alternative termination exist in the literature.³⁹ The exact role of the alternative 3'UTRs length is unclear since it doesn't affect the coding sequence, but it could be associated with the relative mRNA stability. For instance, the gene eIF-2a uses at least two polyA sites generating two common transcripts of 1.6 and 4.2 kb.⁴⁰ The smaller transcript is more readily translated *in vitro* while the longer transcript is more stable. The relative ratios of these transcripts vary with each tissue. The majority of the other clusters, if not all, correspond to artefacts that could be

explained by internal or genomic polyT priming during the reverse transcription reaction. Only two small clusters in our analysis, represented by only one or two ESTs, could not be ruled out as artefacts using our criteria, indicating that we probably identified all the genes in the locus.

The construction of a transcription map of this locus and the identification of 7 candidate genes provide the molecular framework to identify the tumour suppressor gene(s) at 12p12, a locus frequently deleted in haematological malignancies as well as in many different solid neoplasias.

Acknowledgements

We thank Zeina Saikali for critical reading of this manuscript. This work was supported by the Canadian Institutes of Health Research (CIHR). Daniel Sinnott is a scholar of the Fonds de la Recherche en Santé du Québec (FRSQ). Alexandre Montpetit and Gino Boily have received studentships from the CIHR and from National Science and Engineering Research Council, respectively.

Sequence data from this article have been deposited with the EMBL/GenBank Libraries under accession numbers AY040274, AY037865, AY037866, AY037867 and AY038927.

References

- 1 Berger R, Bernheim A, Le Coniat M *et al.* Abnormalities of the short arm of chromosome 12 in acute nonlymphocytic leukemia and dysmyelopoietic syndrome. *Cancer Genet Cytogenet* 1986; **19**: 281-9.
- 2 Raimondi SC, Williams DL, Callihan T, Peiper S, Rivera GK and Murphy SB. Nonrandom involvement of the 12p12 breakpoint in chromosome abnormalities of childhood acute lymphoblastic leukemia. *Blood* 1986; **68**: 69-75.
- 3 Chan LC, Kwong YL, Liu HW, Lee CP, Lie KW and Chan AY. Deletion 12p in de novo acute myeloid leukemia. An association with early progenitor cell. *Cancer Genet Cytogenet* 1992; **62**: 47-9.
- 4 Raimondi SC. Current status of cytogenetic research in childhood acute lymphoblastic leukemia. *Blood* 1993; **81**: 2237-51.
- 5 Stegmaier K, Pendse S, Barker GF *et al.* Frequent loss of heterozygosity at the TEL gene locus in acute lymphoblastic leukemia of childhood. *Blood* 1995; **86**: 38-44.
- 6 Takeuchi S, Bartram CR, Miller CW *et al.* Acute lymphoblastic leukemia of childhood: identification of two distinct regions of deletion on the short arm of chromosome 12 in the region of TEL and KIP1. *Blood* 1996; **87**: 3368-74.
- 7 Baccichet A, Qualman SK and Sinnett D. Allelic loss in childhood acute lymphoblastic leukemia. *Leuk Res* 1997; **21**: 817-23.
- 8 Baccichet A and Sinnett D. Frequent deletion of chromosome 12p12.3 in children with acute lymphoblastic leukaemia. *Br J Haematol* 1997; **99**: 107-14.

- 9 Aissani B and Sinnett D. Fine physical and transcript mapping of a 1.8 Mb region spanning the locus for childhood acute lymphoblastic leukemia on chromosome 12p12.3. *Gene* 1999; **240**: 297-305.
- 10 Baens M, Wlodarska I, Corveleyn A, Hoornaert I, Hagemeyer A and Marynen P. A physical, transcript, and deletion map of chromosome region 12p12.3 flanked by ETV6 and CDKN1B: hypermethylation of the LRP6 CpG island in two leukemia patients with hemizygous del(12p). *Genomics* 1999; **56**: 40-50.
- 11 Aissani B, Bonan C, Baccichet A and Sinnett D. Childhood acute lymphoblastic leukemia: is there a tumor suppressor gene in chromosome 12p12.3? *Leuk Lymphoma* 1999; **34**: 231-9.
- 12 Kibel AS, Freije D, Isaacs WB and Bova GS. Deletion mapping at 12p12-13 in metastatic prostate cancer. *Genes Chromosomes Cancer* 1999; **25**: 270-6.
- 13 Brown SD, Twells RC, Hey PJ *et al.* Isolation and characterization of LRP6, a novel member of the low density lipoprotein receptor gene family. *Biochem Biophys Res Commun* 1998; **248**: 879-88.
- 14 Pinson KI, Brennan J, Monkley S, Avery BJ and Skarnes WC. An LDL-receptor-related protein mediates Wnt signalling in mice. *Nature* 2000; **407**: 535-8.
- 15 Tamai K, Semenov M, Kato Y *et al.* LDL-receptor-related proteins in Wnt signal transduction. *Nature* 2000; **407**: 530-5.
- 16 Wehrli M, Dougan ST, Caldwell K *et al.* Arrow encodes an LDL-receptor-related protein essential for Wingless signalling. *Nature* 2000; **407**: 527-30.

- 17 Golub TR, Barker GF, Lovett M and Gilliland DG. Fusion of PDGF receptor beta to a novel ets-like gene, tel, in chronic myelomonocytic leukemia with t(5;12) chromosomal translocation. *Cell* 1994; **77**: 307-16.
- 18 Van Rompaey L, Dou W, Buijs A and Grosveld G. Tel, a frequent target of leukemic translocations, induces cellular aggregation and influences expression of extracellular matrix components. *Neoplasia* 1999; **1**: 526-36.
- 19 Fenrick R, Wang L, Nip *et al.* TEL, a putative tumor suppressor, modulates cell growth and cell morphology of ras-transformed cells while repressing the transcription of stromelysin-1. *Mol Cell Biol* 2000; **20**: 5828-39.
- 20 Stegmaier K, Takeuchi S, Golub TR, Bohlander SK, Bartram CR and Koeffler HP. Mutational analysis of the candidate tumor suppressor genes TEL and KIP1 in childhood acute lymphoblastic leukemia. *Cancer Res* 1996; **56**: 1413-7.
- 21 Thompson JD, Higgins DG and Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 1994; **22**: 4673-80.
- 22 Baens M, Peeters P, Guo C, Aerssens J and Marynen P. Genomic organization of TEL: the human ETS-variant gene 6. *Genome Res* 1996; **6**: 404-13.
- 23 Roest Crollius H, Jaillon O, Bernot *et al.* Estimate of human gene number provided by genome-wide analysis using Tetraodon nigroviridis DNA sequence. *Nat Genet* 2000; **25**: 235-8.
- 24 Guo B, Godzik A and Reed JC. Bcl-G, a novel pro-apoptotic member of the Bcl-2 family. *J Biol Chem* 2001; **276**: 2780-5.

- 25 Tanoue T, Yamamoto T, Maeda R and Nishida E. A Novel MAPK phosphatase MKP-7 acts preferentially on JNK/SAPK and p38 alpha and beta MAPKs. *J Biol Chem* 2001; **276**: 26629-39.
- 26 Martell KJ, Seasholtz AF, Kwak SP, Clemens KK and Dixon JE. hVH-5: a protein tyrosine phosphatase abundant in brain that inactivates mitogen-activated protein kinase. *J Neurochem* 1995; **65**: 1823-33.
- 27 Denu JM and Dixon JE. Protein tyrosine phosphatases: mechanisms of catalysis and regulation. *Curr Opin Chem Biol* 1998; **2**: 633-41.
- 28 Poirel H, Oury C, Carron *Cet al.* The TEL gene products: nuclear phosphoproteins with DNA binding properties. *Oncogene* 1997; **14**: 349-57.
- 29 Wang LC, Kuo F, Fujiwara Y, Gilliland DG, Golub TR and Orkin SH. Yolk sac angiogenic defect and intra-embryonic apoptosis in mice lacking the Ets-related factor TEL. *Embo J* 1997; **16**: 4374-83.
- 30 Wang LC, Swat W, Fujiwara *Yet al.* The TEL/ETV6 gene is required specifically for hematopoiesis in the bone marrow. *Genes Dev* 1998; **12**: 2392-402.
- 31 Lopez RG, Carron C, Oury C, Gardellin P, Bernard O and Ghysdael J. TEL is a sequence-specific transcriptional repressor. *J Biol Chem* 1999; **274**: 30132-8.
- 32 Chakrabarti SR and Nucifora G. The leukemia-associated gene TEL encodes a transcription repressor which associates with SMRT and mSin3A. *Biochem Biophys Res Commun* 1999; **264**: 871-7.
- 33 Polakis P. Wnt signaling and cancer. *Genes Dev* 2000; **14**: 1837-51.

- 34 Dahia PL. PTEN, a unique tumor suppressor gene. *Endocr Relat Cancer* 2000; **7**: 115-29.
- 35 Bonneau D and Longy M. Mutations of the human PTEN gene. *Hum Mutat* 2000; **16**: 109-22.
- 36 Debant A, Serra-Pages C, Seipel K *et al*. The multidomain protein Trio binds the LAR transmembrane tyrosine phosphatase, contains a protein kinase domain, and has separate rac- specific and rho-specific guanine nucleotide exchange factor domains. *Proc Natl Acad Sci U S A* 1996; **93**: 5466-71.
- 37 Seipel K, Medley QG, Kedersha N *et al*. Trio amino-terminal guanine nucleotide exchange factor domain expression promotes actin cytoskeleton reorganization, cell migration and anchorage-independent cell growth. *J Cell Sci* 1999; **112**: 1825-34.
- 38 Rapraeger AC. Molecular interactions of syndecans during development. *Semin Cell Dev Biol* 2001; **12**: 107-16.
- 39 Edwalds-Gilbert G, Veraldi KL and Milcarek C. Alternative poly(A) site selection in complex transcription units: means to an end? *Nucleic Acids Res* 1997; **25**: 2547-61.
- 40 Miyamoto S, Chiorini JA, Urcelay E and Safer B. Regulation of gene expression for translation initiation factor eIF-2 alpha: importance of the 3' untranslated region. *Biochem J* 1996; **315**: 791-8.

TABLE 1 Predicted transcriptional units within the chromosome 12p12 tumour suppressor locus

| Unigene/EST Cluster | No. ESTs | Gene (acc.no.) | Spliced sequence | Mouse Ortholog (acc. no.) | Exofish Hits (ecores) | Trapped Exons | PolyA signal |
|---------------------|----------|---|------------------|---------------------------|-----------------------|-----------------|--------------|
| Hs.169081 | 40 | ETV6 (NM_001987) | yes | NM_007961 | 4 | nd ^a | yes |
| Hs.293972 | 30 | | no | | - | | yes |
| Hs.146381 | 18 | | no | | - | | yes |
| Hs.11962 | 20 | BCL-G _L /G _S (AF281254/NM_030766) | yes | BF139422 | - | 2 | yes |
| Hs.97818 | 2 | BCL-G _M (AY040274) | yes | | - | | yes |
| Hs.23672 | 52 | LRP6 (NM_002336) | yes | NM_008514 | 17 | 2 | yes |
| Hs.41269 | 93 | | no | | - | | yes |
| Hs.5999 | 40 | LOH3CR12 (AK023622/NM_018050) | yes | NM_026345 | - | - | yes |
| Hs.50860 | 5 | | no | | - | | yes |
| yx93f07 | 5 | LOH2CR12 (AY037867) | yes | - | - | - | no |
| Hs.67553 | 7 | | no | | - | | yes |
| Hs.105040 | 43 | LOH1CR12 (AY037865/AY037866) | yes | AK007423 | 2 | - | yes |
| Hs.86330 | 3 | | no | | - | | yes |
| Hs.232681 | 4 | | no | | - | | yes |
| Hs.20281 | 147 | MKP-7 (AB052156/AY038927) | yes | AB052157 | 3 | 3 | yes |
| Hs.7341 | 61 | | no | | - | | yes |
| Hs.291547 | 1 | - | no | - | - | - | yes |
| Hs.171346 | 2 | - | no | - | - | - | yes |

^a not done. BACs used in the exon trapping experiments did not cover the ETV6 region.

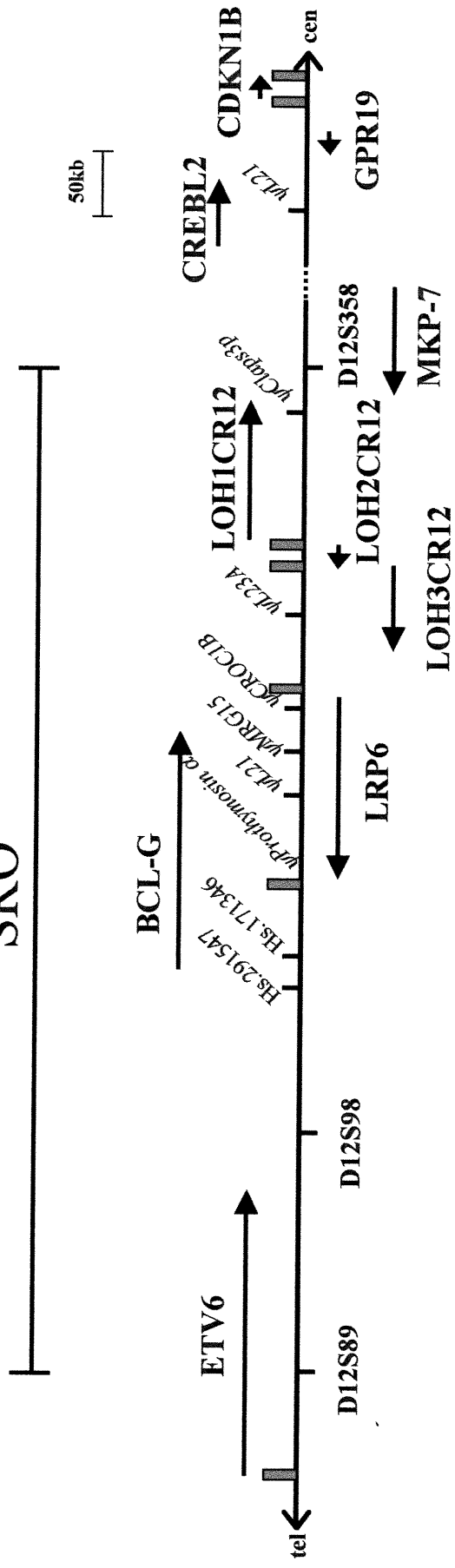
TABLE 2 Gene expression analyses by RT-PCR

| Gene | Brain | Testis | Prostate | Blood | Bone marrow | Placenta | HeLa | SJNB-7 ^b |
|----------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|---------------------|
| BCL-G | - | + ^a | + | - | + | - | - | - |
| LOH3CR12 | + ^a | + ^a | + ^a | + ^a | + ^a | + ^a | + ^a | + ^a |
| LOH2CR12 | - | + | + | - | + | - | - | + |
| LOH1CR12 | + ^a | + ^a | + ^a | + ^a | + ^a | + ^a | + ^a | + |
| MKP7 | + ^a | + ^a | + ^a | + ^a | + ^a | + ^a | + ^a | + ^a |

^apresence of splicing variant(s) (see text). ^bSJNB-7: neuroblastoma-derived cell line

Figure 1 Physical and transcriptional map of the commonly deleted region on chromosome 12p12.3. The shortest region of overlapping deletion (SRO) is delimited by markers *D12S89* and *D12S358*. Filled rectangles indicate the presence of a CpG island. Identified genes are indicated in bold. Arrows indicate the orientation of transcription. Other EST clusters (Hs. 291547 and Hs.171346) found in this study are also indicated on the map. Pseudogenes are indicated in italics. The position and size of the genomic clones, as well as their corresponding accession numbers are shown in the lower section. The dotted lines indicate partial and unordered genomic sequences. Tel, telomere; cen, centromere.

SRO



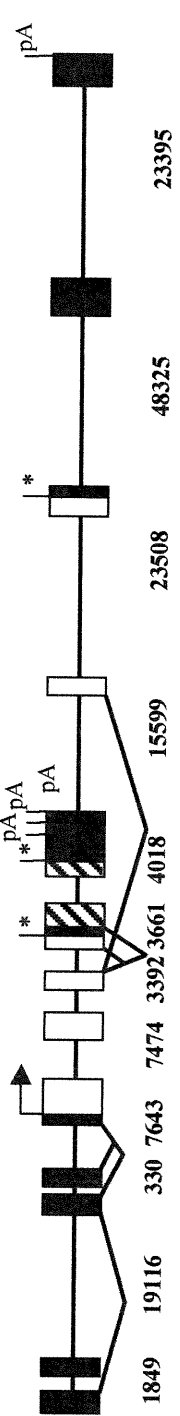
ETV6 contig

| | | | |
|--------------------|--------------------|--------------------|--------------------|
| NT_000600 (281 kb) | NT_000601 (191 kb) | 267J23 | 253I19 |
| ----- | ----- | (AC007537; 191 kb) | (AC007619; 160 kb) |
| ----- | ----- | ----- | ----- |
| AC005989 | 525I13 | 757G14 | 180M15 |
| (0.2 kb) | (AC022222; 200 kb) | (AC007621; 200 kb) | (AC008115; 158kb) |

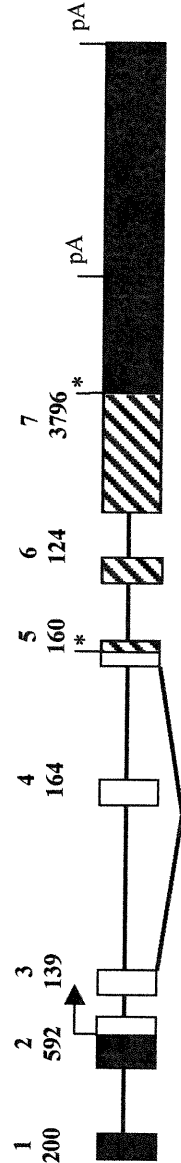
Figure 2 Genomic structure of *BCL-G*, *MKP-7*, *LOH1CR12*, *LOH2CR12* and *LOH3CR12*. Alternative splicing is indicated by solid lines joining two exons. Coding and non-coding exons are represented by open and filled boxes, respectively. Hatched boxes indicate the presence of non-coding or coding sequence depending on the context (see text). Arrows, asterisks and pA indicate initiation sites, stop sites and polyadenylation signals, respectively. The sizes of each exon and intron, when known, are indicated above and below the respective genes, respectively.

1A 1B 1C 1D 2 3 4 5 6 7 8 9 10
 150 44 174 188 440 440 174 71 420 1101 33 144 199 186

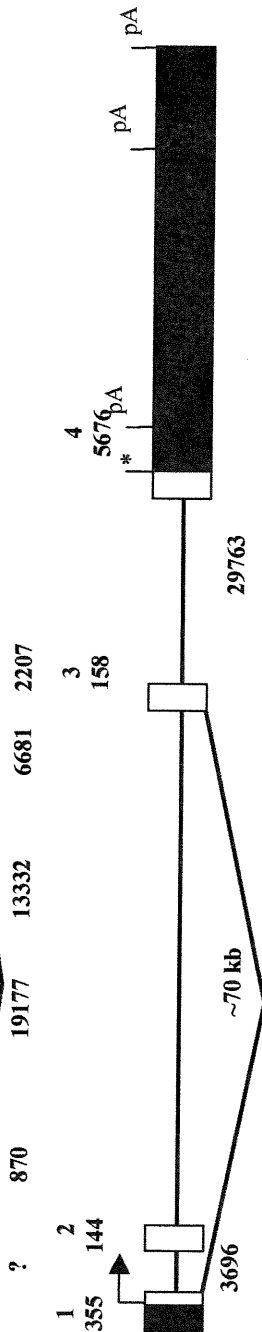
BCL-G



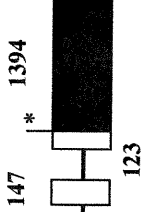
MKP-7



LOH1CR12



LOH2CR12



LOH3CR12

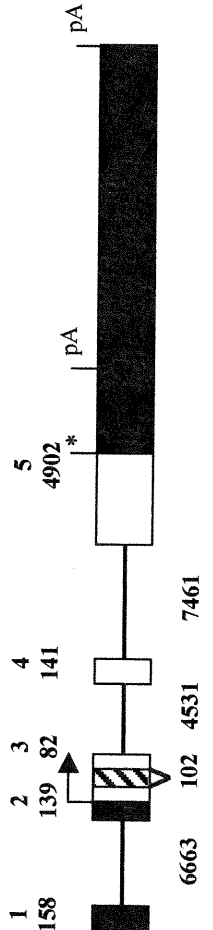


Figure 3 Northern blot analysis of *MKP-7*, *LOH1CR12* and *LOH3CR12* mRNAs. Multiple tissue Northern blots membranes were hybridised with probes specific for each gene. Asterisk indicates *DUSP8/hvh-5* cross-hybridisation (see text).

Figure 4 Amino acid alignment of hypothetical LOH1CR12 proteins from distant species. Human LOH1CR12 (GenBank acc. no. AAK71328), mouse BAB25030, *C.elegans* F59E12.11 (GenBank acc. no. T15266) and *D.melanogaster* CG11802 (GenBank acc. no. AAF48107) hypothetical proteins were aligned using ClustalW and manual refinements. Regions of high similarity with mouse Mekk4 (Swissprot no. O08648) and human Trio (Swissprot no. O75962) are also shown. Residues were shaded using BOXSHADE software according to their relative conservation (dark) or similitude (grey) across all species. Positions in the corresponding proteins are indicated on the right. The shorter human splice variant would lack amino acid 20 to 66.

Figure 5 Amino acid alignment of hypothetical protein LOH3CR12 from mouse and human. Human LOH3CR12 (GenBank acc. no. BAB14621) and mouse RIKEN cDNA 9130403P13 (GenBank acc. no. NP_080621) proteins were aligned with syndecan domain (pfam 01034) using ClustalW and manual refinements. Dashes indicate gaps introduced to maintain optimal alignment. Residues were shaded using BOXSHADE software according to their relative conservation (dark) or similitude (grey) across all species. Positions in the corresponding human protein are indicated on the right. The shorter splice variant would lack amino acid 14 to 47.

Figure 6 The partial cDNA sequence and predicted open reading frame of the *LOH2CR12* gene. The observed exon/intron junction is indicated by a filled triangle. The open triangle indicates the position of the most upstream sequence that was characterized by 5'RACE and RT-PCR. Upstream genomic sequence is indicated in italic. The putative polyA signal is underlined.

gaaccggcagatgacactggctttgccgcacctgcgccggcgggcgccgctttcactcactggttcacg

ctccgcaagctcgctccttccgggacgcagatggg[▽]cggtgccggctacgcaagaccttcagttccggattagg
V P A T Q D L Q F R I R

aggccccgcccccgccccgagggaggggaggagagaccgctcctgcgacttagggcgatgccaccttaaag
R P R P P A R G R G G E T R S C D L G R C H L K G

ggcttgacctcctcgaagccagaactgcggaagaggatggagaaagaaacgctagatagaccaggattcga
L D L L E A R T A E E D G E R N A R *

cccacaacagcttgatgcaaagctcattttgaattctgaagagctgggagtttagcagtgagcaccaaaaca
aaaaataaccagaagacggtgaaagcagcagctggaaatgagagaaaaattaagaaatagaaaatgctgtttc
atgtattagggagtgacatctttaaagagaagtgaagaaccttttggctaggcgcttctggggaccttgg
cagttatcaacaagtcaactctgaaactaacagaaggtgaggagggagtaatccagaaagggaaactgccatttt
tgctagggcaagaaaagtagacctagcaatgcaatgccatcaaggagttagctggctctagcgttctccgaact
ttgcaactcattttatattacattgtctgctcaagaaatttcaagtaaatgccttgagattttatatgtata
aaatgtagtctctggccaggcggggtggctcacgcctgtaatcccagcactttgggaggctgaggcgagtaga
tcacgatgtcaggagatcaagaccatcctggctaacaatggtgaaacccgctctactaaaaatacaaaaaa
ttagccgggcatgatggtgggcgctgtagtcccagctactcgggaggctgagacaggagaaatggtgtgaacc
cgggaggcagagcttgcagtgagccgagatcgcgccactgcaactccagcctgggcgacagagcgagactccgt
ctcaaaaaaaaaaaaaaaaaaaaaaagtctcagaagaagtgtcaagttgactaaatgatttttaagtccttcc
actctactatatcacaataattttcccctattttaccttatttaattttcacaacaaccttgagcctgtaagt
aggtgactcctccatggtcagcaggtggtcagtggtcgaactatttaggactatctgatgctctttcctctgc
tcaccattgcaatacaatgagaaaaatgcaaatgaatctaagtagcagtggtctctgggaatatgaagatagaa
ctaaaaactgcctgagagatcagggcaggttctagaaaagttgtgtctaactagtcttgaggtttacataaac
cacaccaggctgatgggactgggaaggcccactaagcagtgagaccattcccttttgagagctccttgcac
ccctcccagatttctctttgaggggaaggtgagagaggaggtaaaaggggtgaggaatggagaataactcat
tctggttcttgtttcccctttccacataaaaagtataattgtcttgtgtccatacaccagtcatactgatg
tgatggtgtttttatgcttcccttttgaataaacatttct

Chapitre 5

- Article 4 -

Expression and mutational analysis of the 12p candidate tumour suppressor genes in pre-B acute lymphoblastic leukemia.

Article en préparation.

Contribution des auteurs:

A. Montpetit: a effectué la majorité des expériences, toutes les analyses et a écrit l'article.

J. Larose: a effectué quelques expériences de PCR-SSCP.

N. Trudel: a effectué quelques expériences de PCR-SSCP.

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

**Expression and mutational analysis of the 12p candidate tumour
suppressor genes in pre-B acute lymphoblastic leukemia.**

Alexandre Montpetit, Josiane Larose, Nathalie Trudel and Daniel Sinnett². Division of Hematology-Oncology, Charles-Bruneau Cancer Center, Research Center, Sainte-Justine Hospital, Montreal, Canada H3T 1C5 [A.M., J.L., N.T., D.S.], Department of Biochemistry [A.M., D.S.], and Department of Pediatrics [D.S.], University of Montreal, Montreal, Canada.

Running title: 12p candidate genes characterization

Keywords: Tumor suppressor gene, leukemia, chromosome 12, gene expression.

¹ Supported by the Canadian Institutes of Health Research (CIHR). D.S. is a scholar of the Fonds de la Recherche en Santé du Québec (FRSQ). A.M. and G. B. have received studentships from the CIHR and from National Science and Engineering Research Council, respectively.

² To whom request for reprints should be addressed: at Division of Hematology-Oncology, Sainte-Justine Hospital, 3175 Côte Ste-Catherine, Montreal, Quebec, H3T 1C5. Phone: (514) 345-4931 ext. 6142; Fax: (514) 345-4731; E-mail: daniel.sinnett@umontreal.ca

Abstract

Loss of heterozygosity (LOH) on chromosome 12p12 is observed in 26-47% of the patients with acute lymphoblastic leukemia (ALL). The same region is also deleted in a large number of solid tumors. The recent construction of high-resolution physical and transcriptional maps of this putative tumor suppressor locus has enabled the identification of eight candidate genes of which five were previously known: *ETV6*, which encodes an ets-like transcription factor involved in the establishment of the hematopoiesis; *BCL-G*, a new pro-apoptotic gene; *LRP6*, a member of the LDL receptor gene family that encodes a co-receptor of the WNT proteins; *MKP-7*, a new member of the dual-specificity phosphatase gene family; and *CDKN1B* that encodes a member of the CDKI family. The three other candidate genes, *LOH1CR12*, *LOH2CR12* and *LOH3CR12*, have no known functions. To determine whether one (or more) of the candidate genes is the actual target of the 12p12 deletions we have examined the mutational as well as the expressional status of these genes in ALL patients. Although we found nine DNA variants, no inactivating mutation in the eight candidate genes was found in the leukemia cells of 17 ALL patients associated with 12p LOHs by PCR-SSCP. Next, by a semi-quantitative RT-PCR approach, we found that most 12p hemizygotously deleted samples also carried a t(12;21) translocation of which none expressed *ETV6* from the non-translocated allele. Furthermore, we observed two cases of t(12;21) without deletion of *ETV6*, in which the expression of this gene was greatly reduced indicating a different mechanism of inactivation. None of the other genes showed a significant decrease in expression suggesting that *ETV6* is the target of the deletions in ALL patients.

Introduction

Acute lymphoblastic leukaemia (ALL) is the most common pediatric cancer. It is a heterogeneous disease arising from clonal proliferation of lymphoblasts and their progenitors. The frequent loss of genetic material in tumor cells is a hallmark for the presence of tumor suppressor genes on the affected chromosome regions. Recent loss of heterozygosity studies indicated that hemizygous deletions of chromosome 12p12 occurred in 26 to 47 % of pre-B ALL cases, making it one of the most common genetic alteration found in this disease (Stegmaier *et al.*, 1995; Takeuchi *et al.*, 1996; Baccichet *et al.*, 1997). The existence of a putative 12p12 tumor suppressor gene is further substantiated by the observation of hemizygous deletions of the same region in a variety of other hematological malignancies, as well as in certain solid tumors including breast, lung, ovarian and prostate carcinomas (reviewed in Aïssani *et al.*, 1999; Kibel *et al.*, 1999). The construction of high-resolution genetic and transcriptional maps led to the delineation of the shortest commonly deleted region in pre-B ALL patients to a 750 kb interval flanked by markers *DI2S89* and *DI2S358* and includes seven genes (Baccichet and Sinnett 1997; Aïssani and Sinnett, 1999; Baens *et al.*, 1999; Montpetit *et al.*, 2002).

One of the genes, *ETV6*, is an interesting tumor suppressor candidate. It is a repressor of transcription that was found to be essential for angiogenesis and for bone marrow hematopoiesis (Wang *et al.*, 1997; Wang *et al.*, 1998). It has been shown that overexpression of *ETV6* represses the expression of various metalloproteinases such as stromelysin-1 and can counteract transformation of cultured cells by ras (Van Rompaey

et al., 2000; Fenrick *et al.*, 2000; Wang and Hiebert, 2001). Moreover, *ETV6* is frequently involved in translocations in different hematological malignancies as well as in congenital fibrosarcoma, most of these leading to the formation of chimeric fusion products (reviewed in Rubnitz *et al.*, 1999). The most frequent translocation is the t(12;21) resulting in the production of the ETV6-AML1 fusion protein and is observed in about 22 % of ALL cases (Shurtleff *et al.*, 1995). The residual non-rearranged *ETV6* allele is frequently deleted in ALL cases expressing the *ETV6-AML1* transcripts (Cave *et al.*, 1997). The explanation for this frequent association is still unclear. It has been proposed that deletion of the normal *ETV6* allele could provide a further growth advantage to leukemic cells with t(12;21).

Six other genes were characterized in the *D12S89-D12S358* critical region. *BCL-G* encodes a new pro-apoptotic protein of the BCL-2 family (Guo *et al.*, 2001). *LRP6* is a member of the lipoprotein receptor gene family and was recently characterized as the Wnt co-receptor, which activates a pathway frequently involved in human cancers (Pinson *et al.*, 2000; Tamai *et al.*, 2000; Wehrli *et al.*, 2000). *MKP-7* encodes a new dual-specificity phosphatase that acts specifically on JNK and p38MAPK (Tanoue *et al.*, 2001). Finally *LOH1CR12*, *LOH2CR12* and *LOH3CR12* are three recently characterized genes whose functions have not yet been determined (Montpetit *et al.*, 2002). Although it is located just outside the 12p critical region, the cyclin-dependant kinase inhibitor *CDKN1B* is still an excellent candidate since it is hemizygotously deleted in the majority of ALL patients with 12p LOHs, its expression status has been shown to correlate with a poor prognosis in many tumors types and is considered to be one of the few known

haploinsufficient tumor suppressor (reviewed in Philipp-Staheli *et al.*, 2001).

Many studies have focused only on the mutational status of the candidate genes in ALL patients and inactivating somatic mutations were not observed in the non-deleted allele of *ETV6*, *CDKN1B*, *LRP6* or *BCL-G* (Stegmaier *et al.*, 1996; Baens *et al.*, 1999; Guo *et al.*, 2001). In this study, we looked at both the mutation and expression status of eight candidate suppressor genes in 55 ALL patients.

Materials and Methods

Patients.

We studied 40 children who had been diagnosed with ALL by immunophenotyping in the Division of Hematology-Oncology of Ste-Justine Hospital. The patients selected for the study had bone marrow samples taken at diagnosis which contained >80% blast cells. Twenty of them were characterized in a previous study (Baccichet *et al.*, 1997). Written informed consent was obtained from the parents and the study was approved by our Institutional Review Board. Normal bone marrow was obtained from a non-leukemic patient.

DNA and RNA extraction.

Pretreatment bone marrow aspirates were used as a source of tumoural DNA, whereas normal DNA was isolated from either buccal epithelial cells or bone marrow in remission. A modified proteinase K/ phenol protocol was used to isolate genomic DNA

from bone marrow mononuclear cells obtained by centrifugation on a Ficoll-Paque (Pharmacia) step gradient (Sinnott *et al.*, 1988). Total RNA was extracted using TRIzol reagent and using manufacturer's instructions (GibcoBRL-Life Technologies, Burlington, Ontario, Canada).

Microsatellite analysis.

Genomic DNA was amplified using microsatellite markers located on chromosome 12p12: *D12S77*, *D12S89*, *D12S98*, *D12S358* and *D12S320*. Primer pairs were obtained from Research Genetics (Huntsville, Al.). Each PCR reaction contained 5ng of genomic DNA, 1µg of each primer, 100 µM of each dNTP and 1µCi of α -³²P dCTP in buffer containing 60 mM Tris-SO₄ pH 9.0 18 mM of (NH₄)₂SO₄, 8 % of DMSO and 1.5 mM of MgSO₄ in a final volume of 10 µL. After an initial denaturation of 4 min at 94°C, PCR was performed for 32 cycles of 30s at 94°C, 30s at 56-63°C, 1 min at 72°C followed by a final extension of 7 min at 72°C. PCR products were analyzed on a 6% denaturing polyacrylamide:bis-acrylamide (29:1) gel. Each result was confirmed by two independent PCR amplification.

PCR-SSCP analysis and sequencing.

Primer pairs were designed as to amplify all the coding sequence and intron-exon boundaries of *ETV6* (8 exons), *BCL-G* (5 exons), *LOH3CR12* (4 exons), *LOH2CR12* (2 exons), *LOH1CR12* (5 exons), *MKP-7* (6 exons) and *CDKN1B* (2 exons) (Table 1). PCR was performed as described in the previous section. PCR products were denatured and resolved by electrophoresis on a 6% non-denaturing gel at 500V for 20 hours. When

shifts were detected by SSCP, the corresponding normal and tumoral non-labelled PCR products were purified using CONCERT rapid PCR purification system (GibcoBRL-Life Technologies, Burlington, Ontario, Canada) and were directly sequenced.

Semi-quantitative RT-PCR assay.

RNA isolated from leukemic cells was incubated at 37°C for 60 minutes in the presence of 15 µM of random hexamers, 0.2 mM of each dNTPs and 200 U MMLV reverse transcriptase (GibcoBRL-Life Technologies, Burlington, Ontario, Canada). The cDNAs were placed in a solution containing 60 mM Tris-SO₄ pH 9.0, 18 mM of (NH₄)₂SO₄, 8 % of DMSO, 1.5 mM of MgSO₄, 0.2 mM of each dNTP, 1 µCi of α-³²P dCTP and 0.5 U of Platinum Taq. 10 µL of each mix was then amplified in nine different PCR reactions using 1 µM of one of the nine different primer pairs (Table 2). Following a 4 min denaturation at 94°C, PCR was performed 30 s at 94°C, 30 s at 63°C and 1min 30s at 72°C, for the number of cycles indicated in Table 2, followed by a final extension of 7 min at 72°C. PCR products were then separated by electrophoresis on non-denaturing 6% polyacrylamide gels and autoradiographed on X-OMAT films. The intensity of the individual bands was measured on a PhosphorImager. The expression level of the mRNAs was calculated by normalization to the amount of the corresponding GAPDH band. Each expression value is the average of three independent PCR reactions and was converted to an arbitrary scale compared on the median value of the 20 samples for each gene: - (absent: < 10% median value), +/- (decreased: >10% and <50% median value), + (normal: >50% and < 200% median value), ++ (high: >200% median value).

Results

In a previous study, we analysed the LOH status of 20 ALL patients on chromosome 12p with commercially available STS markers. We identified eight patients with a LOH and the commonly deleted region was delimited between markers *DI2S89* and *DI2S358*, separated by a genetic distance of approximately 3cM. In this study, we added 35 new ALL patients and identified nine patients with a chromosome 12p12 LOH using the same polymorphic markers (see Table 3). Combined with the previous samples, we observe the presence of a LOH in 31% of ALL patients (17/55), which is in agreement with previous studies on ALL patients and which is the most common genetic aberration in this disease (reviewed in Aïssani *et al.*, 1999). Of the 35 new patients, 10 had the translocation (28.5 %) including 8 out of the 9 LOH samples (ALL 400, 392, 375, 401, 410, 416, 420, 423), indicating a close association. Two ALL samples have a t(12;21) translocation without any LOH in the 12p12 region (LLA 411 and 413) and one sample has a LOH without the translocation (LLA34)

The 17 patients hemizyously deleted for at least one of the chromosome 12p12 marker tested were further investigated for the presence of mutations in their leukemic cells. The coding regions of the genes *ETV6*, *BCL-G*, *MKP-7*, *CDKN1B*, *LOH1CR12*, *LOH2CR12* and *LOH3CR12* were screened by PCR-SSCP followed by direct sequencing. *LRP6* was not investigated since this gene was analyzed in details in ALL patients in a previous study and that no mutations were observed (Baens *et al.*, 1999). No somatic mutations were observed, but we found nine different polymorphisms (Table

4). Four of these result in a change of an amino acid from the predicted coding sequence, but none were specifically retained or lost in the LOH samples. By using these polymorphisms as additional LOH markers, we were able to reduce the commonly deleted region between marker *DI2S89*, located in *ETV6*, and the third exon of *BCL-G*, thus leaving only these two genes as candidate (Table 3).

To evaluate the effect of an undetected mutation influencing gene expression, we examined mRNA levels of all eight candidate genes in nine LOH patients from which we had RNA samples from the bone marrow at diagnosis. Since total bone marrow contains many different types of cells at various developmental stages, we decided to attribute a "normal" level of expression to each gene by taking the median value of expression of 15 non-LOH leukemia samples. Levels of expression of these samples and of the nine LOH patients were then labelled as no expression (-, less than 10% of normal expression), decreased (+/-, less than 50% of normal expression), normal (+, between 50% and 200% of normal expression), increased (+++, more than 200% of the normal expression) (Table 5).

Expression patterns of *ETV6*, *BCL-G*, *MKP-7* and *CDKN1B* are presented in Figure 3. For *ETV6*, the primers were designed in order to detect only transcripts from the untranslocated allele. For *BCL-G*, *LOH1CR12*, *LOH3CR12* and *MKP-7*, we selected primers in order to discriminate between the different known isoforms (Table 1). Except for *BCL-G* (see below), no variation in the ratio of the levels of expression between the isoforms were observed and thus only the major isoform was analyzed in each case (data

not shown). The ratio of expression of both isoforms of *BCL-G* varies greatly between samples (Figure 3). These isoforms, generating the short and long forms of BCL-G, have pro-apoptotic activities (Guo *et al.*, 2001) and both bands were included to calculate the level of expression of *BCL-G*.

The correlation of the reduced expression of a gene with various clinical and biological factors, such as age at diagnosis, sex, white blood cell counts, high or standard risk of relapse, DNA index, presence of LOH and presence of the t(12;21) translocation was investigated. Statistically significant association between loss of *ETV6* and presence of both t(12;21) or 12p12 LOH was observed (Table 6). Other similar significant associations were observed with *MKP-7* and *CDKN1B* (Table 6), but with no other gene (data not shown). However, reduced expression in the LOH samples, as defined by 50% of the median level of expression of the 20 ALL samples, can be explained by the loss of the deleted allele. In fact, levels of expression of most of the genes in the samples possessing both the translocation and LOH were either below or just above the arbitrary "reduced" threshold (data not shown). But, when only considered the complete absence of expression of a gene, then only *ETV6* show a significant association with the presence of LOH or t(12;21). Interestingly, patient ALL34 is the only one with a deletion, ranging from *D12S89* to *D12S358*, that do not possess the t(12;21) translocation. In this case, some genes in that interval show reduced expression (*LOH2CR12*, *LOH3CR12* and *MKP-7*) but none are completely absent.

Finally, two ALL patients have the t(12;21) translocation and don't show any 12p LOH (ALL 411 and 413; Table 3). They have a reduced level of *ETV6* expression that was expected since only one allele can be amplified with our RT-PCR assay (Table 5). However, the level of expression is very low (in the 10-20 % range compared to normal levels) indicating that another mechanism is responsible for the greater reduction of expression (data not shown).

Discussion

Loss-of-heterozygosity (LOH) of the small arm of chromosome 12 is one of the most common genomic abnormalities found in childhood B-precursor acute lymphoblastic leukemia since it is observed in 26-47 % of the cases and suggests the presence of a tumour suppressor gene (Stegmaier *et al.*, 1995; Takeuchi *et al.*, 1996; Baccichet *et al.*, 1997). It has been recently been defined in a 750 kb interval between markers *D12S89* and *D12S358* (Bacchichet *et al.*, 1997; Baens *et al.*, 1999). Another common genetic defect is the t(12;21)(p13;q22) translocation which is detected in 20-30 % of cases studied and results in the production of the ETV6-AML1 fusion protein also involves the same region (Shurtleff *et al.*, 1995; McLean *et al.*, 1996). The specific role of this fusion protein in the leukemogenesis process is not clear but its presence has been associated with a favourable prognosis (Shurtleff *et al.*, 1995; McLean *et al.*, 1996). Association between the presence of the translocation t(12;21) and the presence of a 12p12 deletion in the other allele has been observed. One study has indicated that more 80 % of the t(12;21) cases also possess a 12p12 LOH (Cave *et al.*, 1997). However,

since this same deletion is observed in some ALL patients without t(12;21) and in many other type of cancers, including breast, lung and prostate, where t(12;21) is not present, many have questioned whether ETV6 is the actual (or only) target of these deletions.

In this study we have analysed the LOH status of the 12p12 region in 55 ALL patients. Thirteen of them had a LOH in the *D12S89-D12S358* critical region. In the leukemic samples, no mutations were observed in the coding exons of seven candidate tumour suppressor genes, but some polymorphisms were observed. We were also able to minimize the deleted region common to all patients to a small interval between D12S89 and the third exon of *BCL-G*. The only known genes present in this interval are *ETV6* and *BCL-G*. It is still possible however that other unidentified genes could be present in the region, but in a previous study, the genes present in this locus were identified by analyzing the published genomic sequence and the EST database and by confirming the results by RT-PCR and sequencing (Montpetit *et al.*, 2002).

If the two-allele inactivation hypothesis for a tumor suppressor holds true and that no mutation alters the coding sequence, then expression of the target gene should be abolished or greatly diminished. Only *ETV6* showed absence of expression in a statistically significant proportion of the patients with LOH or with t(12;21). Indeed all the eight t(12;21) samples which had a corresponding LOH on the intact allele, did not express *ETV6*. Moreover, in these patients, the expression of most of the other genes present in the region defining the LOH was reduced by about 50% in agreement with the presence of an actual true deletion of the region (by opposition to a somatic

recombination or a gene conversion). In addition, in the two samples possessing the translocation but no detectable deletion within *ETV6*, levels of expression of *ETV6* are much lower than if only one allele was missing indicating that possibly another mechanism is responsible for the greater reduction of expression.

Thus, it can be concluded from this study that the intact allele of *ETV6* is clearly the target when translocation t(12;21) occurs and explains most of the LOH observed in pre-B ALL. This is the first report that actually demonstrate that *ETV6* expression is lost in these samples. The actual hypothesis about leukemogenesis is that the first event is the translocation and it probably occurs as early as *in utero* (Ford *et al.*, 1998; Wiemels *et al.*, 1999). This event is probably not sufficient since identical twins have been shown to develop leukemia many years apart from an identical clone. Indeed, ETV6-AML1 was shown not to be sufficient for the transformation of hematopoietic cell lines *in vitro* or to cause hematological disease *in vivo* (Andreasson *et al.*, 2001). Loss of the second allele is thus probably a secondary event giving a selective growth advantage to the t(12;21) cells expressing ETV6-AML1. However, it is not clear whether or not it is enough to cause leukemia. It has been shown that the addition of the ETV6 domains (dimerization and transcriptional repression domains) transforms AML1 into a transcriptional repressor (Hiebert *et al.*, 1996). We do not know whether the loss of the intact ETV6 allele is essential because it segregates ETV6-AML1 by heterodimerization or because the inhibition of the ETV6 pathway is important. Both models though are not mutually exclusive. Although ETV6 has been shown to block ras-induced tumorigenesis

in vivo (Van Rompaey *et al.*, 2000). Functional studies will be needed to prove that ETV6 indeed acts as a tumour suppressor in these ALL samples.

In the only 12p12 LOH case where t(12;21) is not present, the expression analysis was not conclusive. Although expression of some genes were reduced, the lack of other similar samples do not allow us to draw any conclusions. Since this region is a hotspot for rearrangements it is possible that an unspecific deletion occurred (Sato *et al.*, 2001). Alternatively, in this case, another unidentified gene could be the target of the deletion. In the other type of cancers where deletion of chromosome 12p12 has been detected (ovary, breast, prostate, etc.), no translocations involving *ETV6* have ever been observed. For prostate cancer, precise deletion mapping studies have been performed. The commonly deleted region was mapped between *D12S391* and *D12S358* (Kibel *et al.*, 2000), a region including *LOH1CR12*, *LOH2CR12*, *LOH3CR12* and *MKP-7* genes. No mutations were observed in *ETV6* or *CDKN1B*, although in another study decreased expression of *ETV6* has been observed in a significant proportion of the LOH samples (Latil *et al.*, 2000). Also, Takeuchi *et al.* (1996) have observed that a second region, between *D12S358* and *CDKN1B*, was the target of deletions in a small subset of ALL samples. So it seems that chromosome 12p12 has not reveal all its secrets and that probably another tumor suppressor gene is present.

References

1. Aissani B, Bonan C, Baccichet A and Sinnett D. 1999. Childhood acute lymphoblastic leukemia: is there a tumor suppressor gene in chromosome 12p12.3? *Leuk Lymphoma* **34**: 231-239.
2. Aissani B, Sinnett D. 1999. Fine physical and transcript mapping of a 1.8 Mb region spanning the locus for childhood acute lymphoblastic leukemia on chromosome 12p12.
3. *Gene* **240**: 297-305.
3. Andreasson P, Schwaller J, Anastasiadou E, Aster J, Gilliland DG. 2001. The expression of ETV6/CBFA2 (TEL/AML1) is not sufficient for the transformation of hematopoietic cell lines in vitro or the induction of hematologic disease in vivo. *Cancer Genet Cytogenet* **130**: 93-104.
4. Baccichet A and Sinnett D. 1997. Frequent deletion of chromosome 12p12.3 in children with acute lymphoblastic leukaemia. *Br J Haematol* **99**: 107-114.
5. Baens M, Wlodarska I, Corveleyn A, Hoornaert I, Hagemeijer A and Marynen P. 1999. A physical, transcript, and deletion map of chromosome region 12p12.3 flanked by ETV6 and CDKN1B: hypermethylation of the LRP6 CpG island in two leukemia patients with hemizygous del(12p). *Genomics* **56**: 40-50.

6. Cave H, Cacheux V, Raynaud S, Brunie G, Bakkus M, Cochaux P, Preudhomme C, Lai JL, Vilmer E, Grandchamp B. 1997. ETV6 is the target of chromosome 12p deletions in t(12;21) childhood acute lymphocytic leukemia. *Leukemia* **11**: 1459-1464.
7. Fenrick R, Wang L, Nip J *et al.* 2000. TEL, a putative tumor suppressor, modulates cell growth and cell morphology of ras-transformed cells while repressing the transcription of stromelysin-1. *Mol Cell Biol* **20**: 5828-5839.
8. Ford AM, Bennett CA, Price CM, Bruin MC, Van Wering ER, Greaves M. 1998. Fetal origins of the TEL-AML1 fusion gene in identical twins with leukemia. *Proc Natl Acad Sci U S A* **14**: 4584-4588.
9. Guo B, Godzik A and Reed JC. 2001. Bcl-G, a novel pro-apoptotic member of the Bcl-2 family. *J Biol Chem* **276**: 2780-2785.
10. Hiebert SW, Sun W, Davis JN, Golub T, Shurtleff S, Buijs A, Downing JR, Grosveld G, Roussel MF, Gilliland DG, Lenny N, Meyers S. 1996. The t(12;21) translocation converts AML-1B from an activator to a repressor of transcription. *Mol Cell Biol* **16**: 1349-1355.
11. Kibel AS, Faith DA, Bova GS, Isaacs WB. 2000. Loss of heterozygosity at 12P12-13 in primary and metastatic prostate adenocarcinoma. *J Urol* **164**: 192-196.

12. Kibel AS, Freije D, Isaacs WB and Bova GS. 1999. Deletion mapping at 12p12-13 in metastatic prostate cancer. *Genes Chromosomes Cancer* **25**: 270-276.
13. Latil A, Guerard M, Berthon P, Cussenot O. 2001. 12p12-13 deletion in prostate tumors and quantitative expression of CDKN1B and ETV6 candidate genes. *Genes Chromosomes Cancer* **31**: 199-200.
14. McLean TW, Ringold S, Neubergh D, Stegmaier K, Tantravahi R, Ritz J, Koeffler HP, Takeuchi S, Janssen JW, Seriu T, Bartram CR, Sallan SE, Gilliland DG, Golub TR. 1996. TEL/AML-1 dimerizes and is associated with a favorable outcome in childhood acute lymphoblastic leukemia. *Blood* **88**: 4252-4258.
15. Montpetit A, Boily G, Sinnott D. 2002. A detailed transcriptional map of the chromosome 12p12 tumor suppressor locus. Accepted for publication in *Eur J Hum Gen.*
16. Philipp-Staheli J, Payne SR, Kemp CJ. 2001. p27(Kip1): regulation and function of a haploinsufficient tumor suppressor and its misregulation in cancer. *Exp Cell Res* **264**: 148-168
17. Pinson KI, Brennan J, Monkley S, Avery BJ and Skarnes WC. 2000. An LDL-receptor-related protein mediates Wnt signalling in mice. *Nature* **407**: 535-538.

18. Rubnitz JE, Pui CH, Downing JR. 1999. The role of TEL fusion genes in pediatric leukemias. *Leukemia* **13**: 6-13.

19. Sato Y, Kobayashi H, Suto Y, Olney HJ, Davis EM, Super HG, Espinosa R 3rd, Le Beau MM, Rowley JD. 2001. Chromosomal instability in chromosome band 12p13: multiple breaks leading to complex rearrangements including cytogenetically undetectable sub-clones. *Leukemia* **15**: 1193-1202.

20. Shurtleff SA, Buijs A, Behm FG, Rubnitz JE, Raimondi SC, Hancock ML, Chan GC, Pui CH, Grosveld G, Downing JR. 1995. TEL/AML1 fusion resulting from a cryptic t(12;21) is the most common genetic lesion in pediatric ALL and defines a subgroup of patients with an excellent prognosis. *Leukemia* **9**: 1985-1989.

21. Sinnott D, Lavergne L, Melancon SB, Dallaire L, Potier M, Labuda D. 1988. Lesch-Nyhan syndrome: molecular investigation of three French Canadian families using a hypoxanthine-guanine phosphoribosyltransferase cDNA probe. *Hum Genet* **81**: 4-8.

22. Stegmaier K, Pendse S, Barker GF *et al.* 1995. Frequent loss of heterozygosity at the TEL gene locus in acute lymphoblastic leukemia of childhood. *Blood* **86**: 38-44.

23. Stegmaier K, Takeuchi S, Golub TR, Bohlander SK, Bartram CR and Koeffler HP. 1996. Mutational analysis of the candidate tumor suppressor genes TEL and KIP1 in childhood acute lymphoblastic leukemia. *Cancer Res* **56**: 1413-1417.

24. Takeuchi S, Bartram CR, Miller CW *et al.* 1996. Acute lymphoblastic leukemia of childhood: identification of two distinct regions of deletion on the short arm of chromosome 12 in the region of TEL and KIP1. *Blood* **87**: 3368-3374.

25. Tamai K, Semenov M, Kato Y *et al.* 2000. LDL-receptor-related proteins in Wnt signal transduction. *Nature* **407**: 530-535.

26. Tanoue T, Yamamoto T, Maeda R and Nishida E. 2001. A Novel MAPK phosphatase MKP-7 acts preferentially on JNK/SAPK and p38 alpha and beta MAPKs. *J Biol Chem* **276**: 26629-26639.

27. Van Rompaey L, Dou W, Buijs A and Grosveld G. 1999. Tel, a frequent target of leukemic translocations, induces cellular aggregation and influences expression of extracellular matrix components. *Neoplasia* **1**: 526-536.

28. Wang LC, Kuo F, Fujiwara Y, Gilliland DG, Golub TR and Orkin SH. 1997. Yolk sac angiogenic defect and intra-embryonic apoptosis in mice lacking the Ets-related factor TEL. *Embo J* **16**: 4374-4383.

29. Wang LC, Swat W, Fujiwara Y *et al.* 1998. The TEL/ETV6 gene is required specifically for hematopoiesis in the bone marrow. *Genes Dev* **12**: 2392-2402.

30. Wang L, Hiebert SW. 2001. TEL contacts multiple co-repressors and specifically associates with histone deacetylase-3. *Oncogene* **20**: 3716-3725.

31. Wehrli M, Dougan ST, Caldwell K *et al.* 2000. Arrow encodes an LDL-receptor-related protein essential for Wingless signalling. *Nature* **407**: 527-530.

32. Wiemels JL, Ford AM, Van Wering ER, Postma A, Greaves M. 1999. Protracted and variable latency of acute lymphoblastic leukemia after TEL-AML1 gene fusion in utero. *Blood* **94**: 1057-1062.

Table 1 Primers used in the mutational analysis

| Gene-exon | Forward | Reverse | Product Size (bp) |
|-----------|---------------------------|--------------------------|-------------------|
| ETV6-1 | TCTTAAATGACCGCGTCTGG | GAACAGAGCAACTGCAACAGC | 243 |
| ETV6-2 | CACCCCGAGATGGTCTCATA | ATCAATCCTCTGTCCCTCTCG | 250 |
| ETV6-3 | TCTCTTTCTTTCTGCCTCATGC | GGGTGTTAAAGACCAACCAC | 271 |
| ETV6-4 | CTGCTCCGTAGATCGTTGTTG | TCAGGCAAGAAAAGAGAAACCTC | 236 |
| ETV6-5a | GTCTTTCCCTCTGCTCCACAG | GATCAGAGGGTGCATGATGG | 392 |
| ETV6-5b | TCCCATGGAGAATAATCACTGC | GGAGAGTGGGACTTTGTCAGG | 357 |
| ETV6-6 | TGGTTAGTGCCTCAACAAGAAAC | CCAATCCAGGGATAAGGACTTC | 269 |
| ETV6-7 | GCAGTGCCTTTTCTGAGGTTT | ACAGCATCCTCCCAAAGAAG | 219 |
| ETV6-8 | TGGAATCTCTTACCTCCTCCAC | TCCTTTTCCACTCTCCTCAGC | 248 |
| Bcl-G-2 | GGCACAGTGTGGACTTTGTTGT | AGGAAGCGGGAGAGAAAGAAAT | 502 |
| Bcl-G-3 | ACCAATGAGCCAGATGAAGTTCT | AAACACTCCACCCAGGGAAG | 264 |
| Bcl-G-4 | CTTACCAAGCCACAAGCATTCT | CAGAAATCAATCGCATCAGTTCA | 199 |
| Bcl-G-5 | TCGTGGAGCCAAAGCTACTGT | TGGCCACTCAAGTGTCTCTAA | 446 |
| Bcl-G-6 | GGTCACGCCTGGGATAGAG | ACGTGCACAGGACCACAGTA | 189 |
| LOH3-2,3 | TGTTTCTTCCCTTGTTTGTTAAAGA | CACAAAACACATGGAACAAAGGA | 430 |
| LOH3-4 | CCAGTCTTAAATTATTCCAATCACA | AAAATCGCATTTCAAATCATGC | 237 |
| LOH3-5a | TTTTCAGGCAGCAGATGAGCTA | GCATGGAGTTACATAAGCCTTCC | 347 |
| LOH3-5b | TGCATTTGGGCTTCTGGTTACTA | GGACTCGAAAGGCAGCTTAGAA | 416 |
| LOH3-5c | CCCTGTGTTCAAAGTTAGGTTGG | AAGGCCATTCTCAGAGTTCACAA | 398 |
| LOH2 | CTGCTTTCACCGTCTTCTGGTATTT | CCGCTTTCACTCACTGGTTTAC | 465 |
| LOH1-1 | GTCAGCCCCACACATTAGCC | GGAGGAGGGTGGGAGGTTA | 231 |
| LOH1-2 | TGCCTAATATGCATCTCCTAAGCA | AAACGTCTCCAGCAAGTTCAGAT | 231 |
| LOH1-3 | CCCTGTCTTCAGTGAAGTTGTTAAA | AGAATCACAACAACCAGTCAGCA | 295 |
| LOH1-4 | CTGCGGAGAGGACTTGATTA | TGATATGACCTTAGCTCCACACG | 363 |
| MKP7-1 | TTCCTCATTCAAGTGTGTCCATTT | TTTGAGCACACCACCATTACATC | 354 |
| MKP7-2 | GCTCAGCAAGATTCTGAGTCCTTT | CCTTCTTCCTAACAAGCCTCATGT | 219 |
| MKP7-3 | AAACCCAGCAATGAAGGATATTTT | ATTTTGTGCGTGGCATAATTTT | 249 |
| MKP7-4 | CACCTACACATGAGACGATCACC | TCAAGGAAACTGGGCATTTT | 323 |
| MKP7-5 | TGAAAAGCCAATCACAGAAGACA | TGGAAACAAGATTTCACTGCTGA | 229 |
| MKP7-6a1 | GTGGAGATAAAAGGGACCTCTGG | AGGCTGGAAGACAGCAATAAGC | 343 |
| MKP7-6a2 | ATCCAGAGAGAAGGAACGCTTG | TGTAATAATTCGGGTTTTTGT | 432 |
| MKP7-6b1 | GATCCACCTGTTGCTTTTACACC | TGCGGAGACCAAGTCTATTCTG | 372 |
| MKP7-6b2 | GTCAGCTCTGTCACTTGGCTTC | ATTGCATTTCGGTCAGAACCAG | 376 |
| CDKN1B-1a | ATCCTGGCTCTCCT | ACATCTTTCTCCCG | 334 |
| CDKN1B-1b | ATTCTATGGTTGGGAAAGGGT | CACCTCTTGCCACTCGTACT | 285 |
| CDKN1B-2 | TCCAGTGCGTGCTCCTTTAG | TATGGGGCCAACCTTCTGC | 241 |

Table 2 Primers used in the RT-PCR analysis

| Gene | Forward | Reverse | Product size (bp) ^a | # Cycles |
|----------|-------------------------|---------------------------|--------------------------------|----------|
| ETV6 | GCACTCCGTGGATTTCAAAC | CTGAGGTGGACTGTTGGTTC | 536 | 25 |
| BCL-G | GCAGGAGGCTTCAAGTCCAA | CGGGTGGTTGTCAATAGCTG | 354 (507) | 31 |
| LRP6 | CCCTTGTGCTCAGGATAATG | TAGCTCCTCTCTGTGTATGG | 1176 | 29 |
| LOH3CR12 | CCTTGACCTTTGAAGACCAAAA | TCCTCGTTGGGACAGAAAAAT | 350 (248) | 25 |
| LOH2CR12 | GGCTACGCAAGACCTTCAGTT | CTGCTTTCACCGTCTTCTGGTATTT | 278 | 33 |
| LOH1CR12 | CCGACCCCAACGATCTGAACT | GGCGGACATCTCGTTCACTT | 430 (286) | 25 |
| MKP-7 | CGATCAAAGCTCCCCAAGATGT | CACGGCAAATAATTTCTCACAA | 402 (238) | 25 |
| CDKN1B | CGGGAGAAAAGATGTCAAACGTG | CGTTTGACGTCTTCTGAGGCC | 604 | 25 |
| GAPDH | CTCAGACACCATGGGGAAGGTGA | ATGATCTTGAGGCTGTTGTCATA | 450 | 19 |

^aThe number in parentheses reflects the size of an alternative less abundant transcript

Table 3 Deletion mapping of chromosome 12p12.3 in 19 cases of ALL

| Informative PCR | 3 | 5 | 6 | 8 | 15 | 24 | 26 | 27 | 400 | 392 | 375 | 34 | 401 | 410 | 411 | 413 | 416 | 420 | 423 |
|--------------------|------|------|------|------|------|------|------|------|------|------|-----|------|------|------|-----|-----|------|------|------|
| D12S77 | + | LOH | [ni] | LOH | + | + | [ni] | LOH | + | + | + | + | [ni] | LOH | ni | ni | + | LOH | LOH |
| D12S89 | LOH | LOH | LOH | LOH | [ni] | + | [ni] | LOH | [ni] | LOH | LOH | LOH | LOH | LOH | + | + | [ni] | LOH | LOH |
| D12S98 | LOH | [ni] | [ni] | LOH | [ni] | [ni] | [ni] | LOH | LOH | LOH | LOH | LOH | [ni] | [ni] | ni | ni | LOH | LOH | [ni] |
| BCLG-3 | [ni] | [ni] | + | LOH | LOH | [ni] | [ni] | LOH | LOH | + | + | [ni] | LOH | [ni] | ni | ni | [ni] | LOH | [ni] |
| BCLG-5 | LOH | [ni] | + | LOH | [ni] | [ni] | [ni] | [ni] | LOH | [ni] | + | + | [ni] | LOH | ni | ni | + | [ni] | [ni] |
| LOH2 | [ni] | [ni] | + | [ni] | LOH | [ni] | [ni] | [ni] | [ni] | [ni] | ni | ni | [ni] | [ni] | ni | ni | + | [ni] | [ni] |
| MKP7-6 | [ni] | [ni] | ni | [ni] | [ni] | LOH | [ni] | [ni] | [ni] | LOH | ni | ni | [ni] | LOH | + | ni | ni | [ni] | [ni] |
| D12S358 | LOH | LOH | ni | + | LOH | [ni] | [ni] | LOH | LOH | [ni] | ni | + | [ni] | [ni] | + | + | + | [ni] | LOH |
| CDKN1B-1b | [ni] | [ni] | ni | ni | [ni] | LOH | [ni] | LOH | [ni] | LOH | ni | ni | [ni] | [ni] | ni | ni | + | LOH | [ni] |
| D12S320 | LOH | + | + | ni | + | + | LOH | LOH | [ni] | LOH | + | + | LOH | LOH | + | + | + | LOH | [ni] |

ni, not informative; +: heterozygote; LOH: loss-of-heterozygosity. Unconfirmed deletion extents are shown in brackets.

Table 4 Polymorphisms identified

| Gene | Polymorphism^a | Amino acid change |
|-------------|---------------------------------|--------------------------|
| ETV6 | G258A | Thr → Thr |
| | G642A | Pro → Pro |
| BCL-G | C486T | Tyr → Tyr |
| | T697C | Leu → Leu |
| | A874T | Arg → Trp |
| LOH3CR12 | T542C | Leu → Val |
| LOH2CR12 | G149A | Glu → Lys |
| MKP-7 | C1419T | Thr → Thr |
| CDKN1B | G326C | Gly → Val |

^a The polymorphism indicates variation from the published sequence.

The number indicates the position from the start codon.

Table 5 Expression of 12p12.3 genes in 26 pre-B ALL patients

| Patient | 12pLOH | t(12;21) | ETV6 | BCL-G | LRP6 | LOH3CR12 | LOH2CR12 | LOH1CR12 | MKP-7 | CDKN1B |
|---------|--------|----------|------|-------|------|----------|----------|----------|-------|--------|
| 83 | - | - | + | +/- | ++ | + | + | + | + | +/- |
| 362 | - | - | + | + | ++ | + | + | + | + | + |
| 365 | - | - | + | +/- | ++ | + | + | + | ++ | + |
| 372 | - | - | + | ++ | + | + | + | + | + | + |
| 381 | - | - | + | + | ++ | +/- | +/- | + | + | + |
| 383 | - | - | + | ++ | +/- | ++ | ++ | + | + | + |
| 384 | - | - | ++ | + | +/- | + | + | + | +/- | + |
| 387 | - | - | + | ++ | + | +/- | + | + | + | + |
| 388 | - | - | + | ++ | ++ | + | + | + | + | + |
| 389 | - | - | + | ++ | + | + | + | + | ++ | + |
| 390 | - | - | + | +/- | ++ | + | + | + | + | + |
| 391 | - | - | ++ | +/- | +/- | +/- | + | +/- | + | + |
| 394 | - | - | + | +/- | +/- | +/- | +/- | +/- | + | + |
| 396 | - | - | + | + | ++ | ++ | + | + | + | + |
| 397 | - | - | + | + | +/- | ++ | + | + | + | + |
| 411 | - | + | +/- | + | + | + | + | + | + | + |
| 413 | - | + | +/- | + | + | + | + | + | + | + |
| 34 | + | - | + | ++ | + | +/- | +/- | + | +/- | + |
| 375 | + | + | - | +/- | +/- | + | +/- | + | + | + |
| 392 | + | + | - | +/- | + | +/- | + | + | +/- | + |
| 400 | + | + | - | + | + | +/- | + | +/- | +/- | +/- |
| 401 | + | + | - | +/- | + | + | +/- | +/- | +/- | +/- |
| 410 | + | + | - | + | +/- | - | + | - | +/- | +/- |
| 416 | + | + | - | +/- | +/- | + | + | + | + | + |
| 420 | + | + | - | +/- | + | +/- | - | +/- | +/- | + |
| 423 | + | + | - | - | + | + | + | +/- | +/- | +/- |
| N | - | - | +/- | +/- | +/- | ++ | +/- | + | +/- | +/- |

a Only the most abundant isoform of each gene was evaluated

b ++ high expression, + normal expression, +/- low expression, - no expression

N: normal bone marrow

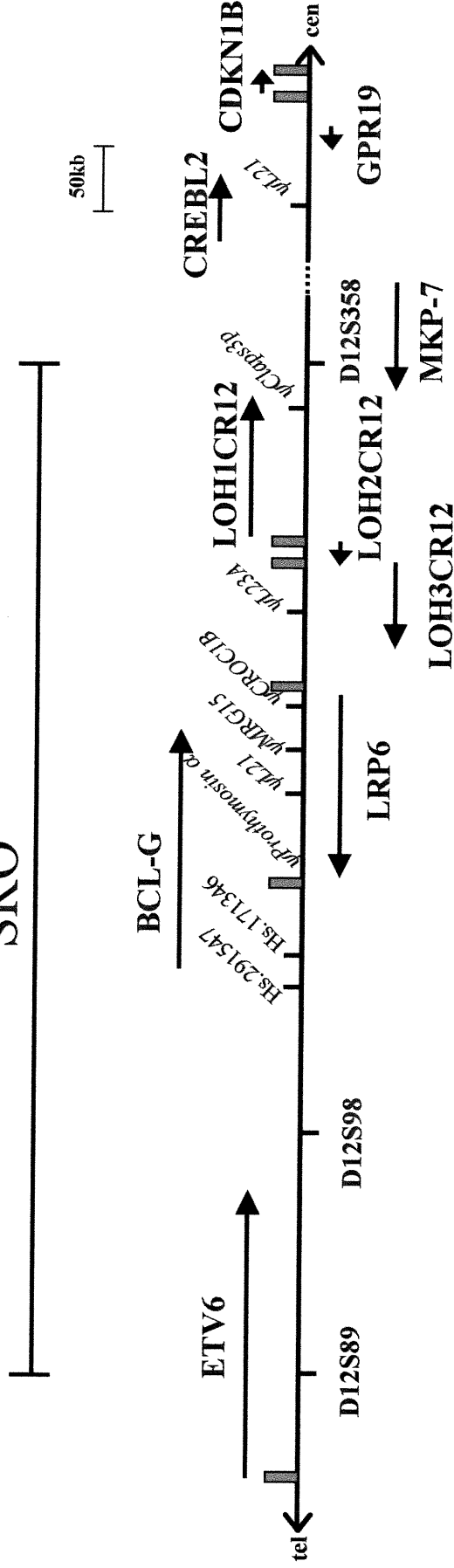
Table 6 Correlation of ETV6, BCL-G_L, MKP-7 and CDKN1B mRNA expression with clinical and biological findings in 26 pre-B LLA patients

| Findings | No. of patients examined | No. of patients with decreased expression (%) | | | |
|------------------------|--------------------------|---|----------------|----------------|----------------|
| | | ETV6 | Bcl-G | MKP-7 | CDKN1B |
| Age | | | | | |
| <1 yr, >10yr | 3 | 0 (0%) | 0 | 1 | 0 |
| >1 yr, <10yr | 23 | 10 (38%) | 11 | 7 | 5 |
| Sex | | | | | |
| M | 16 | 7 | 7 | 7 | 4 |
| F | 10 | 3 | 4 | 1 | 1 |
| WBC | | | | | |
| >50X10 ⁹ /L | 6 | 4 | 3 | 3 | 2 |
| <50X10 ⁹ /L | 20 | 6 | 8 | 5 | 3 |
| Risk group | | | | | |
| High | 13 | 5 | 5 | 5 | 2 |
| Low | 13 | 5 | 6 | 3 | 3 |
| DNA index | | | | | |
| >1.00 | 5 | 0 | 2 | 0 | 0 |
| 1.00 | 19 | 9 | 9 | 7 | 4 |
| LOH status | | | | | |
| + | 9 | 8 ^a | 6 | 7 ^b | 4 ^c |
| - | 17 | 2 | 5 | 1 | 1 |
| t(12;21) | | | | | |
| + | 10 | 10 ^d | 6 ^c | 6 ^e | 4 |
| - | 16 | 0 | 5 | 2 | 1 |

^a P = 0.002, ^b P = 0.004, ^c P = 0.03, ^d P < 0.0001, ^e P = 0.02

Figure 1 Physical and transcriptional map of the commonly deleted region on chromosome 12p12.3. The shortest region of overlapping deletion (SRO) is delimited by markers *DI2S89* and *DI2S358*. Filled rectangles indicate the presence of a CpG island. Identified genes are indicated in bold. Arrows indicate the orientation of transcription. Other EST clusters (Hs. 291547 and Hs.171346) found in this study are also indicated on the map. Pseudogenes are indicated in italics. The position and size of the genomic clones, as well as their corresponding accession numbers are shown in the lower section. The dotted lines indicate partial and unordered genomic sequences. Tel, telomere; cen, centromere.

SRO



ETV6 contig

NT_000600 (281 kb) NT_000601 (191 kb)

267J23
(AC007537; 191 kb)

253H9
(AC007619; 160 kb)

AC005989
(0.2 kb)

525I13
(AC022222; 200 kb)

180M15
(AC008115; 158kb)

Figure 2 SSCP analysis of exon 3 of BCL-G in 13 ALL patients with 12p12 LOH. Each PCR reaction contained 5ng of genomic DNA, 1 μ Ci of α -³²P dCTP and 1 μ M primers described in Table 1. PCR products were denatured and resolved by electrophoresis on a 6% non-denaturing gel at 500V for 20 hours. The numbers refer to the patients from which DNA was obtained. N and T correspond to normal and tumoral bone marrow DNA respectively.

3 5 6 8 15 24 26 27 400 392 375 34 401

N T N T N T N T N T N T N T N T N T N T N T N T

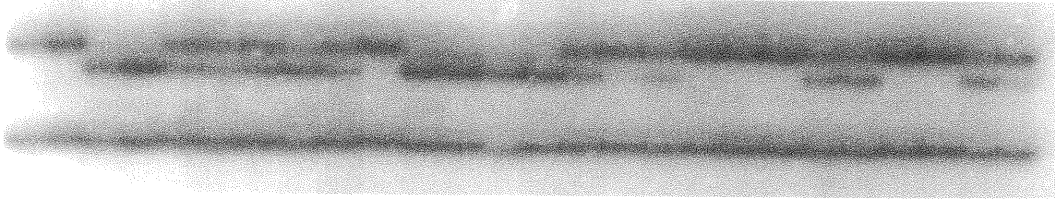


Figure 3 ETV6, BCL-G, MKP-7 and CDKN1B mRNA expression in 20 pre-B ALL patients. RT-PCR was performed as described in Materials and Methods and with primers described in Table 2. PCR products were electrophoresed on a 6% denaturing gel for 4 h at 1000V. The numbers refer to the patients from which DNA was obtained. Short and long indicate the transcripts corresponding to the short and long BCL-G protein products respectively.

Chapitre 6

-Discussion-

6.1 État de la question

Nos travaux ont surtout porté sur l'identification et la caractérisation d'un locus suppresseur de tumeurs qui semble jouer un rôle important dans la leucémie lymphoblastique aiguë de l'enfant. L'identification de gènes suppresseurs de tumeurs est une étape importante pour la compréhension du cancer au niveau moléculaire ainsi que pour le développement de molécules thérapeutiques. Il existe très peu d'information sur les gènes suppresseurs de tumeurs impliqués dans le développement de la leucémie. Par contre, il est connu que dans les leucémies, comme dans tous les cancers, les mécanismes moléculaires de leur développement sont les mêmes, y compris l'activation d'oncogènes et l'inactivation de gènes suppresseurs de tumeurs.

La région 12p12 est apparue comme étant la plus intéressante puisqu'elle est délétée chez 35 % des patients atteints de LLA de type pré-B (Bacchichet *et al.* 1997a). D'autres groupes ont également observé la délétion de cette région chez les patients leucémiques dans 26-40 % des cas (Takeuchi *et al.*, 1996; Cavé *et al.*, 1997; Chambon-Pautas *et al.*, 1998). Cette aberration chromosomique est la plus fréquente observée chez ces patients. Cette même région est également affectée dans de nombreux autres cancers hématologiques et tumeurs solides, y compris le cancer du poumon, de la prostate et des ovaires, le taux y variant de 26 à 33 % (Tableau I-IV). Le fait que la région 12p12.3 soit si fréquemment délétée dans plusieurs types de cellules cancéreuses indique probablement qu'un important gène suppresseur de tumeur s'y trouve. Dans notre laboratoire, nous avons observé que la région minimale de délétion, qui est commune à

tous les patients affectés, se situait dans une région d'environ 750 kb entre les marqueurs *D12S89* et *D12S358* (Figure 1-6). Cette région coïncide avec celles observées par la plupart des autres études. Nous avons donc considéré tout au long de cette étude que cette région incluait le gène suppresseur de tumeurs recherché. Cependant, il faut noter qu'une seconde région de délétion a été observée par deux groupes entre les marqueurs *D12S358* et *D12S320*. Ce point sera discuté plus en détail à la Section 6.7.

Afin d'identifier le ou les gènes suppresseurs de tumeurs impliqués dans le développement de cette leucémie et/ou d'autres cancers, l'étape limitante est l'identification des gènes présents dans la région commune de délétion. En 1998, selon les cartes physiques disponibles, seulement un gène, *ETV6*, impliqué dans les processus d'angiogenèse et d'hématopoïèse, était présent entre les marqueurs *D12S89* et *D12S358* (Figure 1-6). *LRP6*, codant pour un membre de la famille des récepteurs de lipoprotéines et agissant comme un co-récepteur de protéines WNT, y a été assigné par la suite (Baens *et al.*, 1999).

6.2 Analyse comparative du locus 12p12.3 chez *Fugu rubripes*

La recherche de gènes dans une région d'environ 750 kb chez l'humain est une tâche très complexe si l'on ne connaît pas *a priori* sa séquence. À partir de 1998, plusieurs approches ont été utilisées dans notre laboratoire (telles que décrites à la Section 1.6) dans ce but. Nous avons, entre autres, proposé comme hypothèse que l'utilisation du poisson-lune *Fugu rubripes* comme modèle compact d'un génome de vertébré faciliterait l'identification des gènes. *Fugu rubripes* a déjà été proposé comme

organisme modèle pour le clonage positionnel de gènes impliqués dans des maladies (Section 1.6.4).

Ce choix est basé sur le fait que *Fugu* possède un génome beaucoup moins complexe que celui de l'humain (son génome est 7,5 fois plus petit), mais avec une quantité similaire de gènes et qu'il possède de grandes régions synténiques avec le génome humain (i.e. régions chromosomiques possédant les mêmes gènes chez les deux espèces) (Section 1.6.4). Le génome de *Fugu* contient moins de 10 % de séquences répétées et près de 20 % est codant, comparativement à moins de 3 % pour celui des mammifères. On croit que *Fugu* possède un mécanisme qui restreint la propagation des séquences mobiles, mais aussi qui permet leur élimination du génome (revu par Venkatesh *et al.*, 2000). Notre hypothèse était que l'analyse de la région synténique à 12p12.3 chez *Fugu* serait beaucoup plus simple à caractériser et permettrait ainsi d'identifier les gènes présents plus facilement et à un moindre coût. On estimait que la région critique de 750 kb, délimitée par les marqueurs *D12S89* et *D12S358*, pourrait être contenue, avec tous les gènes, dans une région de moins de 100 kb chez *Fugu*, donc à l'intérieur d'un seul clone BAC.

Nous avons donc construit une carte physique d'environ 200 kb couvrant le locus *ETV6* du génome de *Fugu*. Suite à sa caractérisation, nous avons pu confirmer le clonage de l'orthologue de *ETV6* chez *Fugu* (*fETV6*). Cependant, malgré plusieurs essais, nous n'avons pas réussi à identifier des homologues correspondants aux autres gènes liés à *ETV6* chez l'humain, tels *LRP6* ou *CDKN1B* (Figure 1-6). Il faut souligner cependant

que certains gènes ont une identité trop faible pour pouvoir les identifier par hybridation. Même les bases de données de séquences de *Fugu rubripes* ou de *Tetraodon nigroviridis* les plus récentes ne contiennent pas actuellement d'homologues pour ces gènes. Grâce à une collaboration avec le laboratoire de Ben F. Koop de Victoria, via le réseau canadien CGDN (Canadian Genetic Disease Network), environ 80 % de la séquence nucléotidique du locus a été obtenue par la méthode de séquençage shotgun. Après l'assemblage de cette séquence, l'utilisation de programmes informatiques comparant la séquence d'ADN avec les bases de données de séquences exprimées (ESTs) et les séquences traduites avec des protéines connues a mené à l'identification de sept gènes en plus de *ETV6*.

Quatre de ces sept gènes possèdent des orthologues connus chez l'humain, soit *Timp3*, *SynIII*, *SLC16A7* et *KUB3*. *Timp3*, encodant un inhibiteur de métalloprotéinases, est retrouvé dans un des introns de *SynIII*, encodant une synapsine, protéine régulant l'exocytose des vésicules synaptiques, sur le brin opposé. Cette même organisation est observée dans les trois autres "couples" *Syn-Timp* homologues des vertébrés ainsi que chez l'unique orthologue de la drosophile (Pohar *et al.*, 1999; Kao *et al.* 1999). Cet exemple appuie l'hypothèse, toujours controversée, de l'existence de deux événements de duplication complète du génome lors de la transition invertébrés-vertébrés. Puisque des cartes génétiques existent pour zebrafish, contrairement à *Fugu*, nous avons cartographié *ETV6* ainsi que *CDKN1B* chez ce téléoste avec l'hypothèse que l'organisation chromosomique serait similaire à *Fugu*. *ETV6* a été identifié sur le groupe de linkage 4 (LG4) tandis que *CDKN1B* a été identifié sur LG25, soit deux chromosomes différents. L'absence de synténie entre ces deux gènes chez zebrafish indique que c'est

probablement le cas chez *Fugu*. De plus, chez zebrafish, *ETV6* a été cartographié à proximité d'une séquence ayant une forte identité avec *SynIII* renforçant l'hypothèse précédente. Cependant, chez l'humain, les orthologues de *Timp3* et de *SynIII* se retrouvent sur le chromosome 22 indiquant un bris de synténie à cet endroit chez *Fugu*.

En analysant les données de séquence du génome humain, nous avons identifié pour au moins trois des cinq autres gènes un homologue sur le chromosome 12 indiquant une conservation de la synténie entre les deux espèces. Les orthologues humains de *SLC16A7* et de *KUB3* ont été facilement identifiés à cause de leur grande identité de séquence. Pour l'orthologue de *LIG-2*, la situation est plus difficile à cause de l'existence d'au moins trois homologues chez l'humain. La plus forte identité de séquence cependant est observée dans une région du chromosome 12 où aucun gène n'a été caractérisé. La présence d'une séquence très conservée non-codante située dans un des introns ne se retrouve que dans l'orthologue du chromosome 12, enlevant tout doute quant à l'identité de l'orthologue réel. L'importance de cette région conservée sera discutée dans la prochaine section. Chez *Fugu*, les trois gènes, *SLC16A7*, *KUB3* et *LIG-2*, sont situés dans une région d'environ 50 kb tandis que chez l'humain, ceux-ci sont séparés par environ 1,5 Mb. Cependant, après analyse, nous avons observé qu'il existe d'autres gènes chez l'humain qui ne sont pas présents dans la région de 50 kb chez *Fugu*. Ceci peut suggérer que ces gènes ne sont pas importants chez *Fugu* et qu'ils ont été perdus, ou qu'à la suite d'un réarrangement, ils se retrouvent sur un autre chromosome ou bien que le séquençage n'ait pas permis leur identification. Cette dernière hypothèse est peu probable puisqu'environ 75-80 % de cette région a été couverte de façon aléatoire.

Si l'on compare la position chromosomique correspondante chez l'humain de ces trois gènes plus *ETV6* avec les résultats d'une autre étude, effectuée chez *Fugu* par Power *et al.* (2000), comprenant quatre autres gènes localisés sur le chromosome 12, l'hypothèse des réarrangements devient la plus plausible. En effet, après analyse de la position exacte des huit orthologues sur le chromosome 12 (la résolution des cartes physiques est maintenant de 200 kb soit la taille d'un BAC), on observe que chacun des deux groupes de quatre gènes sont distribués sur les deux bras du chromosome 12 et avec plusieurs inversions, indiquant de nombreux réarrangements et pertes ou gains de gènes entre les deux espèces (voir Figure 1 dans le Chapitre 2).

Finalement, les deux autres gènes identifiés dans notre étude possèdent également des homologues humains, mais de fonction inconnue. Le premier possède deux homologues situés sur les chromosomes 1 et 3 (*LOC128320* et *BAB71606* respectivement), tandis que le second possède un homologue sur le chromosome 7 et un sur le chromosome 12, soit *C12ORF6*. Cependant, cet homologue ne possède que 7 exons comparativement à 12 pour le gène de *Fugu* et de l'homologue du chromosome 7. L'analyse de la séquence en amont de *C12ORF6* n'a pas permis de révéler d'autres exons codants (données non-montrées). Par contre, étant donné que l'identité de séquence au niveau protéique est très faible comparativement aux autres gènes identifiés dans la région, soit environ 45 à 47 %, il est possible que le véritable orthologue n'ait pas été identifié. Dans ces deux cas, aucune conclusion quant à la conservation de la synténie avec les autres gènes ne peut être avancée.

Cette analyse démontre bien que *Fugu* n'est pas un modèle de choix pour le clonage positionnel de gènes, soit les mêmes conclusions avancées par Gilley et Fried (1999). De nombreux réarrangements ont eu lieu depuis la divergence des derniers ancêtres communs aux humains et aux téléostes. Ceci a plusieurs conséquences: 1) présence de plusieurs brisures de synténie entre les chromosomes humains et de *Fugu*; 2) même s'il y a conservation de la synténie, il est probable que l'ordre des gènes soit différente chez l'humain. La plupart des études faisant mention d'une conservation de la synténie entre *Fugu* et l'humain, dont l'étude de Power *et al.* (2000), n'ont pu vérifier que l'ordre des gènes était bel et bien conservé puisqu'ils n'avaient pas accès à des cartes physiques ayant une assez haute résolution, soit d'environ 200 kb. Celles-ci ne sont apparues que vers le début de 2001. Une récente analyse utilisant la position génomique d'une vingtaine de gènes a révélé que entre l'humain et la souris environ 180 réarrangements avaient dû se produire contre environ environ 8000 à 32 000 entre l'humain et *Fugu*, ce qui est beaucoup plus qu'une estimation faite par l'extrapolation des données souris-humain (McLysaght *et al.*, 2000). L'hypothèse la plus probable est qu'il y aurait eu une duplication additionnelle du génome des téléostes, par rapport aux autres vertébrés, suivie d'une perte massive de gènes. Cette hypothèse est appuyée par plusieurs exemples, dont la présence de huit locus HOX chez tous les téléostes dont zebrafish et *Fugu*, contrairement à quatre chez tous les autres vertébrés (Armores *et al.*, 1998). La perte de gènes redondants subséquente aurait eu de grandes conséquences sur l'organisation des chromosomes, en conservant une certaine synténie, mais en changeant considérablement l'ordre des gènes.

Le génome de *Fugu* est d'une grande utilité pour l'identification de séquences associées à des fonctions. La comparaison de la séquence complète du chromosome 22 humain avec seulement 33 % de la séquence génomique du poisson-lune *Tetraodon nigroviridis*, possédant plus de 98 % d'identité de séquence avec *Fugu*, a permis d'identifier 66 % des gènes humains connus en plus d'en identifier plusieurs autres en comparant seulement les séquences traduites et avec une spécificité de plus de 95 % (Roest-Croellius *et al.*, 2000). Dans notre étude, nous avons pu caractériser un gène inconnu, soit *LIG-2*, indiquant que l'utilisation du génome des poissons-lune est un outil important pour la caractérisation du génome humain.

6.3 Analyse comparative de ETV6 chez Fugu

Puisque *ETV6* est un gène important dans le développement du système hématopoïétique et que son implication dans la leucémogénèse a été démontrée, nous avons caractérisé en détail son orthologue chez *Fugu*, *fETV6*. Comme mentionné précédemment, à cause du fait que *Fugu* et l'humain soient séparés par plus de 430 millions d'années d'évolution divergente, en principe seules les séquences fonctionnellement importantes devraient être conservées. Notre hypothèse était que la caractérisation de *ETV6* par analyse comparative chez *Fugu* nous permettrait de caractériser plus précisément les régions importantes pour la fonction de ce gène.

fETV6, du codon initiateur ATG jusqu'au codon stop, mesure 15 kb, soit une réduction d'environ 16 fois par rapport au gène humain. Les tailles des huit exons ainsi

que les jonctions exon/intron sont très bien conservés chez *Fugu*, indiquant que c'est bel et bien l'orthologue de *ETV6*. Les introns sont quant à eux réduits de 6 à 96 fois par rapport à leur contrepartie humaine (Table 1 dans le Chapitre 3). L'intron 5 est remarquablement petit chez *Fugu*, 119 pb comparativement à 11 456 pb chez l'humain. Cependant, certains introns sont plus grands, notamment l'intron 2 qui, avec une taille de 8185 pb, est plus long que la moyenne de 150 pb observée chez *Fugu* (Elgar *et al.*, 1998). On observe d'ailleurs dans cet intron une séquence très conservée entre les deux espèces. Cette région a probablement une importance fonctionnelle puisqu'elle a été maintenue au cours de plus de 400 millions d'années d'évolution. Une telle séquence conservée avait été observée également dans le second intron de *LIG-2* (voir section 6.2). On pense que ces séquences conservées non-codantes pourraient jouer un rôle au niveau de la transcription du gène. Alternativement, il pourrait s'agir de séquences contrôlant l'épissage ou la stabilité du messager. Ceci demeure à être confirmé expérimentalement dans le cas de *ETV6* et de *LIG-2*, mais il existe certains exemples dans la littérature qui appuient la première hypothèse. Il a été démontré que deux séquences conservées entre *Fugu* et la souris dans une région située en 3' du gène *Hoxb-1* jouaient le rôle d'éléments activateurs de la transcription ou "enhancers" et assuraient son expression dans le neur ectoderme lorsque introduites dans des souris transgéniques (Marshall *et al.*, 1994). Aussi, Aparicio *et al.* (1995) ont identifié une petite région homologue avec *Fugu* (environ 90 pb) dans un élément activateur situé dans un intron de *Hoxb-4*. La délétion de cette région a aboli l'expression dans le mésoderme et l'ectoderme indiquant l'utilité de *Fugu* pour délimiter de façon précise des régions fonctionnelles. Le remplacement du promoteur ou des éléments activateurs murins de

Hox-b4 par ceux de *Fugu* ne change pas le profil d'expression facilitant ainsi l'identification des séquences régulatrices (Aparicio *et al.*, 1995).

La comparaison du promoteur de *ETV6* de *Fugu* et de l'humain nous a permis d'identifier des sites de liaison putatifs pour les facteurs de transcription GATA, STAT, AML-1, ETS et MZF1. Quoique les protéines activant la transcription de *ETV6* ne sont pas connues actuellement, tous ces facteurs sont actifs dans les processus d'angiogenèse et/ou d'hématopoïèse. Les facteurs GATA, ETS, AML1, MZF1 et la voie JAK/STAT sont tous importants dans le contrôle de la différenciation et de la croissance des précurseurs hématopoïétiques (revu dans Nagamura-Inoue *et al.*, 2001). De plus, la voie JAK/STAT est active durant l'angiogenèse (Valdembri *et al.*, 2001). Cette comparaison permet de plus facilement délimiter le nombre de séquences consensus de liaison à considérer, ce qui est normalement un très grand désavantage de ce type d'analyse informatique. Le promoteur de *Fugu*, contrairement à celui de l'humain, ne semble pas contenir de sites de liaison pour le facteur Sp1, normalement présents lorsque le gène ne possède pas de boîte TATA (Baens *et al.*, 1996). Il est possible que *Fugu* utilise une autre voie afin de recruter la machinerie transcriptionnelle de base.

Au niveau de la protéine, 58 % d'identité et 72 % de similarité sont observées entre les deux espèces. On note que les domaines PNT, d'homo- et d'hétéro-dimérisation, et ETS, de liaison à l'ADN, sont également très conservés chez *Fugu* avec 81 et 95 % d'identité respectivement. L'oligomérisation de *ETV6* via son domaine PNT est essentielle à toutes ses activités, y compris la liaison à l'ADN et la répression de la

transcription. Cependant, le domaine responsable de cette dernière activité est beaucoup moins bien caractérisé. Ce domaine a été cartographié dans la section centrale de la protéine humaine entre les acides aminés 171-285 par des études de délétions quoiqu'une seconde région, incluant le domaine ETS ainsi que les 50 acides aminés précédents, a également été identifiée (Lopez *et al.*, 1999). Il est connu que le domaine ETS peut effectuer des contacts protéine-protéine en plus de lier l'ADN. Il a été montré également que la section centrale peut lier SMRT, mSin3A et N-CoR tous formant un complexe de répression de la transcription en s'associant avec des histones déacétylases (Chakrabarti et Nucifora, 1999; Guidez *et al.*, 2000). En comparant les séquences de *ETV6* provenant de l'humain, la souris, le poulet, *Fugu* et zebrafish, nous avons mis en évidence trois régions bien conservées, dénotées boîtes A, B et C, chez ces espèces suggérant une importance fonctionnelle. D'ailleurs, il aurait été impossible de faire cette observation en utilisant seulement les séquences de souris ou de poulet étant donné la trop grande conservation en général. On estime que les lignées de l'homme et de la souris auraient divergées il y a environ 100 millions d'années (Myr), les mammifères et les oiseaux, 220 Myr et les poissons et les autres vertébrés, 430 Myr (Venkatesh *et al.*, 2000). Les deux téléostes, *Fugu* et zebrafish, auraient divergé il y a plus de 100 Myr, et leurs séquences dans l'exon 5 situées à l'extérieur de ces trois boîtes sont très différentes ajoutant ainsi du poids à l'importance fonctionnelle de ces régions. Une étude récente a permis de délimiter plus précisément la région responsable de la répression de la transcription entre les acides aminés 268 et 303, incluant la boîte C, et s'associant uniquement avec l'histone déacétylase 3 (Wang et Hiebert, 2001). Il est donc possible que *ETV6* fasse des contacts

avec plusieurs co-répresseurs faisant partie d'un même complexe et qu'il possède plusieurs sites d'interaction distincts.

Il est connu que ETV6 est phosphorylé *in vivo* (Poirel *et al.*, 1997). Plusieurs protéines ETS, dont Ets-1 et YAN, sont phosphorylées via la voie de Ras/MAPK. ETV6 possède trois sites consensus de phosphorylation par les MAPK (P-X₍₁₋₂₎-T/S-P), dont un situé à une position homologue au site de phosphorylation de Ets-1 tout juste en N-terminal du domaine PNT. On remarque, à la Figure 2 du Chapitre 3, que ce site n'est pas conservé chez *Fugu*. De plus, une forme tronquée de ETV6, dont la traduction est initiée au second AUG du messenger, est toujours phosphorylée (Poirel *et al.*, 1997). Le seul site conservé chez toutes les espèces est situé dans la section centrale. Cependant, il est possible que les autres sites soient fonctionnels du moins chez les espèces supérieures ou que ETV6 soit phosphorylé via d'autres voies de signalisation.

L'analyse comparative avec des espèces vertébrées relativement éloignées de l'homme permet donc d'identifier des régions fonctionnelles, qu'elles soient codantes ou non-codantes, puisque seules celles-ci sont conservées après plusieurs centaines de millions d'années d'évolution. L'utilisation de la souris comme modèle comparatif permet bien d'identifier des gènes humains à partir de séquences génomiques. Cependant, la délimitation de domaines protéiques importants est plus difficile puisque le bruit de fond, c'est-à-dire le degré global de conservation, est trop important. À l'opposé, l'utilisation de modèles invertébrés tels la drosophile ou *C.elegans* sont trop distants de l'homme pour identifier la majorité des domaines protéiques fonctionnels. De

plus, un autre avantage de *Fugu* vis-à-vis d'autres poissons est son petit génome, ce qui diminue les coûts associés à son séquençage. La séquence de son génome sera complétée sous peu en faisant un outil de comparaison important pour la compréhension du génome humain.

6.4 Analyse de la séquence génomique humaine du locus 12p12.3

En juin 2000, deux organisations, soit le Consortium International du Séquençage du Génome Humain, une organisation publique, et Celera, une compagnie privée, ont annoncé la complétion du génome humain, leurs résultats étant publiés au début de 2001 (Lander *et al.*, 2001; Venter *et al.*, 2001). Les données publiques couvrent environ 90-95 % du génome, mais seulement la moitié environ de façon ordonnée et annotée pour la présence de gènes connus. Dans la région de notre suppresseur de tumeurs sur le chromosome 12, environ 90 % de la séquence est maintenant connue et plus de la moitié de façon ordonnée. Comme aucune technique d'analyse, informatique ou moléculaire, n'est assez sensible et précise pour identifier tous les gènes dans une région donnée, nous avons combiné les résultats de diverses approches. Sept gènes ont été identifiés dans la région critique d'environ 750 kb entre les marqueurs *D12S89* et *D12S358*: *ETV6*, *BCL-G*, *LRP6*, *LOH1CR12*, *LOH2CR12*, *LOH3CR12* et *MKP-7*.

L'analyse EXOFISH, soit la comparaison des séquences traduites provenant de l'humain et de Tetraodon (Roest-Croellius *et al.*, 2000), a permis l'identification de quatre des sept gènes. Aucun faux positif, à part tous les pseudogènes, n'a été identifié confirmant l'utilité de cette méthode. Il est à noter que seulement quelques exons de

chaque gène ont été identifiés. Comme la majorité des séquences traduites sont relativement bien conservées pour être identifiées par comparaison informatique (programme BLAST), on devrait s'attendre à retrouver deux tiers des gènes (ce qui est observé) ainsi qu'un tiers du nombre d'exons au total (puisque un tiers du génome du poisson a été séquencé). Le double de ce nombre est observé, mais il faut noter que l'échantillon est petit avec seulement quatre gènes.

L'approche de capture d'exons fut moins efficace que prévu. Seulement trois gènes ont pu être identifiés de cette façon, avec seulement sept exons (sur une possibilité de plus de 40) et avec également plusieurs faux positifs (ne pouvant générer un cadre de lecture ouvert). Les faux positifs sont souvent observés dans cette technique puisque le contexte génomique est différent (présence de sites accepteurs et donneurs du vecteur) et que certaines régions clonées ne sont normalement jamais transcrites incitant la cellule à utiliser de "faux" sites d'épissage. La faible détection d'exons pourrait s'expliquer par une représentation incomplète de toute la région génomique dans les clones, puisqu'il n'y a pas de sélection pour les séquences normalement transcrites en plus faible quantité. Le nombre de clones de 2 kb nécessaires à une couverture adéquate d'une région de près de 1 Mb est d'au moins 2000 ce qui rend la tâche d'analyse longue et ardue même après avoir combiné ensemble plusieurs de ceux-ci lors des transfections.

L'analyse informatique seule de séquences génomiques doit être effectuée avec précaution en étant conscient des limites de cette approche. Une analyse récente démontre que les différents logiciels d'analyse, qui se basent sur la présence de sites

consensus d'épissage et de cadres de lecture ouverts, ne permettent de détecter qu'environ 40 à 50 % des gènes avec une spécificité similaire (Bursset et Guigo, 1996). Par contre, l'utilisation combinée de plusieurs outils d'analyse informatique permet d'augmenter la validité des résultats obtenus. Nous avons utilisé le programme NIX qui permet de comparer les résultats de six différents logiciels d'analyse génomique. Par cette approche, nous avons pu identifier la majorité des exons de chaque gène. L'utilisation combinée de ces logiciels a permis également de définir avec précision toutes les jonctions exon-intron et ainsi définir un cadre de lecture ouvert complet, ce qui a éventuellement été confirmé par RT-PCR et le séquençage. La plupart des faux-positifs ne sont prédits que par une minorité des logiciels contrairement aux vrais-positifs, ce qui a facilité la tâche d'identification des gènes.

Cependant, sans l'analyse BLAST, sur la base de données de protéines et sur la base de données de séquences exprimées (EST), l'identification complète des gènes aurait été presque impossible. La comparaison avec des domaines protéiques connus a bien sûr permis l'identification des gènes connus, soit *ETV6* et *LRP6*, en plus d'identifier certains domaines de *BCL-G* et de *MKP-7* avant que ceux-ci ne soient caractérisés par d'autres équipes (Guo *et al.*, 2001; Tanoue *et al.*, 2001). Cependant, cette analyse ne permet pas d'identifier les familles de gènes ne possédant aucun domaine connu, comme *LOH1CR12*, *LOH2CR12* et *LOH3CR12*, retrouvés dans la région critique. On estime que près de la moitié des gènes du génome humain entrent dans cette catégorie (Lander *et al.*, 2001). La comparaison avec la base de données EST, provenant non seulement de l'humain, mais aussi d'autres espèces telles la souris, a généré beaucoup d'informations.

Dans notre analyse, nous avons observé plusieurs groupes d'ESTs n'ayant aucun potentiel codant ou signal de polyadénylation et dont leur queue de polyA se retrouvait également dans la séquence génomique correspondante. Ceci indique essentiellement que ce sont des artefacts générés suite à l'hybridation d'un oligo polyT sur une suite de A répétés présents dans l'ADN génomique ou de l'ARN non mature. Ces contaminants sont toujours présents dans les préparations cytoplasmiques même lorsque traitées à la DNase. Il est possible que la séquence répétée riche en A se retrouve à l'intérieur de l'ARN mature. Cependant, il est très rare qu'elle se retrouve dans la région codante et on en retrouve plutôt dans les régions 3'UTR où elle peut être facilement alignée avec d'autres ESTs.

Après avoir éliminé ce type de séquence, soit près de 80 % du total, seulement 20 groupes d'ESTs ont été conservés (Table 1 dans le Chapitre 4). Huit de ces groupes possédaient des séquences épissées par rapport à l'ADN génomique et correspondent aux régions codantes des sept gènes (deux des groupes correspondaient à deux variants d'épissage du même gène). Ces séquences épissées sont très révélatrices puisqu'elles ne peuvent correspondre à des artefacts de préparation et que la plupart du temps elle correspondent à des séquences codantes. Aussi, le fait que la grande majorité des ESTs soient amplifiés à partir de leur queue de polyA crée un large biais en faveur de séquences situées en 3'. Comme la taille de la région 3'UTR peut être très longue, soit plusieurs kb, il arrive souvent qu'aucune séquence codante ne soit identifiée par cette approche ou que plusieurs exons soient manquants. Il est alors nécessaire d'utiliser le 5'RACE pour identifier ces séquences. Cette technique a permis avec succès d'identifier

les extrémités 5' de *BCL-G*, de *MKP-7*, mais pas celle de *LOH2CR12*. Dans ce cas, la transcriptase réverse MMLV ne semble pas pouvoir traverser une région riche en GC correspondant à l'îlot CpG du gène *LOH1CR12*, situé sur le brin complémentaire. L'utilisation d'une transcriptase réverse thermostable comme la rTth pourrait permettre de contourner ce problème en permettant à la réaction de se dérouler à une plus haute température et ainsi dénaturer les structures secondaires. Ni l'analyse informatique, ni l'analyse comparative n'ont permis d'identifier d'autres exons situés en 5' et ainsi la première méthionine de *LOH2CR12*.

Dix autres groupes d'ESTs ont été identifiés comme étant des extrémités 3' alternatives. En effet, par analyse Northern nous avons observé que *MKP-7* (donnée non-montrée), *LOH1CR12* et *LOH3CR12* possédaient au moins deux transcrits qui différaient de plusieurs kb, écartant essentiellement la possibilité que ce soit dû à de l'épissage alternatif. Il avait déjà été démontré, par protection à la RNase, que *ETV6* possédaient trois transcrits qui ne différaient que par la longueur de leur 3'UTR, donc possédant trois sites de terminaison de la transcription distincts (Baens *et al.*, 1996). Après analyse de l'ADN génomique correspondante pour chaque gène et amplification par PCR des régions situées entre les groupes d'ESTs (comparé à un contrôle sans transcriptase réverse et traité à la DNase), nous avons pu démontré l'existence de ces régions 3'UTR alternatives pour *BCL-G*, *MKP-7*, *LRP6*, *LOH1CR12* et *LOH3CR12*. Il semble que ce type de terminaison alternative soit très courant, mais on ne connaît pas très bien son utilité pour la cellule ni si celle-ci est régulée (revu dans Edwalds-Gilbert, 1997). En ce sens, on observe sur les Northern blots que la proportion relative de

chacune des bandes varie pour certains gènes, dont *LOH1CR12*, dans quelques tissus. Il est connu que eIF2A possède deux transcrits, de 1.6 et 2.4 kb, qui ne diffèrent que par la longueur de leur région 3'UTR. Le transcrit le plus court est traduit plus rapidement du moins *in vitro*, tandis que le plus long est plus stable (Miyamoto *et al.*, 1996). Le ratio de chacun des transcrits pourrait donc être régulé et correspondre aux besoins de la cellule.

L'expression des sept régions codantes putatives a été vérifiée par RT-PCR et hybridation Northern confirmant nos prédictions. Plusieurs variants d'épissage ont pu être identifiés suite au séquençage des différents transcrits. Seuls deux groupes d'ESTs, ne correspondant à aucun gène, n'ont pas été rejetés car ils possédaient un signal de polyadénylation en 3'. Le fait qu'ils ne soient représentés que par un ou deux ESTs suggèrent qu'ils sont en fait des artéfacts, ou qu'ils sont du moins très faiblement exprimés et que dans quelques tissus. Seuls des RT-PCRs effectués sur plusieurs tissus pourront le démontrer. Comme, il y a un biais dans la base de données EST pour les séquences les plus abondantes, il est possible que tous les gènes n'aient pas encore été identifiés de cette façon. Dans une cellule normale, la différence entre les transcrits les plus abondants et les moins abondants peut atteindre plusieurs facteurs de 10 et d'autres sont totalement absents dans certains tissus. Il est également possible que certains gènes n'aient pas encore été séquencés dans la région non-assemblée située entre *ETV6* et *BCL-G*. Cependant, cette région a été relativement bien couverte par séquençage shotgun et aucun EST susceptible de coder pour un gène n'y a été identifié. Nous croyons donc que nous avons identifié tous les gènes importants associés au locus suppresseur de tumeurs 12p12.3 qui est délété chez les patients atteints de leucémie lymphoblastique aiguë.

Cette étude est la première à tenter de déterminer la composition complète en gènes dans une région génomique humaine en utilisant une combinaison d'outils d'analyse bio-informatiques et moléculaires. La plupart des études récentes utilisent la séquence de l'ADN génomique humain, certaines bases de données et d'autres approches expérimentales, identifiant ainsi plusieurs gènes, mais en laissant plusieurs informations de côté (ex.: Wang *et al.*, 2000; Smith *et al.*, 2000; Boulwood *et al.*, 2000). D'après notre expérience, la méthode la plus simple possible afin d'identifier les gènes d'une région donnée serait premièrement de ne conserver que les groupes d'ESTs possédant soit des séquences épissées ou des signaux de polyadénylation en 3', seules garanties que ce sont de vraies séquences exprimées, et de les aligner sur la séquence génomique. On doit également s'assurer de bien aligner les variants d'épissage à cette étape afin de ne pas créer de groupes additionnels. Par la suite, l'utilisation des logiciels de prédiction d'exons permettra de définir les jonctions exons-introns et un cadre de lecture ouvert. Ensuite, il faudra ajouter une étape d'analyse comparative, avec EXOFISH et avec des gènes ou des domaines protéiques connus, afin d'identifier d'autres gènes ou de les compléter. On doit s'assurer à cette étape d'éliminer les pseudogènes qui normalement comportent plusieurs mutations abolissant le cadre de lecture ouvert et ne possèdent pas d'introns comparativement au gène fonctionnel. Toutes ces étapes pourraient se faire automatiquement. D'ailleurs, le site Web d'analyse du génome du consortium public a été amélioré récemment en présentant les données d'EXOFISH, de la comparaison des données de séquence génomique de la souris ainsi que l'alignement des séquences exprimées épissées (<http://genome.ucsc.edu/goldenPath/hgTracks.html>). À la suite de

cette analyse informatique, les prédictions devront être confirmées par RT-PCR, dans plusieurs tissus, puis séquençage, et les gènes complétés par 5'RACE du moins jusqu'à la première méthionine. L'analyse par hybridation Northern sera également utile permettant d'identifier la taille de tous les transcrits. Après cette analyse, seuls les gènes faiblement exprimés ou d'expression restreinte ainsi que les gènes encodant des ARNs non traduits pourraient être manqués.

Les nouveaux gènes identifiés dans cette étude sont des candidats comme suppresseur de tumeurs dans les leucémies en raison de leur fonction putative ou de leur profil d'expression. *BCL-G* est un nouveau membre de la famille BCL-2. Il a été démontré qu'il possédait une activité pro-apoptotique dans les cellules en culture (Guo *et al.*, 2001). L'isoforme long (*BCL-G_L*) a un potentiel apoptotique moins grand que l'isoforme court (*BCL-G_S*) puisqu'il possède un domaine conservé supplémentaire BH2 ayant un effet inhibiteur sur cette fonction. L'isoforme médian (*BCL-G_M*), identifié dans notre étude, devrait par conséquent avoir une activité similaire à l'isoforme court puisqu'il aucun domaine connu supplémentaire n'est encodé par les exons 7 à 10. MKP-7 est une phosphatase à spécificité double ciblant les tyrosine ainsi que les sérine/thréonine. Elle cible spécifiquement les kinases Jnk/Sapk et p38 Mapk (Tanoue *et al.*, 2001). Ces kinases sont activées durant la croissance cellulaire. *BCL-G* et MKP-7 sont donc des candidats très prometteurs puisque les activités apoptotique ou phosphatase sont souvent diminuées ou supprimées dans les cancers (Tableau I-III). LOH1CR12 possède un domaine ayant une certaine homologie, quoique faible, avec un

domaine spectrine de la protéine TRIO. Ces domaines permettent normalement l'association au cytosquelette (Seipel *et al.*, 1999). LOH3CR12 quant à elle possède un domaine d'homologie syndecan. Les syndecans sont des proteoglycans membranaires, impliqués dans le contrôle de la croissance cellulaire et la différenciation (revu par Rapraeger, 2001). Aussi ce domaine serait impliqué dans la liaison avec la protéine kinase C et PIP2 pour la signalisation intracellulaire (e.g. Lee *et al.*, 1998). L'analyse de la séquence protéique à l'aide du programme PSORT (Nakai et Horton, 1999), qui détecte les signaux de localisation subcellulaire, permet de prédire que LOH1CR12 est cytoplasmique et que LOH3CR12 est membranaire. Pour LOH2CR12, l'absence d'une séquence complète (on ne connaît que 30 acides aminés) ne nous a pas permis de faire de prédictions. Par contre, tous ces gènes sont au moins exprimés dans la moelle osseuse. Des études avec des anticorps pourraient être effectuées afin d'en savoir plus sur la localisation sub-cellulaire de ces protéines. Pour ce qui est de leur fonction réelle, étant donné que très peu d'indices sont disponibles, un knock-out chez la souris, par recombinaison homologue, fournirait probablement plusieurs informations utiles. Étant donné que *LOH1CR12* possède des homologues chez *S.cerevisiae* et la drosophile, il pourrait être plus avantageux de travailler avec ces organismes dans ce cas.

6.5 Étude mutationnelle des gènes candidats chez les patients

Selon la théorie des "two-hits" de Knudson (1971), des mutations inactivantes sur les deux allèles d'un gène suppresseur de tumeurs devraient être présentes. Les premières études faites sur les patients atteints de leucémie lymphoblastique aiguë n'ont pu révéler de mutations inactivatrices dans *ETV6* ou *CDKN1B* (Stegmaier *et al.*, 1996). Des études

similaires chez des patients atteints du cancer de la prostate ont constaté également une absence de mutations dans l'allèle intact de ces deux gènes (Kibel *et al.*, 2000, 2001). Trois hypothèses peuvent expliquer ces observations: 1- un ou d'autres gènes sont les responsables; 2- les mutations inactivatrices ne sont pas des mutations ponctuelles présentes dans la région codante (exemple: méthylation de la région promotrice comme pour *p16* dans certains cancers); 3- le gène ciblé n'est pas un gène suppresseur de tumeur classique (un seul allèle perdu est nécessaire). Dans le premier cas, l'analyse de la présence de mutations dans les gènes identifiés dans la région critique devrait permettre de l'identifier. Dans le second cas, ce type de mutations devrait entraîner une diminution significative ou une absence du transcrit du gène et peut se vérifier par RT-PCR. Dans le troisième cas, il sera impossible de le prouver directement. Ce point sera discuté plus en détails à la prochaine section.

Nous avons donc analysé les régions codantes des sept gènes candidats, *ETV6*, *BCL-G*, *LOH3CR12*, *LOH2CR12*, *LOH1CR12*, *MKP-7* ainsi que *CDKN1B*, pour la présence de mutations somatiques chez des patients atteints de LLA. Le gène *LRP6* n'a pas été choisi initialement puisque tous les 23 exons de ce gène ont été analysés par SSCP par Baens *et al.* (1999) chez des patients atteints de LLA et qu'aucune mutation n'avait été identifiée. Neuf différents polymorphismes ont été observés, mais aucune mutation somatique n'a été identifiée, c'est-à-dire une séquence différente d'un contrôle provenant également de l'ADN extrait de la moelle chez le même individu, mais en rémission. Quatre de ces polymorphismes provoquent un changement d'un acide aminé dans leur séquence protéique respective, mais aucun de ceux-ci ne sont conservés ou

perdus spécifiquement dans les cas de délétion. Il est connu cependant que la technique de SSCP ne détecte que 80-90 % des variations. La sensibilité peut être plus faible lorsque les fragments sont plus longs que 300 pb (Hayashi et Yandell, 1993). Une autre technique de détection de mutations est le dHPLC, soit une chromatographie à haute pression dénaturante, dans laquelle les fragments possédant un ou plusieurs mauvais appariements, donc un polymorphisme, migrent différemment (revu par Xiao et Oefner, 2001). Nous l'avons utilisé dans quelques cas, mais nous n'avons pas identifié de polymorphismes supplémentaires (données non-montrées). Cette étude confirme les résultats des études précédentes et démontre qu'aucune mutation somatique dans une région codante, dans *ETV6*, *CDKN1B* ou tout autre gène, n'est associée à l'allèle non-déléte chez les patients hémizygotés pour la région 12p12.3. Nous avons utilisé ces nouveaux variants d'ADN afin d'augmenter la densité de marqueurs polymorphiques dans la région. De cette façon, nous avons pu réduire la région commune de délétion à environ 200 kb, entre *DI2S89* et le troisième exon de *BCL-G*, n'incluant plus que deux gènes: *ETV6* et *BCL-G*. Ce résultat est en ligne avec l'observation de Takeuchi *et al.*, (1996b) qui délimitait la région de délétion commune chez ces patients entre *DI2S77* et *DI2S98*, n'incluant que *ETV6*.

Devant la possibilité qu'un gène puisse être inactivé via la répression de son expression, nous avons développé un essai de RT-PCR semi-quantitatif afin de vérifier l'expression des gènes candidats dans les cellules leucémiques. Si un de ces gènes est le gène suppresseur de tumeurs recherché, son expression devrait être abolie ou fortement diminuée par un mécanisme quelconque puisqu'aucune mutation inactivante n'a été

identifiée dans les régions codantes. Nous avons évalué l'expression de chacun de ces gènes chez 26 patients leucémiques, dont neuf possédaient une délétion du chromosome 12p12. Étant donné qu'il est impossible d'obtenir un contrôle correspondant au niveau normal d'expression des cellules pré-B (la moelle osseuse normale contient beaucoup de précurseurs différents), le niveau d'expression des gènes a été normalisé en utilisant des patients leucémiques, mais ne possédant pas de délétion dans la région 12p12. Nous faisons donc la supposition que l'expression des gènes de la région qui nous intéresse n'est pas affectée dans ces cas. Nous avons établi qu'une absence d'expression correspondait à moins de 10 % du contrôle normal afin de tenir compte de la présence de cellules normales dans les échantillons. On peut alors observer que *ETV6* est absent dans huit des neuf patients possédant la délétion, mais dans aucun autre échantillon. D'autres gènes sont également absents dans quelques échantillons mais ne sont pas significativement associés aux délétions. Plusieurs gènes montrent une diminution d'expression, définie par moins de 50 % du niveau contrôle, associée de façon statistiquement significative aux délétions (Table 5 dans le Chapitre 5). Cependant, il est très probable que ces baisses soient dues à la perte d'un allèle via une délétion les englobant plutôt que d'être la cible directe de ces délétions. De plus aucune autre association significative ne peut être faite entre la baisse ou l'absence d'un gène et l'âge au diagnostic, le sexe, le groupe de risque ou le compte de cellules leucémiques (Table 6 dans le Chapitre 5). Toutes ces observations pointent vers le fait que *ETV6* est la cible des délétions. Par contre, dans les huit échantillons où *ETV6* est absent, on observe également la présence de la translocation t(12;21) générant le gène de fusion *ETV6-AML1*, indiquant que ces deux événements sont étroitement associés. Il est à noter que

les amorces utilisées pour le PCR ont été positionnés de part et d'autre de l'exon 5 de *ETV6* afin de ne pas amplifier le gène de fusion. De plus, dans deux autres échantillons possédant la translocation, mais pas la délétion, l'expression de *ETV6* y est fortement réduite, à environ 10-20 % du contrôle normal seulement. Ce résultat a été obtenu trois fois de façon indépendante. Malgré le fait qu'un allèle est nécessairement absent, nous pensons que ce niveau très faible d'expression reflète une répression du moins partielle de son expression par un mécanisme inconnu pour le moment. L'analyse mutationnelle du promoteur d'*ETV6* ou l'analyse de sa méthylation, par l'utilisation d'enzymes de restriction sensibles ou non à la présence de méthylation ou par la technique de séquençage au bisulfite, pourrait nous renseigner davantage sur les mécanismes en cause (revu par Oakeley, 1999).

Ces données rejoignent les observations de Cavé *et al.* (1997) au sujet de l'étroite association entre les translocations et les délétions sur le chromosome 12. Dans cette étude, 80 % des délétions étaient associées à la translocation t(12;21) et 77 % pour la relation inverse. Notre étude va plus loin en ce sens que nous démontrons que c'est bel et bien *ETV6* qui est la cible des délétions et non pas un autre gène dans la région. De plus, nous montrons que même s'il n'y a pas de délétion, la cellule leucémique trouve un moyen d'abolir ou du moins de diminuer fortement l'expression de *ETV6*. Dans notre étude, donc, la translocation est la mutation inactivante que nous recherchions, sauf dans un cas, que nous discuterons dans la prochaine section. Cependant, la preuve que *ETV6* est bel et bien un gène suppresseur de tumeurs n'est pas faite. Il faudra effectuer des études fonctionnelles afin de démontrer que l'absence de *ETV6* cause le cancer ou la leucémie.

6.6 Modèle proposé

Notre interrogation à la base de ce projet portait sur le ou les rôle(s) du locus 12p12.3 dans l'étiologie de la leucémie lymphoblastique aiguë de l'enfant. Il est possible, en combinant nos données à celles déjà disponibles dans la littérature, d'élaborer un modèle de l'implication de ce locus dans le développement de cette leucémie. Le gène de fusion *ETV6-AML1* provenant de la translocation t(12;21) est un événement fréquent car il est retrouvé dans environ 25 % des cas de LLA (Figure 1-2) et même chez près de 5 % de la population générale (Eguchi-Ishimae *et al.*, 2001). L'irradiation de cellules en culture entraîne des bris double brin au niveau de *ETV6* sur le chromosome 12 et de *AML1* sur le chromosome 21 dans des cellules en culture indiquant une instabilité de ces régions (Eguchi-Ishimae *et al.*, 2001). D'ailleurs le ratio de la distance génétique, 3 cM, sur la distance physique, 750 kb, suppose un taux élevé de recombinaison de la région 12p12.3. Aussi, *ETV6* et *AML1* sont impliqués dans de nombreuses autres translocations. L'analyse de chromosomes de cellules leucémiques ou de lymphomes par FISH a révélé la présence de multiples bris ou réarrangements affectant seulement la région 12p12 indiquant encore que cette région est extrêmement fragile (Sato *et al.*, 2001). On ne connaît cependant pas les causes de cette fragilité.

Pourquoi cette translocation est-elle associée à la leucémie, et ce seulement chez les enfants? On croit que la translocation t(12;21) se produit durant la gestation puisque des études faites sur des jumeaux identiques ayant développé la leucémie démontrent que le même clone était à l'origine de celle-ci. Ce clone n'aurait pu se transmettre que via un placenta commun. De plus, comme ces jumeaux développent la leucémie souvent à

plusieurs années d'intervalle, ceci indique que d'autres mutations plus longues à acquérir sont absolument nécessaires (Ford *et al.*, 1998; Wiemels *et al.*, 1999). D'ailleurs, contrairement aux autres translocations impliquant *ETV6*, *ETV6-AML1* n'est pas suffisant pour causer des tumeurs chez la souris (Andreasson *et al.*, 2001). Il est donc possible que les enfants soient plus sujets à développer une leucémie étant donné le nombre de divisions cellulaires plus élevé, la fragilité de leur ADN, *in utero* ou suivant leur naissance, ou les mécanismes de contrôles moins serrés durant ces étapes du développement. Il a été montré que les translocations impliquant *MLL* se produisent également *in utero*, mais dans ces cas la leucémie se développe la plupart du temps dans la première année de vie (revu par Rowley, 1999).

Une étude a identifié deux patients possédant la translocation t(12;21), mais n'ayant la délétion de la région 12p12 que dans certaines de leurs cellules cancéreuses, indiquant donc que la délétion est un événement secondaire (Romana *et al.*, 1996). Dans notre étude, il est possible que la diminution d'expression de *ETV6* observée dans les deux échantillons possédant la translocation, mais non la délétion soit due justement à une hétérogénéité de l'échantillon. Une étude cytogénétique pourrait nous permettre de répondre à cette question en séparant les cellules individuellement, si bien sûr la délétion est assez grande pour être observée. La délétion 12p12 serait donc un événement secondaire et possiblement aléatoire, mais très avantageux pour les cellules exprimant *ETV6-AML1*.

AML1 peut agir comme activateur ou répresseur de la transcription selon les cas, mais son association avec la portion N-terminale de ETV6, qui possède toujours un domaine de répression, force la protéine de fusion à agir comme un répresseur de la transcription (Hiebert *et al.*, 1996; Fenrick *et al.*, 1999). Comme seule AML1 contient un domaine de liaison à l'ADN, les gènes normalement régulés par cette dernière, comme *IL-3*, *GM-CSF* et *CSF-1*, devraient être affectés par la protéine de fusion. Cependant, l'absence essentielle de l'allèle normal de *ETV6* entraîne deux possibilités: 1- ETV6 en se dimérisant avec ETV6-AML1 empêche l'action de cette dernière; 2- ETV6-AML1 n'a aucun effet, et c'est l'absence de ETV6 qui est important. On sait que AML1 contrôle la croissance et la différenciation des précurseurs hématopoïétiques. Le blocage de la différenciation cellulaire est une étape essentielle pour le développement de la leucémie (Section 1.3.4.1.5). ETV6 est également exprimé dans ces cellules et est essentiel pour l'établissement de l'hématopoïèse dans la moelle osseuse, mais son implication directe dans le contrôle de la différenciation cellulaire n'a pas été démontrée. Les seules cibles connues de ETV6 sont les métalloprotéinases, dont la *stromélysine-1*, qui est nécessaire pour la digestion de la matrice et la migration cellulaire (Van Rompaey *et al.*, 1999; Fenrick *et al.*, 2000). De plus, l'expression de ETV6 bloque la croissance et le potentiel cancérigène de cellules transformées par Ras indiquant que ETV6 a un effet négatif sur la croissance cellulaire (Fenrick *et al.*, 2000; Van Rompaey *et al.*, 2000). La Figure 6-1 résume les deux modèles proposés. Nous ne pouvons pas pour l'instant choisir entre ces deux modèles et ceux-ci ne sont pas mutuellement exclusifs. Le fait que le domaine de répression de ETV6 et le domaine de liaison à l'ADN de AML1 soient toujours conservés et le fait que des délétions homozygotes de

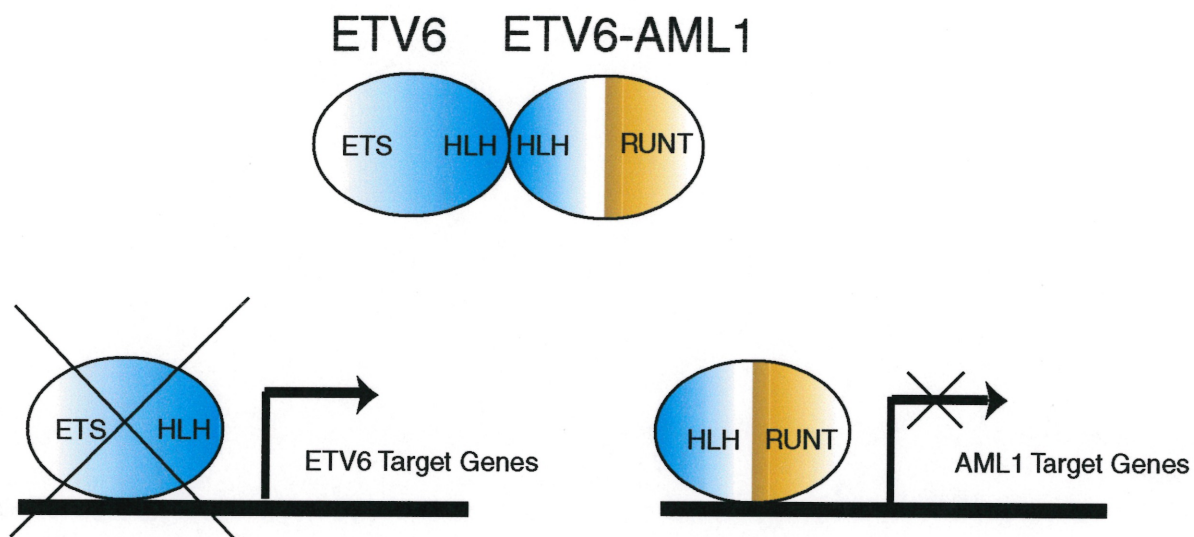


Figure 6.1. Modèle décrivant les participations possibles de ETV6-AML1 et de ETV6 à la leucémogénèse. La protéine de fusion ETV6-AML1 pourrait causer la répression des gènes transcrits normalement par AML1. L'absence de ETV6 pourrait être nécessaire pour empêcher l'hétérodimérisation ETV6/ETV6-AML1 ou pour activer l'expression de gènes normalement réprimés par ETV6.

ETV6 ne soient jamais observées pointent cependant vers le premier modèle ou une combinaison des deux.

6.7 Perspectives

Afin de comprendre la progression de ce type de leucémie, l'analyse des gènes différentiellement exprimés dans des cellules leucémiques ou des lignées hématopoïétiques où *ETV6*, *AML1* et/ou *ETV6-AML1* auront été introduits nous permettrait de mieux comprendre les voies de signalisation impliquées. Cette étude pourrait se faire par l'utilisation de microréseaux d'ADNc (cDNA microarrays) ou de puces d'ADNc (cDNA microchips). Ces méthodes permettent d'étudier l'expression globale de plusieurs milliers de gènes simultanément pour ensuite être en mesure d'établir des profils d'expression liés aux phénotypes étudiés (revu par Devaux *et al.*, 2001). Nous possédons également une lignée cellulaire leucémique qui possède la translocation t(12;21) et qui n'exprime pas *ETV6*. Son analyse ainsi que celle où *ETV6* est réintroduit dans ces cellules (et/ou de *CDKN1B*; voir plus loin) pourrait nous permettre de discriminer entre les différents modèles. De plus, l'analyse du potentiel oncogénique de toutes ces cellules *in vitro*, par le taux de croissance et la formation de colonies dans l'agar mou, et *in vivo*, mesuré par l'apparition d'une leucémie chez la souris, pourra nous permettre de raffiner notre modèle en définissant mieux le rôle de *ETV6-AML1* et la perte de *ETV6* dans le processus leucémique.

La preuve définitive que *ETV6* est un gène suppresseur de tumeurs proviendrait de l'utilisation de cellules ou de souris dont *ETV6* a été éliminé spécifiquement (sans la

présence de ETV6-AML1). Chez la souris, le knock-out de *ETV6* est léthal (Wang *et al.*, 1997). L'étude des souris chimériques a permis de constater que les cellules ETV6 *-/-* ne s'établissaient pas dans la moelle osseuse (Wang *et al.*, 1998). Il serait possible d'utiliser le système de recombinaison Cre-LOX afin d'effectuer le knock-out de *ETV6* à un moment précis du développement de la souris et alors d'évaluer son effet suppresseur de tumeurs (revu par Sauer, 1998). Il est aussi possible d'éliminer ou de réduire l'expression d'un gène, comme *ETV6*, à l'aide d'ARN antisens ou par la technique de RNAi (ou interférence d'ARN). Il est difficile avec l'utilisation d'ARN antisens d'obtenir une inhibition complète de l'expression d'un gène et de façon soutenue (revu par Giles, 2000). La technique de RNAi a été développée initialement chez *C.elegans* (Fire *et al.*, 1998; Montgomery *et al.*, 1998). La présence de molécules d'ARN double brin dirige la dégradation d'ARNm spécifiques par un mécanisme encore inconnu (revu dans Zamore, 2001). Il semble que cette technique pourrait être applicable à des cellules mammifères en culture (Elbashir *et al.*, 2001).

Takeuchi *et al.* (1996b) ont identifié chez les patients atteints de LLA une autre région de délétion située entre les marqueurs D12S358 et D12S320. Cette région correspond à la région critique identifiée chez les patients atteints du cancer du poumon (Takeuchi *et al.*, 1996a). Tous les gènes de cette région ne sont pas identifiés, mais une analyse sommaire révèle la présence, en plus de *CDKN1B*, *GPR19* et *CREBL2*, de *RAI3*, codant pour un récepteur couplé aux protéines G induit par l'acide rétinoïque, de *EMPI*, codant pour une protéine membranaire et de *7-HT*, encodant un récepteur sérotoninergique. *CDKN1B* demeure donc de loin le candidat le plus intéressant. Il est

probable par contre, comme nombre d'études l'ont démontré, qu'aucune mutation ne soit observée dans l'allèle non-déléte de ce gène (Section 1.5.2). À cause de ce type d'observation et le fait que la perte d'un seul allèle chez la souris cause une augmentation importante du nombre des tumeurs, *CDKN1B* est considéré comme un suppresseur de tumeurs haploinsuffisant. On sait que la diminution de la protéine est directement reliée à l'agressivité de certains cancers et leucémies, mais il ne s'agit alors que d'une corrélation. La relation entre la quantité de protéines, mesurée par Western blot ou tout autre essai permettant la quantification, et la corrélation avec une telle délétion, si elle est statistiquement significative, pourrait nous permettre de conclure que *CDKN1B* est effectivement la cible de ce type de délétions. Aussi, le fait que *CDKN1B* soit inclus dans la majorité des délétions accompagnant les translocations pourrait ne pas être que le fruit du hasard. Sa délétion hémizygotique pourrait donner un avantage sélectif important à la cellule pré-leucémique. Dans notre étude, nous avons identifié un cas possédant la délétion 12p12.3, mais qui n'était pas associé à la translocation t(12;21) ni à une diminution de l'expression de *ETV6*. La région de la délétion dans ce cas particulier s'étend de *D12S89* à *LOH1CR12*, englobant six gènes, *ETV6*, *BCL-G*, *LRP6*, *LOH3CR12*, *LOH2CR12* et *LOH1CR12*. Il faudrait analyser un plus grand nombre d'échantillons similaires afin d'identifier le gène ciblé dans ces cas. Par contre, il pourrait s'agir d'une délétion totalement aléatoire sans relation avec la leucémie. Aussi, il faut tenir compte du fait que concrètement une délétion est équivalente à une région homozygote et il pourrait s'agir aussi bien d'un événement de recombinaison somatique ou de conversion génique que d'une perte nette de matériel génétique. Dans les cas de translocations, la délétion comme telle ou un événement de recombinaison entraîne

automatiquement la perte de l'allèle intact de *ETV6*, si l'on considère que la translocation est l'événement primaire.

Récemment, Latil *et al.* (2001) ont analysé l'expression de *ETV6* et de *CDKN1B* dans le cancer de la prostate et une association significative entre la diminution de l'expression de *ETV6*, mais non de *CDKN1B*, et la présence de délétions a été observée. Dans les tumeurs solides où la délétion 12p a été caractérisée, aucune translocation impliquant *ETV6* n'est présente. *ETV6* pourrait donc agir comme gène suppresseur de tumeurs ailleurs que dans le système hématopoïétique, lui conférant un rôle beaucoup plus important dans la cellule. Cependant, du moins chez le cancer de la prostate, aucune mutation n'a été identifiée dans l'allèle intact de *ETV6*, ou de *CDKN1B* (Kibel *et al.*, 2000 et 2001).

6.8 Conclusion

Dans cette étude, nous avons démontré que le téléoste *Fugu rubripes*, en raison de son génome compact, était un excellent modèle, non pas pour le clonage positionnel de gènes chez l'humain, mais pour l'identification des gènes et la caractérisation de domaines fonctionnels importants, codants ou non-codants, dans le génome humain. Nous avons également développé avec succès une approche intégrative permettant d'identifier la grande majorité des gènes présents dans une région génomique donnée, en combinant l'utilisation d'approches informatiques et moléculaires. Ainsi, nous avons identifié et caractérisé sept gènes présents dans la région de délétion minimale sur le chromosome 12 chez les patients atteints de leucémie lymphoblastique aiguë de type

pré-B. Nous avons également démontré que le gène *ETV6* était ciblé par ces délétions, et accompagnés par la translocation t(12;21), causant ainsi la perte totale de son expression dans 89 % des cas. Nous avons présenté deux modèles, non exclusifs, pouvant expliquer ces observations: 1- l'inhibition de la voie d'AML1 par la protéine ETV6-AML1 est la cause principale de l'apparition de ce type leucémie; 2- l'inhibition de la voie de ETV6 est la cause principale de l'apparition de ce type de leucémie. Des expériences futures seront nécessaires afin de discriminer entre ces deux modèles et afin de démontrer que *ETV6* est un gène suppresseur de tumeurs.

Références

- Ahlbom A, Day N, Feychting M, Roman E, Skinner J, Dockerty J, Linet M, McBride M, Michaelis J, Olsen JH, Tynes T, Verkasalo PK. 2000. A pooled analysis of magnetic fields and childhood leukaemia. *Br J Cancer* **83**: 692-698.
- Aissani B, Bonan C, Baccichet A and Sinnett D. 1999. Childhood acute lymphoblastic leukemia: is there a tumor suppressor gene in chromosome 12p12.3? *Leuk Lymphoma* **34**: 231-239.
- Aissani B, Sinnett D. 1999. Fine physical and transcript mapping of a 1.8 Mb region spanning the locus for childhood acute lymphoblastic leukemia on chromosome 12p12.3. *Gene* **240**: 297-305.
- Amores A, Force A, Yan YL, Joly L, Amemiya C, Fritz A, Ho Langeland J, Prince V, Wang YL, Westerfield M, Ekker M, Postlethwait JH. 1998. Zebrafish hox clusters and vertebrate genome evolution. *Science* **282**: 1711-1714.
- Andersen MK, Christiansen DH, Jensen BA, Ernst P, Hauge G, Pedersen-Bjergaard J. 2001. Therapy-related acute lymphoblastic leukaemia with MLL rearrangements following DNA topoisomerase II inhibitors, an increasing problem: report on two new cases and review of the literature since 1992. *Br J Haematol* **114**: 539-543.
- Andreasson P, Schwaller J, Anastasiadou E, Aster J, Gilliland DG. 2001. The expression of ETV6/CBFA2 (TEL/AML1) is not sufficient for the transformation of hematopoietic cell lines in vitro or the induction of hematologic disease in vivo. *Cancer Genet Cytogenet* **130**: 93-104.

- Aparicio S, Morrison A, Gould A, Gilthorpe J, Chaudhuri C, Rigby P, Krumlauf R, Brenner S. 1995. Detecting conserved regulatory elements with the model genome of the Japanese puffer fish, *Fugu rubripes*. *Proc Natl Acad Sci U S A* **92**: 1684-1688.
- Aparicio S, Hawker K, Cottage A, Mikawa Y, Zuo L, Venkatesh B, Chen E, Krumlauf R, Brenner S. 1997. Organization of the *Fugu rubripes* Hox clusters: evidence for continuing evolution of vertebrate Hox complexes. *Nat Genet* **16**: 79-83.
- Aplan PD, Lombardi DP, Ginsberg AM, Cossman J, Bertness VL, Kirsch IR. 1990. Disruption of the human SCL locus by "illegitimate" V-(D)-J recombinase activity. *Science* **250**: 1426-1429.
- Baccichet A, Qualman SK and Sinnett D. 1997a. Allelic loss in childhood acute lymphoblastic leukemia. *Leuk Res* **21**: 817-823.
- Baccichet A and Sinnett D. 1997b. Frequent deletion of chromosome 12p12.3 in children with acute lymphoblastic leukaemia. *Br J Haematol* **99**: 107-114.
- Bachand F, Autexier C. 1999. Le rôle de la télomérase et des télomères dans l'immortalisation cellulaire. *Médecine/sciences* **15**: 1286-1291.
- Baens M, Aerssens J, van Zand K, Van den Berghe H, Marynen P. 1995. Isolation and regional assignment of human chromosome 12p cDNAs. *Genomics* **29**: 44-52.
- Baens M, Peeters P, Guo C, Aerssens J, Marynen P. 1996. Genomic organization of TEL: the human ETS-variant gene 6. *Genome Res* **6**: 404-413.
- Baens M, Wlodarska I, Corveleyn A, Hoornaert I, Hagemeyer A and Marynen P. 1999. A physical, transcript, and deletion map of chromosome region 12p12.3 flanked

by ETV6 and CDKN1B: hypermethylation of the LRP6 CpG island in two leukemia patients with hemizygous del(12p). *Genomics* **56**: 40-50.

Bartsch H, Nair U, Risch A, Rojas M, Wikman H, Alexandrov K. 2000. Genetic polymorphism of CYP genes, alone or in combination, as a risk modifier of tobacco-related cancers. *Epidemiol Biomarkers Prev* **9**: 3-28.

Baxendale S, Abdulla S, Elgar G, Buck D, Berks M, Micklem G, Durbin R, Bates G, Brenner S, Beck S. 1995. Comparative sequence analysis of the human and pufferfish Huntington's disease genes. *Nat Genet* **10**: 67-76.

Barton K, Muthusamy N, Fischer C, Ting CN, Walunas TL, Lanier LL, Leiden JM. 1998. The Ets-1 transcription factor is required for the development of natural killer cells in mice. *Immunity* **9**: 555-563.

Bastians H, Townsley FM, Ruderman JV. 1998. The cyclin-dependent kinase inhibitor p27(Kip1) induces N-terminal proteolytic cleavage of cyclin A. *Proc Natl Acad Sci USA* **95**: 15374-15381.

Baxendale S, Abdulla S, Elgar G, Buck D, Berks M, Micklem G, Durbin R, Bates G, Brenner S, Beck S. 1995. Comparative sequence analysis of the human and pufferfish Huntington's disease genes. *Nat Genet* **10**: 67-76.

Baylin SB, Herman JG, Graff JR, Vertino PM and Issa JP. 1998. Alterations in DNA methylation: a fundamental aspect of neoplasia. *Adv Cancer Res* **72**: 141-196.

Berger R, Bernheim A, Le Coniat M *et al.* 1986. Abnormalities of the short arm of chromosome 12 in acute nonlymphocytic leukemia and dysmyelopoietic syndrome. *Cancer Genet Cytogenet* **19**: 281-289.

- Beverloo HB, Panagopoulos I, Isaksson M, van Wering E, van Drunen E, de Klein A, Johansson B, Slater R. 2001. Fusion of the homeobox gene HLXB9 and the ETV6 gene in infant acute myeloid leukemias with the t(7;12)(q36;p13). *Cancer Res* **61**: 5374-5377.
- Bird A, Taggart M, Frommer M, Miller OJ, Macleod D. 1985. A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA. *Cell* **40**: 91-99.
- Bird AP. 1986. CpG-rich islands and the function of DNA methylation. *Nature* **321**: 209-213.
- Bishop JM. 1995. Cancer: the rise of the genetic paradigm. *Genes Dev* **9**: 1309-1315.
- Boehm T, Mengle-Gaw L, Kees UR, Spurr N, Lavenir I, Forster A, Rabbitts TH. 1989. Alternating purine-pyrimidine tracts may promote chromosomal translocations seen in a variety of human lymphoid tumours. *EMBO J* **8**: 2621-2631.
- Boorman GA, Rafferty CN, Ward JM, Sills RC. 2000. Leukemia and lymphoma incidence in rodents exposed to low-frequency magnetic fields. *Radiat Res* **153**: 627-636.
- Bonneau D and Longy M. 2000. Mutations of the human PTEN gene. *Hum Mutat* **16**: 109-122.
- Bories JC, Willerford DM, Grevin D, Davidson L, Camus A, Martin P, Stehelin D, Alt FW. 1995. Increased T-cell apoptosis and terminal B-cell differentiation induced by inactivation of the Ets-1 proto-oncogene. *Nature* **377**: 635-638.
- Bos JL. 1989. ras oncogenes in human cancer: a review. *Cancer Res* **49**: 4682-4689.

- Boultwood J, Fidler C, Strickson AJ, Watkins F, Kostrzewa M, Jaju RJ, Muller U, Wainscoat JS. 2000. Transcription mapping of the 5q- syndrome critical region: cloning of two novel genes and sequencing, expression, and mapping of a further six novel cDNAs. *Genomics* **66**: 26-34.
- Brenner S, Elgar G, Sandford R, Macrae A, Venkatesh B, Aparicio S. 1993. Characterization of the pufferfish (Fugu) genome as a compact model vertebrate genome. *Nature* **366**: 265-268.
- Brown SD, Twells RC, Hey PJ, Cox RD, Levy ER, Soderman AR, Metzker ML, Caskey CT, Todd JA, Hess JF. 1998. Isolation and characterization of LRP6, a novel member of the low density lipoprotein receptor gene family. *Biochem Biophys Res Commun* **248**: 879-888.
- Broeker PL, Harden A, Rowley JD, Zeleznik-Le N. 1996. The mixed lineage leukemia (MLL) protein involved in 11q23 translocations contains a domain that binds cruciform DNA and scaffold attachment region (SAR) DNA. *Curr Top Microbiol Immunol* **211**: 259-268.
- Buckler AJ, Chang DD, Graw SL, Brook JD, Haber DA, Sharp PA, Housman DE. 1991. Exon amplification: a strategy to isolate mammalian genes based on RNA splicing. *Proc Natl Acad Sci USA* **88**: 4005-4009.
- Buckley JD, Buckley CM, Breslow NE, Draper GJ, Roberson PK, Mack TM. 1996. Concordance for childhood cancer in twins. *Med Pediatr Oncol* **26**: 223-229.
- Buick RN and Tannock IF. Properties of malignant cells. In: Tannock IF and Hill RP, editors. *The Basic Science of Oncology*. McGraw-Hill, Inc. New York, 1992.
- Buijs A, van Rompaey L, Molijn AC, Davis JN, Vertegaal AC, Potter MD, Adams C, van Baal S, Zwarthoff EC, Roussel MF, Grosveld GC. 2000. The MN1-TEL

- fusion protein, encoded by the translocation (12;22)(p13;q11) in myeloid leukemia, is a transcription factor with transforming activity. *Mol Cell Biol* **20**: 9281-9293.
- Burel SA, Harakawa N, Zhou L, Pabst T, Tenen DG, Zhang DE. 2001. Dichotomy of AML1-ETO functions: growth arrest versus block of differentiation. *Mol Cell Biol* **21**: 5577-5590.
- Burset M, Guigo R. 1996. Evaluation of gene structure prediction programs. *Genomics* **34**: 353-367.
- Caldas C, Kim MH, MacGregor A, Cain D, Aparicio S, Wiedemann LM. 1998. Isolation and characterization of a pufferfish MLL (mixed lineage leukemia)-like gene (fMll) reveals evolutionary conservation in vertebrate genes related to *Drosophila trithorax*. *Oncogene* **16**: 3233-3241.
- Carroll M, Tomasson MH, Barker GF, Golub TR, Gilliland DG. 1996. The TEL/platelet-derived growth factor beta receptor (PDGF beta R) fusion in chronic myelomonocytic leukemia is a transforming protein that self-associates and activates PDGF beta R kinase-dependent signaling pathways. *Proc Natl Acad Sci USA* **93**: 14845-14850.
- Carron C, Cormier F, Janin A, Lacronique V, Giovannini M, Daniel MT, Bernard O, Ghysdael J. 2000. TEL-JAK2 transgenic mice develop T-cell leukemia. *Blood* **95**: 3891-3899.
- Castellanos A, Pintado B, Weruaga E, Arevalo R, Lopez A, Orfao A, Sanchez-Garcia I. 1997. A BCR-ABL(p190) fusion gene made by homologous recombination causes B-cell acute lymphoblastic leukemias in chimeric mice with independence of the endogenous bcr product. *Blood* **90**: 2168-2174.

- Cave H, Cacheux V, Raynaud S, Brunie G, Bakkus M, Cochaux P, Preudhomme C, Lai JL, Vilmer E, Grandchamp B. 1997. ETV6 is the target of chromosome 12p deletions in t(12;21) childhood acute lymphocytic leukemia. *Leukemia* **11**: 1459-1464.
- Chakrabarti SR, Sood R, Ganguly S, Bohlander S, Shen Z, Nucifora G. 1999. Modulation of TEL transcription activity by interaction with the ubiquitin-conjugating enzyme UBC9. *Proc Natl Acad Sci USA* **96**: 7467-7474.
- Chakrabarti SR, Nucifora G. 1999. The leukemia-associated gene TEL encodes a transcription repressor which associates with SMRT and mSin3A. *Biochem Biophys Res Commun* **264**: 871-877.
- Chakrabarti SR, Sood R, Nandi S, Nucifora G. 2000. Posttranslational modification of TEL and TEL/AML1 by SUMO-1 and cell-cycle-dependent. *Proc Natl Acad Sci USA* **97**: 13281-13285.
- Chambon-Pautas C, Cave H, Gerard B, Guidal-Giroux C, Duval M, Vilmer E, Grandchamp B. 1998. High-resolution allelotyping analysis of childhood B-lineage acute lymphoblastic leukemia. *Leukemia* **12**: 1107-1113.
- Chan LC, Kwong YL, Liu HW, Lee CP, Lie KW and Chan AY. 1992. Deletion 12p in de novo acute myeloid leukemia. An association with early progenitor cell. *Cancer Genet Cytogenet* **62**: 47-49.
- Charrad RS, Li Y, Delpech B, Balitrand N, Clay D, Jasmin C, Chomienne C, Smadja-Joffe F. 1999. Ligation of the CD44 adhesion molecule reverses blockage of differentiation in human acute myeloid leukemia. *Nat Med* **5**: 669-676.
- Chen YC, Chen PJ, Yeh SH, Tien HF, Wang CH, Tang JL, Hong RL. 1990. Deletion of the human retinoblastoma gene in primary leukemias. *Blood* **76**: 2060-2064.

- Clark MS. 1999. Comparative genomics: the key to understanding the Human Genome Project. *Bioessays* **21**: 121-130.
- Corn PG, Kuerbitz SJ, van Noesel MM, Esteller M, Compitello N, Baylin SB, Herman JG. 1999. Transcriptional silencing of the p73 gene in acute lymphoblastic leukemia and Burkitt's lymphoma is associated with 5' CpG island methylation. *Cancer Res* **59**: 3352-3356.
- Corral J, Lavenir I, Impey H, Warren AJ, Forster A, Larson TA, Bell S, McKenzie AN, King G, Rabbitts TH. 1996. An Mll-AF9 fusion gene made by homologous recombination causes acute leukemia in chimeric mice: a method to create fusion oncogenes. *Cell* **85**: 853-861.
- Cortes F, Labastie MC. 2000. Contrôle du développement du système hématopoïétique chez les vertébrés. *Médecine/sciences* **16**: 198-204.
- Cox DR, Burneister M, Price ER, Kim S, Myers RM. 1990. Radiation hybrid mapping: a somatic cell genetic method for constructing high-resolution maps of mammalian chromosomes. *Science* **250**: 245-250.
- Cox LS, Lane DP. 1995. Tumour suppressors, kinases and clamps: how p53 regulates the cell cycle in response to DNA damage. *Bioessays* **17**: 501-508.
- Dahia PL. 2000. PTEN, a unique tumor suppressor gene. *Endocr Relat Cancer* **7**: 115-29.
- Dang J, Inukai T, Kurosawa H, Goi K, Inaba T, Lenny NT, Downing JR, Stifani S, Look AT. 2001. The E2A-HLF oncoprotein activates Groucho-related genes and suppresses Runx1. *Mol Cell Biol* **21**: 5935-5945.
- Debant A, Serra-Pages C, Seipel K *et al.* 1996. The multidomain protein Trio binds the LAR transmembrane tyrosine phosphatase, contains a protein kinase domain,

and has separate rac- specific and rho-specific guanine nucleotide exchange factor domains. *Proc Natl Acad Sci U S A* **93**: 5466-5471.

Dedera DA, Waller EK, LeBrun DP, Sen-Majumdar A, Stevens ME, Barsh GS, Cleary ML. 1993. Chimeric homeobox gene E2A-PBX1 induces proliferation, apoptosis, and malignant lymphomas in transgenic mice. *Cell* **74**: 833-843.

Denu JM and Dixon JE. 1998. Protein tyrosine phosphatases: mechanisms of catalysis and regulation. *Curr Opin Chem Biol* **2**: 633-641.

Deppert W, Gohler T, Koga H, Kim E. 2000. Mutant p53: "gain of function" through perturbation of nuclear structure and function? *J Cell Biochem* **35**(suppl.): 115-122.

Devaux F, Marc P, Jacq C. 2001. Transcriptomes, transcription activators and microarrays. *FEBS Lett* **498**: 140-144.

Di Cristofano A, Pesce B, Cordon-Cardo C, Pandolfi PP. 1998. Pten is essential for embryonic development and tumour suppression. *Nat Genet* **19**: 348-355.

Donaldson LW, Petersen JM, Graves BJ, McIntosh LP. 1996. Solution structure of the ETS domain from murine Ets-1: a winged helix-turn-helix DNA binding motif. *EMBO J* **15**: 125-134.

Duncan BK, Miller JH. 1980. Mutagenic deamination of cytosine residues in DNA. *Nature* **287**: 560-561.

Duyk GM, Kim SW, Myers RM, Cox DR. 1990. Exon trapping: a genetic screen to identify candidate transcribed sequences in cloned mammalian genomic DNA. *Proc Natl Acad Sci U S A* **87**: 8995-8999.

- Edel MJ. 1998. The ETS-related factor TEL is regulated by angiogenic growth factor VEGF in HUVE-cells. *Anticancer Res* **18**: 4505-4509.
- Edwards-Gilbert G, Veraldi KL and Milcarek C. 1997. Alternative poly(A) site selection in complex transcription units: means to an end? *Nucleic Acids Res* **25**: 2547-2561.
- Eguchi-Ishimae M, Eguchi M, Ishii E, Miyazaki S, Ueda K, Kamada N, Mizutani S. 2001. Breakage and fusion of the TEL (ETV6) gene in immature B lymphocytes induced by apoptogenic signals. *Blood* **97**: 737-743.
- Elbashir SM, Harborth J, Lendeckel W, Yalcin A, Weber K, Tuschl T. 2001. Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature* **411**: 494-498.
- Elgar G, Clark MS, Meek S, Smith S, Warner S, Edwards YJ, Bouchireb N, Cottage A, Yeo GS, Umrana Y, Williams G, Brenner S. 1999. Generation and analysis of 25 Mb of genomic DNA from the pufferfish *Fugu rubripes* by sequence scanning. *Genome Res* **9**: 960-971.
- Elgar G, Sandford R, Aparicio S, Macrae A, Venkatesh B, Brenner S. 1996. Small is beautiful: comparative genomics with the pufferfish (*Fugu rubripes*). *Trends Genet* **12**: 145-150.
- Felix CA. 2001. Leukemias related to treatment with DNA topoisomerase II inhibitors. *Med Pediatr Oncol* **36**: 525-535.
- Fenrick R, Amann JM, Lutterbach B, Wang L, Westendorf JJ, Downing JR, Hiebert SW. 1999. Both TEL and AML-1 contribute repression domains to the t(12;21) fusion protein. *Mol Cell Biol* **19**: 6566-6574.

- Fenrick R, Wang L, Nip J, Amann JM, Rooney RJ, Walker-Daniels J, Crawford HC, Hulboy DL, Kinch MS, Matrisian LM, Hiebert SW. 2000. TEL, a putative tumor suppressor, modulates cell growth and cell morphology of ras-transformed cells while repressing the transcription of stromelysin-1. *Mol Cell Biol* **20**: 5828-5839.
- Fero ML, Rivkin M, Tasch M, Porter P, Carow CE, Firpo E, Polyak K, Tsai LH, Broudy V, Perlmutter RM, Kaushansky K, Roberts JM. 1996. A syndrome of multiorgan hyperplasia with features of gigantism, tumorigenesis, and female sterility in p27(Kip1)-deficient mice. *Cell* **85**: 733-744.
- Fero ML, Randel E, Gurley KE, Roberts JM, Kemp CJ. 1998. The murine gene p27Kip1 is haplo-insufficient for tumour suppression. *Nature* **396**: 177-180.
- Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC. 1998. Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* **391**: 806-811.
- Ford AM, Bennett CA, Price CM, Bruin MC, Van Wering ER, Greaves M. 1998. Fetal origins of the TEL-AML1 fusion gene in identical twins with leukemia. *Proc Natl Acad Sci U S A* **14**: 4584-4588.
- Gallie BL, Phillips RA. 1984. Retinoblastoma: a model of oncogenesis. *Ophthalmology* **9**: 666-672.
- Gebow D, Miselis N, Liber HL. 2000. Homologous and nonhomologous recombination resulting in deletion: effects of p53 status, microhomology, and repetitive DNA length and orientation. *Mol Cell Biol* **20**: 4028-4035.
- Gellner K, Brenner S. 1999. Analysis of 148 kb of genomic DNA around the wnt1 locus of *Fugu rubripes*. *Genome Res* **9**: 251-258.

- Giles RV. 2000. Antisense oligonucleotide technology: from EST to therapeutics. *Curr Opin Mol Ther* **2**: 238-252.
- Gillespie GA, Somlo S, Germino GG, Weinstat-Saslow D, Reeders ST. 1991. CpG island in the region of an autosomal dominant polycystic kidney disease locus defines the 5' end of a gene encoding a putative proton channel. *Proc Natl Acad Sci U S A* **88**: 4289-4293.
- Gilley J, Fried M. 1999. Extensive gene order differences within regions of conserved synteny between the Fugu and human genomes: implications for chromosomal evolution and the cloning of disease genes. *Hum Mol Genet* **8**: 1313-1320.
- Gilley J, Armes N, Fried M. 1997. Fugu genome is not a good mammalian model. *Nature* **385**: 305-306.
- Gishizky ML, Johnson-White J, Witte ON. 1993. Efficient transplantation of BCR-ABL-induced chronic myelogenous leukemia-like syndrome in mice. *Proc Natl Acad Sci U S A* **90**: 3755-3759.
- Golub TR, Barker GF, Lovett M and Gilliland DG. 1994. Fusion of PDGF receptor beta to a novel ets-like gene, tel, in chronic myelomonocytic leukemia with t(5;12) chromosomal translocation. *Cell* **77**: 307-316.
- Golub TR, Barker GF, Bohlander SK, Hiebert SW, Ward DC, Bray-Ward P, Morgan E, Raimondi SC, Rowley JD, Gilliland DG. 1995. Fusion of the TEL gene on 12p13 to the AML1 gene on 21q22 in acute lymphoblastic leukemia. *Proc Natl Acad Sci U S A* **92**: 4917-4921.
- Greaves M. 1999. Molecular genetics, natural history and the demise of childhood leukaemia. *Eur J Cancer* **35**: 173-185.

- Grignani F, De Matteis S, Nervi C, Tomassoni L, Gelmetti V, Ciocce M, Fanelli M, Ruthardt M, Ferrara FF, Zamir I, Seiser C, Grignani F, Lazar MA, Minucci S, Pelicci PG. 1998. Fusion proteins of the retinoic acid receptor-alpha recruit histone deacetylase in promyelocytic leukaemia. *Nature* **391**: 815-818.
- Gu X, Shin BH, Akbarali Y, Weiss A, Boltax J, Oettgen P, Libermann TA. 2001. Tel-2 is a novel transcriptional repressor related to the Ets factor Tel/ETV-6. *J Biol Chem* **276**: 9421-9436.
- Guidez F, Petrie K, Ford AM, Lu H, Bennett CA, MacGregor A, Hannemann J, Ito Y, Ghysdael J, Greaves M, Wiedemann LM, Zelent A. 2000. Recruitment of the nuclear receptor corepressor N-CoR by the TEL moiety of the childhood leukemia-associated TEL-AML1 oncoprotein. *Blood* **96**: 2557-2561.
- Guo B, Godzik A and Reed JC. 2001. Bcl-G, a novel pro-apoptotic member of the Bcl-2 family. *J Biol Chem* **276**: 2780-2785.
- Gupta PK, Sahota A, Boyadjiev SA, Bye S, Shao C, O'Neill JP, Hunter TC, Albertini RJ, Stambrook PJ, Tischfield JA. 1997. High frequency in vivo loss of heterozygosity is primarily a consequence of mitotic recombination. *Cancer Res* **57**: 1188-1193.
- Haber DA, Fearon ER. 1998. The promise of cancer genetics. *Lancet* **351**: (suppl II) 1-8.
- Hahn WC, Counter CM, Lundberg AS, Beijersbergen RL, Brooks MW and Weinberg RA. 1999. Creation of human tumour cells with defined genetic elements. *Nature* **400**: 464-468.
- Hannemann JR, McManus DM, Kabarowski JH, Wiedemann LM. 1998. Haemopoietic transformation by the TEL/ABL oncogene. *Br J Haematol* **102**: 475-485.

- Hanrahan J, Pan L, Greenbaum S, Bradney C, Hjelmeland M, Dai M, Zhuang Y. 2001. A genetic investigation of E2A function in lymphocyte development. *Immunol Res* **22**: 211-222.
- Hanson RD, Hess JL, Yu BD, Ernst P, van Lohuizen M, Berns A, van der Lugt NM, Shashikant CS, Ruddle FH, Seto M, Korsmeyer SJ. 1999. Mammalian Trithorax and polycomb-group homologues are antagonistic regulators of homeotic development. *Proc Natl Acad Sci U S A* **96**: 14372-14377.
- Harras A, Edwards BK, Blot WJ and Gloeckler Ries, LA, Editors. *Cancer: Rates and Risks*. 4th edition, 1996. NIH publication No.96-691.
- Hatta Y, Takeuchi S, Yokota J, Koeffler HP. 1997. Ovarian cancer has frequent loss of heterozygosity at chromosome 12p12.3-13.1 (region of TEL and Kip1 loci) and chromosome 12q23-ter: evidence for two new tumour-suppressor genes. *Br J Cancer* **75**: 1256-1262.
- Hayashi K, Yandell DW. 1993. How sensitive is PCR-SSCP? *Hum Mutat* **2**: 338-346.
- Hayflick L and Moorhead PS. 1961. The serial cultivation of human diploid cell strains. *Exp Cell Res* **25**: 585-621.
- Hebert J, Cayuela JM, Berkeley J, Sigaux F. 1994. Candidate tumor-suppressor genes MTS1 (p16INK4A) and MTS2 (p15INK4B) display frequent homozygous deletions in primary cells from T- but not from B-cell lineage acute lymphoblastic leukemias. *Blood* **84**: 4038-4044.
- Heisterkamp N, Stam K, Groffen J, de Klein A, Grosveld G. 1985. Structural organization of the bcr gene and its role in the Ph' translocation. *Nature* **315**: 758-761.

- Hendrich B, Hardeland U, Ng HH, Jiricny J, Bird A. 1999. The thymine glycosylase MBD4 can bind to the product of deamination at methylated CpG sites. *Nature* **401**: 301-304.
- Hiebert SW, Sun W, Davis JN, Golub T, Shurtleff S, Buijs A, Downing JR, Grosveld G, Rousell MF, Gilliland DG, Lenny N, Meyers S. 1996. The t(12;21) translocation converts AML-1B from an activator to a repressor of transcription. *Mol Cell Biol* **16**: 1349-1355.
- Hinegardner. 1968. Evolution of cellular DNA content in teleost fishes. *Am Nat* **102**: 517-523
- Hoornaert I, Marynen P, Baens M. 1998. CREBL2, a novel transcript from the chromosome 12 region flanked by ETV6 and CDKN1B. *Genomics* **51**: 154-157.
- How GF, Venkatesh B, Brenner S. 1996. Conserved linkage between the puffer fish (Fugu rubripes) and human genes for platelet-derived growth factor receptor and macrophage colony-stimulating factor receptor. *Genome Res* **6**: 1185-1191
- Hsiao MH, Yu AL, Yeargin J, Ku D, Haas M. 1994. Nonhereditary p53 mutations in T-cell acute lymphoblastic leukemia are associated with the relapse phase. *Blood* **83**: 2922-2930.
- Hudson TJ, Stein LD, Gerety SS, Ma J, Castle AB, Silva J, Slonim DK, Baptista R, Kruglyak L, Xu SH, et al. 1995. An STS-based map of the human genome. *Science* **270**: 1945-1954.
- Hunter T. 1993. Braking the cycle. *Cell* **75**: 839-841.

- Hussussian CJ, Struewing JP, Goldstein AM, Higgins PA, Ally DS, Sheahan MD, Clark WH Jr, Tucker MA, Dracopoli NC. 1994. Germline p16 mutations in familial melanoma. *Nat Genet* **8**: 15-21.
- Imai Y, Kurokawa M, Tanaka K, Friedman AD, Ogawa S, Mitani K, Yazaki Y, Hirai H. 1998. TLE, the human homolog of groucho, interacts with AML1 and acts as a repressor of AML1-induced transactivation. *Biochem Biophys Res Commun* **252**: 582-589.
- Iwai T, Yokota S, Nakao M, Nakazawa N, Taniwaki M, Kimura T, Sonoda Y, Kaneko H, Okuda T, Azuma H, Oka T, Takeda T, Watanabe A, Kikuta A, Asami K, Sekine I, Matsushita T, Tsuchiya T, Mimaya J, Koizumi S, Ohta S, Miyake M, Takaue Y, Iwai A, Fujimoto T. 1998. Frequent aberration of FHIT gene expression in acute leukemias. *Cancer Res* **58**: 5182-5187.
- Jiang W, Ananthaswamy HN, Muller HK, Ouhtit A, Bolshakov S, Ullrich SE, El-Naggar AK, Kripke ML. 2001. UV irradiation augments lymphoid malignancies in mice with one functional copy of wild-type p53. *Proc Natl Acad Sci U S A* **98**: 9790-9795.
- John A, Tuszynski G. 2001. The role of matrix metalloproteinases in tumor angiogenesis and tumor metastasis. *Pathol Oncol Res* **7**: 14-23.
- Kamb A, Gruis NA, Weaver-Feldhaus J, Liu Q, Harshman K, Tavitgian SV, Stockert E, Day RS 3rd, Johnson BE, Skolnick MH. 1994. A cell cycle regulator potentially involved in genesis of many tumor types. *Science* **264**: 436-440.
- Kamps MP, Baltimore D. 1993. E2A-Pbx1, the t(1;19) translocation protein of human pre-B-cell acute lymphocytic leukemia, causes acute myeloid leukemia in mice. *Mol Cell Biol* **13**: 351-357.

- Kanno Y, Kanno T, Sakakura C, Bae SC, Ito Y. 1998. Cytoplasmic sequestration of the polyomavirus enhancer binding protein 2 (PEBP2)/core binding factor alpha (CBFalpha) subunit by the leukemia-related PEBP2/CBFbeta-SMMHC fusion protein inhibits PEBP2/CBF-mediated transactivation. *Mol Cell Biol* **18**: 4252-4261.
- Kao HT, Porton B, Hilfiker S, Stefani G, Pieribone VA, DeSalle R, Greengard P. 1999. Molecular evolution of the synapsin gene family. *J Exp Zool* **285**: 360-377.
- Kibel AS, Schutte M, Kern SE, Isaacs WB, Bova GS. 1998. Identification of 12p as a region of frequent deletion in advanced prostate cancer. *Cancer Res* **58**: 5652-5655.
- Kibel AS, Freije D, Isaacs WB and Bova GS. 1999. Deletion mapping at 12p12-13 in metastatic prostate cancer. *Genes Chromosomes Cancer* **25**: 270-276.
- Kibel AS, Faith DA, Bova GS, Isaacs WB. 2000. Loss of heterozygosity at 12P12-13 in primary and metastatic prostate adenocarcinoma. *J Urol* **164**: 192-196.
- Kibel AS, Christopher M, Faith DA, Bova GS, Goodfellow PJ, Isaacs WB. 2001. Methylation and mutational analysis of p27(kip1) in prostate carcinoma. *Prostate* **48**: 248-253.
- Kimura C, Takeda N, Suzuki M, Oshimura M, Aizawa S, Matsuo I. 1997. Cis-acting elements conserved between mouse and pufferfish Otx2 genes govern the expression in mesencephalic neural crest cells. *Development* **124**: 3929-3941.
- Kiyokawa H, Kineman RD, Manova-Todorova KO, Soares VC, Hoffman ES, Ono M, Khanam D, Hayday AC, Frohman LA, Koff A. 1996. Enhanced growth of mice lacking the cyclin-dependent kinase inhibitor function of p27(Kip1). *Cell* **85**: 721-732.
- Kiyono T, Foster SA, Koop JI, McDougall JK, Galloway DA, Klingelutz AJ. 1998. Both Rb/p16INK4a inactivation and telomerase activity are required to immortalize human epithelial cells. *Nature* **396**: 84-88.

- Kleinerman RA, Kaune WT, Hatch EE, Wacholder S, Linet MS, Robison LL, Niwa S, Tarone RE. 2000. Are children living near high-voltage power lines at increased risk of acute lymphoblastic leukemia? *Am J Epidemiol* **151**: 512-515.
- Knudson AG. 1971. Mutation and cancer: statistical study of retinoblastoma. *Proc Natl Acad Sci U S A* **68**: 820-823.
- Knudson AG. 1993. Antioncogenes and human cancer. *Proc Natl Acad Sci USA* **90**: 10914-10921.
- Kobayashi H, Montgomery KT, Bohlander SK, Adra CN, Lim BL, Kucherlapati RS, Donis-Keller H, Holt MS, Le Beau MM, Rowley JD. 1994. Fluorescence in situ hybridization mapping of translocations and deletions involving the short arm of human chromosome 12 in malignant hematologic diseases. *Blood* **84**: 3473-3482.
- Koike M, Tasaka T, Spira S, Tsuruoka N, Koeffler HP. 1999. Allelotyping of acute myelogenous leukemia: loss of heterozygosity at 7q31.1 (D7S486) and q33-34 (D7S498, D7S505). *Leuk Res* **23**: 307-310.
- Koop BF, Nadeau JH. 1996. Pufferfish and new paradigm for comparative genome analysis. *Proc Natl Acad Sci U S A* **93**: 1363-1365.
- Kwiatkowski BA, Bastian LS, Bauer TR Jr, Tsai S, Zielinska-Kwiatkowska AG, Hickstein DD. 1998. The ets family member Tel binds to the Fli-1 oncoprotein and inhibits its transcriptional activity. *J Biol Chem* **273**: 17525-17530.
- Lacronique V, Boureux A, Valle VD, Poirel H, Quang CT, Mauchauffe M, Berthou C, Lessard M, Berger R, Ghysdael J, Bernard OA. 1997. A TEL-JAK2 fusion protein with constitutive kinase activity in human leukemia. *Science* **278**: 1309-1312.

- Lacronique V, Boureux A, Monni R, Dumon S, Mauchauffe M, Mayeux P, Gouilleux F, Berger R, Gisselbrecht S, Ghysdael J, Bernard OA. 2000. Transforming properties of chimeric TEL-JAK proteins in Ba/F3 cells. *Blood* **95**: 2076-2083.
- Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J et al. 2001. Initial sequencing and analysis of the human genome. *Nature* **409**: 860-921.
- Latil A, Guerard M, Berthon P, Cussenot O. 2001. 12p12-13 deletion in prostate tumors and quantitative expression of CDKN1B and ETV6 candidate genes. *Genes Chromosomes Cancer* **31**: 199-200.
- Laudet V, Hanni C, Stehelin D, Duterque-Coquillaud M. 1999. Molecular phylogeny of the ETS gene family. *Oncogene* **18**: 1351-1359.
- Lee D, Oh ES, Woods A, Couchman JR, Lee W. 1998. Solution structure of a syndecan-4 cytoplasmic domain and its interaction with phosphatidylinositol 4,5-bisphosphate. *J Biol Chem* **275**: 13022-13029.
- Leonhardt H, Cardoso MC. 2000. DNA methylation, nuclear structure, gene expression and cancer. *J Cell Biochem* **35** (Suppl.): 78-83.
- Leprince D, Gégonne A, Coll J, de Taisne C, Schneeberger A, Lagrou C, Stehelin D. 1983. A putative second cell-derived oncogene of the avian leukaemia retrovirus E26. *Nature* **306**: 395-397.
- Levanon D, Goldstein RE, Bernstein Y, Tang H, Goldenberg D, Stifani S, Paroush Z, Groner Y. 1998. Transcriptional repression by AML1 and LEF-1 is mediated by the TLE/Groucho corepressors. *Proc Natl Acad Sci U S A* **95**: 11590-11595.
- Lezon-Geyda K, Najfeld V, Johnson EM. 2001. Deletions of PURA, at 5q31, and PURB, at 7p13, in myelodysplastic syndrome and progression to acute myelogenous leukemia. *Leukemia* **15**: 954-962.

- Liu Q, Schwaller J, Kutok J, Cain D, Aster JC, Williams IR, Gilliland DG. 2000. Signal transduction and transforming properties of the TEL-TRKC fusions associated with t(12;15)(p13;q25) in congenital fibrosarcoma and acute myelogenous leukemia. *EMBO J* **19**: 1827-1838.
- Lloyd RV, Erickson LA, Jin L, Kulig E, Qian X, Cheville JC, Scheithauer BW. 1999. p27kip1: a multifunctional cyclin-dependent kinase inhibitor with prognostic significance in human cancers. *Am J Pathol* **154**: 313-323.
- Look AT. 1997. Oncogenic transcription factors in human acute leukemias. *Science* **278**: 1059-1064.
- Lopez RG, Carron C, Oury C, Gardellin P, Bernard O, Ghysdael J. 1999. TEL is a sequence-specific transcriptional repressor. *J Biol Chem* **274**: 30132-30138.
- Lorsbach RB, Downing JR. 2001. The role of the AML1 transcription factor in leukemogenesis. *Int J Hematol* **74**: 258-265.
- Lovett M, Kere J, Hinton LM. 1991. Direct selection: a method for the isolation of cDNAs encoded by large genomic regions. *Proc Natl Acad Sci U S A* **88**: 9628-9632.
- Lu L, Osmond DG. 2000. Apoptosis and its modulation during B lymphopoiesis in mouse bone marrow. *Immunol Rev* **175**: 158-174.
- Lutterbach B, Westendorf JJ, Linggi B, Patten A, Moniwa M, Davie JR, Huynh KD, Bardwell VJ, Lavinsky RM, Rosenfeld MG, Glass C, Seto E, Hiebert SW. 1998. ETO, a target of t(8;21) in acute leukemia, interacts with the N-CoR and mSin3 corepressors. *Mol Cell Biol* **18**: 7176-7184.

- MacGrogan D, Alvarez S, DeBlasio T, Jhanwar SC, Nimer SD. 2001. Identification of candidate genes on chromosome band 20q12 by physical mapping of translocation breakpoints found in myeloid leukemia cell lines. *Oncogene* **20**: 4150-4160.
- Maloney K, McGavran L, Murphy J, Odom L, Stork L, Wei Q, Hunger S. 1999. TEL-AML1 fusion identifies a subset of children with standard risk acute lymphoblastic leukemia who have an excellent prognosis when treated with therapy that includes a single delayed intensification. *Leukemia* **13**: 1708-1712.
- Maroulakou IG, Bowe DB. 2000. Expression and function of Ets transcription factors in mammalian development: a regulatory network. *Oncogene* **19**: 6432-6442.
- Marshall H, Studer M, Popperl H, Aparicio S, Kuroiwa A, Brenner S, Krumlauf R. 1994. A conserved retinoic acid response element required for early expression of the homeobox gene Hoxb-1. *Nature* **370**: 567-571.
- Martell KJ, Seasholtz AF, Kwak SP, Clemens KK and Dixon JE. 1995. hVH-5: a protein tyrosine phosphatase abundant in brain that inactivates mitogen-activated protein kinase. *J Neurochem* **65**: 1823-1833.
- Martin GS. 1970. Rous sarcoma virus: a function required for the maintenance of the transformed state. *Nature* **227**: 1021-1023.
- Maru Y. 2001. Molecular biology of chronic myeloid leukemia. *Int J Hematol* **73**: 308-322.
- Mason PJ, Stevens DJ, Luzzatto L, Brenner S, Aparicio S. 1995. Genomic structure and sequence of the Fugu rubripes glucose-6-phosphate dehydrogenase gene (G6PD). *Genomics* **26**: 587-591.

- Mavrothalassitis G, Ghysdael J. 2000. Proteins of the ETS family with transcriptional repressor activity. *Oncogene* **19**: 6524-6532.
- McBride ML. 1998. Childhood cancer and environmental contaminants. *Can J Public Health* **89** (Suppl 1): S53-62, S58-68.
- McBride ML, Gallagher RP, Theriault G, Armstrong BG, Tamaro S, Spinelli JJ, Deadman JE, Fincham S, Robson D, Choi W. 1999. Power-frequency electric and magnetic fields and risk of childhood leukemia in Canada. *Am J Epidemiol* **149**: 831-842.
- McLean TW, Ringold S, Neuberg D, Stegmaier K, Tantravahi R, Ritz J, Koeffler HP, Takeuchi S, Janssen JW, Seriu T, Bartram CR, Sallan SE, Gilliland DG, Golub TR. 1996. TEL/AML-1 dimerizes and is associated with a favorable outcome in childhood acute lymphoblastic leukemia. *Blood* **88**: 4252-4258.
- McLysaght A, Enright AJ, Skrabanek L, Wolfe KH. 2000. Estimation of synteny conservation and genome compaction between pufferfish (Fugu) and human. *Yeast* **17**: 22-36.
- Merlo A, Herman JG, Mao L, Lee DJ, Gabrielson E, Burger PC, Baylin SB, Sidransky D. 1995. 5' CpG island methylation is associated with transcriptional silencing of the tumour suppressor p16/CDKN2/MTS1 in human cancers. *Nat Med* **1**: 686-692.
- Meyer A, Scharl M. 1999. Gene and genome duplications in vertebrates: the one-to-four (-to-eight in fish) rule and the evolution of novel gene functions. *Curr Opin Cell Biol* **11**: 699-704.
- Miles C, Elgar G, Coles E, Kleinjan DJ, van Heyningen V, Hastie N. 1998. Complete sequencing of the Fugu WAGR region from WT1 to PAX6: dramatic

- compaction and conservation of synteny with human chromosome 11p13. *Proc Natl Acad Sci U S A* **95**: 13068-13072.
- Mitelman F. 1994. Chromosomes, genes, and cancer. *CA Cancer J Clin* **44**: 133-135.
- Miyamoto S, Chiorini JA, Urcelay E, Safer B. 1996. Regulation of gene expression for translation initiation factor eIF-2 alpha: importance of the 3' untranslated region. *Biochem J* **315**: 791-798.
- Monk M. 1995. Epigenetic programming of differential gene expression in development and evolution. *Dev Genet* **17**: 188-197.
- Montgomery MK, Xu S, Fire A. 1998. RNA as a target of double-stranded RNA-mediated genetic interference in *Caenorhabditis elegans*. *Proc Natl Acad Sci U S A* **95**: 15502-15507.
- Montpetit A, Sinnett D. 1999. Physical mapping of the G-protein coupled receptor 19 (GPR19) in the chromosome 12p12.3 region frequently rearranged in cancer cells. *Hum Genet* **105**: 162-164.
- Montpetit A, Sinnett D. 2001. Comparative analysis of the ETV6 gene in vertebrate genomes from pufferfish to human. *Oncogene* **20**: 3437-3442
- Montpetit A, Boily G, Sinnett D. 2002. A detailed transcriptional map of the chromosome 12p12 tumor suppressor locus. Accepted pour publication dans *Eur J Hum Gen.*
- Moynahan ME, Jasin M. 1997. Loss of heterozygosity induced by a chromosomal double-strand break. *Proc Natl Acad Sci U S A* **94**: 8988-8993.

- Nagamura-Inoue T, Tamura T, Ozato K. 2001. Transcription factors that regulate growth and differentiation of myeloid cells. *Int Rev Immunol* **20**: 83-105.
- Nakai K, Horton P. 1999. PSORT: a program for detecting sorting signals in proteins and predicting their subcellular localization. *Trends Biochem Sci* **24**: 34-36
- Nakayama K, Ishida N, Shirane M, Inomata A, Inoue T, Shishido N, Horii I, Loh DY, Nakayama K. 1996. Mice lacking p27(Kip1) display increased body size, multiple organ hyperplasia, retinal dysplasia, and pituitary tumors. *Cell* **85**: 707-720.
- Nathanson KN, Wooster R, Weber BL. 2001. Breast cancer genetics: what we know and what we need. *Nat Med* **7**: 552-556.
- Neglia JP, Friedman DL, Yasui Y, Mertens AC, Hammond S, Stovall M, Donaldson SS, Meadows AT, Robison LL. 2001. Second malignant neoplasms in five-year survivors of childhood cancer: childhood cancer survivor study. *J Natl Cancer Inst* **93**: 618-629.
- Nowell PC, Hungerford. 1960. A minute chromosome in human granulocytic leukemia. *Science* **132**: 1497.
- Nuchprayoon I, Meyers S, Scott LM, Suzow J, Hiebert S, Friedman AD. 1994. PEBP2/CBF, the murine homolog of the human myeloid AML1 and PEBP2 beta/CBF beta proto-oncoproteins, regulates the murine myeloperoxidase and neutrophil elastase genes in immature myeloid cells. *Mol Cell Biol* **14**: 5558-5568.
- Nunn MF, Seeburg PH, Moscovici C, Duesberg PH. 1983. Tripartite structure of the avian erythroblastosis virus E26 transforming gene. *Nature* **306**: 391-395.

- Nusse R, van Ooyen A, Cox D, Fung YK, Varmus H. 1984. Mode of proviral activation of a putative mammary oncogene (int-1) on mouse chromosome. *Nature* **307**: 131-136.
- Oakeley EJ. 1999. DNA methylation analysis: a review of current methodologies. *Pharmacol Ther* **84**: 389-400.
- Odero MD, Carlson K, Calasanz MJ, Lahortiga I, Chinwalla V, Rowley JD. 2001. Identification of new translocations involving ETV6 in hematological malignancies by fluorescence in situ hybridization and spectral karyotyping. *Genes Chromosomes Cancer* **31**: 134-142.
- O'Dowd BF, Nguyen T, Lynch KR, Kolakowski LF Jr, Thompson M, Cheng R, Marchese A, Ng G, Heng HH, George SR. 1996. A novel gene codes for a putative G protein-coupled receptor with an abundant expression in brain. *FEBS Lett* **394**: 325-329.
- Okuda K, Golub TR, Gilliland DG, Griffin JD. 1996. p210BCR/ABL, p190BCR/ABL, and TEL/ABL activate similar signal transduction pathways in hematopoietic cell lines. *Oncogene* **13**: 1147-1152.
- Oshimura M, Gilmer TM, Barrett JC. 1985. Nonrandom loss of chromosome 15 in Syrian hamster tumours induced by v-Ha-ras plus v-myc oncogenes. *Nature* **316**: 636-639.
- Palmer S, Brouillet JP, Kilbey A, Fulton R, Walker M, Crossley M, Bartholomew C. 2001. Evi-1 transforming and repressor activities are mediated by CtBP co-repressor proteins. *J Biol Chem* **276**: 25834-25840.

- Parimoo S, Patanjali SR, Shukla H, Chaplin DD, Weissman SM. 1991. cDNA selection: efficient PCR approach for the selection of cDNAs encoded in large chromosomal DNA fragments. *Proc Natl Acad Sci U S A* **88**: 9623-9627.
- Peifer M, Polakis P. 2000. Wnt signaling in oncogenesis and embryogenesis--a look outside the nucleus. *Science* **287**: 1606-1609.
- Perkins A, Kongsuwan K, Visvader J, Adams JM, Cory S. 1990. Homeobox gene expression plus autocrine growth factor production elicits myeloid leukemia. *Proc Natl Acad Sci U S A* **87**: 8398-8402.
- Peto R, Roe FJC, Lee PN, Levy L and Clack J. 1975. Cancer and aging in mice and men. *Br J Cancer* **32**: 411-426.
- Philipp-Staheli J, Payne SR, Kemp CJ. 2001. p27(Kip1): regulation and function of a haploinsufficient tumor suppressor and its misregulation in cancer. *Exp Cell Res* **264**: 148-168
- Pietenpol JA, Bohlander SK, Sato Y, Papadopoulos N, Liu B, Friedman C, Trask BJ, Roberts JM, Kinzler KW, Rowley JD, et al. 1995. Assignment of the human p27Kip1 gene to 12p13 and its analysis in leukemias. *Cancer Res* **55**: 1206-1210.
- Pinson KI, Brennan J, Monkley S, Avery BJ and Skarnes WC. 2000. An LDL-receptor-related protein mediates Wnt signalling in mice. *Nature* **407**: 535-538.
- Pohar N, Godenschwege TA, Buchner E. 1999. Invertebrate tissue inhibitor of metalloproteinase: structure and nested gene organization within the synapsin locus is conserved from *Drosophila* to human. *Genomics* **15**: 293-296.

- Poirel H, Oury C, Carron C, Duprez E, Laabi Y, Tsapis A, Romana SP, Mauchauffe M, Le Coniat M, Berger R, Ghysdael J, Bernard OA. 1997. The TEL gene products: nuclear phosphoproteins with DNA binding properties. *Oncogene* **14**: 349-357.
- Poirel H, Lacronique V, Mauchauffe M, Le Coniat M, Raffoux E, Daniel MT, Erickson P, Drabkin H, MacLeod RA, Drexler HG, Ghysdael J, Berger R, Bernard OA. 1998. Analysis of TEL proteins in human leukemias. *Oncogene* **16**: 2895-2903.
- Poirel H, Lopez RG, Lacronique V, Della Valle V, Mauchauffe M, Berger R, Ghysdael J, Bernard OA. 2000. Characterization of a novel ETS gene, TELB, encoding a protein structurally and functionally related to TEL. *Oncogene* **19**: 4802-4806.
- Polakis P. 2000. Wnt signaling and cancer. *Genes Dev* **14**: 1837-1851.
- Potter MD, Buijs A, Kreider B, van Rompaey L, Grosveld GC. 2000. Identification and characterization of a new human ETS-family transcription factor, TEL2, that is expressed in hematopoietic tissues and can associate with TEL1/ETV6. *Blood* **95**: 3341-3348.
- Power DM, Ingleton PM, Flanagan J, Canario AV, Danks J, Elgar G, Clark MS. 2000. Genomic structure and expression of parathyroid hormone-related protein gene (PTHrP) in a teleost, *Fugu rubripes*. *Gene* **30**: 67-76.
- Pui CH, Crist WM, Look AT. 1990. Biology and clinical significance of cytogenetic abnormalities in childhood acute lymphoblastic leukemia. *Blood* **76**: 1449-1463.
- Raimondi SC, Williams DL, Callihan T, Peiper S, Rivera GK and Murphy SB. 1986. Nonrandom involvement of the 12p12 breakpoint in chromosome abnormalities of childhood acute lymphoblastic leukemia. *Blood* **68**: 69-75.

- Raimondi SC. 1993. Current status of cytogenetic research in childhood acute lymphoblastic leukemia. *Blood* **81**: 2237-2251.
- Rapraeger AC. 2001. Molecular interactions of syndecans during development. *Semin Cell Dev Biol* **12**: 107-116.
- Rego EM, Wang ZG, Peruzzi D, He LZ, Cordon-Cardo C, Pandolfi PP. 2001. Role of promyelocytic leukemia (PML) protein in tumor suppression. *J Exp Med* **193**: 521-529.
- Roest Crollius H, Jaillon O, Bernot A *et al.* 2000. Estimate of human gene number provided by genome-wide analysis using Tetraodon nigroviridis DNA sequence. *Nat Genet* **25**: 235-238.
- Romana SP, Mauchauffe M, Le Coniat M, Chumakov I, Le Paslier D, Berger R, Bernard OA. 1995. The t(12;21) of acute lymphoblastic leukemia results in a tel-AML1 gene fusion. *Blood* **85**: 3662-3670.
- Romana SP, Le Coniat M, Poirel H, Marynen P, Bernard O, Berger R. 1996. Deletion of the short arm of chromosome 12 is a secondary event in acute lymphoblastic leukemia with t(12;21). *Leukemia* **10**: 167-170.
- Rowley JD. 1998. The critical role of chromosome translocations in human leukemias. *Annu Rev Genet* **32**: 495-519.
- Rowley JD. 1999. The role of chromosome translocations in leukemogenesis. *Semin Hematol* **36** (Suppl 7): 59-72.
- Rowitch DH, Echelard Y, Danielian PS, Gellner K, Brenner S, McMahon AP. 1998. Identification of an evolutionarily conserved 110 base-pair cis-acting regulatory

- sequence that governs Wnt-1 expression in the murine neural plate. *Development* **125**: 2735-2746.
- Rubnitz JE, Look AT. 1998a. Molecular basis of leukemogenesis. *Curr Opin Hematol* **5**: 264-270.
- Rubnitz JE, Look AT. 1998b. Molecular genetics of childhood leukemias. *J Pediatr Hematol Oncol* **20**: 1-11.
- Rubnitz JE, Pui CH, Downing JR. 1999. The role of TEL fusion genes in pediatric leukemias. *Leukemia* **13**: 6-13.
- Sathasivam K, Baxendale S, Mangiarini L, Bertaux F, Hetherington C, Kanazawa I, Lehrach H, Bates GP. 1997. Aberrant processing of the Fugu HD (FrHD) mRNA in mouse cells and in transgenic mice. *Hum Mol Genet* **6**: 2141-2149.
- Sato Y, Kobayashi H, Suto Y, Olney HJ, Davis EM, Super HG, Espinosa R 3rd, Le Beau MM, Rowley JD. 2001. Chromosomal instability in chromosome band 12p13: multiple breaks leading to complex rearrangements including cytogenetically undetectable sub-clones. *Leukemia* **15**: 1193-1202.
- Sauer B. 1998. Inducible gene targeting in mice using the Cre/lox system. *Methods* **14**: 381-392.
- Schwab M, Alitalo K, Klempnauer KH, Varmus HE, Bishop JM, Gilbert F, Brodeur G, Goldstein M, Trent J. 1983. Amplified DNA with limited homology to *myc* cellular oncogene is shared by human neuroblastoma cell lines and a neuroblastoma tumor. *Nature* **305**: 245-248.

- Schindler T, Bornmann W, Pellicena P, Miller WT, Clarkson B, Kuriyan J. 2000. Structural mechanism for STI-571 inhibition of abelson tyrosine kinase. *Science* **289**: 1938-1942.
- Schwaller J, Parganas E, Wang D, Cain D, Aster JC, Williams IR, Lee CK, Gerthner R, Kitamura T, Frantsve J, Anastasiadou E, Loh ML, Levy DE, Ihle JN, Gilliland DG. 2000. Stat5 is essential for the myelo- and lymphoproliferative disease induced by TEL/JAK2. *Mol Cell* **6**: 693-704.
- Schuz J, Grigat JP, Brinkmann K, Michaelis J. 2001. Residential magnetic fields as a risk factor for childhood acute leukaemia: results from a German population-based case-control study. *Int J Cancer* **91**: 728-735.
- Schwaller J, Frantsve J, Aster J, Williams IR, Tomasson MH, Ross TS, Peeters P, Van Rompaey L, Van Etten RA, Ilaria R Jr, Marynen P, Gilliland DG. 1998. Transformation of hematopoietic cell lines to growth-factor independence and induction of a fatal myelo- and lymphoproliferative disease in mice by retrovirally transduced TEL/JAK2 fusion genes. *EMBO J* **17**: 5321-5333
- Seipel K, Medley QG, Kedersha NL *et al.* 1999. Trio amino-terminal guanine nucleotide exchange factor domain expression promotes actin cytoskeleton reorganization, cell migration and anchorage-independent cell growth. *J Cell Sci* **112**: 1825-1834.
- Sementchenko VI, Watson DK. 2000. Ets target genes: past, present and future. *Oncogene* **19**: 6533-6548.
- Serrano M, Hannon GJ, Beach D. 1993. A new regulatory motif in cell-cycle control causing specific inhibition of cyclin D/CDK4. *Nature* **366**: 704-707.

- Serrano M, Gomez-Lahoz E, DePinho RA, Beach D, Bar-Sagi D. 1995. Inhibition of ras-induced proliferation and cellular transformation by p16INK4. *Science* **267**: 249-252.
- Shiohara M, el-Deiry WS, Wada M, Nakamaki T, Takeuchi S, Yang R, Chen DL, Vogelstein B, Koeffler HP. 1994. Absence of WAF1 mutations in a variety of human malignancies. *Blood* **84**: 3781-3784.
- Shurtleff SA, Buijs A, Behm FG, Rubnitz JE, Raimondi SC, Hancock ML, Chan GC, Pui CH, Grosveld G, Downing JR. 1995. TEL/AML1 fusion resulting from a cryptic t(12;21) is the most common genetic lesion in pediatric ALL and defines a subgroup of patients with an excellent prognosis. *Leukemia* **9**: 1985-1989.
- Simons M, Horowitz A. 2001. Syndecan-4-mediated signalling. *Cell Signal* **13**: 855-862.
- Sinn E, Muller W, Pattengale P, Tepler I, Wallace R, Leder P. 1987. Coexpression of MMTV/v-Ha-ras and MMTV/c-myc genes in transgenic mice: synergistic action of oncogenes in vivo. *Cell* **49**: 465-475.
- Sinnett D, Lavergne L, Melancon SB, Dallaire L, Potier M, Labuda D. 1988. Lesch-Nyhan syndrome: molecular investigation of three French Canadian families using a hypoxanthine-guanine phosphoribosyltransferase cDNA probe. *Hum Genet* **81**: 4-8.
- Slupsky CM, Gentile LN, Donaldson LW, Mackereth CD, Seidel JJ, Graves BJ, McIntosh LP. 1998. Structure of the Ets-1 pointed domain and mitogen-activated protein kinase phosphorylation site. *Proc Natl Acad Sci USA* **95**:12129-12134.

- Smith JS, Tachibana I, Pohl U, Lee HK, Thanarajasingam U, Portier BP, Ueki K, Ramaswamy S, Billings SJ, Mohrenweiser HW, Louis DN, Jenkins RB. 2000. A transcript map of the chromosome 19q-arm glioma tumor suppressor region. *Genomics* **64**: 44-50.
- Spirin KS, Simpson JF, Takeuchi S, Kawamata N, Miller CW, Koeffler HP. 1996. p27/Kip1 mutation found in breast cancer. *Cancer Res* **56**: 2400-2404.
- Stegmaier K, Pendse S, Barker GF, Bray-Ward P, Ward DC, Montgomery KT, Krauter KS, Reynolds C, Sklar J, Donnelly M, et al. 1995. Frequent loss of heterozygosity at the TEL gene locus in acute lymphoblastic leukemia of childhood. *Blood* **86**: 38-44.
- Stegmaier K, Takeuchi S, Golub TR, Bohlander SK, Bartram CR and Koeffler HP. 1996. Mutational analysis of the candidate tumor suppressor genes TEL and KIP1 in childhood acute lymphoblastic leukemia. *Cancer Res* **56**: 1413-1417.
- Takeuchi S, Bartram CR, Wada M, Reiter A, Hatta Y, Seriu T, Lee E, Miller CW, Miyoshi I, Koeffler HP. 1995. Allelotype analysis of childhood acute lymphoblastic leukemia. *Cancer Res* **55**: 5377-5382.
- Takeuchi S, Mori N, Koike M, Slater J, Park S, Miller CW, Miyoshi I, Koeffler HP. 1996a. Frequent loss of heterozygosity in region of the KIP1 locus in non-small cell lung cancer: evidence for a new tumor suppressor gene on the short arm of chromosome 12. *Cancer Res* **56**: 738-740.
- Takeuchi S, Bartram CR, Miller CW, Reiter A, Seriu T, Zimmerann M, Schrappe M, Mori N, Slater J, Miyoshi I, Koeffler HP. 1996b. Acute lymphoblastic leukemia of childhood: identification of two distinct regions of deletion on the short arm of chromosome 12 in the region of TEL and KIP1. *Blood* **87**: 3368-3374.

- Tamai K, Semenov M, Kato Y, Spokony R, Liu C, Katsuyama Y, Hess F, Saint-Jeannet JP, He X. 2000. LDL-receptor-related proteins in Wnt signal transduction. *Nature* **407**: 530-535.
- Tanoue T, Yamamoto T, Maeda R and Nishida E. 2001. A Novel MAPK phosphatase MKP-7 acts preferentially on JNK/SAPK and p38 alpha and beta MAPKs. *J Biol Chem* **276**: 26629-26639.
- Taylor JS, Van de Peer Y, Braasch I, Meyer A. 2001. Comparative genomics provides evidence for an ancient genome duplication event in fish. *Philos Trans R Soc Lond B Biol Sci* **356**: 1661-1679.
- Thandla SP, Ploski JE, Raza-Egilmez SZ, Chhalliyil PP, Block AW, de Jong PJ, Aplan PD. 1999. ETV6-AML1 translocation breakpoints cluster near a purine/pyrimidine repeat region in the ETV6 gene. *Blood* **93**: 293-299.
- Thompson JD, Higgins DG and Gibson TJ. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**: 4673-4680.
- Tomasson MH, Williams IR, Hasserjian R, Udomsakdi C, McGrath SM, Schwaller J, Druker B, Gilliland DG. 1999. TEL/PDGFBetaR induces hematologic malignancies in mice that respond to a specific tyrosine kinase inhibitor. *Blood* **93**:1707-1714.
- Trower MK, Orton SM, Purvis IJ, Sanseau P, Riley J, Christodoulou C, Burt D, See CG, Elgar G, Sherrington R, Rogaev EI, St George-Hyslop P, Brenner S, Dykes CW. 1996. Conservation of synteny between the genome of the pufferfish (*Fugu rubripes*) and the region on human chromosome 14 (14q24.3) associated

- with familial Alzheimer disease (AD3 locus). *Proc Natl Acad Sci U S A* **20**: 1366-1369.
- Tsujimoto Y, Finger LR, Yunis J, Nowell PC, Croce CM. 1984. Cloning of the chromosome breakpoint of neoplastic B cells with the t(14;18) chromosome translocation. *Science* **226**: 1097-1099.
- Valdembri D, Serini G, Vacca A, Ribatti D, Bussolino F. 2001. In vivo activation of JAK2/STAT-3 pathway during angiogenesis induced by GM-CSF. *FASEB J*.
- Van Rompaey L, Dou W, Buijs A and Grosveld G. 1999. Tel, a frequent target of leukemic translocations, induces cellular aggregation and influences expression of extracellular matrix components. *Neoplasia* **1**: 526-536.
- Van Rompaey L, Potter M, Adams C, Grosveld G. 2000. Tel induces a G1 arrest and suppresses ras-induced transformation. *Oncogene* **19**: 5244-5250.
- Venkatesh B, Si-Hoe SL, Murphy D, Brenner S. 1997. Transgenic rats reveal functional conservation of regulatory controls between the Fugu isotocin and rat oxytocin genes. *Proc Natl Acad Sci U S A* **94**: 12462-12466.
- Venkatesh B, Gilligan P, Brenner S. 2000. Fugu: a compact vertebrate reference genome. *FEBS Lett* **476**: 3-7.
- Venter DJ, Bevan KL, Ludwig RL, Riley TEW, Jat PS, Thomas DGT, Noble MD. 1991. Retinoblastoma gene deletions in human glioblastomas. *Oncogene* **6**: 445-448.
- Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO et al. 2001. The sequence of the human genome. *Science* **291**: 1304-1351.

- Villard L, Tassone F, Crnogorac-Jurcevic T, Clancy K, Gardiner K. 1998. Analysis of pufferfish homologues of the AT-rich human APP gene. *Gene* **210**: 17-24.
- Vogelstein B and Kinzler KW. 1993. The multistep nature of cancer. *Trends Gen* **9**: 138-141.
- Wada M, Bartram CR, Nakamura H, Hachiya M, Chen DL, Borenstein J, Miller CW, Ludwig L, Hansen-Hagge TE, Ludwig WD, et al. 1993. Analysis of p53 mutations in a large series of lymphoid hematologic malignancies of childhood. *Blood* **82**: 3163-3169.
- Wai DH, Knezevich SR, Lucas T, Jansen B, Kay RJ, Sorensen PH. 2000. The ETV6-NTRK3 gene fusion encodes a chimeric protein tyrosine kinase that transforms NIH3T3 cells. *Oncogene* **19**: 906-915.
- Wasylyk B, Hagman J, Gutierrez-Hartmann A. 1998. Ets transcription factors: nuclear effectors of the Ras-MAP-kinase signaling pathway. *Trends Biochem Sci* **23**: 213-216.
- Wang LC, Kuo F, Fujiwara Y, Gilliland DG, Golub TR and Orkin SH. 1997. Yolk sac angiogenic defect and intra-embryonic apoptosis in mice lacking the Ets-related factor TEL. *Embo J* **16**: 4374-4383.
- Wang LC, Swat W, Fujiwara Y, Davidson L, Visvader J, Kuo F, Alt FW, Gilliland DG, Golub TR, Orkin SH. 1998. The TEL/ETV6 gene is required specifically for hematopoiesis in the bone marrow. *Genes Dev* **12**: 2392-2402.
- Wang JY. 2000. Regulation of cell death by the Abl tyrosine kinase. *Oncogene* **19**: 5643-5650.

- Wang PW, Eisenbart JD, Espinosa R 3rd, Davis EM, Larson RA, Le Beau MM. 2000. Refinement of the smallest commonly deleted segment of chromosome 20 in malignant myeloid diseases and development of a PAC-based physical and transcription map. *Genomics* **67**: 28-39.
- Wang L, Hiebert SW. 2001. TEL contacts multiple co-repressors and specifically associates with histone deacetylase-3. *Oncogene* **20**: 3716-3725.
- Warrell RP, HE LZ, Richon V, Calleja E, Pandolfi PP. 1998. Therapeutic targeting of transcription in acute promyelocytic leukemia by use of an inhibitor of histone deacetylase. *J Natl Cancer Inst* **90**: 1621-1625.
- Wehrli M, Dougan ST, Caldwell K, O'Keefe L, Schwartz S, Vaizel-Ohayon D, Schejter E, Tomlinson A, DiNardo S. arrow encodes an LDL-receptor-related protein essential for Wingless signalling. *Nature* **407**: 527-530.
- Weis K, Rambaud S, Lavau C, Jansen J, Carvalho T, Carmo-Fonseca M, Lamond A, Dejean A. 1994. Retinoic acid regulates aberrant nuclear localization of PML-RAR alpha in acute promyelocytic leukemia cells. *Cell* **76**: 345-356.
- Wiemels JL, Greaves M. 1999. Structure and possible mechanisms of TEL-AML1 gene fusions in childhood acute lymphoblastic leukemia. *Cancer Res* **59**: 4075-4082.
- Wiemels JL, Ford AM, Van Wering ER, Postma A, Greaves M. 1999. Protracted and variable latency of acute lymphoblastic leukemia after TEL-AML1 gene fusion in utero. *Blood* **94**: 1057-1062.
- Wiese C, Gauny SS, Liu WC, Cherbonnel-Lasserre CL, Kronenberg A. 2001. Different mechanisms of radiation-induced loss of heterozygosity in two human lymphoid cell lines from a single donor. *Cancer Res* **61**: 1129-1137.

- Wilbanks AM, Mahajan S, Frank DA, Druker BJ, Gilliland DG, Carroll M. 2000. TEL/PDGFBetaR fusion protein activates STAT1 and STAT5: a common mechanism for transformation by tyrosine kinase fusion proteins. *Exp Hematol* **28**: 584-593.
- Worm J, Bartkova J, Kirkin AF, Straten P, Zeuthen J, Bartek J, Guldberg P. 2000. Aberrant p27Kip1 promoter methylation in malignant melanoma. *Oncogene* **19**: 5111-5115.
- Xiao W, Oefner PJ. 2001. Denaturing high-performance liquid chromatography: A review. *Hum Mutat* **17**: 439-474.
- Yokozawa T, Towatari M, Iida H, Takeyama K, Tanimoto M, Kiyoi H, Motoji T, Asou N, Saito K, Takeuchi M, Kobayashi Y, Miyawaki S, Kidera Y, Ohno R, Saito H, Naoe T. 2000. Prognostic significance of the cell cycle inhibitor p27Kip1 in acute myeloid leukemia. *Leukemia* **14**: 28-33.
- Zamore PD. 2001. RNA interference: listening to the sound of silence. *Nat Struct Biol* **8**: 746-750.
- Zhang Y, Matthiesen P, Harder S, Siebert R, Castoldi G, Calasanz MJ, Wong KF, Rosenwald. 2000. A 3-cM commonly deleted region in 6q21 in leukemias and lymphomas delineated by fluorescence in situ hybridization. *Genes Chromosomes Cancer* **27**: 52-58.

Curriculum vitae

FORMATION ACADÉMIQUE

sept.1998 - hiver 2002

Ph.D. en Biochimie. Université de Montréal.
 Directeur de recherche: Dr. Daniel Sinnett.
 Titre de la thèse: "Caractérisation du locus *ETV6-CDKN1B* impliqué dans la leucémie lymphoblastique aiguë".

mai 1996 - août 1998

M.Sc. en Biochimie. Université de Montréal.
 Directrice de recherche: Dr. Léa Brakier-Gingras.
 Titre du mémoire: "Caractérisation de l'architecture du domaine 3' majeur du RNA ribosomique 16S d'*Escherichia coli*."

sept.1993 - mai 1996

B.Sc. en Biochimie. Université de Montréal.

PUBLICATIONS

Montpetit A, Larose J, Sinnett D. 2002. Expression and mutational analysis of the 12p12 candidate tumor suppressor genes in pre-B acute lymphoblastic leukemia. En préparation.

Montpetit A, Wilson MD, Chevrette M, Koop B, Sinnett D. 2002. Absence of synteny at the *ETV6* locus between human and Fugu genomes. En préparation.

Montpetit A, Boily G, Sinnett D. 2002. A detailed transcriptional map of the chromosome 12p12 tumor suppressor locus. *Eur J Hum Gen* **10**: 62-71

Montpetit A, Sinnett D. 2001. Comparative analysis of the *ETV6* gene in vertebrate genomes from pufferfish to human. *Oncogene* **20**: 3437-3442.

Sinnett D, Aïssani B, Bonan C, Montpetit A. 1999. A chromosome 12p12.3 tumor suppressor locus involved in hematological malignancies and solid neoplasias. *Recent Res Devel Cancer* **1**: 213-227.

Montpetit A, Sinnett D. 1999. Physical mapping of the G-protein coupled receptor 19 (GPR 19) in the chromosome 12p12.3 region frequently rearranged in cancer cells. *Hum Genet* **105**: 162-164.

Montpetit A, Payant C, Nolan JM, Brakier-Gingras L. 1998. Analysis of the conformation of the 3' major domain of *Escherichia coli* 16S ribosomal RNA using site-directed photaffinity crosslinking. *RNA* 4: 1455-1466.

COMMUNICATIONS

Larose J, Montpetit A, Sinnett D. Expression and mutational analysis of candidate genes in the chromosome 12p tumor suppressor locus. AACR meeting, San Francisco, april 2002.

Montpetit A, Sinnett D. Identification de gènes candidats supresseurs de tumeurs sur le chromosome 12 . Rencontre scientifique Louis-Philippe Bouthillier, Orford, october 2001.

Montpetit A, Trudel N, Sinnett D. Une analyse détaillée du locus 12p12.3 impliqué dans la leucémie lymphoblastique aiguë. XVIe congrès annuel de la recherche des étudiants gradués et post-gradués, Centre de Recherche de l'Hôpital Ste-Justine, juin 2001.

D. Sinnett, A. Montpetit, N. Trudel. A transcript map of the human chromosome 12p12.3 tumor suppressor locus. 10th International Congress of Human Genetics, Vienna, Austria, May 2001.

Montpetit A, Trudel N, Sinnett D. A detailed transcript map of the human chromosome 12p12.3 tumor suppressor locus: The usefulness of an integrative approach. Oncogenomics conference: Dissecting cancer through genome research, Tucson, Az. Jan. 2001.

Montpetit A, Sinnett D. Comparative analysis in *Fugu rubripes* of the 12p12.3 locus, a region frequently affected in cancer. HUGO meeting 2000, Vancouver, BC.

Montpetit A, Sinnett D. Comparative analysis in *Fugu rubripes* of the 12p12.3 locus, a region frequently affected in cancer. CGDN Annual Scientific Meeting 2000, Vancouver, BC.

Montpetit A, Sinnett D. *Fugu rubripes* as a model for comparative study of the *ETV6* locus, frequently rearranged in various hematological disorders. 9th International Workshop: Beyond the Identification of Transcribed Sequences: Functional and Expression Analysis. Reston, Va. Oct. 1999.

Montpetit A, Payant C, Nolan JM, Brakier-Gingras L. 1998. Site-directed crosslinking within the 3' major domain of *E.coli* 16S ribosomal RNA. RNA'98, Madison, Wi. May 1998.

Montpetit A, Nolan JM, Brakier-Gingras L. Mapping the three-dimensional structure of the 3' major domain of *Escherichia coli* 16S ribosomal RNA using site-directed cross-linking. RNA Structure Symposium, Santa Cruz, Ca. June 1997.