

Université de Montréal

**Étude de l'organisation fonctionnelle du génome humain par un modèle  
d'interactions entre les séquences répétitives de type LINE-1**

Par

**Hélène D'Anjou**

Programme de Biologie Moléculaire

Faculté des Études Supérieures

Thèse présentée à la Faculté des Études Supérieures

en vue de l'obtention du grade de

Philosophiae Docteur (Ph.D.)

en Biologie Moléculaire

Avril 2003

© Hélène D'Anjou, 2003



QH  
506  
U54  
2004  
v.002

**Direction des bibliothèques**

**AVIS**

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

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

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

**NOTICE**

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

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

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

## **IDENTIFICATION DU JURY**

Université de Montréal

Faculté des études supérieures

**Cette thèse intitulée :**

Étude de l'organisation fonctionnelle du génome humain par un modèle  
d'interactions entre les séquences répétitives de type LINE-1

**présentée par :**

Hélène D'Anjou

**a été évaluée par un jury composé des personnes suivantes :**

**Abdellah Belmaaza:** *président rapporteur*

**Pierre Chartrand:** *directeur de recherche*

**Edward Bradley:** *membre du jury*

**Mark Featherstone:** *examineur externe*

**Benoît Coulombe:** *représentant du doyen de la FES*

## RÉSUMÉ

Au-delà de son séquençage, l'établissement de l'organisation tridimensionnelle du génome humain est essentiel à la pleine compréhension de son fonctionnement. Nous savons maintenant qu'à l'interphase, les chromosomes sont dispersés sous forme de territoires chromosomiques à l'intérieur desquels la chromatine est organisée selon un patron de compaction hautement conservé. Bien qu'un certain dynamisme semble exister à l'intérieur de cette organisation, sa conservation pourrait être cruciale au contrôle de l'accessibilité génomique entre les différentes régions du génome.

Pour étudier cette accessibilité, nous avons élaboré un modèle basé sur les interactions homologues entre les éléments LINES du génome humain. Ces séquences répétitives naturelles sont dispersées en 500,000 copies sur l'ensemble des chromosomes et l'homologie qu'elles partagent leur permet d'interagir par recombinaison homologue (RH), un mécanisme important dans la réparation des cassures double-brin d'ADN (CDBs). Dans notre système, l'induction ciblée d'une CDB à l'intérieur d'un élément LINE intégré à un locus donné du génome permet de stimuler la production d'événements de réparation par RH avec des LINES endogènes. Le séquençage des événements de recombinaison permet ensuite de retracer dans le génome entier les partenaires endogènes impliqués dans la réparation de la cassure et de comprendre par quel mécanisme ils ont interagit.

Pour 2 loci distincts, nous avons démontré que le patron d'accessibilité génomique était contrôlé par l'organisation tridimensionnelle de la chromatine et conservé d'une cellule à l'autre. Nos résultats démontrent que l'accessibilité aux sites intrachromosomiques est dictée par l'organisation interne des territoires, et que la proximité physique linéaire d'un site n'est pas la considération première. Ils suggèrent également l'existence d'une proximité particulière entre des territoires distincts, qui favoriserait des interactions interchromosomiques.

Notre étude de la recombinaison ectopique entre les séquences répétitives a révélé qu'une vaste gamme de partenaires endogènes pouvaient être impliqués, et que leur niveau de divergence ne constituait pas un obstacle majeur. Cependant, ces interactions pourraient être stimulées dans une région transcriptionnellement active augmentant l'accessibilité à la machinerie de recombinaison, ainsi que dans les régions fragiles plus sensibles aux nucléases.

Pour restreindre les effets délétères de la RH ectopique entre les séquences répétitives du génome, la cellule pourrait exercer un contrôle au niveau du mécanisme de recombinaison. Nous avons démontré que la RH induite par une CDB dans un élément LINE s'effectuait principalement par un mécanisme conservatif de type conversion génique, et que la prédominance de segments de conversion unidirectionnels suggérait l'utilisation d'un mécanisme d'invasion homologue par une seule extrémité. Dans une lignée cellulaire déficiente pour le gène ATM qui contrôle la recombinaison, nous avons observé quelques événements de RH délétères de type *crossing-over* ainsi que plusieurs événements de réarrangements chromosomiques non-homologues menaçant l'intégrité du génome.

Constamment affligé de CDBs, le génome humain dont le tiers de l'espace est occupé par des séquences répétitives dispersées doit finement contrôler les interactions homéologues ectopiques pouvant menacer sa stabilité. En restreignant l'accessibilité génomique entre ces séquences et en privilégiant un mécanisme d'interaction conservatif, la cellule s'assure de préserver son intégrité. Conséquemment, la perte de ce contrôle pourrait significativement contribuer à l'instabilité des cellules tumorales.

## **MOTS CLÉS**

Accessibilité, cassure double-brin, conversion génique, instabilité génomique, recombinaison homologue, réparation d'ADN, territoires chromosomiques.

## ABSTRACT

Beyond the linear sequence of DNA, establishment of a three-dimensional organization for the human genome is important to fully understand its function. In the interphase nucleus, we know that chromosomes occupy distinct territories inside of which a chromatin compaction pattern is highly conserved. Despite dynamism of this structure, conservation of certain aspects of this pattern would be essential to control the access between different genomic loci.

To study this genomic access, we used a system based on homologous interactions between LINE elements of the human genome. These natural repetitive sequences are highly dispersed on all chromosomes in 500,000 copies. They share a high degree of homology allowing interactions by homologous recombination (HR), an important pathway of double-strand break (DSB) repair. In our system, we induced a DSB in an integrated LINE element to stimulate the repair by HR with endogenous LINE elements. Sequencing of repair events allows identification and localization of endogenous partners, and characterization of the repair mechanisms used.

For two distinct loci, we showed that genomic access is controlled by the three-dimensional organization pattern, and is conserved by each cell. Our results show that intrachromosomal access is restricted by the internal organization of chromosomal territories, and that linear proximity does not stimulate interaction. Our study also suggests that a preferential proximity could exist between given territories, favouring precise interchromosomal interactions.

The study of ectopic recombination between repetitive elements reveals that a large range of endogenous partners could be implicated in interactions without sharing precise homology. However, we demonstrated that HR is enhanced by the transcriptional activity of chromatin due to higher access to recombination machinery. HR could also be enhanced in fragile sites, where DNA is more sensitive to nucleases.

Deleterious effects of ectopic HR between repetitive sequences could be restricted by the control of the recombination mechanism. When a DSB is induced in a LINE element, HR repair is mostly performed by a conservative gene conversion mechanism, and the unidirectionality of conversion tracts suggest the use of a one-sided invasion mechanism. In a human cell line deficient for ATM, a gene controlling recombination, we observed few deleterious crossing-over HR events and many non-homologous chromosomal rearrangements, representing serious threats for genomic integrity.

A third of the human genome is occupied by dispersed repeat elements. Highly afflicted by numerous DSBs, cells must control homeologous ectopic interactions affecting its genomic stability. By restricting genomic access between these elements and favouring a conservative mechanism of interaction, cells can preserve their integrity. Consequently, the loss of this control could significantly contribute to the genetic instability of tumour cells.

### **KEY WORDS**

Access, double-strand break, chromosomal territories, DNA repair, gene conversion, genomic instability, homologous recombination.

## TABLE DES MATIÈRES

|   |              |
|---|--------------|
| <b>RÉSUMÉ.....</b>  | <b>III</b>   |
| <b>ABSTRACT .....</b>   | <b>V</b>     |
| <b>TABLE DES MATIÈRES .....</b>   | <b>VII</b>   |
| <b>LISTE DES FIGURES.....</b>   | <b>X</b>     |
| <b>LISTE DES TABLEAUX .....</b>   | <b>XIII</b>  |
| <b>LISTE DES SIGLES ET ABRÉVIATIONS .....</b>                                   | <b>XV</b>    |
| <b>REMERCIEMENTS.....</b>   | <b>XVIII</b> |
| <b>I. INTRODUCTION.....</b>   | <b>19</b>    |
| <b>I.1. Objectifs de recherche .....</b>  | <b>19</b>    |
| <b>I.2. Organisation du génome humain.....</b>                                  | <b>21</b>    |
| I.2.1. Territoires chromosomiques et espace interchromatique.....               | 21           |
| I.2.2. Organisation intraterritoriale.....                                      | 23           |
| I.2.3. Conservation de l'organisation tridimensionnelle du génome.....          | 24           |
| <b>I.3. Modèle d'étude : les séquences répétitives dispersées .....</b>         | <b>25</b>    |
| <b>I.4. Les cassures double-brin (CDBs).....</b>                                | <b>28</b>    |
| <b>I.5. Détection et signalement des CDBs.....</b>                              | <b>29</b>    |
| <b>I.6. Réparation des CDBs .....</b>   | <b>30</b>    |
| <b>I.7. Recombinaison non-homologue .....</b>                                   | <b>33</b>    |
| I.7.1. Mécanisme de <i>non-homologous end-joining</i> (NHEJ).....               | 33           |
| I.7.2. Réarrangements chromosomiques non-homologues (NHCR).....                 | 33           |
| I.7.3. Déficience de la recombinaison non-homologue.....                        | 36           |
| <b>I.8. Recombinaison homologue (RH) .....</b>                                  | <b>37</b>    |
| I.8.1. Mécanisme de réparation des CDBs par RH ( <i>DSB repair model</i> )..... | 37           |
| I.8.2. Mécanisme de <i>one-sided invasion</i> .....                             | 41           |
| I.8.3. Mécanisme de <i>single-strand annealing</i> (SSA) .....                  | 43           |
| I.8.4. RH et instabilité génomique .....  | 43           |
| <b>I.9. Étude de la RH chez les cellules mammifères .....</b>                   | <b>47</b>    |
| I.9.1. Système d'induction de CDBs par <i>I-SceI</i> .....                      | 47           |
| I.9.2. Modèles d'étude de la RH intrachromosomique.....                         | 49           |
| I.9.3. Étude de la RH méiotique dans les cellules germinales.....               | 50           |

|  |            |
|--|------------|
| 1.9.3. Étude de la RH mitotique dans les cellules somatiques.....  | 51         |
| 1.9.4. Étude de la RH ectopique .....  | 51         |
| 1.9.5. Influence de la divergence sur la RH .....  | 52         |
| 1.9.6. Implication du système de MMR dans le contrôle de la RH .....   | 54         |
| 1.9.7. Influence de la structure de la chromatine sur la RH .....  | 55         |
| <br>   |            |
| <b>II. MATÉRIEL ET MÉTHODE .....</b>   | <b>59</b>  |
| <br>   |            |
| <b>II.1. Construction du vecteur pL1HsScel .....</b>   | <b>59</b>  |
| <b>II.2. Lignée cellulaire.....</b>  | <b>59</b>  |
| <b>II.3. Transfection stable du vecteur pL1HsScel.....</b>   | <b>59</b>  |
| <b>II.4. Extraction de l'ADN génomique .....</b>   | <b>60</b>  |
| <b>II.5. Analyse des sites d'intégration.....</b>  | <b>60</b>  |
| <b>II.6. Induction de la cassure double-brin.....</b>  | <b>61</b>  |
| <b>II.7. Plasmid rescue.....</b>   | <b>61</b>  |
| <b>II.8. Analyse des recombinants.....</b>   | <b>62</b>  |
| <br>   |            |
| <b>III. RÉSULTATS.....</b>   | <b>63</b>  |
| <br>   |            |
| <b>III.1. Stratégie expérimentale .....</b>  | <b>63</b>  |
| <b>III.2. Génération et analyse des clones.....</b>  | <b>66</b>  |
| <b>III.3. Induction d'une cassure double-brin au site I-SceI.....</b>  | <b>70</b>  |
| <b>III.4. Récupération et analyse des événements de réparation.....</b>  | <b>73</b>  |
| III.4.1. Fréquence relative des différents mécanismes de réparation de CDB .....                                       | 76         |
| Mécanisme .....  | 77         |
| <b>III.5. Analyse des événements de recombinaison non-homologue.....</b>   | <b>80</b>  |
| III.5.1. Analyse des événements de réparation de CDB par NHEJ .....  | 80         |
| III.5.2. Analyse des événements de réparation de CDB par NHCR.....   | 82         |
| <b>III.6. Analyse des événements de recombinaison homologue (RH).....</b>  | <b>87</b>  |
| III.6.1. Analyse des événements de RH du clone ATA34 .....   | 87         |
| III.6.2. Analyse des événements de RH du clone ATA67. ....   | 91         |
| III.6.2.1. Identification des partenaires par la méthode du BLAST.....   | 91         |
| III.6.2.2. Identification des partenaires par alignement des segments de conversion.....                               | 98         |
| III.6.2.3. Description détaillée des événements ayant impliqué le même partenaire dans le clone<br>ATA67 .....         | 102        |
| III.6.2.4. Analyse des événements de crossing-over .....   | 109        |
| III.6.2.5. Analyse de l'événement de one-sided invasion 67A10-240.....   | 111        |
| III.6.3. Analyse des segments de conversion .....  | 113        |
| <b>III.7. Environnement détaillé des locus d'intégration ATA34 et ATA67.....</b>                                       | <b>117</b> |
| III.7.1. Distribution des partenaires intrachromosomiques de RH et de NHCR du clone ATA34 sur<br>le chromosome 6q..... | 117        |

|   |            |
|---|------------|
| III.7.2. Distribution des partenaires intrachromosomiques de RH et de NHCR du clone ATA67 sur le chromosome 2p .....                | 117        |
| III.7.3. Environnement immédiat détaillé du locus ATA34 .....   | 120        |
| III.7.4. Environnement immédiat détaillé du locus ATA67 .....   | 120        |
| <b>III.8. Distribution des gènes et des séquences répétitives sur les chromosomes d'intégration des clones ATA34 et ATA67 .....</b> | <b>123</b> |
| <b>IV. DISCUSSION .....</b>   | <b>129</b> |
| IV.1. Contribution du projet.....   | 129        |
| IV.2. Nature et fréquence des événements de recombinaison .....   | 130        |
| IV.2.1. Induction d'une cassure double-brin (CDB).....  | 130        |
| IV.2.2. Statuts ATM et p53 de la lignée GM05849 .....   | 132        |
| IV.2.3. Différence de fréquence de RH entre les clones .....  | 134        |
| IV.3. Événements de réparation non-homologue.....   | 135        |
| IV.3.1. Événements de <i>non-homologous end-joining</i> (NHEJ).....   | 135        |
| IV.3.2. Événements de <i>non-homologous chromosomal rearrangement</i> (NHCR) .....  | 136        |
| IV.4. Événements de RH .....  | 137        |
| IV.4.1. Identification des partenaires LI endogènes de RH.....  | 137        |
| IV.4.2. Mécanisme de RH.....  | 140        |
| IV.4.2.1. Effet du statut ATM sur la longueur du segment de conversion .....  | 140        |
| IV.4.2.2. Relation entre le mécanisme et la directionnalité de la conversion.....   | 141        |
| IV.4.2.3. Implication du système de réparation des mismatches (MMR) .....   | 144        |
| IV.4.3. Choix du partenaire de RH .....   | 145        |
| IV.4.3.1. Effet de la divergence.....   | 145        |
| IV.4.3.2. Effet de la longueur de l'homologie.....  | 147        |
| IV.5. Structure de la chromatine.....   | 149        |
| IV.5.1. Influence de l'activité transcriptionnelle.....   | 149        |
| IV.5.2. Influence de la méthylation.....  | 149        |
| IV.5.3. Position télomérique/ centromérique .....   | 150        |
| IV.5.4. Sites fragiles.....   | 152        |
| IV.6. Organisation tridimensionnelle du génome.....   | 154        |
| <b>V. CONCLUSION.....</b>   | <b>158</b> |
| <b>VI. RÉFÉRENCES .....</b>   | <b>161</b> |

## LISTE DES FIGURES

|  |    |
|--|----|
| Figure 1 : Structure des séquences répétitives LINEs (a) et SINEs (b).....   | 26 |
| Figure 2 : Rôles multiples de la protéine kinase ATM.....  | 31 |
| Figure 3 : Réarrangements chromosomiques délétères pouvant résulter de la RH<br>ectopique entre des séquences répétitives .....  | 32 |
| Figure 4 : Voie de réparation des CDBs par NHEJ.....   | 34 |
| Figure 5 : Mécanismes de réarrangements chromosomiques non-homologues.....   | 35 |
| Figure 6 : Mécanisme RH ( <i>DSB repair model</i> ).....   | 38 |
| Figure 7 : Produits de réparation des CDBs par RH .....  | 40 |
| Figure 8 : Mécanisme de RH par one-sided invasion .....  | 42 |
| Figure 9 : Mécanisme de RH par single-strand annealing (SSA) .....   | 44 |
| Figure 10 : Mécanisme du " <i>intron homing</i> " de l'endonucléase I-SceI dans les<br>mitochondries de la levure <i>Saccharomyces cerevisiae</i> <sup>140</sup> ..... | 48 |
| Figure 11 : Système d'étude de la réparation des CDBs chez les cellules mammifères<br>.....  | 49 |
| Figure 12 : Structure du vecteur pL1HsSce1 .....   | 64 |
| Figure 13 : Schéma de la stratégie expérimentale.....  | 65 |
| Figure 14 : Analyse des clones par hybridation Southern.....   | 67 |
| Figure 15 : Structure du pL1HsSce1 intégré dans les clones ATA34 et ATA67.....   | 68 |
| Figure 16 : Site d'intégration du pL1HsSce1 dans le clone ATA67 .....  | 69 |
| Figure 17 : Efficacité d'induction d'une CDB au site I-SceI dans les clones ATA34<br>(a) et ATA67 (b).....   | 71 |

|   |     |
|---|-----|
| Figure 18 : Classement des plasmides recombinants selon le mécanisme de réparation de la CDB .....  | 74  |
| Figure 19 : Séquençage des événements de réparation de CDB par NHCR .....   | 83  |
| Figure 20 : Classement des événements de RH du clone ATA34 par alignement des segments de conversion.....   | 90  |
| Figure 21 : Classement des événements de RH du clone ATA67 par alignement des segments de conversion.....   | 99  |
| Figure 22 : Hypothèse d'une conversion génique spontanée entre les partenaires L1 endogènes proximaux 67-B et 67-C dans la lignée cellulaire GM05849..... | 103 |
| Figure 23 : Alignement des segments de conversion des événements de RH impliquant le même partenaire 67-A(a), 67-B(b), 67-C(c) et 67-D(d) .....           | 104 |
| Figure 24 : Structure détaillée des événements de RH par crossing-over a) 67-R101, b) 67-232, c) 67-nt8.....  | 110 |
| Figure 25: Description de l'événement de RH 67A10-240.....  | 111 |
| Figure 26 : Alignement et classement des segments de conversion en fonction de leur longueur et de leur directionnalité.....                              | 115 |
| Figure 27: Localisation intrachromosomique des partenaires de RH et des sites de NHCR sur le chromosome 6q (clone ATA34).....                             | 118 |
| Figure 28: Localisation intrachromosomique des partenaires de RH et des sites de NHCR sur le chromosome 2p (clone ATA67).....                             | 119 |
| Figure 29 : Environnement immédiat détaillé du locus ATA34 sur le chromosome 6q23.3 .....   | 121 |
| Figure 30 : Environnement immédiat détaillé du locus ATA67 sur le chromosome 2p22.1 .....   | 122 |

|  |     |
|--|-----|
| Figure 31 : Distribution relative des L1s et des gènes dans le génome.....   | 124 |
| Figure 32 : Distribution des gènes, des séquences répétitives et des regroupements de partenaires du clone ATA34 sur le chromosome 6 .....         | 127 |
| Figure 33 : Distribution des gènes, des séquences répétitives et des regroupements de partenaires du clone ATA67 sur le chromosome 2 .....         | 128 |
| Figure 34: Corrélation entre la position centromérique relative des sites d'interaction interchromosomiques et leur fréquence d' interaction ..... | 152 |

## LISTE DES TABLEAUX

|   |     |
|---|-----|
| Tableau I : Phénotypes associés à la déficience des protéines impliquées dans le contrôle de la RH.....   | 45  |
| Tableau II : Fréquences relatives des mécanismes de réparation de CDBs obtenus avec les clones ATA34 et ATA67.....  | 77  |
| Tableau III : Comparaison des fréquences de recombinaison relatives obtenues pour le clone ATA67 et le sous-clone ATA67A10 avec ou sans enrichissement de la ligation par une digestion I-SceI..... | 79  |
| Tableau IV : Description des événements de réparation par NHEJ suite à une CDB au site I-SceI des clones ATA34 et ATA67.....  | 81  |
| Tableau V : Description des événements de réparation par NHCR suite à une CDB au site I-SceI du clone ATA34.....  | 84  |
| Tableau VI : Description des événements de réparation par NHCR suite à une CDB au site I-SceI du clone ATA67.....   | 85  |
| Tableau VII : Partenaires de RH identifiés pour le clone ATA34.....   | 88  |
| Tableau VIII: Partenaires intrachromosomiques de RH identifiés par la méthode du BLAST pour le clone ATA67.....   | 92  |
| Tableau IX: Partenaires interchromosomiques de RH identifiés par la méthode du BLAST pour le clone ATA67.....   | 94  |
| Tableau X :Événements de RH du clone ATA67 impliquant des partenaires endogènes non-identifiables par la méthode du BLAST.....  | 96  |
| Tableau XI : Proportion de l'unidirectionnalité et de la bidirectionnalité chez les événements de conversion génique.....   | 116 |

|  |     |
|--|-----|
| Tableau XII : Localisation des partenaires L1s endogènes dans des séquences<br>codantes .....  | 125 |
| Tableau XIII : Influence de la longueur d'homologie ininterrompue au site de CDB<br>sur le choix de la directionnalité de la conversion .....                          | 143 |
| Tableau XIV : Influence de degré d'homologie et de la longueur sur le choix d'un<br>partenaire L1 endogène de RH .....   | 146 |
| Tableau XV : Nombre d'aberrations chromosomiques associées à des néoplasies<br>humaines pour chaque site de translocation (a) et de RH interchromosomique (b)<br>..... | 153 |

## LISTE DES SIGLES ET ABRÉVIATIONS

|                 |   |
|-----------------|---|
| %               | pourcentage                                     |
| Ω               | ohm(s)  |
| °C              | degré(s) celcius                                |
| ADN             | acide désoxyribonucléique                       |
| ADNc            | ADN complémentaire                              |
| APRT            | adenine phosphoribosyltransferase               |
| ARN             | acide ribonucléique                             |
| A+T             | adénine+thymine                                 |
| ATM             | ataxia-telangiectasia mutated                   |
| ATP             | adenosine triphosphate                          |
| ATR             | ATM and Rad3 related                            |
| BLAST           | basical local alignment search tool             |
| BRCA1,2         | breast cancer susceptibility gene 1, 2          |
| CDB             | cassure double-brin                             |
| CHO             | Chinese hamster ovary                           |
| chr             | chromosome                                      |
| cm              | centimètre(s)                                   |
| CO <sub>2</sub> | dioxyde de carbone                              |
| DNA-PKc         | DNA-dependant protein kinase, catalytic subunit |
| DSB             | double-strand break                             |
| <i>E.coli</i>   | <i>Escherichia coli</i>                         |
| EMEM            | eagle's minimal essential medium                |
| EN              | endonucléase                                    |
| FBS             | fetal bovine serum                              |
| FISH            | fluorescence in situ hybridization              |
| G+C             | guanine+cytosine                                |
| G/Q             | Giemsa/Quinacrine                               |

|                 |  |
|-----------------|--|
| I-SceI          | intron encoded endonuclease I from <i>Saccharomyces cerevisiae</i> |
| kb              | kilopaire(s) de base   |
| K.-O.           | knock-out  |
| L1              | LINE-1   |
| L1Hs            | LINE-1 human specific  |
| LINE            | long interspersed nuclear element                                  |
| LTR             | long terminal repeat   |
| M               | molaire  |
| Mars            | methylated ascobolus repeated sequences                            |
| Mb              | mégapaire(s) de base   |
| MEPS            | minimal efficient processing segment                               |
| mg              | milligramme(s)   |
| min             | minute(s)  |
| mL              | millilitres(s)   |
| mM              | millimolaire   |
| MMR             | mismatch repair  |
| MSH1, 2         | MutS homologous gene 1, 2  |
| NBS1            | Nijmegen breakage syndrome   |
| NCBI            | national center for biotechnology information                      |
| néo             | néomycine  |
| NHCR            | non-homologous chromosomal   |
| rearrangement   |  |
| NHEJ            | non-homologous end-joining   |
| <sup>32</sup> P | phosphate 32   |
| pb              | paire(s) de base   |
| PCR             | polymerase chain reaction  |
| PMA1, 2         | plasma membrane H <sup>+</sup> -ATPase gene 1, 2                   |
| RNase           | enzyme ribonucléase  |
| RH              | recombinaison homologue  |

|               |                                      |
|---------------|--------------------------------------|
| RPA           | replication protein A                |
| RT-PCR        | réverse transcriptase PCR            |
| SDR           | short direct repeat                  |
| sec           | secondes(s)                          |
| SINE          | short interspersed nuclear element   |
| SSA           | single-strand annealing              |
| SV40          | simian virus 40                      |
| SV40Ag-T      | antigène grand T du virus SV40       |
| Ta            | transcribed, subset a                |
| TE            | tris-EDTA                            |
| tk            | thymidine kinase                     |
| $\mu\text{F}$ | microfarad(s)                        |
| $\mu\text{g}$ | microgramme(s)                       |
| $\mu\text{L}$ | microlitre(s)                        |
| UTR           | untranslated region                  |
| XRCC2,3,4     | X-ray cross-complementing gene 2,3,4 |

## REMERCIEMENTS

Ici prend fin mon aventure dans l'apprentissage de la recherche, qui m'a présenté de grands défis, d'innombrables embûches mais autant de satisfactions. Ces années d'acharnement à vouloir atteindre un but en ayant parfois l'impression d'une inaccessibilité m'ont donné l'occasion de comprendre bien des choses. Enfin arrivée au fil d'arrivée, j'en retire maintenant les profits avec fierté.

Dans un laboratoire accueillant, libre et très humain, j'ai eu la chance d'être formée par un directeur de thèse impliqué et inspirant, pour qui la recherche est une vocation. Je le remercie de son enthousiasme à ce projet et de ses encouragements à mon égard. Je tiens également à formuler de sincères remerciements à Roxane, qui a toujours été là avec son soutien technique et moral, ainsi qu'à Nadine qui avait toujours une pensée du jour réconfortante. Je ne pourrais passer sous silence tous les moments scientifiques et amicaux passés en compagnie de Jacinte, Catherine, Annie, Hugo et Kevin, des étudiants avec qui j'ai partagé beaucoup de choses. J'adresserais une reconnaissance particulière à Catherine, avec qui j'ai partagé mon projet et mes plus beaux moments scientifiques.

Je voudrais finalement remercier ma famille, mes amis et Christian, qui m'ont soutenue et encouragée au cours de ces nombreuses années d'étude et qui ont accepté de parfois passer au second rang.

Merci à l'Institut du Cancer de Montréal, au CRSNG et au FRSQ pour leur soutien financier.

# I. INTRODUCTION

## I.1. Objectifs de recherche

La compréhension moléculaire des principes d'organisation de la chromatine dans le noyau est un sujet de pointe dans le domaine de la biologie moléculaire. Un patron de compartimentalisation des chromosomes dans la cellule en interphase commence à prendre forme, suggérant que cette structure tridimensionnelle puisse avoir un rôle fonctionnel important au sein du noyau. Certaines questions demeurent toutefois obscures. Il n'est pas clair si la structure du génome est fixe ou dynamique tout au long du cycle cellulaire, si elle est conservée d'une cellule à une autre, d'un type cellulaire à l'autre et d'une espèce à l'autre. Nous ne savons pas comment elle est contrôlée par la cellule, quels en sont les avantages et quels effets peuvent avoir la perte de ce contrôle sur l'intégrité génomique.

Pour élucider ce mystère, nous avons voulu étudier le contrôle de l'accessibilité génomique entre les différentes régions du génome. En capturant les événements d'interaction entre différentes séquences chromosomiques appartenant à un même territoire ou à des territoires distincts, nous avons tenté de déterminer les principes de base de l'accessibilité et l'importance de son contrôle sur la stabilité génomique. L'élucidation des mécanismes moléculaires gouvernant la stabilité génétique dans les cellules normales et l'instabilité génétique dans les cellules néoplasiques est à la base même de la compréhension du processus de tumorigénèse.

Pour représenter l'accessibilité génomique dans le contexte le moins artificiel possible, nous avons choisi un modèle d'interactions homologues entre les séquences répétitives LINEs du génome humain. Conservés chez tous les mammifères et dispersés en centaines de milliers de copies à travers tout le génome, ces éléments naturels représentent un modèle idéal pour étudier à grande échelle l'accessibilité libre d'un locus donné face à l'immensité du génome. Les séquences répétitives dispersées sont également un modèle idéal pour l'étude de la recombinaison

homéologue ectopique chez les cellules mammifères, dont le contrôle pourrait être crucial au maintien de la stabilité génomique. Dans un contexte cellulaire instable, les interactions massives et incontrôlées entre les séquences répétitives qui occupent près de 35% du génome pourraient gravement menacer l'intégrité cellulaire et contribuer significativement au processus de la tumorigénèse.

La recombinaison homologue est fortement stimulée par les cassures double-brin d'ADN, et elle en est le mécanisme de réparation le plus fiable. Compte tenu de la proportion des séquences répétitives dans le génome, près d'une cassure sur 3 surviendrait à l'intérieur d'une répétition, qui aurait alors la possibilité de se réparer par recombinaison homéologue avec un partenaire ectopique. Face à l'attaque constante de l'ADN par de multiples agents endommageants, il est donc indispensable que ces échanges soient restreints et que le mécanisme par lequel ils s'effectuent ne soit pas délétère. D'un autre côté, les échanges homéologues conservatifs entre les séquences répétitives LINEs et Alus pourraient jouer un rôle important dans la création d'une diversité génétique et dans l'évolution du génome. Si elles ont été préservées depuis si longtemps chez l'ensemble des mammifères, elles ne sont certainement pas de simples parasites envahisseurs et doivent apporter des effets bénéfiques à leur hôte.

Par notre étude de la recombinaison homéologue entre les séquences LINEs du génome humain, nous avons donc voulu éclaircir les règles de l'accessibilité entre les différentes régions du génome, étudier les mécanismes de recombinaison ectopique, définir l'importance du contrôle de la recombinaison homéologue ectopique entre les séquences répétitives, étudier les mécanismes de réparation des cassures double-brin et déterminer le rôle des séquences répétitives LINEs dans notre génome.

## **I.2. Organisation du génome humain**

Il y a plus de 100 ans, bien avant la découverte de la double-hélice<sup>1</sup> et le séquençage complet du génome humain<sup>2</sup>, des pionniers de l'étude des chromosomes et du noyau spéculaient déjà sur la capacité des chromosomes à conserver leur structure et leur identité tout au long du cycle cellulaire<sup>3,4</sup>. Une telle évidence de configuration ne fût apportée par les études de Cremer qu'au début des années 1980<sup>5</sup>. Surgissait alors la notion de compartimentalisation des chromosomes à l'interphase, sous forme de territoires chromosomiques distincts. Depuis cette découverte, plusieurs règles de base ont pu être établies concernant l'organisation du génome dans le noyau.

### **I.2.1. Territoires chromosomiques et espace interchromatique**

Selon le modèle de Cremer<sup>6</sup>, les chromosomes occuperaient à l'interphase des territoires distincts dans le noyau<sup>7,8</sup>, séparés par un espace interchromatique contenant les complexes macromoléculaires nécessaires aux processus nucléaires tels que la réplication, la transcription, l'épissage et la réparation<sup>6,9</sup>. Des analyses à haute résolution ont démontré que les territoires ne présentaient pas de barrière à toute cette machinerie protéique<sup>10</sup>. L'espace interchromatique ne serait effectivement pas limité à la périphérie des territoires. Prenant leur source aux pores nucléaires, les canaux interchromatiques s'étendraient non seulement entre les territoires, mais également à l'intérieur de ceux-ci, permettant ainsi aux processus nucléaires de fonctionner partout dans les territoires et non seulement à leur surface<sup>11</sup>. Les territoires chromosomiques et l'espace interchromatique sont les composantes fondamentales de l'architecture nucléaire fonctionnelle.

La disposition des territoires dans le noyau ne serait pas aléatoire. Il fût en effet observé que les grands chromosomes étaient plus souvent localisés en périphérie du noyau, près de la membrane nucléaire, alors que les petits chromosomes se retrouvaient plus souvent vers le centre du noyau<sup>12</sup>. Cette préférence pour l'intérieur

du noyau a également été observée pour les chromosomes plus actifs en transcription<sup>13</sup>. Une étude récente a démontré que le mouvement de la chromatine était restreint par l'attachement aux compartiments nucléaires comme la périphérie nucléaire et le nucléole, qui imposeraient une plus grande restriction de mouvement à la chromatine, résultant en un environnement hétérochromatique plus condensé<sup>14,15</sup>. Une hypothèse de protection avait déjà été proposée pour expliquer la fonction d'une hétérochromatine en périphérie nucléaire, qui pourrait protéger l'euchromatine centrale contre les agents mutagènes<sup>16</sup>. Plusieurs études ont cependant indiqué une distribution non-aléatoire des cassures double-brin et des sites d'aberration chromosomiques induits par les endonucléases et les radiations, localisés préférentiellement dans les bandes G-light riches en gènes localisées à l'intérieur du compartiment nucléaire et non en périphérie<sup>17,18</sup>.

Les études de la dynamique de la chromatine à l'interphase ont démontré un mouvement assez restreint des territoires dans le noyau de plusieurs types cellulaires étudiés, pour lesquels la position semble assez conservée<sup>19</sup>. De grands mouvements ont toutefois été observés durant la différenciation des cellules neuronales<sup>20</sup> et dans certaines situations pathologiques<sup>21</sup>. La mobilité de la chromatine pourrait être finement régulée par la densité des sites d'attachement à la matrice nucléaire qui fournirait suffisamment l'élasticité pour de petits mouvements, mais les mouvements majeurs des territoires requerraient une réorganisation locale ou même la dissolution de la matrice<sup>22</sup>.

Il fût démontré que même si les chromosomes occupaient des territoires distincts, ils pouvaient parfois être structurellement associés. Des études de microscopie électronique ont en effet permis de visualiser que les territoires n'étaient pas toujours bordés de tous les côtés par l'espace interchromatique, mais que certains domaines pouvaient parfois se toucher, ouvrant l'opportunité d'une interaction entre l'ADN de domaines appartenant à des territoires chromosomiques distincts<sup>11</sup>. De telles interactions ont été observées entre des gènes appartenant à des territoires chromosomiques distincts, entre lesquels une colocalisation particulière est observée

à l'interphase. Par exemple, la proximité particulière entre les gènes PML et RAR $\alpha$  pourrait être responsable des translocations (15;17) conduisant à la leucémie promyélocytaire aiguë<sup>24</sup>.

Il a été observé que la plupart des territoires chromosomiques étaient séparés entre eux par une masse de chromatine non-homologue, qui pourrait empêcher l'association de territoires homologues<sup>25</sup>. Bien que les territoires ne subissent pas de réarrangements majeurs au cours du cycle cellulaire ou lors d'une irradiation<sup>25-27</sup>, il n'est pas exclu que sous certaines conditions, des boucles de chromatine homologues puissent diffuser à la surface des territoires et interagir entre elles par recombinaison homologue.

Il n'existe pas de cartographie relative des différents territoires dans le noyau en interphase, mais il se pourrait que certaines régions du génome partagent une accessibilité particulière entre elles qui pourrait favoriser leur interaction. Pour mieux comprendre cette accessibilité, il est impératif de comprendre les principes de base de l'organisation de l'ADN à l'intérieur des territoires, qui ne représentent que le premier niveau d'une organisation tridimensionnelle très complexe.

### **I.2.2. Organisation intraterritoriale**

À l'intérieur d'un territoire, il fut observé que chacun des bras chromosomiques occupait son propre domaine globulaire compact et distinct, qui ne se superpose pas avec celui de l'autre bras mais qui demeure tout de même flexible par rapport au centromère<sup>28,29</sup>. Les territoires chromosomiques de 50 à 244 Mb seraient séparés en petits foci de l'ordre de 1 Mb dont la structure pourrait persister tout au long du cycle cellulaire<sup>30-32</sup>. Associés à des complexes protéiques spécifiques fournis par les

canaux interchromatiques, ces foci pourraient constituer des unités stables et distinctes de réplication ou de transcription<sup>7,32</sup>.

Selon le modèle de Cremer<sup>7</sup>, les domaines de chromatine pauvres en gènes se localiseraient préférentiellement à la périphérie des territoires, près de la membrane nucléaire ou autour du nucléole, alors que les domaines plus riches en gènes se retrouveraient plus souvent vers le centre du territoire. Des études récentes ont également démontré que des régions très denses en gènes ayant un niveau transcriptionnel élevé pouvaient se décondenser en surface du territoire chromosomique et s'étendre dans l'espace interchromatique suite à leur activation transcriptionnelle<sup>29,32,33</sup>. Cette extériorisation pourrait créer un environnement plus permissif à leur transcription.

Selon le modèle MLS (*multiloop subcompartment*) de Münkel<sup>35</sup>, les domaines de l'ordre de 1 Mb seraient sous-compartmentalisés en une dizaine de rosettes de l'ordre de 30 à 150 kb comprenant chacune 150 à 1000 nucléosomes et de un à plusieurs gènes. Ces compartiments seraient connectés entre eux par de courts segments de chromatine et séparés par des canaux interchromatiques envahissant le territoire chromosomique. Selon ce même modèle, les gènes actifs en transcription se retrouveraient en bordure des boucles, près de l'espace interchromatique fournissant la machinerie transcriptionnelle et dans lesquels les ARNm naissants seraient relargués pour rejoindre les pores nucléaires. Inversement, les gènes inactifs seraient localisés vers l'intérieur des boucles, où ils seraient réprimés<sup>7</sup>.

### **I.2.3. Conservation de l'organisation tridimensionnelle du génome**

La compilation de plusieurs études indique que la position radiale d'un territoire donné dans le noyau est corrélée avec sa taille, sa densité en gènes et son moment de réplication. Récemment, des études comparative chez différentes espèces de vertébrés ont démontré une conservation évolutive du positionnement radial des territoires chromosomiques sur plus de 300 millions d'années<sup>36,37</sup>. Une étude a

également démontré la conservation de l'organisation intraterritoriale du domaine subchromosomique 11p13 d'une cellule à l'autre, d'un type cellulaire à l'autre, et même entre les génomes humain et murin<sup>38</sup>. La conservation des différents niveaux d'organisation génomique suggère par conséquent un véritable rôle fonctionnel de cette configuration dans le contrôle des processus cellulaires et le maintien de l'intégrité génomique.

### **I.3. Modèle d'étude : les séquences répétitives dispersées**

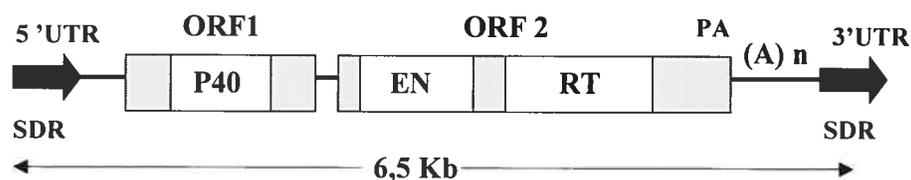
Nous savons maintenant que les séquences codantes représentent moins de 5% de l'espace du génome humain, alors que les séquences répétitives en occupent plus de 50%<sup>2</sup>. La plupart de cet ADN répétitif est dérivé d'éléments transposables<sup>39,40</sup>, regroupant les transposons à ADN (groupes hAT, Tc-1, PiggyBac-like), les éléments LTR (ERV classes I à III, MalR) et les rétrotransposons à ARN (LINEs, SINEs). Les rétrotransposons sont dispersés en de multiples copies et occupent à eux-seuls 34% du génome humain<sup>2</sup>.

Les éléments SINEs (*short interspersed nuclear elements*) (figure 1b, p.26) sont des éléments de 100 à 300 pb présents en 1,5 millions de copies occupant 13% de l'espace du génome. Chez l'humain, la famille de SINEs la plus importante en termes de copies et d'activité est la famille Alu. Bien qu'ils possèdent un promoteur interne pour l'ARN polymérase III, ces éléments ne codent pour aucune des protéines nécessaires à leur rétrotransposition et ne sont donc pas autonomes. La machinerie enzymatique nécessaire leur serait fournie par les éléments LINEs (*long interspersed nuclear elements*) (figure 1a, p.26), qui encodent toutes les protéines nécessaires à leur propre rétrotransposition, c'est-à-dire une protéine de transport cytoplasmique, une endonucléase et une reverse transcriptase<sup>41</sup>.

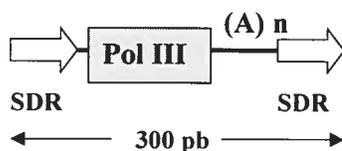
Les LINEs sont présents en 850,000 copies dispersées occupant 21% du génome humain<sup>2</sup>. D'une longueur maximale de 6 à 7 kb, la plupart sont tronqués par divers

processus à leur extrémité 5' et sont donc inactifs en rétrotransposition. Parmi les 3000 à 4000 LINEs pleine longueur résiduels, l'accumulation de mutations dans les cadres de lecture les a rendus inactifs et on estime à seulement 30 à 60 le nombre d'éléments toujours actifs dans le génome humain<sup>42</sup>. D'une taille moyenne de 900 pb, les LINEs de la famille L1 ont subi beaucoup de modifications au cours de l'évolution et se classent en 55 sous-familles différentes partageant en moyenne 95% d'homologie entre eux<sup>2</sup>. Dans les sous-familles les plus récentes L1PA1 à L1PA5 (25 millions d'années), on estime à 8,5% la proportion de L1s pleine longueur, qui diminue en fonction de l'âge de la sous-famille<sup>43</sup>.

a)



b)



**Figure 1 : Structure des séquences répétitives LINEs (a) et SINEs (b)**

- a) Les LINEs codent pour 3 protéines nécessaires au mécanisme de rétrotransposition : une protéine de transport cytoplasmique (P40), une endonucléase (EN) et une reverse transcriptase (RT). La région 5' non-transcrite (UTR) contient un promoteur interne reconnu par l'ARN polymérase III et la région 3' non-transcrite (UTR) contient un signal de polyadénylation (PA) ainsi qu'une queue de polyA.
- b) Les SINEs possèdent un promoteur interne pour l'ARN polymérase III (Pol III) et une queue de polyA.

De courtes répétitions directes (SDR) de 5 à 15 pb résultent de l'insertion d'un élément rétrotransposé vers un nouveau site.

Les LINEs et les SINEs sont dispersés sur tous les chromosomes mais ont chacun leur distribution personnelle qui demeure conservée chez l'humain<sup>44</sup> et la souris<sup>45</sup>. Les éléments SINEs sont concentrés dans les bandes R (reverse) qui sont plus riches en G+C et en gènes, de réplication plus précoce et de condensation plus lente en mitose. Complètement à l'opposé, les éléments LINEs sont plus concentrés dans les bandes G/Q (Giemsa/Quinacrine) qui sont plus riches en A+T et plus pauvres en gènes, de réplication plus tardive et de condensation plus rapide en mitose. Les LINEs sont particulièrement exclus des bandes C comprenant les régions centromériques et les constriction secondaires de certains chromosomes.

Le patron de distribution inverse des Alus et des L1s leur suggère des rôles différents, qui demeurent encore mal compris et controversés. Toutefois, il apparaît de plus en plus évident que ces éléments mobiles affectent significativement l'évolution et l'intégrité du génome humain depuis des centaines de milliards d'années<sup>2</sup>, de 4 principales façons. Premièrement, leur insertion par rétrotransposition peut altérer l'expression ou la régulation des gènes. L'insertion récente de rétrotransposons à l'intérieur de séquences codantes a été associée à plus de 35 cas de pathologies humaines comme la dystrophie musculaire de Duchenne<sup>46</sup>, l'hémophilie A<sup>47</sup> ainsi que des cas de cancer du sein<sup>48</sup> et du colon<sup>49</sup>.

Deuxièmement, il semble que les L1s puissent représenter un véhicule pour la mobilisation de séquences existantes comme des exons et des promoteurs de gènes<sup>50</sup>. En effet, il peut arriver que le faible signal de polyadénylation des L1s soit surpassé en faveur d'un signal plus fort en 3' appartenant à un gène proximal. Ce phénomène de transduction pourrait augmenter la plasticité du génome et constituer un mécanisme évolutif pour la création de nouveaux gènes. Une étude a rapporté que près de 15% des L1s complets du génome portaient une évidence d'ADN flanquant transduit<sup>51</sup>. Par extrapolation aux 500,000 copies de L1s du génome, la quantité d'ADN transduit par les L1s représenterait environ 1% du génome, une fraction comparable à celle occupée par les exons.

Troisièmement, ces séquences répétitives partagent une homologie suffisante pour pouvoir interagir entre eux par recombinaison homologue (RH). La RH entre les éléments Alus, qui résident surtout dans les régions du génome riches en G+C, est plus susceptible d'avoir un effet délétère détectable. Des événements de RH inégale entre des éléments Alus auraient été responsables d'une trentaine de cas de maladies génétiques germinales et de 16 cas de cancers<sup>52</sup>. Bien que moins fréquents, des événements de RH entre des éléments L1s responsables de maladies humaines ont également été rapportés<sup>53-58</sup>.

La RH entre les L1s pourrait finalement contribuer significativement à l'évolution du génome humain. Par exemple, la RH inégale entre 2 éléments L1s du gène  $\gamma$ -globine serait à l'origine de la duplication du gène, un événement responsable de la conversion du gène embryonnaire vers le gène fœtal<sup>59</sup>. Un événement de RH entre des L1s du bras chromosomique Yp pourrait aussi être à l'origine de l'inversion responsable de la séparation évolutive de l'hominidé et de l'humain<sup>60</sup>. La conversion génique entre les séquences répétitives pourrait jouer un rôle évolutif constant dans la sculpture de la diversité génomique<sup>61-63</sup>. Chez les séquences Alu récemment intégrées, 10 à 20% de la variation de séquence pourrait avoir été générée par la conversion génique<sup>64</sup>. De plus, nous avons démontré au laboratoire que les LINEs du génome variaient d'un individu à l'autre, et que leur évolution concertée impliquait leurs interactions homologues par conversion génique<sup>63</sup>.

#### **I.4. Les cassures double-brin (CDBs)**

Nos cellules sont constamment la cible des agents chimiques et physiques qui endommagent notre ADN par des cassures simple-brin (CSBs) ou double-brin (CDBs) qui menacent l'intégrité de notre génome. Contrairement aux CSBs où l'un des 2 brins demeure intact, les CDBs sont des sites où l'intégrité physique et l'information sont perdues sur les 2 brins, et représentent des dommages très sérieux pour la cellule. Si elle n'est pas correctement réparée dans les plus brefs délais, une

seule CDB est suffisante pour engendrer la mort cellulaire ou pire, induire la tumorigénèse. Les CDBs peuvent être induites par des agents exogènes comme les radiations ionisantes, la bléomycine et les radicaux libres oxydatifs, mais elles sont aussi produites constamment et naturellement à l'intérieur des cellules par l'action de la topoisomérase II et des radicaux libres, ainsi qu'au cours de processus cellulaires importants comme la recombinaison méiotique, la recombinaison V(D)J, la recombinaison *class-switch* des immunoglobulines et aux fourches de réplication<sup>65,66</sup>.

### I.5. Détection et signalement des CDBs

L'un des premiers événements détectables suite au traitement des cellules par des radiations ionisantes est la phosphorylation rapide de l'histone H2AX, qui survient également à la méiose et la recombinaison V(D)J, suggérant qu'il puisse représenter un signal général de détection des CDBs<sup>67-69</sup>. La modification des histones pourrait altérer la structure de la chromatine pour permettre l'accès et l'action des facteurs de réparation, ou encore servir de signal d'amplification. La transformation de l'histone H2AX sous sa forme phosphorylée H2AX $\gamma$  serait importante pour le recrutement de plusieurs facteurs de réparation aux sites de CDBs, tels que RAD50, RAD51 et BRCA1<sup>70</sup>.

Les protéines kinases ATM et ATR jouent également un rôle crucial et central dans la détection et le signalement des dommages à l'ADN pour l'enclenchement des processus de réparation. Alors qu'ATM répond exclusivement aux CDBs induites par les radiations ionisantes, ATR répond aussi aux dommages causés par les ultraviolets et l'arrêt aux fourches de réplication<sup>71</sup>. Partageant également un certain nombre de substrats de phosphorylation<sup>72</sup>, ces 2 protéines ont des rôles chevauchants mais non redondants.

Jusqu'à maintenant, la fonction d'ATM est beaucoup plus étudiée et mieux connue. La déficience du gène est associée à l'ataxie-télangiectasique, une maladie

autosomale récessive très grave caractérisée principalement par une dysfonction neuromotrice progressive, une télangiectasie, une dégénérescence du cerveau et du thymus, une déficience immunitaire, un vieillissement prématuré, une prédisposition extrême aux leucémies et lymphomes et une sensibilité accrue aux radiations ionisantes<sup>73</sup>.

Essentiellement localisée au noyau<sup>74</sup>, la protéine ATM est retrouvée le long de l'axe synaptique<sup>75</sup> et son activité kinase serait activée par les dommages à l'ADN<sup>76</sup>. Cependant, les résultats concernant la liaison physique et l'activation directe d'ATM par les extrémités d'ADN brisées demeurent contradictoires<sup>77,78</sup>. L'activité kinase d'ATM cible des protéines d'importance majeure dans le contrôle du cycle cellulaire (P53, CHK1/CHK2, RPA) et dans la recombinaison (RAD50/NBS1/MRE11, BRCA1, c-ABL). ATM est également impliquée dans le transport des vésicules synaptiques neuronales par son interaction avec la  $\beta$ -adaptine<sup>79</sup>, ainsi que dans le contrôle du vieillissement par le maintien des télomères<sup>80,81</sup> (figure 2, p.31). L'implication d'ATM dans ces multiples voies cellulaires explique les nombreuses facettes de la maladie associée à sa déficience.

## **I.6. Réparation des CDBs**

Pour se protéger contre les nombreuses sources de CDBs qui lui sont imposées, notre génome a dû se doter d'excellents systèmes de détection des CDBs, de signalement des dommages et de réparation de l'ADN, qui sont conservés chez tous les eucaryotes de la levure aux mammifères. Chez la levure, dont le génome est essentiellement composé de séquences codantes uniques, les CDBs sont principalement réparées par un mécanisme de recombinaison homologue (RH) au cours duquel une séquence homologue intacte est utilisée comme modèle de réparation, de façon à conserver toute l'information génétique et éviter les erreurs<sup>82</sup>.

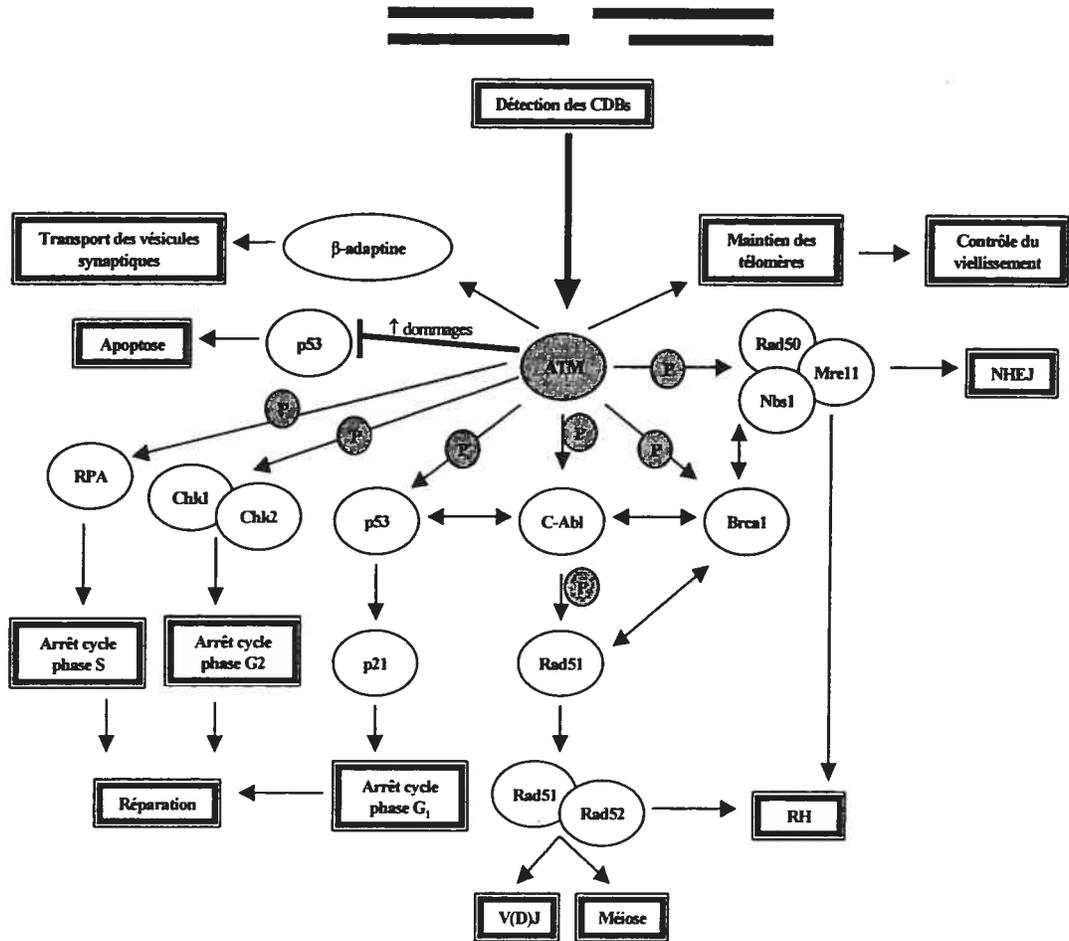
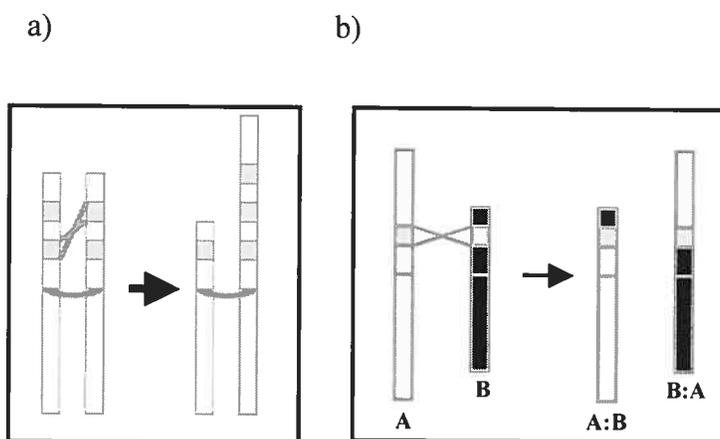


Figure 2 : Rôles multiples de la protéine kinase ATM

Chez les eucaryotes supérieurs comme les mammifères, ce mécanisme est moins utilisé et est surtout limité aux phases du cycle cellulaire où les chromatides sœurs ou les chromosomes homologues sont alignés. Puisque 50% du génome humain est composé de séquences répétitives, la RH ectopique entre elles pourrait conduire à des réarrangements chromosomiques délétères comme des translocations, délétions, expansions et pertes d'hétérozygotie (figure 3, p.32). Chez les cellules mammifères, les CDBs sont donc la plupart du temps réparées par un mécanisme de recombinaison non-homologue, même s'il en résulte une perte d'information génétique locale.

Les 2 voies de réparation des CDBs sont donc importantes chez les cellules mammifères, mais leur contribution relative dépend du cycle cellulaire. Dans un modèle cellulaire mammifère étudiant le mécanisme de réparation d'une CDB induite dans un gène de sélection néomycine<sup>83</sup>, 30 à 50% des événements de réparation impliquaient un mécanisme de RH utilisant une répétition directe adjacente et 50 à 70% des événements impliquaient un mécanisme de NHEJ accompagné d'une petite délétion et/ou insertion au site de cassure.



**Figure 3 : Réarrangements chromosomiques délétères pouvant résulter de la RH ectopique entre des séquences répétitives**

- a) Échange inégal entre 2 chromatides sœurs (délétions / expansions)
- b) Recombinaison interchromosomique (translocations)

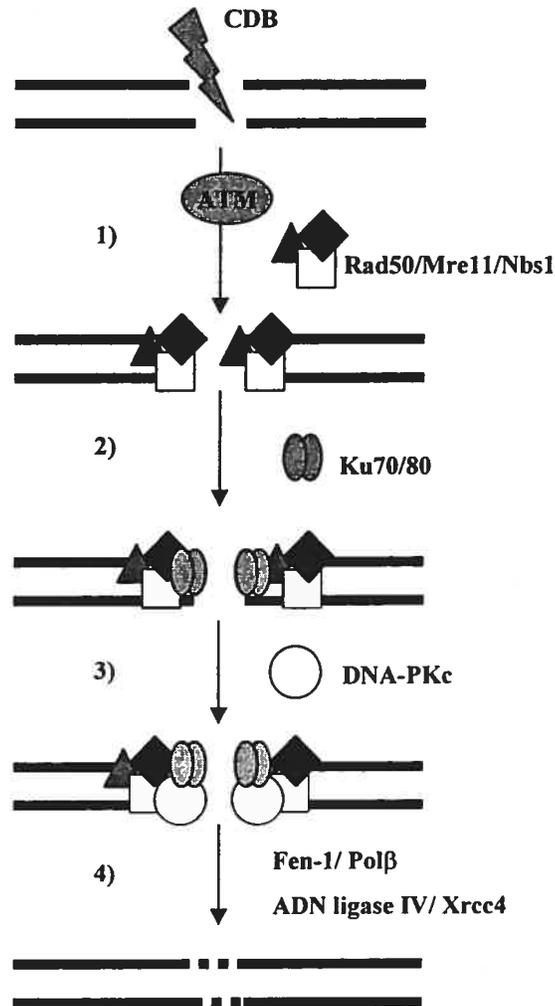
## **I.7. Recombinaison non-homologue**

### **I.7.1. Mécanisme de *non-homologous end-joining* (NHEJ)**

La plupart des CDBs survenant chez les cellules mammifères sont réparées par un mécanisme de recombinaison non-homologue (NHEJ) au cours duquel les extrémités brisées d'ADN sont rejointes sans base d'homologie, ou par des microhomologies de moins de 10 pb. Dans cette voie de réparation (figure 4, p.34), l'activation d'ATM recruterait le complexe RAD50/MRE11/NBS1 qui resterait alors en place jusqu'à la fin de la réparation<sup>84</sup>. Ce complexe recruterait ensuite l'hétérodimère Ku70/80 qui aurait pour fonction de protéger les extrémités brisées par sa liaison, et de dérouler l'ADN par son activité hélicase. Suite à l'activation des extrémités, Ku70/80 recruterait et activerait la DNA-PKc qui pourrait ensuite procéder à l'alignement des extrémités à refermer. Avant d'être religaturées par le complexe ADN ligase IV/XRCC4, les extrémités seraient modifiées par la nucléase FEN-1 et la polymérase pol $\beta$ , entraînant l'insertion et/ou la délétion de quelques nucléotides au site de cassure réparé.

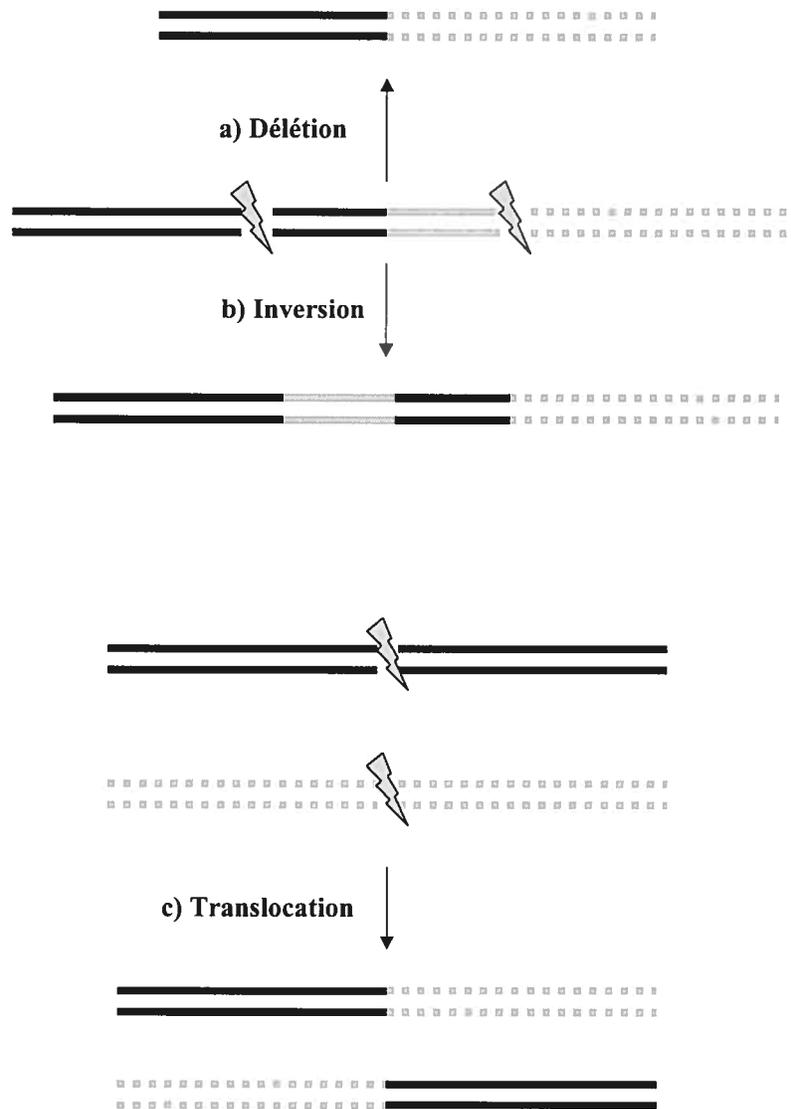
### **I.7.2. Réarrangements chromosomiques non-homologues (NHCR)**

Lorsque plusieurs CDBs sont simultanément présentes dans la cellule, les extrémités brisées d'une cassure ont la possibilité de se religaturer avec celles d'une autre. Si les cassures se retrouvent sur le même chromosome, il pourra en résulter un réarrangement chromosomique majeur comme une inversion ou une délétion. Si les cassures sont localisées sur des chromosomes différents, il en résultera une translocation (figure 5, p.35). Ces événements délétères doivent être évités à tout prix par la cellule, pour laquelle les aberrations chromosomiques ne peuvent être tolérées pour assurer l'intégrité génomique.



**Figure 4 : Voie de réparation des CDBs par NHEJ**

- 1) Détection des CDBs par ATM et recrutement du complexe RAD50/MRE11/NBS1.
- 2) Recrutement de l'hétérodimère Ku70/80 et activation des extrémités.
- 3) Recrutement et activation de la DNA-PKc et alignement des extrémités.
- 4) Modification des extrémités par FEN-1/ Polβ et ligation par le complexe ADN ligase IV/ XRCC4.



**Figure 5 : Mécanismes de réarrangements chromosomiques non-homologues**

- a) Délétion entre 2 CDBs intrachromosomiques
- b) Inversion entre 2 CDBs intrachromosomiques
- c) Translocation entre 2 CDBs interchromosomiques

### I.7.3. Déficience de la recombinaison non-homologue

Chez les mammifères, la déficience de la voie du NHEJ est associée à des aberrations chromosomiques majeures ainsi qu'à une augmentation marquée de la fréquence de transformation maligne<sup>85</sup>. Chez des lignées primaires leucémiques, plusieurs sites de cassure de translocations ont été décrits comme des jonctions typiques issues du NHEJ, suggérant une contribution majeure de cette voie de réparation de CDBs au maintien de la stabilité du génome humain<sup>86</sup>.

L'inactivation des gènes codant pour les protéines impliquées dans la voie du NHEJ conduisent à une hypersensibilité aux radiations ionisantes associée à une déficience de réparation des CDBs et de recombinaison V(D)J<sup>87,88</sup>. Chez des modèles de souris K.-O. pour DNA-Pk, Ku70 ou Ku80, les embryons sont viables mais sont hypersensibles aux radiations ionisantes, présentent une déficience immune combinée sévère et ont un taux élevé de lymphomes. Les modèles de souris K.-O. pour les gènes RAD50, MRE11, XRCC4 ou ADN ligase IV sont quant à eux létaux au niveau embryonnaire. Par contre, la létalité et l'apoptose neuronale associées à XRCC4 ou à l'ADN ligase IV peuvent être sauvées par une déficience additionnelle des gènes ATM ou P53<sup>89-91</sup>. Chez des souris double K.-O. pour P53 et XRCC4, les translocations observées dans les lymphomes pro-B ne sont cependant pas réciproques<sup>89</sup>, c'est-à-dire que seules 2 des 4 extrémités se recollent entre elles.

Chez des fibroblastes embryonnaires de souris, la double déficience des gènes P53 et ADN ligase IV est également associée à une fragmentation massive de l'ADN, ce qui suggère un rôle important pour la voie du NHEJ dans la suppression des translocations<sup>92</sup>. La déficience du gène ATM dans des extraits cellulaires A-T a été associée à une baisse de fidélité de refermeture des CDBs par NHEJ<sup>93-99</sup>, qui pourrait découler d'un rôle protecteur des extrémités brisées par la protéine ATM. La déficience de la recombinaison V(D)J dans les lymphocytes A-T serait également à l'origine d'une fréquence élevée de translocations impliquant les chromosomes 2, 7, 14 et 22 porteurs des gènes d'immunoglobulines et de récepteurs de cellules T<sup>100</sup>.

## I.8. Recombinaison homologue (RH)

### I.8.1. Mécanisme de réparation des CDBs par RH (*DSB repair model*)

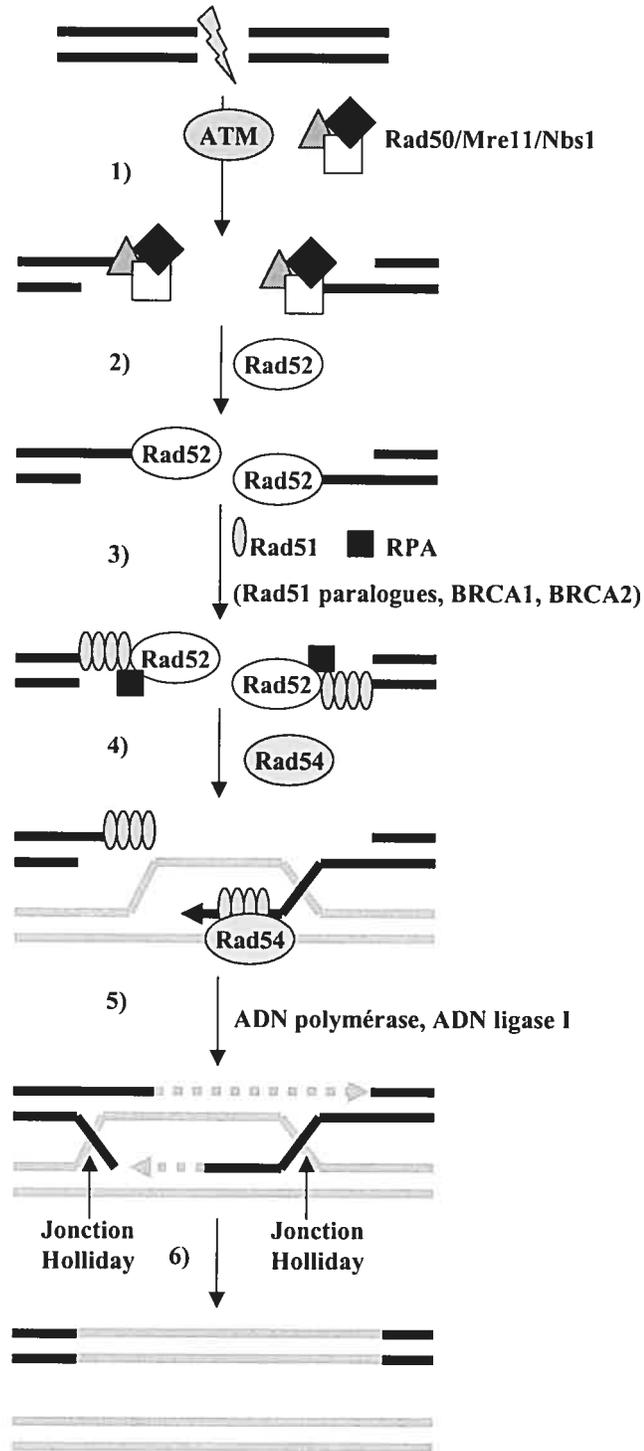
Les bases moléculaires de la RH ont été initialement définies par des études chez la bactérie et la levure, mais semblent aussi être hautement conservées chez les eucaryotes supérieurs<sup>101</sup>. D'abord identifiés chez la levure, le groupe de gènes RAD50 jouant des rôles importants dans le processus de RH ont tous trouvés leur homologue chez l'humain<sup>102</sup>.

Comme dans la voie du NHEJ, le signalement des CDBs par ATM est d'abord transmis au complexe RAD50/MRE11/NBS1 qui s'associe au site de cassure pour générer les extrémités 3' nécessaires à la recherche de l'homologie et à l'invasion de l'hétéroduplex homologue<sup>88,103</sup> (figure 6, pp.38-39). La protéine RAD52 est ensuite recrutée au site de cassure et sa liaison pourrait compétitionner avec celle de l'hétérodimère Ku70/Ku80, déterminant ainsi la voie de réparation utilisée<sup>104</sup>. Tout comme Ku, RAD52 aurait un rôle protecteur des extrémités brisées contre l'attaque des exonucléases et maintiendrait l'association des extrémités. RAD52 interagit avec RAD51 et la sous-unité p34 de la protéine RPA, pour permettre à RAD51 de s'assembler en filament nucléoprotéique sur les extrémités.

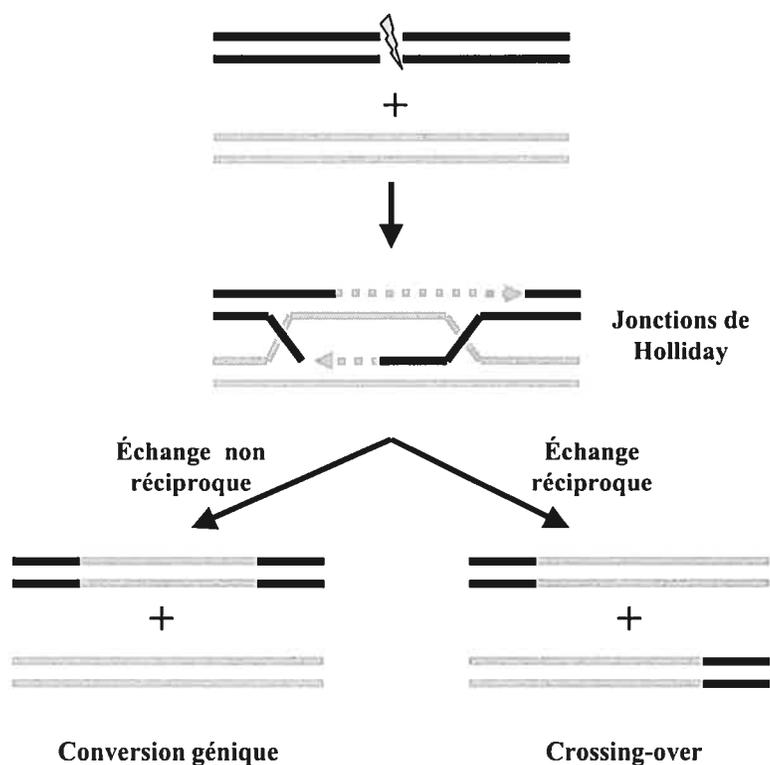
Par leur liaison à RAD51<sup>105</sup>, les protéines BRCA1 et BRCA2 pourraient contribuer significativement au processus de RH, qui est grandement affecté par la perte de fonction de ces gènes<sup>106-108</sup>. Il semble que les 5 paralogues de RAD51 (RAD51B, C, D, XRCC2 et XRCC3) collaborent aussi avec Rad51 dans son rôle central<sup>109</sup>. Avec l'aide de tous ces facteurs et de RAD54, RAD51 pourrait promouvoir le pairage homologue dépendant de l'ATP, l'envahissement de l'hétéroduplex et l'échange de brin. Au terme de cet échange, les extrémités terminales sont ensuite réparées par l'ADN polymérase, qui copie l'information du partenaire intact sur le brin à réparer. Les extrémités sont ensuite refermées par l'ADN ligase I.

**Figure 6 : Mécanisme de RH (*DSB repair model*)**

- 1) Détection des CDBs par ATM et recrutement du complexe RAD50/MRE11/NBS1 qui dégrade les extrémités 5'.
- 2) Recrutement de RAD52 qui protège les extrémités.
- 3) Recrutement de RAD51, RPA1, paralogues RAD51, BRCA1 et BRCA2. Assemblage du filament nucléoprotéique RAD51 aux extrémités.
- 4) Recrutement de RAD54 et catalyse du pairage homologue ATP-dépendant, de l'invasion de l'hétéroduplex et de l'échange de brin.
- 5) Synthèse d'ADN sur les extrémités par l'ADN polymérase et refermeture par l'ADN ligase I.
- 6) Migration d'embranchement et résolution des jonctions de Holliday.



Suite à la migration d'embranchement, les jonctions de Holliday sont résolues par un clivage et une ligation pour générer 2 molécules intactes. La résolution des jonctions peut conduire à 2 types de produits (figure 7, p.40). Un échange non-réciproque de type conservatif appelé conversion génique implique la réparation de la molécule brisée par acquisition de séquences du partenaire homologue, sans modification de ce dernier. L'étendue des séquences acquises correspond au segment de conversion. Alternativement, un produit de type non-conservatif appelé *crossing-over* implique un échange réciproque de séquences entre les partenaires homologues.



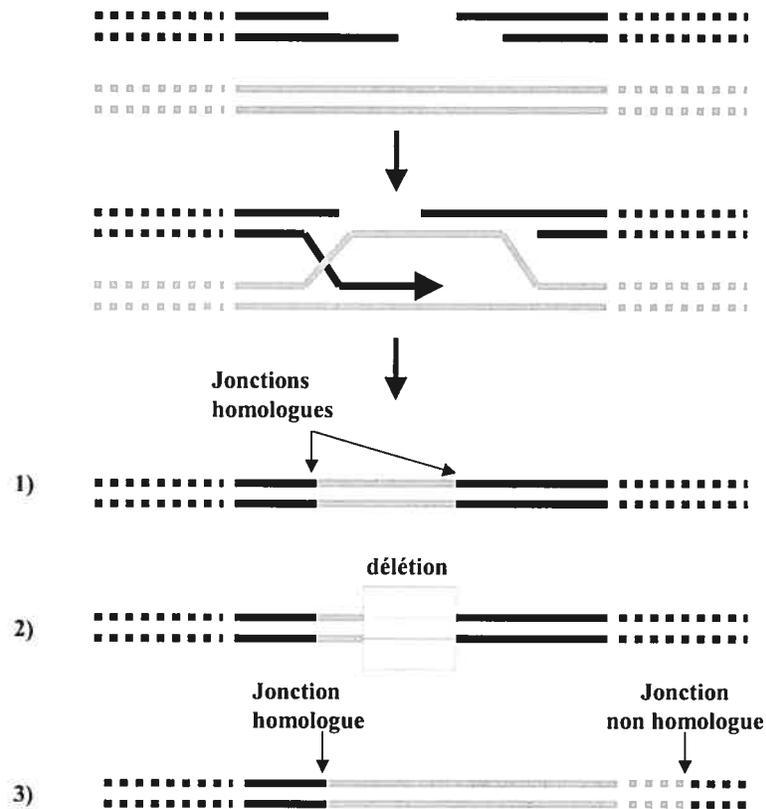
**Figure 7 : Produits de réparation des CDBs par RH**

Selon la résolution des jonctions de Holliday, le mécanisme de RH conduira à un produit conservatif de type conversion génique (échange non réciproque) ou à un produit non-conservatif de *crossing-over* (échange réciproque).

Le modèle standard de réparation des CDBs par RH a été proposé par Szostak<sup>110</sup>. Dans ce modèle, la cassure est d'abord élargie par une dégradation exonucléotidique des 2 extrémités et les jonctions de Holliday sont formées suite à l'invasion de l'hétéroduplex par ces 2 extrémités. Dans le produit de recombinaison final, l'étendue du segment de conversion de part et d'autre du site de CDB dépendra de chacune des étapes du mécanisme (dégradation exonucléotidique, invasion, pairage, extension de synthèse, migration d'embranchement et résolution des jonctions). L'étude des mécanismes de RH par des systèmes cellulaires a révélé une majorité de segments de conversion unidirectionnels chez la levure<sup>111-114</sup>, et une majorité de segments bidirectionnels chez les mammifères<sup>115-118</sup>. Une différence majeure pourrait donc exister au niveau du mécanisme de RH chez la levure et les eucaryotes supérieurs.

### **I.8.2. Mécanisme de *one-sided invasion***

Un autre mécanisme a été proposé par notre laboratoire pour expliquer la génération de produits de recombinaison homologue conservatifs et non-conservatifs observés lors de l'étude de la recombinaison extrachromosomique dans les cellules mammifères<sup>119,120</sup>. Dans ce modèle appelé *one-sided invasion* (figure 8, p.42), une seule des 2 extrémités brisées envahit l'hétéroduplex et initie la synthèse d'ADN, dans une seule direction. La longueur et la nature des séquences acquises par la molécule brisée sera alors dépendante de l'extension de la polymérase. Si la synthèse d'ADN atteint la fin de la région dégradée et s'arrête, le produit de réparation sera identique à celui obtenu par le mécanisme de conversion génique décrit par Szostak. Par contre, si la synthèse d'ADN ne couvre pas toute la région dégradée de la cassure, la réparation sera partielle et une délétion en résultera. Finalement, si la synthèse s'étend au-delà de la région d'homologie, des séquences non-homologues seront acquises par la molécule brisée et on observera une jonction homologue et une jonction non-homologue.



**Figure 8 : Mécanisme de RH par one-sided invasion**

- 1) La synthèse complète d'ADN dans la zone d'homologie génère un produit identique à la conversion génique standard.
- 2) La synthèse incomplète d'ADN aux extrémités brisées génère une délétion.
- 3) La synthèse d'ADN au-delà de la zone d'homologie implique l'acquisition de séquences non-homologues et la formation d'une jonction non-homologue.

Séquences homologues : —————

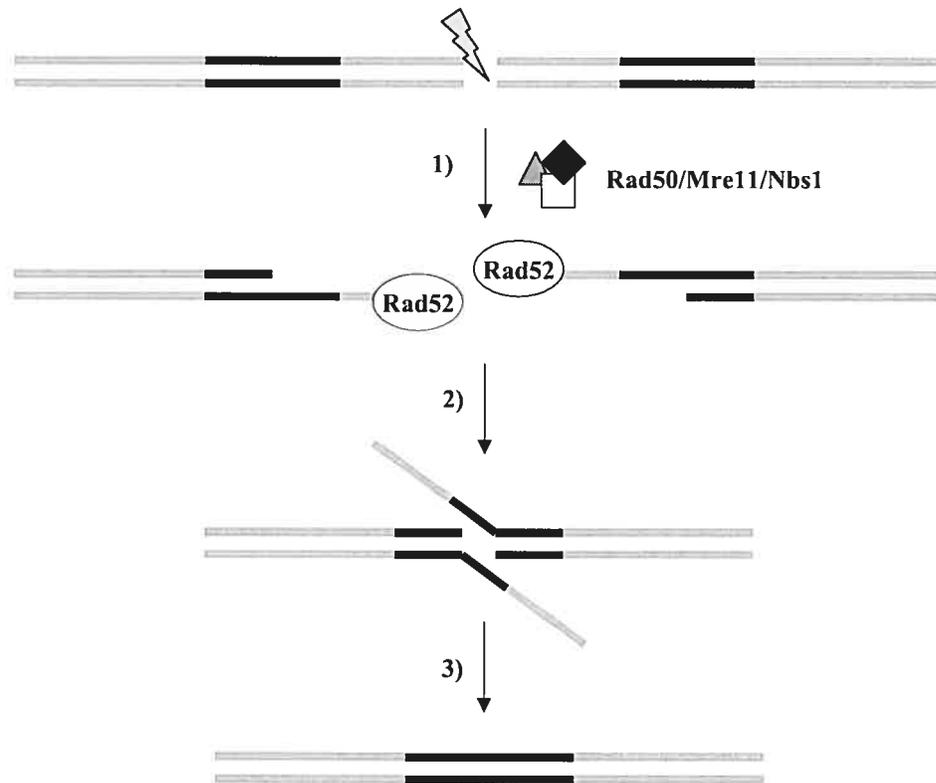
Séquences non-homologues : .....

### I.8.3. Mécanisme de *single-strand annealing* (SSA)

Le voie de réparation des CDBs par SSA est un mécanisme de RH non-conservatif. Il implique la formation d'un hétéroduplex homologue, mais n'implique pas d'invasion de brin, de jonctions de Holliday ni de migration d'embranchement<sup>121</sup>. Dans ce mécanisme, le complexe RAD50/MRE11/NBS1 est recruté à la cassure et exerce son activité exonucléotidique pour créer des extrémités complémentaires simple-brin et exposer des séquences homologues flanquantes situées à l'origine de part et d'autre du site de cassure<sup>122</sup> (figure 9, p.44). Rad52 aurait aussi un rôle crucial dans le SSA en protégeant les extrémités au cours de la recherche d'homologie. Les séquences homologues exposées sont alors pairées entre elles, et les extrémités non-homologues sont enlevées. Une étape finale de synthèse d'ADN et de ligation fournit un produit non-conservatif dans lequel les séquences situées entre les 2 répétitions sont délétées.

### I.8.4. RH et instabilité génomique

Sous un contrôle cellulaire serré, la RH peut agir en toute fidélité pour promouvoir la diversité génétique en méiose, assurer le développement des gènes immunitaires, préserver l'intégrité génomique des cellules au cours de la réplication, réparer adéquatement les nombreuses CDBs subies par l'ADN et restreindre les interactions ectopiques entre les séquences répétitives. Par contre, la déficience d'un gène impliqué de loin ou de près dans le contrôle de la RH peut avoir un effet dramatique sur la stabilité du génome. Plusieurs syndromes associés à une perte de contrôle de la RH présentent une fréquence élevée de RH, des produits de RH aberrants, une forte instabilité génomique ainsi qu'une prédisposition marquée au développement de tumeurs<sup>123</sup>. Le phénotype associé à la déficience de plusieurs protéines impliquées dans le processus de RH a été décrit dans plusieurs modèles cellulaires et animaux (tableau I, p.45).



**Figure 9 : Mécanisme de RH par *single-strand annealing* (SSA)**

- 1) Recrutement du complexe RAD50/MRE11/NBS1 au site de CDB, résection des extrémités et protection par RAD52.
- 2) Alignement des extrémités simple brin dans les séquences homologues flanquantes.
- 3) Résection des extrémités non-homologues, synthèse d'ADN et ligation.

| Protéine          | Fonctions dans la RH  | Phénotype associé à la déficience  |
|-------------------|---|--|
| ATM               | Détection et signalement des CDBs<br>Contrôle de l'arrêt du cycle cellulaire<br>Activation des effecteurs de la RH  | K.-O. viable, infertile, retard de croissance<br>Augmentation fréquence de RH spontanée<br>Aberrations chromosomiques<br>Leucémies et lymphomes                        |
| P53               | Contrôle de l'arrêt du cycle cellulaire<br>Interfère avec la liaison et l'activité de RAD51<br>Suppression de la RH | K.-O. normal mais susceptible aux tumeurs<br>et instabilité génomique<br>Fréquence élevée de RH<br>Cancers variés  |
| RAD50             | Stimule l'activité nucléase de MRE11  | K.-O. léthal embryonnaire<br>Cellules ES non viables<br>Augmentation fréquence de RH spontanée   |
| MRE11             | Activité nucléase   | Pas de modèle K.-O.<br>Cellules ES non viables<br>Augmentation fréquence de RH spontanée   |
| RAD51             | Promouvoit le pairage homologue<br>Promouvoit l'échange de brin   | K.-O. léthal embryonnaire<br>Hypersensibilité aux R.I.<br>Aberrations chromosomiques<br>Diminution de la RH conservative   |
| RAD52             | Liaison et protection des extrémités brisées<br>Stimule l'activité de RAD51   | K.-O. viable et fertile<br>Cellules ES: réduction modérée RH   |
| RAD54             | Promouvoit le pairage homologue<br>Stabilisation de l'hétéroduplex<br>Stimule l'activité de RAD51                   | K.-O. viable et fertile<br>Cellules ES: réduction modérée RH<br>Hypersensibilité aux CDBs et aux R.I.<br>Diminution de la RH<br>Lymphomes, cancers du colon et du sein |
| BRCA1<br>et BRCA2 | Contrôle du cycle cellulaire<br>Collaboration avec RAD50<br>Assemblage de RAD51                                     | K.-O. léthal embryonnaire<br>Hypersensibilité aux R.I.<br>Aberrations chromosomiques spontanées<br>Diminution de la RH<br>Cancers du sein et de l'ovaire               |

**Tableau I : Phénotypes associés à la déficience des protéines impliquées dans le contrôle de la RH**

La déficience du gène ATM est associée à des taux élevés de RH et d'aberrations chromosomiques spontanées. Chez des lignées cellulaires humaines A-T, la fréquence de RH intrachromosomale spontanée entre des répétitions néo est stimulée de 40 à 100 fois<sup>124</sup>. Chez des souris K.-O. pour le gène, on observe une augmentation plus modeste de l'ordre de 2 fois du taux de RH<sup>125</sup>. Les individus homozygotes pour la mutation ont un risque 500 fois plus élevé de développer une leucémie ou un lymphome, des tumeurs impliquant le réarrangement chromosomique des gènes d'immunoglobulines. La maladie prédispose également à d'autres types de cancers comme le cancer du sein, de l'ovaire, du colon et du col de l'utérus<sup>126-130</sup>.

Une augmentation de la fréquence de RH est également associée à la déficience de P53. Impliqué à la fois dans l'arrêt du cycle cellulaire, l'induction de l'apoptose, la réparation des *mismatches*, la réparation des dommages causés par les ultraviolets (*nucleotide excision repair*) et la RH, P53 joue le rôle du gardien du génome. Sa déficience est donc associée à une multitude de cancers. Par son association physique à RAD51<sup>131</sup>, P53 jouerait un rôle clé dans la suppression de la RH. La liaison de P53 bloquerait l'association de RAD51 à d'autres cibles, inhiberait le pairage homologue et le transfert de brin<sup>132</sup>. La mutation de P53 pourrait donc induire un échange incontrôlé de matériel génomique via la RH et contribuer à l'instabilité génomique. Chez des lignées cellulaires tumorales déficientes pour P53, la fréquence de RH spontanée était stimulée par un facteur de 20 à 10 000 fois par rapport à une lignée normale<sup>133-135</sup>. Les souris K.-O. pour P53 se développent normalement, mais sont ensuite très susceptibles au développement de tumeurs spontanées<sup>136</sup>. Ces souris présentent aussi une fréquence élevée d'aberrations chromosomiques et une forte instabilité génomique.

Les protéines RAD50 et MRE11, impliquées à la fois dans la RH et le NHEJ, sont très importantes et leur déficience ne permet pas la viabilité des modèles de souris K.-O. ou des cellules souches embryonnaires. Comme pour ATM et P53, leur absence est associée à une augmentation de la fréquence de RH spontanée<sup>137</sup>. Chez les modèles déficients pour les gènes RAD51, RAD52, RAD54, BRCA1 ou BRCA2, on

observe également une forte instabilité génomique associée à des aberrations chromosomiques et à une prédisposition au cancer mais on observe plutôt une diminution de la fréquence de RH<sup>137-139</sup>. RAD51, BRCA1 et BRCA2 pourraient jouer un rôle crucial unique dans la RH, puisque leur déficience ne permet pas la viabilité des souris K.-O. Par contre, la déficience des gènes RAD52 et RAD54 permet la viabilité et la fertilité, suggérant un rôle secondaire moins important.

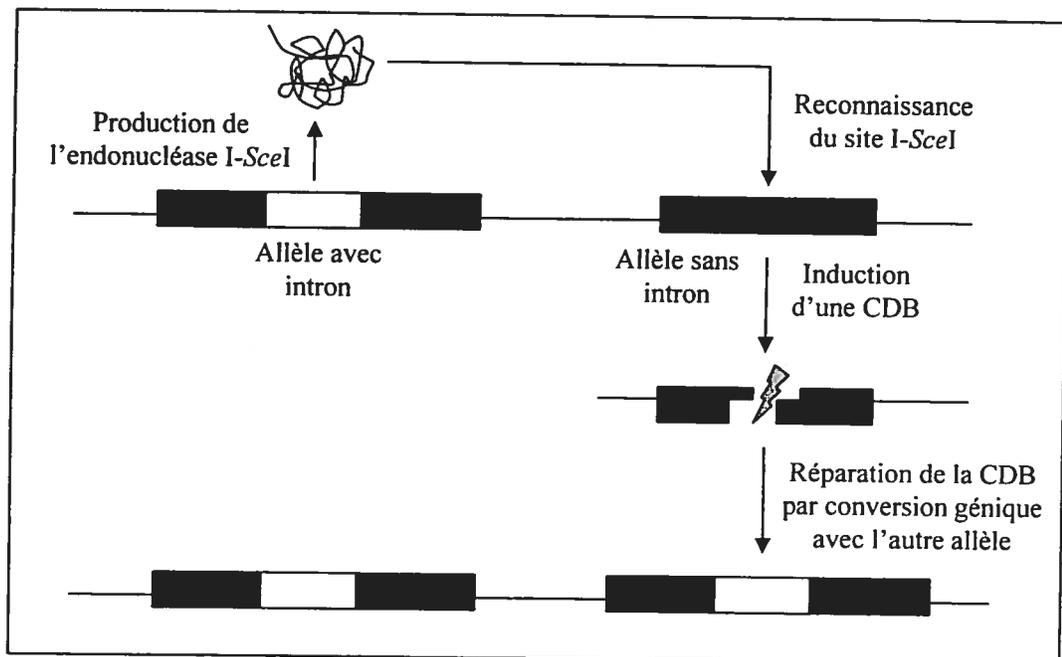
## I.9. Étude de la RH chez les cellules mammifères

### I.9.1. Système d'induction de CDBs par I-SceI

Les radiations ionisantes ont été utilisées pour caractériser l'implication de plusieurs gènes dans la détection, le signalment et la réparation des CDBs. Cependant, les dommages causés par ces radiations ne sont pas limités aux CDBs et elles ne peuvent être utilisées pour cibler un site particulier du génome pour analyser sa réparation au niveau moléculaire. Cette limitation a été contournée par l'utilisation des *rare-cutting homing endonucleases*, un groupe d'endonucléases agissant sur de longs sites de reconnaissance<sup>140</sup>. L'une des mieux caractérisée est l'endonucléase I-SceI de *Saccharomyces cerevisiae*, initialement découverte lors de l'analyse du système génétique  $\omega$  des mitochondries de levure.

Lors du mécanisme de *intron homing*<sup>140</sup> (figure 10, p.48), l'endonucléase est produite par l'intron d'un allèle de gène. Cette endonucléase reconnaît et clive une version sans intron du même gène au site d'insertion de l'intron, initiant un événement de conversion génique entre les 2 allèles. L'endonucléase I-SceI reconnaît un site de 18 pb, qu'il clive pour produire des extrémités 3' cohésives de 4 bases. Dans le génome mammifère, cette séquence de 18 pb est très rare, n'étant retrouvée que dans 1 génome sur 20. Ce site de reconnaissance peut donc être introduit dans le génome et devenir un site unique de CDB. En faisant exprimer le gène de l'endonucléase dans les cellules par la transfection d'un vecteur contenant l'intron mitochondrial de levure

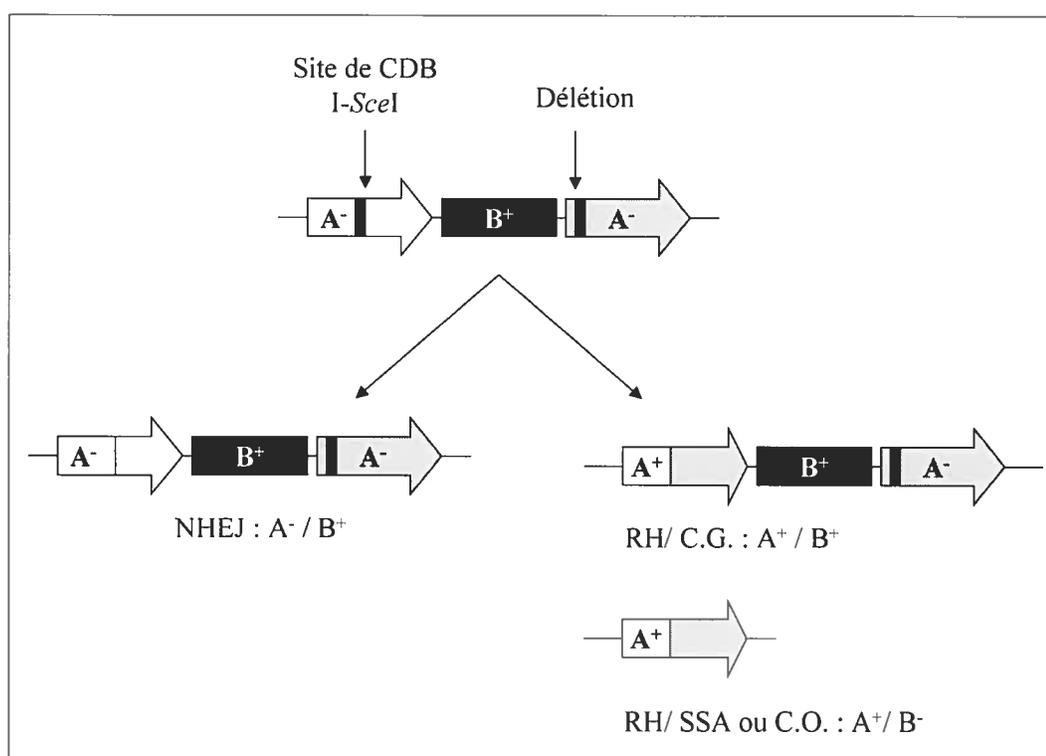
sous le contrôle d'un fort promoteur, on peut induire une CDB à un site unique et connu du génome<sup>141</sup>. Le système I-SceI est largement utilisé pour étudier les mécanismes de réparation des CDBs. Chez les cellules mammifères, l'induction d'une CDB par I-SceI stimule la fréquence de RH<sup>142-144</sup>, de *gene targeting*<sup>141,144</sup> et de recombinaison illégitime (non-homologue)<sup>143</sup> par un facteur de 100 à 1000 fois.



**Figure 10 : Mécanisme du "intron homing" de l'endonuclease I-SceI dans les mitochondries de la levure *Saccharomyces cerevisiae*<sup>140</sup>**

### I.9.2. Modèles d'étude de la RH intrachromosomique

Chez les cellules mammifères, les modèles d'études courants de la RH impliquent l'induction d'une CDB à l'intérieur d'un plasmide intégré au génome<sup>83,143</sup> (figure 11, p.49). Ce plasmide contient normalement 2 copies parfaitement homologues d'un gène de sélection, l'une étant inactivée par une mutation et l'autre par l'introduction d'un site *I-SceI*. Suite à l'induction de la cassure, les événements de conversion génique sont détectés par la génération d'un gène de sélection fonctionnel, qui restera par contre inactif lors des événements de NHEJ. Entre les 2 répétitions, la perte d'un second gène de sélection permet de détecter les événements de SSA ou de *crossing-overs* intrachromosomiques.



**Figure 11 : Système d'étude de la réparation des CDBs chez les cellules mammifères**

A : Gène de sélection en 2 copies (répétitions directes).

B : Gène de sélection en copie unique.

Chez des lignées cellulaires CHO, de tels systèmes d'étude ont généré des résultats assez disparates. Dans une étude, l'induction d'une CDB dans une répétition néo a généré 50 à 70% d'événements de réparation non-homologues (NHEJ) et 30 à 50% d'événements de RH entre les répétitions directes, de type conversion génique ou SSA<sup>83</sup>. Dans un système semblable utilisant des répétitions APRT, l'induction d'une CDB a généré 21% de NHEJ contre 79% de RH<sup>143</sup>. Ce type de modèle ne permet d'étudier que la RH intrachromosomique limitée entre 2 séquences artificielles parfaitement homologues et juxtaposées à un site choisi du génome. Ils peuvent donc difficilement présenter une image réelle de la réparation des CDBs dans le génome mammifère, qui contient 34% de séquences répétitives divergentes hautement dispersées.

### I.9.3. Étude de la RH méiotique dans les cellules germinales

Dans les cellules diploïdes en méiose, le *crossing-over* est un mécanisme essentiel à la bonne ségrégation des chromosomes et à la création d'une diversité génétique entre les individus d'une population<sup>145</sup>. La RH méiotique implique le pairage des chromosomes homologues et l'échange d'information génétique entre les chromatides non-sœurs. Pour initier la RH, des CDBs sont induites par la topoisomérase SPOII à des endroits précis et à une fréquence contrôlée<sup>146</sup>.

Puisque l'information génétique portée par les cellules germinales servira à générer un nouvel individu, il n'est pas désirable que ces CDBs méiotiques soient réparés par une voie prônant les erreurs. Consistant avec l'hypothèse que la voie du NHEJ soit réprimée en méiose, il a été démontré que l'expression de la protéine Ku70 était fortement supprimée en phase précoce de la spermatogénèse<sup>147</sup>.

### I.9.3. Étude de la RH mitotique dans les cellules somatiques

Contrairement aux CDBs méiotiques, les CDBs survenant chez les cellules somatiques ne sont pas contrôlées par la cellule. Résultant de dommages induits accidentellement lors de la réplication, de la ségrégation ou de l'effet d'agents exogènes, elles surviennent à une fréquence et à des sites aléatoires, et sont de ce fait plus dangereuses pour l'intégrité génomique.

Même si la machinerie protéique utilisée pour réparer ce type de CDB semble être la même qu'en méiose, le mécanisme de RH utilisé serait différent. D'abord, il a été démontré que la RH interallélique entre des chromosomes homologues était beaucoup moins fréquente en mitose qu'en méiose<sup>148</sup>. Pour éviter la perte d'hétérozygotie pouvant contribuer à l'instabilité génomique, la RH mitotique entre des chromatides interchromosomiques est réprimée par un facteur de  $10^4$ , la RH intrachromosomique étant beaucoup plus fréquente<sup>149</sup>. Dans ce même optique de maintien d'intégrité, le mécanisme de *crossing-over* pourrait être réprimé dans les cellules mitotiques en faveur d'un mécanisme de conversion génique plus conservatif<sup>150</sup>.

### I.9.4. Étude de la RH ectopique

Chez la levure, plusieurs études ont rapporté que la RH ectopique entre des séquences artificielles dupliquées<sup>151-154</sup> ou des séquences répétitives naturelles<sup>155,156</sup> était aussi fréquente que la RH allélique, autant chez les cellules méiotiques que mitotiques. Toutefois, il semblerait que les effets délétères de la recombinaison ectopique soient restreints dans les cellules somatiques. En méiose, il a été démontré que 35 à 60% de la recombinaison ectopique était associée à un mécanisme de *crossing-over* résultant en translocations réciproques<sup>157</sup>, alors que la recombinaison ectopique associée à ce mécanisme était grandement supprimée en mitose<sup>154</sup>.

Chez les mammifères, la recombinaison ectopique entre des séquences situées à des positions non-alléliques ne se produit pas à une fréquence très élevée mais a tout de même été observée. D'abord, plusieurs cas de translocations ont été associés à un mécanisme de NHEJ, suggérant que des séquences d'ADN situées à des sites ectopiques puissent interagir les unes avec les autres<sup>158</sup>. Des événements de RH ectopiques entre des séquences homologues ont également été observés dans le génome de souris<sup>142,159-161</sup>. Dans un système de RH ectopique entre 2 répétitions néo localisées sur des chromosomes différents, l'induction d'une seule CDB a généré principalement des événements de conversion génique, le mécanisme de *crossing-over* étant supprimé<sup>142</sup>. Par contre, l'induction d'une CDB dans les 2 répétitions du même système a permis de détecter des translocations à une fréquence 80 fois plus élevée<sup>161</sup>.

Au laboratoire, nous avons démontré dans un système cellulaire qu'une CDB induite dans un élément répétitif L1 intégré au génome de souris pouvait être réparée par conversion génique avec des séquences L1s endogènes ectopiques à une fréquence de 1% (1 événement de conversion génique par 100 CDBs réparées)<sup>125</sup>. La RH entre les séquences ectopiques pourrait donc être essentiellement limitée à un mécanisme de conversion génique conservatif et les événements délétères de *crossing-over* pourraient être évités pour préserver l'intégrité génomique. Dans le génome humain, des événements de RH ectopiques délétères entre des séquences répétitives LINEs<sup>53-58</sup> et SINEs<sup>52</sup> ont échappé à ce contrôle et ont conduit à des translocations chromosomiques responsables de plusieurs pathologies.

### **I.9.5. Influence de la divergence sur la RH**

Les interactions ectopiques entre les séquences répétitives dispersées du génome mammifère sont également contrôlées par la divergence de séquence, qui joue un rôle majeur dans la suppression de la RH. La machinerie enzymatique du processus de RH requiert en premier lieu des substrats d'ADN homologues. Même si la stabilité du génome s'en trouve menacée, des échanges homéologues entre des séquences

similaires mais non identiques peuvent tout de même avoir lieu puisque le transfert de brin peut s'effectuer sur des régions d'homologie imparfaite. Toutefois, il semble que la fréquence de recombinaison soit fortement dépendante du taux d'homologie et de la longueur des partenaires homologues.

L'effet de la divergence de séquence sur l'efficacité de la RH a d'abord été étudié chez les procaryotes. Chez *E. coli*, une divergence de 10% réduisait de 40 fois la fréquence de RH, et le segment minimal d'homologie (*minimal efficient processing segment* ou MEPS) requis pour une RH efficace a été évalué à environ 50 pb<sup>162</sup>. Chez la levure, la divergence de séquence semble imposer une barrière moins haute à la RH. Dans une variété de systèmes de recombinaison, la conversion génique entre des séquences homéologues (de 73 à 85% homologues) était diminuée de 10 à 20 fois par rapport à celle survenant entre des séquences parfaitement homologues<sup>163</sup>. Dans un système de RH interchromosomique entre les gènes PMA1 et PMA2 homologues à 85% et localisés à des positions ectopiques, la divergence avait un impact majeur sur la recombinaison en diminuant de 75 fois sa fréquence<sup>164</sup>. Chez la levure, l'impact de la divergence semble être le même sur la RH mitotique ou méiotique<sup>165</sup>.

Contrairement à la RH procaryote qui ne requiert qu'un MEPS de 50 pb pour être efficace<sup>162</sup>, la RH mammifère est grandement réduite sous un MEPS de 200 pb<sup>166,167</sup>. Les fréquences de RH intrachromosomique<sup>167</sup>, de RH extrachromosomique<sup>117</sup> et de *gene targeting*<sup>168,169</sup> sont réduites de façon exponentielle avec une diminution de la longueur de l'homologie.

Chez les cellules de souris, la RH intrachromosomique entre les gènes thymidine kinase du virus herpes simplex (HSV-1 et HSV-2) partageant 81% d'homologie est réduite de 1000 fois par rapport à la fréquence de RH spontanée entre 2 séquences homologues HSV-1<sup>170</sup>. Une analyse précise des recombinants homéologues générés a cependant révélé une prédominance marquée de *crossing-overs* et une absence complète de conversion génique<sup>116</sup>. Dans le même sens, nos analyses de réparation

homologue d'une CDB par des séquences répétitives naturelles de type LINEs ont démontré qu'une grande divergence de 15% entre les partenaires de RH conduisait exclusivement à des produits imprécis de recombinaison (*one-sided invasion*), alors que la RH entre des séquences plus homologues générerait 50% de produits précis<sup>171</sup>.

Il est important de considérer que la RH entre des séquences uniques ou des séquences faiblement répétées du génome pourrait s'effectuer selon des règles complètement différentes que celles s'appliquant lors de la RH homéologue entre des séquences hautement répétitives. Chez les mammifères, la forte densité en séquences répétitives dispersées pourrait nécessiter une contrainte majeure au mécanisme de RH de type *crossing-over*, de façon à éviter les événements de RH délétères entre des séquences divergentes ectopiques. Chez la levure, cette contrainte ne semble pas être nécessaire, puisque l'augmentation de la divergence ne diminue pas la proportion d'événements de *crossing-over*<sup>164</sup>.

#### I.9.6. Implication du système de MMR dans le contrôle de la RH

Le système de *mismatch repair* (MMR) est une voie de réparation d'ADN utilisée lors du processus de réplication. Conservé chez les procaryotes et les eucaryotes, ce système semble également être impliqué dans le processus de RH où il pourrait jouer un rôle important dans l'enlèvement de l'ADN non-homologue durant les mécanismes de conversion génique et de SSA<sup>172</sup>. Chez la levure, la protéine MSH2 se lie *in vitro* aux jonctions de Holliday, des structures intermédiaires de la RH<sup>173</sup>. Chez *E. coli*, MutS et MutL inhiberaient le transfert de brin catalysé par recA<sup>174</sup>.

La mutation des gènes MSH2 et MSH3 augmente la fréquence de RH entre des séquences identiques<sup>175</sup> et lève partiellement la barrière de RH intrachromosomique et de *gene targeting* imposée par la divergence de séquence, sans toutefois l'éliminer<sup>176,177</sup>. Des études *in vivo* ont également démontré que les souris K.-O. pour le gène MSH2 développaient des tumeurs hématologiques initiées par des réarrangements

chromosomiques<sup>178</sup>. Le système de MMR pourrait donc partiellement contrôler la RH entre les séquences homéologues.

La compilation des études du sujet ont démontré que le taux de RH était davantage déterminé par la longueur d'homologie parfaite disponible que par l'homologie globale entre 2 séquences. L'hétérologie pourrait donc affecter l'initiation de la recombinaison, qui pourrait ensuite se propager librement sur une hétérologie considérable. Consistant avec cette hypothèse, il semblerait que les *mismatches* inhibent uniquement la RH lorsqu'ils sont situés près des extrémités des fragments<sup>177</sup>. Ces *mismatches* pourraient empêcher la formation d'un hétéroduplex stable entre les partenaires homéologues et faire avorter le processus de RH. La liaison directe des protéines du système du MMR aux *mismatches* ou aux facteurs de la machinerie de recombinaison pourrait interférer avec la formation de l'hétéroduplex ou son extension.

Chez des cellules mitotiques de levure, la déficience du système de MMR génère des hétéroduplexs 65% plus longs que dans des cellules où il est fonctionnel<sup>165</sup>. Par contre, les recombinants dérivés de mutants MSH2 présentent plus de segments de conversion hybrides mal corrigés<sup>176</sup> et génèrent moins d'événements délétères de *crossing-over*<sup>174,179</sup>. La RH homéologue pourrait donc fréquemment impliquer la correction des *mismatches* de l'hétéroduplex, plutôt que la dégradation des extrémités en *gap*. Pour éviter l'implication du système de MMR dans leur processus de RH, des partenaires homéologues doivent partager une homologie consécutive minimale établie à 610 pb chez la levure<sup>165,175,180</sup>.

### **1.9.7. Influence de la structure de la chromatine sur la RH**

Le caryotype humain est défini par un patron de bandes associées à diverses caractéristiques fonctionnelles et biochimiques. Tout au long des chromosomes, la structure de la chromatine varie en fonction de plusieurs facteurs comme la

compaction, le contenu en gènes et en séquences répétitives, l'activité transcriptionnelle, l'acétylation des histones et la méthylation des îlots CpG. Ces variations pourraient directement influencer le taux de recombinaison qui semble être très variable. La comparaison des distances physiques et génétiques des chromosomes humains suggère en effet une suppression du taux de recombinaison près des centromères et des constriction secondaires de certains chromosomes, ainsi qu'une augmentation marquée de ce taux dans la portion distale des chromosomes, dans les 20 à 30 derniers Mb vers les télomères<sup>2</sup>. Il semblerait également que le taux moyen de recombinaison varie en fonction de la longueur du bras chromosomique, de façon à augmenter les chances de *crossing-over* sur les plus petits bras lors du brassage des gènes à la méiose<sup>2</sup>. La fréquence de recombinaison interchromosomique semble également être affectée par la position relative des partenaires homologues sur leur chromosome respectif. Chez la levure et les mammifères, des études ont démontré qu'une séquence interagissait plus souvent avec un partenaire situé sur un chromosome non-homologue à une distance équivalente de son centromère<sup>142,181</sup>.

Dans la structure compacte de la chromatine, l'ADN est associé aux protéines histones pour former des nucléosomes qui bloquent l'accessibilité des facteurs de transcription à leurs séquences de liaison<sup>182</sup>. La chromatine doit donc être temporairement remodelée dans les régions nécessitant la transcription des gènes actifs, pour fournir une structure d'ADN ouverte accessible à la machinerie transcriptionnelle. Le mécanisme de recombinaison, qui nécessite également un déroulement de la structure d'ADN pour permettre l'invasion de l'hétéroduplex homologue, pourrait donc être couplé à la transcription. La première évidence d'une corrélation entre l'accessibilité de l'ADN, la transcription et la recombinaison a été décrite chez la levure. Dans cette étude, l'activation du promoteur HOTT1 stimulait de 25 à 100 fois la recombinaison d'un gène adjacent<sup>151</sup>. Une telle stimulation de la RH interchromosomale a été observée chez des cellules mitotiques et méiotiques<sup>184</sup>.

Chez les eucaryotes supérieurs, il fût d'abord observé que la transcription des gènes de région variable de la chaîne lourde d'immunoglobuline  $V_H$  survenait au moment de la recombinaison  $V_H(D)J_H$  des cellules B, suggérant une corrélation entre la transcription et la recombinaison<sup>185</sup>. Des évidences directes ont ensuite démontré que la transcription stimulait la RH extrachromosomale<sup>186</sup> et intrachromosomale<sup>187</sup> spontanée chez les cellules mammifères. Ces études ont également démontré que seule la fréquence de RH était affectée par la transcription, le mécanisme utilisé étant indépendant du niveau transcriptionnel. Contrairement aux observations chez la levure où seule la fréquence des événements délétères de SSA était stimulée par un niveau transcriptionnel élevé<sup>142</sup>, les résultats chez les cellules mammifères ont démontré une stimulation équivalente de la fréquence de conversion génique<sup>187</sup>. La transcription ne semble cependant pas affecter la fréquence ni le mécanisme de la RH induite par les CDBs<sup>115</sup>, suggérant que la stimulation par la transcription agit au niveau de l'initiation de la RH. Une fois initiée par une CDB qui stimule la fréquence de RH par un facteur de l'ordre de 100 à 1000 fois<sup>140,142,143</sup>, l'effet de la transcription ne serait plus perceptible.

L'effet de la transcription pourrait s'expliquer par un changement torsionnel induit par les topoisomérases, qui fournirait une structure de chromatine à la fois plus accessible à la machinerie de recombinaison et à l'attaque par les nucléases. Des études *in vivo* de l'ADN eucaryote ont effectivement révélé que l'accessibilité de la chromatine aux différentes nucléases différait grandement entre les gènes transcriptionnellement actifs et inactifs<sup>189</sup>. Les régions transcriptionnellement actives pourraient donc être plus exposées aux dommages causés par les carcinogènes<sup>190</sup>. De la même façon, les sites fragiles du génome, qui sont souvent la cible de CDBs, pourraient être plus souvent impliqués dans la recombinaison<sup>191,192</sup>. Puisque plusieurs oncogènes codent pour des facteurs de transcription ou des produits régulant la transcription par les voies de transduction de signal, on pourrait également suggérer la possibilité qu'une augmentation de la transcription médiée par les oncogènes puisse augmenter la RH ectopique entre les séquences répétitives du génome et contribuer à son instabilité.

La structure de la chromatine est également affectée par la méthylation des cytosines de l'ADN, pour laquelle plusieurs conséquences biophysiques ont été démontrées. La méthylation joue un rôle critique dans la régulation transcriptionnelle et serait impliquée dans l'inactivation du chromosome X et l'établissement des patrons d'*imprinting* génomique. L'observation d'une différence élevée de fréquence de RH à des régions chromosomiques identiques lors des méioses mâle ou femelle suggère un effet de la méthylation<sup>193</sup>. Chez l'humain, des translocations chromosomiques résultant d'un échange entre des satellites (courtes répétitions simples en tandem) décondensés et déméthylés ont fréquemment été associées à une immunodéficiência, une instabilité centromérique et le syndrome d'anomalies faciales<sup>194,195</sup>. *In vitro*, des études ont démontré que la méthylation des îlots CpG réduisait l'accessibilité des recombinaisons aux sites de recombinaison V(D)J<sup>196,197</sup> et qu'elle inhibait la RH réciproque (*crossing-overs*) à des *hot-spots* de recombinaison<sup>198</sup>. Par contre, la méthylation est demeurée sans effet sur la RH et le NHEJ extrachromosomiques des cellules mammifères<sup>199</sup>.

Contrairement à la transcription, la méthylation pourrait affecter la recombinaison à une étape subséquente à celle de l'initiation de la cassure. Elle pourrait prévenir la formation des intermédiaires de recombinaison en affectant leur stabilité. La méthylation pourrait contribuer à stabiliser le génome en empêchant la RH entre les séquences répétitives du génome, qui sont les cibles majeures de méthylation chez les mammifères<sup>200</sup>. Consistant avec cette hypothèse, il est intéressant de constater que le génome de la levure ne subit pas de méthylation, et n'a pas à se préoccuper de la recombinaison ectopique entre les séquences répétitives qui sont peu nombreuses dans son génome.

## II. MATÉRIEL ET MÉTHODE

### II.1. Construction du vecteur pL1HsSceI

Le vecteur pL1HsSceI a été généré à partir du vecteur pL1Hs1V1.2<sup>201</sup> contenant la portion 3' du LINE cd11 (cd11B). Le LINE cd11 appartient à la sous-famille Ta (*transcribed, subset a*) active en rétrotransposition dans le génome humain et son ADNc a été isolé de cellules de tératocarcinome<sup>202</sup>. La portion conservée de 23 pb *NheI-BamHI* de cet ADNc a été remplacée par le site de restriction *I-SceI* de 18 pb (figure 12, p.64).

### II.2. Lignée cellulaire

La lignée cellulaire GM05849 a été obtenue par le NIGMS Human Mutant Cell Repository (Coriell Institute for Medical Research). Cette lignée de fibroblastes déficiente pour le gène ATM provient d'un patient atteint d'ataxie-télangiectasique<sup>203</sup> et a été immortalisée par le virus défectif SV-40. Ces cellules présentent un niveau élevé de recombinaison homologue<sup>124</sup>. La croissance des cellules a été maintenue dans un milieu EMEM (*Eagle's Minimal Essential Medium*, Wisent) complété de 10% de sérum bovin fœtal (FBS, Wisent) et de 50 µg/mL de gentamicine (Wisent), dans un environnement à 37°C et 5% de CO<sub>2</sub>.

### II.3. Transfection stable du vecteur pL1HsSceI

Le vecteur pL1HsSceI a été linéarisé par une digestion *BglIII* (New England Biolabs) puis transfecté par électroporation dans la lignée GM05849. Une quantité de  $5 \times 10^5$  cellules ont été transfectées par 1 µg de vecteur dans une cuvette de 0,4 cm avec des conditions de 960 µF et 270 volts (électroporateur *gene zapper*, Biorad). Les cellules ayant intégré de façon stable le vecteur ont été sélectionnées 48 heures après la

transfection dans un milieu EMEM sélectif (200 µg/µL G418, Wisent). Les clones résistants ont été individuellement repiqués et amplifiés.

#### II.4. Extraction de l'ADN génomique

L'ADN génomique a été extrait des cellules dans une solution de lyse (50 mM Tris-HCl pH=7.6, 100 mM EDTA, 100 mM NaCl, 1% SDS et 1,6 mg/mL protéinase K) 18 heures à 60°C avec faible agitation. L'ADN a été purifié par un traitement à la RNase (0,3 mg/mL) 1 heure à 37°C et une extraction au phénol-chloroforme<sup>204</sup>. L'ADN a ensuite été précipité à l'isopropanol<sup>204</sup> et resuspendu dans du TE 1X.

#### II.5. Analyse des sites d'intégration

Une première analyse des clones a été effectuée par hybridation Southern<sup>204</sup> de 13 µg d'ADN génomique digéré par *SacI* (ATA34) ou *BglIII* (ATA67, GM05849) avec une sonde néomycine marquée au <sup>32</sup>P. Une sonde correspondant au fragment *EagI-BamHI* du vecteur pMC1neo (Stratagene) a été utilisée.

La localisation génomique des intégrations a été caractérisée par le séquençage des jonctions chromosomiques à l'aide du *Thermo Sequenase Radiolabeled Terminator Cycle Sequencing kit* (Amersham). Les jonctions du clone ATA34 ont été séquencées par les amorces TK4415R (5'- TTAACAGCGTCAACAGCGT-3') et TK3688 (5'- GGTCGATGTGTCTGTCCT-3'). Celles du clone ATA67 ont été séquencées par les amorces TK4415R et TK2883 (5'- AACCCACGAACCATAAACC-3'). La séquence des sites d'intégration a été identifiée par le programme *BLAST Human Genome* de GenBank ([www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)).

## II.6. Induction de la cassure double-brin

Une cassure double-brin a été induite au site I-SceI des clones ATA34 et ATA67 par la transfection transitoire d'un vecteur d'expression de l'enzyme méganucléase I-SceI, le pCBASceTcr. Le pCBASceTcr est dérivé du pCBASce<sup>141</sup>. Le pCBASce a été modifié par le clonage du gène de résistance à la tétracycline (issu de l'amplification PCR du vecteur pBR322, New England Biolabs) à l'intérieur du site *ScaI* du gène de résistance à l'ampicilline. Une quantité de  $5 \times 10^5$  cellules ont été électroporées par 100 µg de vecteur pCBASceTcr circulaire dans les mêmes conditions qu'à la section II.3. Les cellules viables ont été récoltées 72 heures après l'induction de la cassure et l'ADN a été extrait (section II.4, p.60).

L'efficacité de coupure au site I-SceI a été analysée par la digestion I-SceI (New England Biolabs) des produits d'amplification PCR de la portion cd11B du vecteur pL1HsSceI intégré. Un fragment de 3,5 kb a été amplifié par le système *Expand High Fidelity* (Boehringer Mannheim) avec les amorces néo7890 (5'-ACTGGAGCGGGAAGGGACTG-3') et tk2904 (5'-GGTTTATGGTTCGTGGGGGTT-3') dans les conditions suivantes : 1 cycle (94°C/ 2 min), 10 cycles (94°C/ 1 min; 65°C/ 1 min; 72°C/ 2 min), 20 cycles (94°C/ 1 min; 65°C/ 1 min; 72°C/ 2 min; extension de 20 sec/cycle). La proportion du produit PCR résistant à la coupure par I-SceI a été quantifiée par densitométrie.

## II.7. Plasmid rescue

L'ADN génomique a été digéré durant 18 heures par des enzymes (New England Biolabs) permettant la récupération des plasmides intégrés au génome, soit *SacI* pour le clone ATA34 et *BglII* pour le clone ATA67. Les digestions ont été nettoyées par une extraction au phénol-chloroforme<sup>204</sup> et une précipitation à l'éthanol. Une ligation a été effectuée sur 5 µg d'ADN dans un volume total de 200 µL avec 40 Unités de T4 DNA ligase (New England Biolabs), à 16°C durant 18 heures. La ligation a été

nettoyée par une extraction au phénol-chloroforme<sup>204</sup> et précipitée à l'éthanol pour être resuspendue dans un petit volume de 5  $\mu\text{L}$  d' $\text{H}_2\text{O}$ . Chaque  $\mu\text{L}$  (1  $\mu\text{g}$ ) de ligation a ensuite été électroporé dans 50  $\mu\text{L}$  de bactéries compétentes DH10B (préparation maison) dans une cuvette de 0,2 cm et des conditions de 2500 volts, 21  $\mu\text{F}$  et 200  $\Omega$  (électroporateur *gene zapper*, Biorad). Les bactéries transformées ont été étalées sur du milieu LB agar (50  $\mu\text{g}/\mu\text{L}$  ampicilline) et les colonies ont été individuellement repiquées dans 2 mL de milieu LB (50  $\mu\text{g}/\mu\text{L}$  ampicilline) après une incubation de 16 heures à 37°C. L'ADN plasmidique a été récupéré par une méthode d'extraction rapide<sup>205</sup>.

## II.8. Analyse des recombinants

Les plasmides ont été analysés par des digestions caractéristiques avec des enzymes de restriction (New England Biolabs) et par séquençage à l'aide du *Thermo Sequenase Radiolabeled Terminator Cycle Sequencing kit* (Amersham) et de plusieurs amorces différentes hybridant en direction 5' (HDR) ou 3' (HDF) du LINE cd11. Les amorces HDF4683 (5'-GCTTACCTACAACACTATCTGA-3') et HDR4885 (5'-GGGTTTTTATGGTTTTAGGT-3') ont été utilisées pour couvrir la région de la cassure double-brin. Les amorces TK4415R (5'-TTAACAGCGTCAACAGCGT-3') et HDR6q14A (5'-GCATATAGACTCATCCACTTC-3') ont été utilisées pour séquencer les jonctions des événements de NHCR des clones ATA67 et ATA34 respectivement. Dans les cas d'événements de RH ayant généré un long segment de conversion, des amorces de plus en plus distantes du site de cassure ont été utilisées pour couvrir la totalité de la conversion (HDF3602 à HDF5698) et (HDR5903 à HDR3689). Les séquences ont été identifiées par le programme *BLAST Human Genome* du NCBI ([www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)) selon la version de juin 2002.

### III. RÉSULTATS

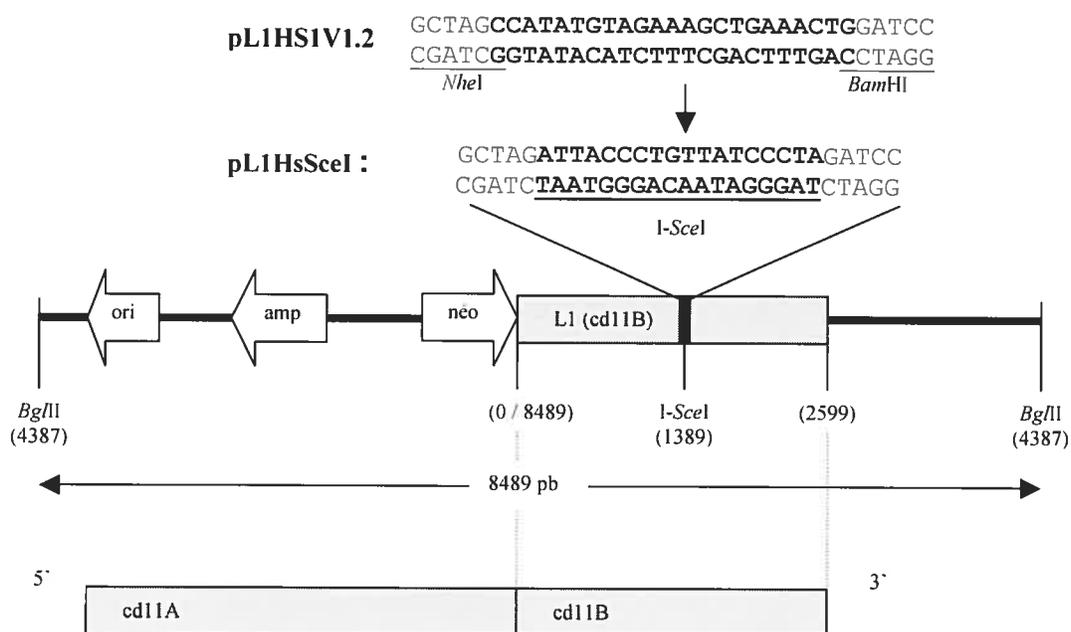
#### III.1. Stratégie expérimentale

Notre travail consistait à mieux comprendre les principes et l'importance de l'organisation du génome humain en étudiant l'accessibilité entre différentes régions du génome et leur possibilité d'interaction. Afin de bien représenter l'ensemble du génome, nous avons choisi d'étudier les interactions entre les séquences répétitives de type L1, dont les  $5 \times 10^5$  copies sont dispersées sur l'ensemble des chromosomes. En induisant une CDB à l'intérieur d'un élément L1 intégré à un locus donné, notre stratégie était d'induire des événements de réparation par RH avec des L1s endogènes et de déterminer lesquels de tous les L1s du génome ont pu avoir accès à la cassure pour la réparer.

Pour induire une CDB à l'intérieur d'un L1 unique et récupérable, nous avons introduit dans le génome des cellules GM05849 le vecteur pL1HsSceI (figure 12, p. 64) contenant un élément L1. Puisque la majorité des L1s sont tronqués à leur extrémité 5', nous avons utilisé la portion 3' plus conservée, permettant ainsi l'interaction avec une plus grande population de L1s endogènes. Le L1 contenu dans le vecteur a été modifié par l'introduction du site de restriction I-SceI qui est absent du génome des mammifères et peut donc être utilisé pour induire une CDB à un site unique.

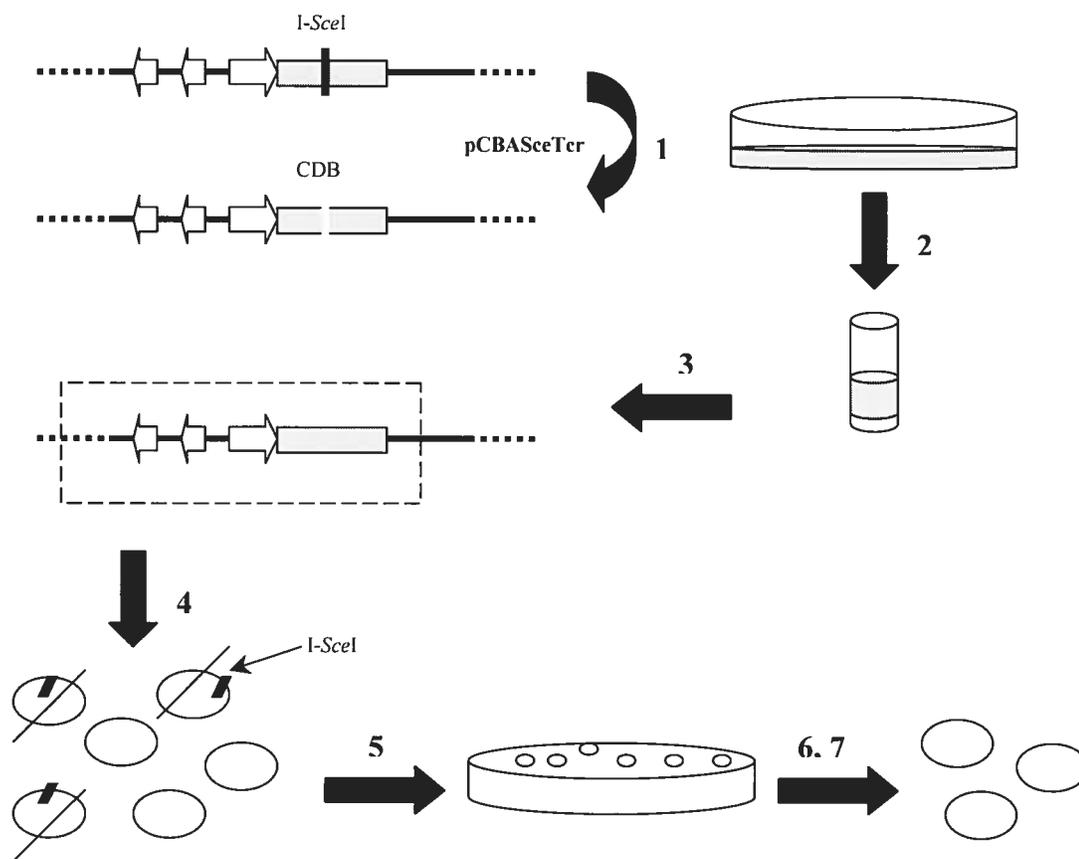
La figure 13 (p. 65) schématise les différentes étapes de la génération, de la récupération et de l'analyse des recombinants. La CDB est d'abord induite dans les clones par la transfection transitoire des cellules par le vecteur d'expression de la méganucléase I-SceI, le pCBASceTcr. Les cellules sont récoltées 48 heures plus tard pour en extraire l'ADN génomique, qui est ensuite digéré par un enzyme de restriction permettant la séparation des plasmides du reste du génome. Une ligation permet ensuite de refermer les plasmides sur eux-mêmes et de faciliter leur introduction dans une bactérie déficiente pour les mécanismes de réparation. Au

préalable, une étape supplémentaire consiste à linéariser par *I-SceI* le pool de plasmides, afin d'éliminer ceux ayant échappé à la CDB ou les événements de réparation par refermeture intacte du site. Chaque plasmide recombinant, correspondant à une colonie bactérienne, est ensuite individuellement extrait pour être analysé par digestion et séquençage.



**Figure 12 : Structure du vecteur pL1HsSce1**

Le vecteur pL1HsSce1 a été construit à partir du vecteur pL1HSIV1.2<sup>201</sup> contenant le gène de sélection néomycine (néo) ainsi que la portion 3' cd11B du L1Hs cd11<sup>202</sup>. La portion de 23 pb *NheI-BamHI* a été enlevée et remplacée par une cassette contenant le site de restriction *I-SceI*. Le vecteur pL1HsSce1 a été linéarisé par le site *BglII* pour être intégré au génome.



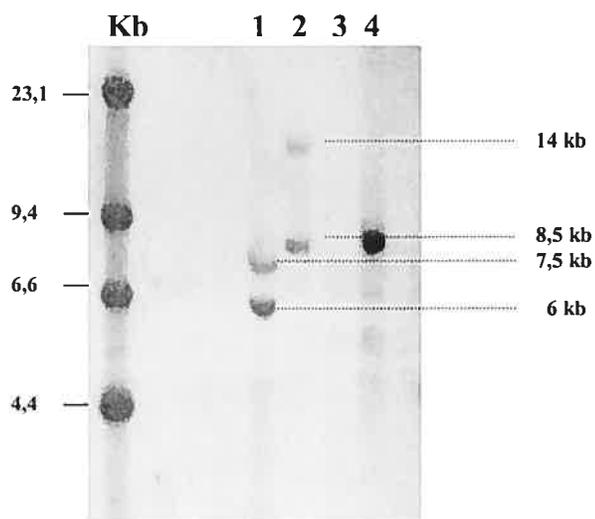
**Figure 13 : Schéma de la stratégie expérimentale**

1. Induction d'une CDB au site I-SceI du pL1HsSceI par la transfection transitoire des clones cellulaires par le vecteur pCBASceTer.
2. Récolte des cellules après 48 heures et extraction de l'ADN génomique.
3. Digestion de l'ADN par une enzyme permettant l'extraction du plasmide (*SceI* pour le clone ATA34 et *BglII* pour le clone ATA67).
4. Ligation des plasmides et enrichissement par une digestion I-SceI.
5. Transformation des plasmides dans les bactéries DH10B.
6. Repiquage des colonies et extraction de l'ADN plasmidique.
7. Analyse des recombinants par digestion et séquençage.

### III.2. Génération et analyse des clones

La toute première étape de notre travail consistait à produire des clones cellulaires par l'intégration stable du vecteur pL1HsSceI dans la lignée GM05849. Suite à la transfection des cellules par le vecteur et la sélection des clones porteurs du gène néomycine, nous avons extrait l'ADN génomique et analysé les intégrations du pL1HsSceI. Une première analyse des clones a été effectuée par l'hybridation Southern avec une sonde néo (figure 14, p.67). Les clones ayant intégré le vecteur en un faible nombre de copies ont été retenus et testés pour la présence des séquences plasmidiques nécessaires à la récupération du plasmide (origine de réplication et ampicilline) ainsi que pour la présence du site I-SceI. Les clones ATA34 et ATA67 répondant à ces critères ont été choisis et la structure de leurs sites d'intégration (figure 15, p.68) a été analysée par séquençage.

Comme le révèle l'hybridation Southern (figure 14, p.67), le clone ATA34 contient 2 copies du vecteur, récupérées respectivement à 8,5 kb et 14 kb suite à une digestion *SacI*. Toutefois, le séquençage des jonctions (figure 15b, p.68) a démontré que ces 2 copies étaient intégrées en tandem au même site sur le contig NT\_034482 du chromosome 6q23.1. Par contre, il est à noter que seule la copie de 8,5 kb contient le site I-SceI, la copie de 14 kb étant tronquée en 3' au début de la séquence L1. Pour ce clone, la CDB a donc pu être induite à un site unique du génome et le plasmide de 8,5 kb a pu être récupéré à l'aide des séquences intactes d'origine de réplication et de résistance à l'ampicilline.



**Figure 14 : Analyse des clones par hybridation Southern**

L'ADN génomique du clone ATA67 (puit 1) a été digéré par l'enzyme *Bgl*III et celui du clone ATA34 (puit 2) par l'enzyme *Sac*I. L'ADN génomique des cellules mères GM05849 a été digéré par *Bgl*III et utilisé comme contrôle négatif (puit 3). Le plasmide pL1HsSceI linéarisé par *Bgl*III a été utilisé comme contrôle positif (puit 4). L'ADN du phage lambda digéré par *Hind*III a été utilisé comme standard de poids moléculaire (kb). La sonde néomycine utilisée pour l'hybridation correspond au fragment *Eag*I-*Bam*HI du vecteur pMC1néo.

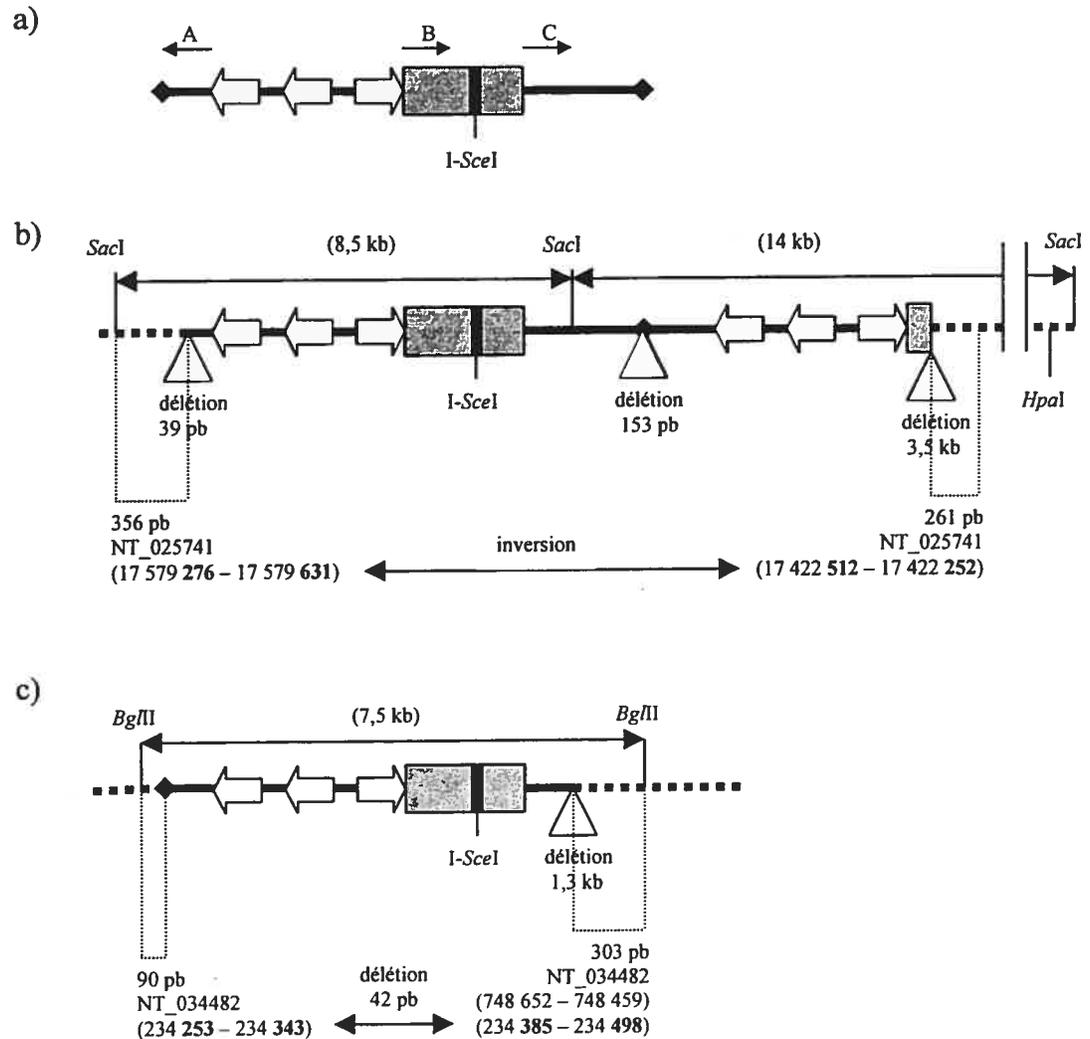
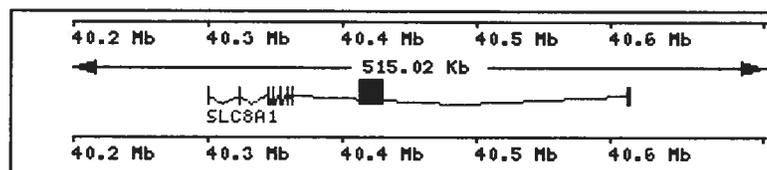


Figure 15 : Structure du pL1HsSce1 intégré dans les clones ATA34 et ATA67

La structure des sites d'intégration du vecteur pL1HsSce1 (a) dans les clones ATA34 (b) et ATA67 (c) a été analysée suite à la digestion de l'ADN génomique par *SacI* (ATA34) ou *BglII* (ATA67), la ligation et la transformation des plasmides dans les bactéries. Les jonctions d'ADN génomique ont été séquencées à l'aide d'amorces situées dans le vecteur (a): A et B pour ATA34, A et C pour ATA67. (A=TK4415R; B=TK3688; C=TK2883). Les positions sur les contigs (NT\_) sont indiqués pour chaque segment de jonction séquencé.

( ..... : ADN génomique ; — : séquences plasmidiques)

Dans le cas du clone ATA67, l'analyse Southern (figure 14, p.67) a également révélé l'existence de 2 copies de vecteur, hybridant à 6 kb et 7,5 kb suite à une digestion *Bgl*III. Cependant, puisque seule la copie de 7,5 kb a pu être récupérée, nous supposons que la copie de 6 kb puisse avoir été tronquée en 5', perdant ainsi les séquences plasmidiques nécessaires à la récupération du plasmide suite à la transformation. Par contre, il nous a été impossible de déterminer si cette copie contenait un site *I-Sce*I. Nos recombinants ont donc été obtenus à partir d'une CDB induite dans la copie fonctionnelle de 7,5 kb, qui bien que tronquée en 3', possède toutes les séquences nécessaires à sa récupération ainsi qu'un site *I-Sce*I (figure 15c, p.68). Cette copie est intégrée sur le contig NT\_034482 du chromosome 2p22.1, à l'intérieur d'un intron du gène *SLC8A1*, un échangeur d'ions  $\text{Na}^+/\text{Ca}^{2+}$  impliqué dans la régulation de l'homéostasie neuronale (figure 16, p.69). Comme il le sera discuté plus loin, l'activité transcriptionnelle d'un site d'intégration du vecteur pourrait éventuellement affecter la fréquence ainsi que le mécanisme de réparation de la CDB induite dans le vecteur intégré.



**Figure 16 : Site d'intégration du pL1HsSceI dans le clone ATA67**

Le vecteur est intégré ( ■ ) en position 40,42 Mb de la bande 2p22.1, à l'intérieur de l'intron 2 du gène *SLC8A1* (*solute carrier family 8 Na<sup>+</sup>/Ca<sup>2+</sup> exchanger member 1*). Le gène *SLC8A1* s'étend de 40,32 à 40,63 Mb et contient 10 exons.

### III.3. Induction d'une cassure double-brin au site I-SceI

Pour chacun des 2 clones, nous avons induit une CDB au site I-SceI au moyen d'une transfection transitoire du vecteur pCBASceTcr, exprimant l'enzyme méganucléase I-SceI. Puisqu'une quantité limitée de cellules peuvent être transformées dans une même expérience et que l'obtention d'un grand nombre de recombinants requiert une grande quantité d'ADN génomique, nous avons du effectuer plusieurs transfactions indépendantes. L'efficacité de coupure au site I-SceI a été évaluée par l'amplification PCR du fragment L1 du vecteur et sa digestion par l'enzyme I-SceI (figure 17, pp.71-72). L'intensité de la bande résistante à la digestion est représentative de la proportion de cellules ayant subi une réparation entraînant la perte du site I-SceI. Cette proportion est toutefois sous-estimée par cette méthode, puisque tous les événements de réparation ayant fait perdre les séquences plasmidiques situées en 3' du L1 ne peuvent être amplifiés par PCR. À cette analyse échappe également une proportion non négligeable de vecteurs qui ont pu être ouverts par une digestion I-SceI puis refermés précisément en reformant un site intact. Finalement, certains événements de recombinaison spontanée ont pu entraîner la perte du site I-SceI sans qu'il n'y ait induction d'une CDB à ce site.

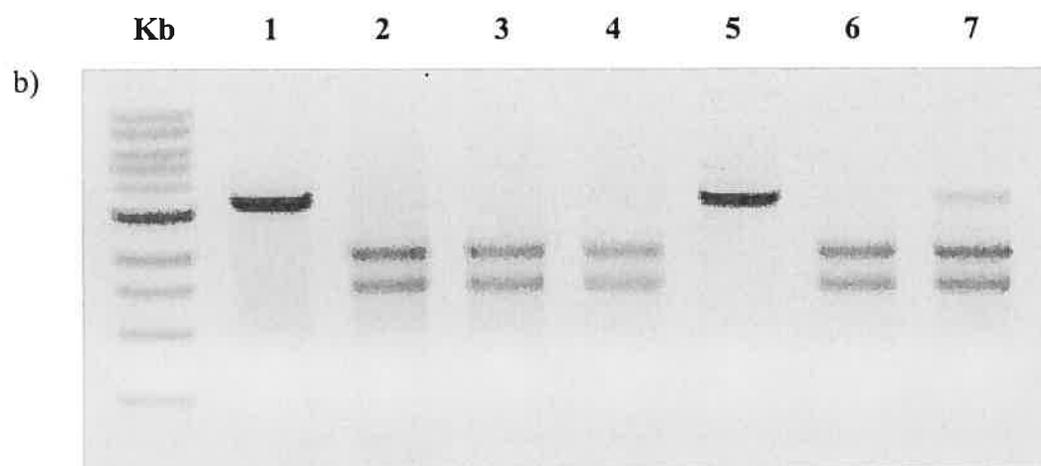
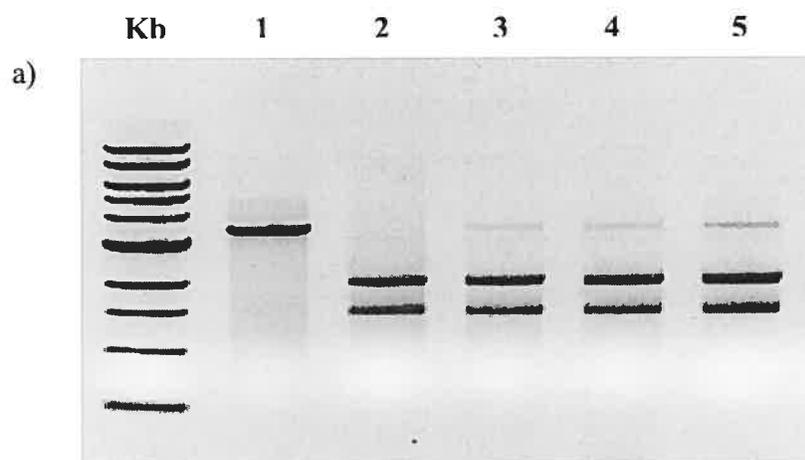
Plusieurs transfactions consécutives des cellules du clone ATA34 (pB2, pB2.1, pB2.3) nous ont permis d'améliorer progressivement la très faible efficacité de coupure jusqu'à 2,4% pour la transfection pB2.3 (figure 17a, pp.71-72). L'efficacité de coupure de la transfection pB1.2 du clone ATA67 était également très faible, soit de 1,4% (figure 17b, pp.71-72). Par contre, l'utilisation du sous-clone ATA67A10 nous a permis de contourner ce problème et d'atteindre une efficacité de 8,2% dès la première transfection (pB1). Ce sous-clone a été obtenu suite à des essais infructueux de rétablissement d'un gène ATM fonctionnel dans le clone ATA67 déficient pour ATM. Le sous-clone ATA67A10 contient donc des séquences plasmidiques intégrées additionnelles, qui à la lumière des résultats obtenus, ne semblent toutefois pas interférer avec la réparation de la CDB. L'augmentation de son efficacité de coupure par I-SceI pourrait simplement s'expliquer par un état différent des cellules

au moment de la transfection, permettant une meilleure introduction du vecteur à l'intérieur de la cellule. Alternativement, l'accessibilité du site *I-SceI* a pu être améliorée dans le sous-clone suite à un réarrangement, la lignée GM05849 ayant un génome très instable.

**Figure 17 : Efficacité d'induction d'une CDB au site *I-SceI* dans les clones ATA34 (a) et ATA67 (b)**

Le produit PCR de 3,5 kb a été amplifié à l'aide des amorces néo7890 et TK2904 situées de part et d'autre du fragment L1 du pL1HsSce1. La digestion du produit PCR par *I-SceI* fournit 2 fragments, de 2 kb et 1,5 kb respectivement. L'intensité de la bande résistante a été évaluée par densitométrie.

- (a) Inductions consécutives d'une CDB dans le clone ATA34: puit 1 (ATA34 non induit, non digéré *I-SceI*) ; puit 2 (ATA34 non induit, complètement digéré par *I-SceI*) ; puits 3, 4 et 5 (ATA34pB2, ATA34pB2.1 et ATA34pB2.3, digérés *I-SceI*). L'efficacité d'induction est progressivement améliorée par 3 transfections consécutives des cellules par le pCBASceTcr, pour atteindre un maximum d'environ 2,4%.
  
- (b) Inductions indépendantes d'une CDB dans le clone ATA67 et le sous-clone ATA67A10 : puit 1 (ATA67 non induit, non digéré *I-SceI*) ; puit 2 (ATA67 non induit, complètement digéré par *I-SceI*) ; puits 3 et 4 (ATA67pB1.2 et ATA67pB2, digérés *I-SceI*) ; puit 5 (ATA67A10 non induit, non digéré *I-SceI*) ; puit 6 (ATA67A10 non induit, complètement digéré par *I-SceI*) ; puit 7 (ATA67A10pB1, digéré *I-SceI*). L'efficacité d'induction n'atteint que 1,4% dans la transfection ATA67pB1.2 et s'élève à 8,2% dans la transfection du sous-clone ATA67A10pB1.



### III.4. Récupération et analyse des événements de réparation

Suite à leur amplification dans la bactérie, les plasmides extraits ont subi un premier classement en fonction de la présence du site *I-SceI* (figure 18, pp.74-75). Les plasmides ayant conservé le site ont pu soit ne jamais avoir subi de CDB, soit avoir été réparés par une religation exacte du site fonctionnel (figure 18a, pp.74-75). Parmi les plasmides ayant perdu le site *I-SceI*, nous avons pu établir un classement en fonction du mécanisme de réparation. Les événements de RH avec un partenaire L1 endogène ont pu être retracés par le retour du fragment de 23 pb *NheI-BamHI* conservé par une grande proportion des L1s (figures 18b et c, pp.74-75). Les recombinants ne pouvant être classés dans cette catégorie ont été coupés au site *SpeI* situé en 3' du site *I-SceI*, qui est perdu lors d'un réarrangement chromosomique non-homologue (NHCR) (figure 18d, pp.74-75). Le reste des recombinants ont finalement été classés dans la catégorie des événements de religation non-homologue des extrémités brisées (NHEJ) (figure 18e, pp.74-75). Par séquençage, nous avons confirmé et précisé le mécanisme de réparation de tous les événements de RH, de la plupart des événements de NHCR et de 13% des événements de NHEJ.

Pour le clone ATA34, la présence d'un second fragment de 14 kb sans site *I-SceI* dans les plasmides récupérés nous a contraint à en effectuer une sélection négative en digérant les ligations au site *HpaI* (figure 15b, p.68), qui linéarise le fragment de 14 kb mais demeure absent de la molécule de 8,5 kb recherchée. Par cet enrichissement supplémentaire, nous avons pu diminuer le nombre de plasmides à analyser suite à la transformation. De plus, il est à noter qu'en raison de la méthode de récupération des plasmides utilisée, tous les événements de réparation ayant entraîné la perte des séquences d'origine de répllication et de résistance à l'ampicilline n'ont pu être analysés. Nos résultats sont donc quelque peu biaisés par la directionnalité de la recombinaison.

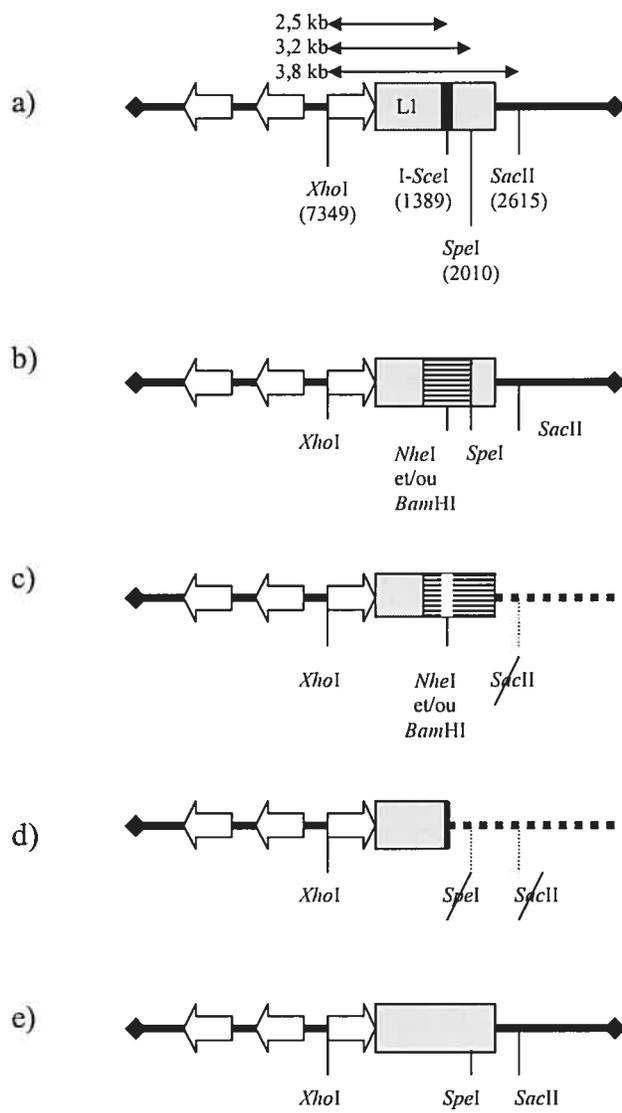
**Figure 18 : Classement des plasmides recombinants selon le mécanisme de réparation de la CDB**

- a) Structure du plasmide pL1HsSceI non recombinant, ou s'étant réparé en reformant un site I-SceI fonctionnel. La taille du plasmide récupéré est de 8,5 kb pour le clone ATA34 et de 7,5 kb pour le clone ATA67.
- b) Structure des RH de type conversion génique caractérisés par la perte du site I-SceI, l'acquisition des sites *NheI* et/ou *BamHI* ainsi que par la conservation des sites en 3' de la CDB (*SpeI* et *SacII*). La taille des plasmides récupérés est la même que celle des plasmides non recombinants.
- c) Structure des RH de type *crossing-over* caractérisés par la perte du site I-SceI, l'acquisition des sites *NheI* et/ou *BamHI* ainsi que la perte des séquences plasmidiques en 3' de la CDB, donc la perte du site *SacII*. La taille des plasmides récupérés peut varier en fonction de la position du nouveau site de récupération *SacI* (ATA34) ou *BglII* (ATA67).
- d) Structure des recombinants non-homologues de type NHCR caractérisés par la perte du site I-SceI et de la portion plasmidique située en 3' de la CDB, donc des sites *SpeI* et *SacII*. La taille des plasmides récupérés peut varier en fonction de la position du nouveau site de récupération *SacI* (ATA34) ou *BglII* (ATA67).
- e) Structure des recombinants non-homologues de type NHEJ caractérisés par la perte du site I-SceI et la conservation de tous les autres sites caractéristiques du pL1HsSceI. La taille des plasmides récupérés peut varier en fonction d'une délétion et/ou d'une insertion au site de CDB.

ADN génomique : .....  
 ADN plasmidique : —————

Site I-SceI : █  
 Fragment 23 pb *NheI*-*BamHI* : ▤

Séquences L1 du pL1HsSceI : █  
 Séquences L1 endogènes : ▨



### III.4.1. Fréquence relative des différents mécanismes de réparation de CDB

Au total, nous avons récupéré 427 plasmides recombinants pour le clone ATA34 et 650 pour le clone ATA67 (tableau II, p.77). La majorité des événements de réparation analysés sont issus d'un mécanisme de recombinaison non-homologue, dans une proportion relative de 96,2% pour le clone ATA34 et de 85,2% pour le clone ATA67. Parmi ces recombinants, la plupart sont le résultat de la fermeture de la cassure par la ligation des extrémités brisées (NHEJ), avec délétion et/ou insertion de nucléotides entraînant la perte du site *I-SceI*. Toutefois, une proportion de 2,3% et de 12,3% des recombinants des clones ATA34 et ATA67 sont issus d'un réarrangement chromosomique non-homologue (NHCR). Dans des proportions semblables de 3,8% et de 14,3%, la réparation de la CDB s'est effectuée par RH avec un partenaire L1 endogène. Dans la très grande majorité des recombinants homologues analysés, un mécanisme conservatif de conversion génique a été utilisé. Étant donné le nombre supérieur de recombinants récupérés avec le clone ATA67 et sa fréquence de RH 3,7 fois plus élevée par rapport à celle du clone ATA34, nous avons également récupéré dans ce clone de rares événements de *crossing-over* et de *one-sided invasion* qui seront analysés en détail un peu plus loin.

Certains contrôles ont été effectués pour confirmer la validité de ces premiers résultats. D'abord, nous avons vérifié que les plasmides étaient bien réparés dans le génome des cellules en culture et non par les bactéries suite à l'enrichissement *I-SceI* de la ligation et à la transformation des bactéries. Pour ce faire, nous avons effectué une ligation de l'ADN génomique de la lignée mère GM05849 digéré *BgIII*, additionné du plasmide pL1HsSceI également digéré *BgIII* en proportion de 10 copies de plasmide par génome. Cette ligation a subi un enrichissement par *I-SceI* et a été transformée dans la bactérie. Cette reconstitution effectuée dans les mêmes conditions que la véritable expérience a démontré qu'aucun recombinant n'a été généré dans la bactérie sur un équivalent de  $3 \times 10^5$  plasmides. La fréquence de faux recombinants dans la bactérie a donc été évaluée inférieure à  $3 \times 10^{-5}$ .

| Mécanisme de réparation   | ATA34      |              | ATA67      |              |
|---------------------------|------------|--------------|------------|--------------|
|                           | nombre     | %            | nombre     | %            |
| <b>Non-homologue</b>      | <b>410</b> | <b>96,2%</b> | <b>554</b> | <b>85,2%</b> |
| NHEJ                      | 400        | 93,9%        | 474        | 72,9%        |
| NHCR                      | 10         | 2,3%         | 80         | 12,3%        |
| <b>Homologue</b>          | <b>16*</b> | <b>3,8%</b>  | <b>96</b>  | <b>14,8%</b> |
| Conversion génique        | 16         | 3,8%         | 93         | 14,3%        |
| <i>Crossing-over</i>      | 0          | 0,0%         | 2          | 0,3%         |
| <i>One-sided invasion</i> | 0          | 0,0%         | 1          | 0,2%         |

**Total des recombinants :**      426                      650

**Tableau II : Fréquences relatives des mécanismes de réparation de CDBs obtenues avec les clones ATA34 et ATA67**

\*Les 16 événements de RH du clone ATA34 classés au tableau ont été récupérés par la technique du *plasmid rescue*. Il est à noter que 3 événements de RH supplémentaires ont été obtenus par une technique alternative sélective ne permettant pas le calcul des fréquences. Un total de 19 événements de RH ont donc été récupérés et analysés pour ce clone.

À titre comparatif, un deuxième contrôle nous a permis d'évaluer à un ordre beaucoup plus élevé de  $10^{-2}$  la fréquence de réparation des plasmides dans les cellules (tableau III, p.79). Ainsi, sur un total de 1623 plasmides récupérés sans enrichissement I-SceI pour l'ADN de la transfection ATA67pB2, un nombre de 22 recombinants ont été récupérés pour une fréquence de 1,4% de recombinaison. Une seconde expérience a été effectuée avec l'ADN de la transfection ATA67A10pB1, pour laquelle la fréquence s'est avérée encore plus élevée, soit de 8,2%.

Ces fréquences sont bien entendu directement dépendantes de l'efficacité d'induction de CDB dans chacune des transfections, et confirment par ailleurs les résultats obtenus par digestion des produits PCR (figure 17, pp.71-72). Par la même occasion, ce dernier contrôle nous a permis de constater que l'étape d'enrichissement de la ligation par une digestion I-SceI ne biaisait pas le mécanisme de recombinaison. Ainsi, nous pouvons constater que pour chacune des transfections analysées, la proportion relative de chaque mécanisme est équivalente avec ou sans enrichissement I-SceI. Une variation plus importante peut toutefois être observée entre les fréquences de NHCR avec la transfection ATA67A10pB1 (10% versus 22%), mais peut être expliquée par un plus faible nombre d'événements dans la méthode non-enrichie. Finalement, nous pouvons remarquer que les proportions relatives sont également conservées entre des inductions de CDB indépendantes.



### **III.5. Analyse des événements de recombinaison non-homologue**

Les événements de réparation de CDB par recombinaison non-homologue (NHEJ et NHCR) générés ont été analysés plus en détail par Catherine Chabot, une étudiante à la maîtrise du laboratoire qui en a fait l'objet détaillé de son mémoire<sup>206</sup>. Pour un portrait complet de la réparation des CDBs dans ce système, ils sont à nouveau présentés ici mais de façon plus simplifiée.

#### **III.5.1. Analyse des événements de réparation de CDB par NHEJ**

Sur le total des 400 et 474 recombinants classés comme NHEJ suite à des digestions caractéristiques (figure 18e, pp.74-75) des plasmides dans les clones ATA34 et ATA67, 57 et 59 recombinants ont respectivement été choisis sur une base aléatoire et ont été plus précisément caractérisés par séquençage du site de cassure réparé (tableau IV, p.81). Une proportion de 70% des sites de cassure ont été refermés suite à la perte de quelques nucléotides à l'une ou aux 2 extrémités générées, accompagnée ou non par l'insertion de quelques nucléotides. Notons que la perte ou l'ajout d'une seule paire de bases est suffisante pour que l'enzyme méganucléase *I-SceI* ne puisse digérer son site de reconnaissance de 18 pb.

Douze pourcent des sites de reconnaissance ont été brisés par la simple insertion de quelques nucléotides, sans qu'il n'y ait de délétion aux extrémités. Nous avons également récupéré quelques événements impliquant une plus grande délétion au site de cassure, pouvant atteindre jusqu'à 688 pb. Finalement, 8 événements récupérés avec le clone ATA34 présentaient une grande insertion au site de cassure. Toutefois, les séquences introduites provenaient dans tous les cas du vecteur de transfection transitoire pCBASceTcr utilisé pour induire la CDB. Nous soupçonnons donc que la préparation de plasmide utilisée pour l'une des transfections du clone ATA34 puisse avoir été partiellement dégradée, générant ainsi dans les cellules transfectées les fragments libres insérés au site de cassure.

| Description du site de cassure I-SceI*                                       | Clone ATA34 |     | Clone ATA67 |     |
|--|-------------|-----|-------------|-----|
|  | n           | %   | n           | %   |
| Petite insertion au site I-SceI<br>(1 à 21 pb)                               | 7           | 12% | 7           | 12% |
| Petite délétion au site I-SceI<br>(1 à 21 pb)<br>(+/- petite insertion)      | 40          | 70% | 40          | 68% |
| Grande insertion au site I-SceI<br>(77 pb à 3,5 kb)<br>(+/- petite délétion) | 8           | 14% | 0           | 0%  |
| Grande délétion au site I-SceI<br>(30 à 688 pb)<br>(+/- petite insertion)    | 2           | 4%  | 12          | 20% |

Total :

|    |      |    |      |
|----|------|----|------|
| 57 | 100% | 59 | 100% |
|----|------|----|------|

**Tableau IV : Description des événements de réparation par NHEJ suite à une CDB au site I-SceI des clones ATA34 et ATA67**

\* La structure du site de cassure réparé a été principalement analysée par le séquençage des recombinants avec les amorces HDF4683 et HDR4885. Les événements impliquant la délétion de la position des amorces ont été séquencés à tâtons à l'aide d'amorces plus éloignées du site de cassure.

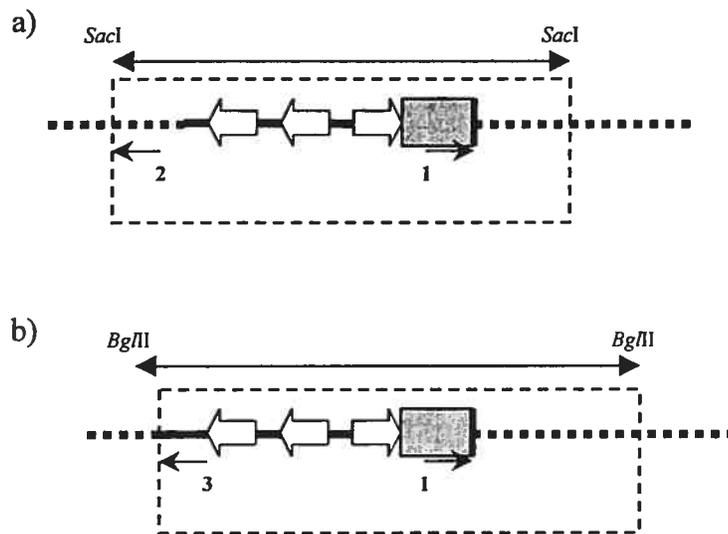
### III.5.2. Analyse des événements de réparation de CDB par NHCR

La structure d'un grand nombre d'événements de réparation par NHCR a été précisée par le séquençage du site de cassure et l'identification des fragments chromosomiques retrouvés à la jonction des réarrangements. La figure 19 (p.83) situe les principales amorces de séquençage utilisées pour les recombinants du clone ATA34 (figure 19a, p.83) et du clone ATA67 (figure 19b, p.83). Parmi les événements du clone ATA34 (tableau V, p.84), un seul a subi une grande délétion au site de cassure (34-H300) et un seul (34-A542) présentait une insertion de 130 pb correspondant à un fragment du pCBASceTcr, du même type que ceux retrouvés dans les 8 événements de NHEJ du clone ATA34 (tableau IV, p.81). Par l'identification des sites de jonction chromosomiques, 2 réarrangements se sont révélés intrachromosomiques (34-B693 et 34-A542) et les autres ont été générés par un mécanisme de translocation avec un autre chromosome. Il est à noter que 2 événements (34-H281 et 34-B876) provenant d'expériences indépendantes ont impliqué exactement le même site de translocation, à la différence d'une base près.

La plupart des événements de NHCR du clone ATA67 présentaient une petite ou une grande délétion au site de jonction du réarrangement (tableau VI, pp.85-86). Parmi les événements récupérés, la moitié se sont révélés interchromosomiques. Ceux-ci ont impliqué 11 chromosomes différents. Parmi ces sites, les régions 9q, 1q et 14q sont représentées plus d'une fois, leur suggérant une accessibilité particulière à la CDB. Dans certains cas (67-A2, 67-F76, 67-R441 et 67-A149), 2 fragments chromosomiques différents ont été retrouvés au site de jonction, suggérant un réarrangement plus complexe.

Parmi les événements de NHCR intrachromosomiques, les sites impliqués sont tous situés près du site de cassure du plasmide intégré en 2p22.1, le plus éloigné étant situé en 2p25.1 à 25,6 Mb (67-t28). Il est à noter que l'événement 67-nt13 a été généré spontanément dans des cellules n'ayant pas subi de CDB. Ces essais de recombinaison spontanée ont été effectués dans le but de comparer les

réarrangements impliqués en dehors d'un contexte de réparation d'une CDB induite. Toutefois, les limites de la technique nous ont permis de ne récupérer qu'un seul événement de NHCR spontané.



**Figure 19 : Séquençage des événements de réparation de CDB par NHCR**

Les sites de jonction chromosomiques des événements de NHCR récupérés avec le clone ATA34 (a) et ATA67 (b) ont été séquencés avec les amorces 1 (HDF4683), 2 (HDR6q14A) et 3 (TK4415R).

| Événement<br>NHCR | Déletion<br>(pb)* | Insertion<br>(pb)* | Site de jonction de NHCR** |       |           |        |       |
|-------------------|-------------------|--------------------|----------------------------|-------|-----------|--------|-------|
|                   |                   |                    | Chr.                       | (Mb)  | Contig    | (kb)   | Nom   |
| 34-H300           | 503               | 0                  | 20p11.1                    | 26,2  | NT_011387 | 26 197 | 34-t1 |
| 34-H281           | 0                 | 0                  | 1q24.3                     | 178,9 | NT_004487 | 6 746  | 34-t2 |
| 34-B876           | 0                 | 0                  | 1q24.3                     | 178,9 | NT_004487 | 6 745  |       |
| 34-B315           | 0                 | 0                  | 3q12.2                     | 107,6 | NT_029256 | 1 514  | 34-t3 |
| 34-B693           | 0                 | 0                  | 6q24.1                     | 141,1 | NT_025741 | 18 133 | 34-t4 |
| 34-A542           | 1                 | 130                | 6q15                       | 85,9  | NT_007299 | 2 639  | 34-t5 |

**Tableau V : Description des événements de réparation par NHCR suite à une CDB au site I-Scel du clone ATA34**

\*La structure du site de cassure réparé a été analysée par le séquençage des recombinants avec l'amorce HDF4683. L'événement 34-H300 impliquant une grande délétion a été séquencé par l'amorce HDF4260.

\*\* Les sites de jonction des événements de réarrangement ont été séquencés avec l'amorce HDR6q14A et ont été localisés sur les chromosomes et leurs contigs par le programme *BLAST human genome* de GenBank.

**Tableau VI : Description des événements de réparation par NHCR suite à une CDB au site I-SceI du clone ATA67**

\* La structure du site de cassure réparé a été principalement analysée par le séquençage des recombinants avec l'amorce HDF4683. Les événements impliquant la délétion de la position de l'amorce ont été séquencés à tâtons à l'aide d'amorces plus éloignées du site de cassure.

\*\* Les sites de jonction des événements de réarrangement ont été séquencés avec l'amorce TK4415R et ont été localisés sur les chromosomes et leurs contigs par le programme *BLAST human genome* de GenBank.

<sup>a</sup> Le recombinant 67-nt13 a été généré spontanément dans des cellules où une CDB n'a pas été induite par le pCBASceTcr.

<sup>b</sup> Dans certains cas, 2 fragments chromosomiques différents ont été séquencés au site de NHCR.

| Événement<br>NHCR    | Déletion<br>(pb)* | Insertion<br>(pb)* | Site de jonction de NHCR** |       |           |        |        |
|----------------------|-------------------|--------------------|----------------------------|-------|-----------|--------|--------|
|                      |                   |                    | Chr.                       | (Mb)  | Contig    | (kb)   | Nom    |
| 67-A167              | 1926              | 0                  | 8q12.1                     | 63,2  | NT_008218 | 1 972  | 67-t1  |
| 67-F255              | 419               | 0                  | 13q21.33                   | 72,6  | NT_024524 | 21 749 | 67-t2  |
| 67-F375              | 4                 | 0                  | 11q14.2                    | 93,0  | NT_008984 | 1 073  | 67-t3  |
| 67-A2 <sup>b</sup>   | 320               | 8                  | 17q21.1                    | 37,4  | NT_010647 | 1 204  | 67-t4  |
|                      |                   |                    | 9q21.33                    | 78,8  | NT_023935 | 7 095  | 67-t5  |
| 67-F402              | n.d.              | 0                  | 9q33.3                     | 113,6 | NT_017568 | 11 921 | 67-t6  |
| 67-K9                | 7                 | 0                  | Yp11.2                     | 9,0   | NT_011878 | 1 010  | 67-t7  |
| 67-F557              | 34                | 0                  | 7q22.1                     | 99,3  | NT_007933 | 26 176 | 67-t8  |
| 67-R440              | 573               | 0                  | 6q25.1                     | 150,3 | NT_023451 | 787    | 67-t9  |
| 67-F241              | 13                | 0                  | 1q31.3                     | 199,0 | NT_034408 | 24     | 67-t10 |
| 67-F76 <sup>b</sup>  | 4                 | 0                  | 21q22.11                   | 30,6  | NT_011512 | 19 567 | 67-t11 |
|                      |                   |                    | 1p12                       | 115,0 | NT_019273 | 4 046  | 67-t12 |
| 67-R1146             | 3                 | 22                 | 14q21.3                    | 45,6  | NT_025892 | 28 666 | 67-t13 |
| 67-A162              | 2                 | 9                  | 14q23.2                    | 62,7  | NT_025892 | 45 690 | 67-t14 |
| 67-F506              | 7                 | 0                  | 2p22.1                     | 39,4  | NT_005367 | 6 028  | 67-t15 |
| 67-F267              | 1118              | 0                  | 2p23.2                     | 29,4  | NT_005204 | 6 418  | 67-t16 |
| 67-R441 <sup>b</sup> | 12                | 0                  | 2p23.3                     | 29,0  | NT_005204 | 6 029  | 67-t17 |
|                      |                   |                    | 1q25.2                     | 183,2 | NT_004487 | 10 782 | 67-t18 |
| 67-nt13 <sup>a</sup> | 699               | 0                  | 2p23.3                     | 28,7  | NT_005204 | 5 670  | 67-t19 |
| 67-F490              | 150               | 0                  | 2p24.1                     | 25,2  | NT_005204 | 2 182  | 67-t20 |
| 67-F346              | 1112              | 0                  | 2p24.1                     | 23,3  | NT_005204 | 339    | 67-t21 |
| 67-R1115             | 24                | 0                  | 2p22.1                     | 40,3  | NT_034482 | 120    | 67-t22 |
| 67-F416              | 0                 | 0                  | 2p22.1                     | 40,4  | NT_034482 | 181    | 67-t23 |
| 67-A149 <sup>b</sup> | 20                | 0                  | 2p22.1                     | 42,5  | NT_034483 | 235    | 67-t24 |
|                      |                   |                    | 2p22.1                     | 40,4  | NT_034482 | 184    | 67-t25 |
| 67-F263              | 6                 | 0                  | 2p21                       | 45,8  | NT_034483 | 3 553  | 67-t26 |
| 67-F282              | 5                 | 0                  | 2p24.3                     | 15,7  | NT_005334 | 10 679 | 67-t27 |
| 67-F259              | 2                 | 0                  | 2p25.1                     | 14,8  | NT_005334 | 9 760  | 67-t28 |

### **III.6. Analyse des événements de recombinaison homologue (RH)**

Les événements de réparation par RH ont été séquencés dans chaque direction à partir du site de CDB jusqu'à l'atteinte de la fin du segment de conversion, c'est-à-dire jusqu'à ce que l'on cesse d'observer des différences avec la séquence du pL1HsScel. La séquence visible (minimale) du segment de conversion a ensuite été recherchée dans l'ensemble des L1s répertoriés dans la base de données GenBank. Pour un segment minimal de conversion suffisamment long, un partenaire précis a pu se démarquer parmi tous les L1s du génome. Étant donné que la séquence du génome varie d'un individu à un autre et d'un allèle à l'autre, il est normal que les L1s de la base de données GenBank soient parfois différents de ceux de la lignée GM05849. Par conséquent, il est normal que le résultat du BLAST n'ait pas toujours été d'identité 100%.

#### **III.6.1. Analyse des événements de RH du clone ATA34**

Avec le clone ATA34, nous avons pu identifier le partenaire L1 endogène de 15 des 19 événements de conversion génique récupérés dont la longueur minimale du segment de conversion variait de 23 à 576 pb, pour une moyenne de 109 pb (tableau VII, pp.88-89). Parmi les 9 partenaires différents utilisés (34-A à 34-I), 8 ont été localisés sur le même bras chromosomique que le site d'intégration du vecteur en 6q. Le partenaire interchromosomique 34-A a été localisé en 4q24, et a été utilisé 3 fois. Les partenaires intrachromosomiques 34-B, C et E semblent aussi avoir été privilégiés, ayant également été utilisés plus d'une fois. Les L1s utilisés comme partenaires appartiennent aux familles récentes<sup>205</sup> (L1Hs, L1PA2-A4), partagent une homologie de 94 à 98% avec le cd11 et leur taille varie de 347 pb à une longueur complète de 5976 pb. À la figure 20 (p.90), l'alignement des segments de conversion de chaque événement nous a permis de vérifier que les partenaires étaient tous différents les uns des autres, portant chacun un patron distinct de différences lorsque comparés au LINE présent dans le vecteur.

### Tableau VII : Partenaires de RH identifiés pour le clone ATA34

<sup>a</sup>Les segments de conversion ont été principalement séquencés avec les amorces HDF4383 et HDR4885. Des amorces supplémentaires plus éloignées ont été utilisés pour les plus longues conversions. Le nombre de bases converties dans le plasmide recombinant est accompagné de la directionnalité de la conversion par rapport au site de cassure et l'étendue minimale et maximale a été déterminée par l'alignement des séquences du partenaire endogène et du cd11.

<sup>b</sup>Les partenaires L1 endogènes ont été identifiés par un BLAST du segment minimal de conversion dans GenBank.

<sup>c</sup>La taille complète (5976 pb) d'un L1 correspond à celle du cd11.

<sup>d</sup>Le % d'homologie du L1 partenaire a été déterminé par l'alignement de la séquence chevauchant celle du cd11.

\*Les événements 34-194 et 34-518 peuvent avoir impliqué à probabilité égale 2 partenaires différents tous deux situés sur le chromosome 6.

\*\*Les événements 34-C32, 34-C9 et 34-D359 ont été récupérés par une technique sélective d'hybridation des colonies par une sonde-oligonucléotide radioactive correspondant à la portion de 23 pb caractéristique des événements de RH<sup>204</sup>.

Les événements de RH suivants n'ont pu être identifiés :

34-B14 (2 conversions en 5' de la CDB)

34-A329, 34-A536, 34-B533 (aucune modification autre que le segment de 23 pb *NheI*-*Bam*HI)

| No. RH             | Conversion <sup>a</sup> |                    | Résultat<br>Blasf <sup>b</sup> | Identification du partenaire L1 endogène |                   |                             |                         |         |       |
|--------------------|-------------------------|--------------------|--------------------------------|--|-------------------|-----------------------------|-------------------------|---------|-------|
|                    | n pb<br>(sens)          | Étendue<br>min-max |                                | Localisation                             |                   | Taille <sup>c</sup><br>(pb) | Homo.<br>% <sup>d</sup> | Famille | Nom   |
|                    |                         |                    |                                | Contig (position kb)                     | Chr (position Mb) |                             |                         |         |       |
| 34-C32**<br>34-407 | 2 (3')                  | 81-234             | 1 x 100% (1 x chr4)            | NT_006383 (2 416-2 422)                  | 4q24 (105)        | complet                     | 97%                     | n.d.    | 34-A* |
| 34-985             | 4 (3')                  | 142-254            | 2 x 99% (chr4, 3)              |  |                   |                             |                         |         |       |
| 34-B250            | 5 (5'-3')               | 245-345            | 1 x 100%                       | NT_025741 (18 581-18587)                 | 6q24.1 (142)      | complet                     | 98%                     | LIPA2   | 34-B  |
| 34-B370            | 3 (5')                  | 118-211            | 2 x 100% (chr6, X)             |  |                   |                             |                         |         |       |
| 34-B19<br>34-121   | 1 (5')                  | 23-89              | 12 x 100% (1 x chr6)           | NT_025741 (17 891-17 892)                | 6q23.3 (141)      | 354pb                       | 96%                     | LIPA3   | 34-C  |
| 34-238             | 3 (3')                  | 51-124             | 1 x 100%                       |  |                   |                             |                         |         |       |
| 34-194             | 1 (3')                  | 37-53              | 13 x 100% (2 x chr6)           | NT_025741 (19 108-19 111)                | 6q24.1 (142)      | 3283pb                      | 94%                     | LIPA4   | 34-D* |
| 34-C9**            | 1 (5')                  | 69-165             | 9 x 100% (1 x chr6)            | NT_007592 (13 413-13 414)                | 6p22.3 (23)       | 347pb                       | 94%                     | LIPA4   |       |
| 34-D359**          |                         |                    |                                | NT_025741 (5 973-5975)                   | 6q22.33 (129)     | 1725pb                      | 97%                     | LIPA3   | 34-E  |
| 34-A693            | 2 (5')                  | 35-99              | 1 à 100%                       | NT_007299 (6 076-6 078)                  | 6q15 (89)         | 2147pb                      | 95%                     | LIPA4   | 34-F  |
| 34-518             | 2 (3')                  | 51-86              | 35 x 100% (2 x chr6)           | NT_007422 (2 189-2 195)                  | 6q26 (160)        | complet                     | 98%                     | LIPA2   | 34-G* |
|                    |                         |                    |                                | NT_034880 (4,8-6,7)                      | 6p24.3 (7)        | 1795pb                      | 97%                     | LIPA2   |       |
| 34-B905            | 4 (5')                  | 35-143             | 1 à 100%                       | NT_025741 (17 177-17 179)                | 6q23.3 (140)      | 1875pb                      | 95%                     | LIPA3   | 34-H  |
| 34-151             | 10 (3')                 | 576-749            | 18 x 99% (1 x chr6)            | NT_033944 (9 876-9 882)                  | 6q22.1 (113)      | complet                     | 98%                     | L1Hs    | 34-I  |

| Position                 | 4685 | 4715 | 4728 | 4734 | 4766 | 4768 | 4773 | 4780 | 4783 | 4787 | 4790 | 4797 | 4817 | 4819 | 4826 | 4833 | 4852 | 4854 | 4860 | 4863 | 4865 | 4877 | 4894 | 4924 | 4973 | 5069 | 5121 | 5174 | 5226 | 5235 | 5358 |  |  |  |
|--------------------------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|--|--|--|
| L1 endogène <sup>a</sup> |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |  |  |  |
| L1 cdl1 <sup>b</sup>     |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |  |  |  |
| 34-B                     |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |  |  |  |
| 34-E                     |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |  |  |  |
| 34-H                     |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |  |  |  |
| 34-F                     |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |  |  |  |
| 34-D                     |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |  |  |  |
| 34-G                     |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |  |  |  |
| 34-C                     |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |  |  |  |
| 34-A                     |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |  |  |  |
| 34-I                     |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |  |  |  |
| 34-B14                   |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |  |  |  |

**Figure 20 : Classement des événements de RH du clone ATA34 par alignement des segments de conversion**

Dans la grille, les positions répertoriées correspondent à toutes les bases du L1 cdl1 qui ont été converties par celle d'un partenaire L1 endogène de RH dans l'un ou l'autre des événements récupérés. Afin de simplifier la figure, les bases nucléotidiques (A,C,G,T) n'ont pas été représentées dans les cases. <sup>a</sup>Les cases ombragées correspondent aux bases modifiées suite à une conversion par un partenaire L1 endogène. Par exemple, à la position 4783 du cdl1 on retrouve un « A », alors que dans le segment de conversion on retrouve un « C » appartenant au partenaire 34-H. <sup>b</sup>Les cases blanches correspondent aux bases du L1 cdl1 du vecteur avant la conversion. Les lignes verticales délimitent les conversions impliquées pour le remplacement du site I-Scel par la portion de 23 pb *MheI-BamHI* (4781 à 4803).

### III.6.2. Analyse des événements de RH du clone ATA67.

#### III.6.2.1. Identification des partenaires par la méthode du BLAST.

La meilleure efficacité d'induction de CDBs dans le clone ATA67 nous a permis de récupérer un total de 96 événements de RH. Dans un premier temps, la recherche d'homologie d'un segment de conversion minimal moyen de 124 pb par le programme de recherche *BLAST human genome* de GenBank a mené à l'identification d'un partenaire particulier pour 36 événements (tableaux VIII et IX, pp.92-95). Ces partenaires font partie des familles récentes de L1 (L1Hs, L1PA2-A6)<sup>207</sup>, partagent une homologie de 92 à 98% avec le cd11 du pL1HsSceI et leur taille varie de 608 pb à une longueur complète de 5976 pb.

Parmi les 12 partenaires identifiés, 6 ont été localisés sur le même bras chromosomique que le locus d'intégration du pL1HsSceI en 2p (tableau VIII, pp.92-93). Les partenaires 67-A, B, C et D ont chacun été impliqués dans plusieurs événements. Hautement préférentiel, le partenaire 67-A aurait été impliqué dans 13 événements. Les 6 autres partenaires identifiés ont été localisés sur d'autres chromosomes (3, 4 et 11) et l'un d'entre eux a été utilisé dans un événement de *crossing-over* (67A10-232, tableau IX, pp.94-95). Les partenaires 67-H et 67-F sont tous deux situés sur le bras 3q, à 41 Mb l'un de l'autre. Les partenaires 67-I et 67-K sont également localisés à proximité sur le chromosome 11p, séparés par seulement 6 Mb. Ces sites pourraient donc avoir une accessibilité particulière à la CDB.

Un nombre de 24 événements récupérés avec le clone ATA67 ne présentaient aucune conversion en dehors de la réacquisition du segment de 23 pb *NheI-BamHI*. Une tentative d'identification par BLAST était alors inutile. S'ajoutant à ce nombre, 35 événements ont généré un segment de conversion visible, mais de longueur et de spécificité insuffisante pour être associé à un L1 endogène particulier par le programme de BLAST (tableau X, pp.96-97).

**Tableau VIII: Partenaires intrachromosomiques de RH identifiés par la méthode du BLAST pour le clone ATA67**

<sup>a</sup>Les segments de conversion ont été principalement séquencés avec les amorces HDF4383 et HDR4885. Des amorces supplémentaires plus éloignées ont été utilisées pour les plus longues conversions. Le nombre de bases converties dans le plasmide recombinant est accompagné de la directionnalité de la conversion par rapport au site de cassure et l'étendue minimale et maximale a été déterminée par l'alignement des séquences du partenaire endogène et du cd11.

<sup>b</sup>Les partenaires L1 endogènes ont été identifiés par un BLAST du segment minimal de conversion dans GenBank.

<sup>c</sup>La taille complète (5976 pb) d'un L1 correspond à celle du L1 cd11.

<sup>d</sup>Le % d'homologie du L1 partenaire a été déterminé par l'alignement de la séquence chevauchant celle du L1 cd11.

<sup>e</sup>Les événements 67-R101 et 67-nt8 sont issus d'un mécanisme de *crossing-over* pour lequel l'étendue 3' du segment de conversion est inconnue (n.a.=non-applicable).

<sup>f</sup>L'événement 67-nt8 a été généré spontanément dans des cellules où une CDB n'a pas été induite par le pCBASceTcr.

| No. RH               | Conversion <sup>A</sup> |                 | Résultat Blast <sup>B</sup> | Identification du partenaire L1 endogène |                   |                          |                      |         |      |
|----------------------|-------------------------|-----------------|-----------------------------|--|-------------------|--------------------------|----------------------|---------|------|
|                      | n pb (gens)             | Étendue min-max |                             | Localisation                             |                   | Taille <sup>C</sup> (pb) | Homo. % <sup>D</sup> | Famille | Nom  |
|                      |                         |                 |                             | Contig (position kb)                     | Chr (position Mb) |                          |                      |         |      |
| 67-F471              | 6 (5)                   | 81-131          | 1 x 100%                    | NT_005367 (6843-6845)                    | 2p22.1 (40.2)     | 1995                     | 94%                  | LIP4    | 67-A |
| 67-F229              | 8 (3)                   | 141-189         | 1 x 100%                    |  |                   |                          |                      |         |      |
| 67-F493              |                         |                 |                             |  |                   |                          |                      |         |      |
| 67-F542              |                         |                 |                             |  |                   |                          |                      |         |      |
| 67A10-25             | 6 (3)                   | 94-149          | 1 x 100%                    |  |                   |                          |                      |         |      |
| 67A10-58             |                         |                 |                             |  |                   |                          |                      |         |      |
| 67A10-92             | 4 (3)                   | 51-105          | 1 x 100%                    |  |                   |                          |                      |         |      |
| 67A10-265            | 14 (3)                  | 338-416         | 1 x 99%                     |  |                   |                          |                      |         |      |
| 67A10-552            | 12 (5)                  | 158-231         | 1 x 100%                    |  |                   |                          |                      |         |      |
| 67A10-638            | 3 (5)                   | 42-79           | 1 x 100%                    |  |                   |                          |                      |         |      |
| 67A10-718            | 2                       | 23-66           | 1 x 100%                    |  |                   |                          |                      |         |      |
| 67A10-735            | 5 (5)                   | 81-131          | 2 x 99% (1 x chr2)          |  |                   |                          |                      |         |      |
| 67-F244              | 9 (3)                   | 212-278         | 1 x 98%                     | NT_034482 (23-29)                        | 2p22.1 (40.2)     | complet                  | 97%                  | LIP2    | 67-B |
| 67-R71               | 9 (5-3)                 | 200-271         | 1 x 98%                     |  |                   |                          |                      |         |      |
| 67-R101 <sup>E</sup> | n.a.                    | n.a.            | 1 x 97%                     |  |                   |                          |                      |         |      |
| 67A10-365            | 7 (3)                   | 140-189         | 1 x 98%                     |  |                   |                          |                      |         |      |
| 67A10-608            | 14 (5-3)                | 294-366         | 5 x 97% (2 x chr2)          |  |                   |                          |                      |         |      |
| 67A10-V304           | 7 (5-3)                 | 143-207         | 1 x 97%                     |  |                   |                          |                      |         |      |
| 67-R295              | 7 (5)                   | 68-121          | 1 x 100%                    | NT_034482 (39-42)                        | 2p22.1 (40.2)     | 3570                     | 94%                  | LIP5    | 67-C |
| 67A10-164            | 4 (3)                   | 77-124          | 1 x 100%                    |  |                   |                          |                      |         |      |
| 67A10-482            | 24 (5-3)                | 215-246         | 1 x 99%                     |  |                   |                          |                      |         |      |
| 67-F325              | 4 (3)                   | 191-343         | 1 x 100%                    | NT_005204 (6237-6240)                    | 2p24.1 (23.0)     | 2893                     | 97%                  | LIP3    | 67-D |
| 67-F480              |                         |                 |                             |  |                   |                          |                      |         |      |
| 67-F290              | 5 (5-3)                 | 181-383         | 14 x 100% (1 x chr2)        |  |                   |                          |                      |         |      |
| 67A10-396            | 3 (3)                   | 142-256         | 1 x 100%                    |  |                   |                          |                      |         |      |
| 67A10-614            |                         |                 |                             |  |                   |                          |                      |         |      |
| 67-47                | 6 (3)                   | 103-130         | 1 x 98%                     | NT_005367 (6854-6856)                    | 2p22.1 (40.2)     | 2017                     | 95%                  | LIP3    | 67-E |
| 67-nl8 <sup>EF</sup> | n.a.                    | n.a.            | 1 x 99%                     | NT_005375 (6350-6356)                    | 2p16.1 (57.5)     | 5744                     | 93%                  | LIP6    | 67-M |

**Tableau IX: Partenaires interchromosomiques de RH identifiés par la méthode du BLAST pour le clone ATA67**

<sup>a</sup>Les segments de conversion ont été principalement séquencés avec les amorces HDF4383 et HDR4885. Des amorces supplémentaires plus éloignées ont été utilisées pour les plus longues conversions. Le nombre de bases converties dans le plasmide recombinant est accompagné de la directionnalité de la conversion par rapport au site de cassure et l'étendue minimale et maximale a été déterminée par l'alignement des séquences du partenaire endogène et du cd11.

<sup>b</sup>Les partenaires L1 endogènes ont été identifiés par un BLAST du segment minimal de conversion dans GenBank.

<sup>c</sup>La taille complète (5976 pb) d'un L1 correspond à celle du L1 cd11.

<sup>d</sup>Le % d'homologie du L1 partenaire a été déterminé par l'alignement de la séquence chevauchant celle du L1 cd11.

<sup>e</sup>L'événement 67A10-232 est issu d'un mécanisme de *crossing-over* pour lequel l'étendue 3' du segment de conversion est inconnue (n.a.=non-applicable).

| No. RH                 | Conversion <sup>A</sup> |                    | Résultat<br>Blast <sup>B</sup> | Identification du partenaire L1 endogène |               |                             |                         |         |       |
|------------------------|-------------------------|--------------------|--------------------------------|--|---------------|-----------------------------|-------------------------|---------|-------|
|                        | n pb<br>(sens)          | Étendue<br>min-max |                                | Localisation                             |               | Taille <sup>C</sup><br>(pb) | Homo.<br>% <sup>D</sup> | Famille | Nom   |
|                        |                         |                    | Contig (position kb)           | Chr (position Mb)                        |               |                             |                         |         |       |
| 67A10-232 <sup>E</sup> | n.a.                    | n.a.               | 1 x 99%                        | NT_025667 (2262-2259)                    | 3q26.32 (174) | 2672                        | 94%                     | 67-H    | L1PA4 |
| 67-205<br>67-F309      | 2 (5'-3')               | 44-74              | 2 x 100%                       | NT_005567 (540-545)                      | 3q22 (133)    | 5188                        | 93%                     | 67-F    | L1PA5 |
| 67-F330                | 2 (5')                  | 24-51              | 1 x 100%                       | NT_009151 (3175-3169)                    | 11q22.1 (103) | complet                     | 92%                     | 67-G    | L1PA6 |
| 67A10-358              | 3 (3')                  | 77-165             | 1 x 99%                        | NT_006344 (1 002-1 008)                  | 4p15.33 (14)  | complet                     | 97%                     | 67-J    | L1PA2 |
| 67A10-356              | 5 (5'-3')               | 76-184             | 1 x 100%                       | NT_009237 (5637-5639)                    | 11p13 (37)    | 1310                        | 95%                     | 67-I    | L1PA3 |
| 67A10-453              | 1 (3')                  | 95-146             | 2 x 100%                       | NT_009237 (595-596)                      | 11p14.1 (31)  | 608                         | 98%                     | 67-K    | L1P   |

**Tableau X :Événements de RH du clone ATA67 impliquant des partenaires endogènes non-identifiables par la méthode du BLAST**

<sup>a</sup>Les segments de conversion ont été principalement séquencés avec les amorces HDF4383 et HDR4885. Des amorces supplémentaires plus éloignées ont été utilisées pour les plus longues conversions. Le nombre de bases converties dans le plasmide recombinant est accompagné de la directionnalité de la conversion par rapport au site de cassure et l'étendue minimale et maximale a été déterminée par l'alignement des séquences du partenaire endogène et du cd11.

<sup>b</sup>Le BLAST du segment de conversion a généré un résultat selon lequel un grand nombre de partenaires endogènes sont possibles. Ces événements ne peuvent donc significativement être affectés à un partenaire particulier du génome.

Les événements de RH suivants n'impliquent que le retour des 23 pb *NheI-BamHI* sans conversion supplémentaire :

67-7, 67-K17,

67-R453, R1153,

67-F236, F279, F306, F310, F508, F510, F515,

67A10-41, 131, 137, 211, 214, 240, 315, 361, 377, 499, 566, 669 et 744.

| No. RH   | Conversion <sup>A</sup>  |                     | Résultat<br>du BLAST <sup>B</sup>     |
|--|--------------------------|---------------------|---------------------------------------|
|  | n pb<br>(sens)           | Étendue<br>min (pb) |                                       |
| 67-90<br>67-K4<br>67A10-3<br>67A10-57<br>67A10-75<br>67A10-185<br>67A10-524<br>67A10-687 | 1                        | 23                  | >100 x 100%                           |
| 67-F17<br>67A10-316<br>67A10-688   | 1 (5')                   | 89                  | >100 x 100%                           |
| 67-F125<br>67A10-194<br>67-F274<br>67A10-742   | 2 (5')                   | 51                  | >100 x 98%                            |
| 67-R334<br>67A10-710<br>67-F301  | 3 (5')<br>5 (5'-3')<br>1 | 125<br>155<br>23    | >100 x 100%<br>30 x 100%<br>46 x 100% |
| 67A10-91<br>67A10-V299   | 3 (5')                   | 89                  | 25 x 99%                              |
| 67A10-295  | 1                        | 23                  | 67 x 100%                             |
| 67A10-512  | 1 (3')                   | 44                  | 30 X 100%                             |
| 67A10-117  | 1 (5')                   | 59                  | 60 x 100%                             |
| 67-F47   | 1 (3')                   | 79                  | >100 x 100%                           |
| 67A10-22   | 1 (5')                   | 46                  | 3 x 100% (chrX,13,16)                 |
| 67A10-V37  | 1                        | 23                  | >100 x 100%                           |
| 67-R1471   | 2 (5')                   | 30                  | 3 x 100% (chr5,5,4)                   |
| 67A10-144  | 2 (5')                   | 35                  | 67 x 98%                              |
| 67-R23   | 3 (5'-3')                | 74                  | 65 x 97%                              |
| 67A10-225  | 3 (3')                   | 194                 | 16 x 99%                              |
| 67-F408  | 4 (5'-3')                | 112                 | 32 x 98%                              |
| 67A10-398  | 4 (5')                   | 55                  | 3 x 97% (chr5,13,19)                  |
| 67A10-602  | 4 (5'-3')                | 109                 | 50 x 97%                              |
| 67A10-679  | 4 (5'-3')                | 91                  | 31 x 97%                              |
| 67A10-V220   | 9 (5'-3')                | 177                 | 13 x 98%                              |

### *III.6.2.2. Identification des partenaires par alignement des segments de conversion*

Puisque la méthode du BLAST ne permettait l'identification que du tiers des partenaires, une deuxième méthode consistant en l'alignement des segments de conversion de tous les événements a permis de faire certains regroupements et de procéder à un classement plus probable (figure 21, pp.99-101). Cet alignement nous a d'abord permis de constater que les partenaires 67-J et 67-F identifiés par BLAST comme étant les plus probables pouvaient aussi correspondre au partenaire 67-B. En effet, leur segment de conversion est parfaitement identique à celui du partenaire 67-B sur la portion 4781 à 4861. Dix-huit événements supplémentaires non-identifiés par la méthode de BLAST pourraient également être reliés à ce partenaire, portant ainsi le nombre d'événements ayant possiblement impliqué le partenaire 67-B à 27.

De la même façon, le nombre d'événements ayant impliqué le partenaire 67-D pourrait être augmenté à 12 par l'ajout de 7 événements ayant des conversions communes. Puisque les premières conversions rencontrées avec ce partenaire sont situées à environ 60 pb d'un côté ou de l'autre de la portion de 23 pb, il est possible qu'un plus grand nombre encore d'événements non identifiés soient reliés à ce partenaire, incluant les 24 événements sans modification de base. Le partenaire 67-C pourrait avoir été impliqué dans l'événement 67A10-V37 dans lequel une seule conversion est observée dans la portion du 23 pb. Finalement, 3 événements peuvent être regroupés ensemble par leurs conversions communes (groupe1, figure 21b, pp.99-101) mais leur partenaire demeure non-identifié. Six événements demeurent sans identification ni regroupement, ayant chacun des conversion uniques.

### **Figure 21 : Classement des événements de RH du clone ATA67 par alignement des segments de conversion**

Dans la grille, les positions répertoriées correspondent à toutes les bases du L1 cd11 qui ont été converties par celle d'un partenaire L1 endogène de RH dans l'un ou l'autre des événements récupérés. Afin de simplifier la figure, les bases nucléotidiques (A,C,G,T) n'ont pas été représentées dans les cases.

<sup>a</sup>Les cases ombragées correspondent aux bases modifiées suite à une conversion par un partenaire L1 endogène. Par exemple, à la position 4740 du cd11 on retrouve un « G », alors que dans le segment de conversion on retrouve un « A » appartenant au partenaire 67-A.

<sup>b</sup>Les cases blanches correspondent aux bases du L1 cd11 du vecteur avant la conversion.

- a) Les partenaires identifiés par la méthode du BLAST (67-A à 67-M) sont représentés sur l'étendue de toutes les conversions présentes dans les recombinants et sont encadrés. Certains partenaires non-identifiables par la méthode du BLAST (événements en italique) pourraient correspondre à des partenaires déjà identifiés dans d'autres événements (67-C et 67-D). Étant donné le chevauchement de leur conversion (section ombragée), les partenaires 67-J et 67-F identifiés par BLAST pourraient aussi correspondre au partenaire 67-B.
- b) Certains événements non-identifiés par la méthode du BLAST peuvent être regroupés (groupe 1) et pourraient impliquer un même partenaire inconnu. Six événements demeurent uniques et sans identification de partenaire.



b)

| Position | LI endogène <sup>a</sup> | LI cd11 <sup>b</sup> |
|----------|--------------------------|----------------------|
|          |                          | 4685                 |
|          |                          | 4686                 |
|          |                          | 4688                 |
|          |                          | 4696                 |
|          |                          | 4706                 |
|          |                          | 4715                 |
|          |                          | 4719                 |
|          |                          | 4721                 |
|          |                          | 4729                 |
|          |                          | 4735                 |
|          |                          | 4740                 |
|          |                          | 4745                 |
|          |                          | 4749                 |
|          |                          | 4751                 |
|          |                          | 4752                 |
|          |                          | 4753                 |
|          |                          | 4756                 |
|          |                          | 4759                 |
|          |                          | 4763                 |
|          |                          | 4766                 |
|          |                          | 4769                 |
|          |                          | 4781                 |
|          |                          | 4782                 |
|          |                          | 4783                 |
|          |                          | 4784                 |
|          |                          | 4787                 |
|          |                          | 4795                 |
|          |                          | 4801                 |
|          |                          | 4803                 |
|          |                          | 4808                 |
|          |                          | 4815                 |
|          |                          | 4817                 |
|          |                          | 4820                 |
|          |                          | 4821                 |
|          |                          | 4826                 |
|          |                          | 4833                 |
|          |                          | 4834                 |
|          |                          | 4839                 |
|          |                          | 4842                 |
|          |                          | 4843                 |
|          |                          | 4853                 |
|          |                          | 4860                 |
|          |                          | 4861                 |

| non-classés |
|-------------|
| 67A10-117   |
| 67A10-22    |
| 67-R1471    |
| 67-F301     |
| 67A10-512   |
| 67A10-V220  |

| groupe I  |
|-----------|
| 67A10-398 |
| 67A10-144 |
| 67A10-295 |

### ***III.6.2.3. Description détaillée des événements ayant impliqué le même partenaire dans le clone ATA67***

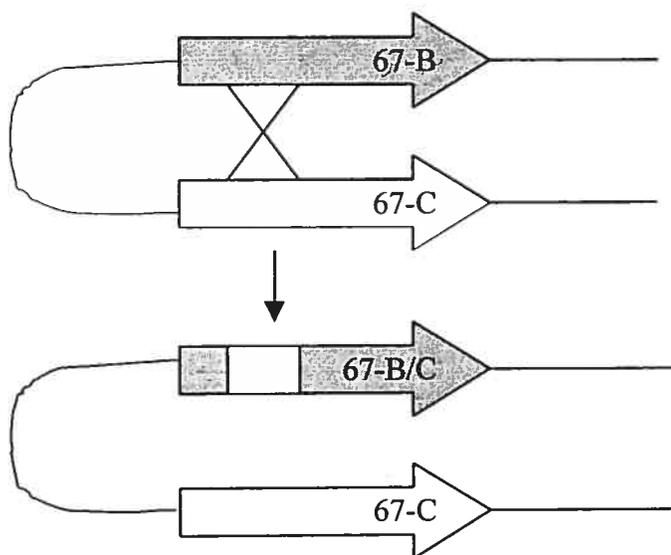
Le partenaire 67-A pourrait avoir été impliqué dans un total de 13 événements de RH, laissant sa trace par 9 segments de conversion différents alignés à la figure 23a (p.105). Par l'événement 67A10-265, nous avons observé que la conversion en position 5040 présente dans le partenaire identifié par GenBank était absente du partenaire de notre lignée cellulaire, confirmant la présence de variation allélique entre les LINEs de différents individus. La conversion située en position 4795 du L1 a été observée dans tous les recombinants, sauf dans l'événement 67A10-735. Cette différence est probablement le résultat d'une réparation de l'hétéroduplex de recombinaison en faveur de la molécule brisée, par le système de réparation des *mismatches*<sup>176</sup>.

En figure 23b (p.106), nous avons aligné les 15 segments de conversion des 27 événements ayant pu impliquer le partenaire 67-B. Encore une fois, 3 conversions (4725, 4783 et 4994) prédites par GenBank pour ce partenaire n'ont pas été observées et sont donc des variations alléliques. Les événements 67A10-679 et 67A10-608 s'alignent bien avec le segment de conversion du partenaire, mais possèdent des conversions supplémentaires non-observées chez les autres événements et le partenaire GenBank. Il se pourrait donc que ces événements aient impliqué un autre partenaire dérivé d'une conversion avec le 67-B.

Le partenaire 67-B est particulièrement intéressant, puisque l'un des événements l'ayant impliqué s'est produit par un mécanisme de *crossing-over* (67-R101, figure 23b, p.106). La comparaison du segment de conversion avec le partenaire de GenBank nous a permis de constater que la portion 5' de cette conversion ne s'alignait pas, plusieurs différences étant observées entre le partenaire réel dans la lignée GM05849 et le partenaire de GenBank. En fait, le BLAST de cette seule portion (4326 à 4569) a résulté par une homologie presque parfaite avec le partenaire 67-C sur cette portion (99%). Il serait donc probable que dans les cellules GM05849,

le partenaire 67-B ait acquis un segment du partenaire 67-C lors d'une conversion génique antérieure à l'événement de CDB (figure 22, p.103). Puisque aucun autre événement de RH ne s'est produit avec cette portion du partenaire 67-B et qu'il nous a été impossible d'amplifier cet élément pour le séquencer, cette explication la plus probable demeure tout de même hypothétique.

Pour le partenaire 67-C, 4 segments de conversion différents peuvent être alignés (figure 23c, p.107) et 2 variations alléliques sont observées en début de conversion, aux positions 4588 et 4590. Finalement, l'alignement des événements de conversion avec le partenaire 67-D est présenté en figure 23d (p.108). Les recombinants 67A10-225 et 67-R334 ont une conversion en moins pouvant une fois de plus être expliquée par la réparation d'un *mismatch* en faveur de la molécule à réparer. Finalement, le recombinant 67A10-710 en a une supplémentaire en position 4833 qui pourrait nous faire douter de l'identification du partenaire pour cet événement.



**Figure 22 : Hypothèse d'une conversion génique spontanée entre les partenaires L1 endogènes proximaux 67-B et 67-C dans la lignée cellulaire GM05849**

**Figure 23 : Alignement des segments de conversion des événements de RH impliquant le même partenaire 67-A(a), 67-B(b), 67-C(c) et 67-D(d)**

<sup>a</sup>Séquence du L1 endogène dans la base de données GenBank.

<sup>b</sup>Séquence du L1 cd11 avant conversion dans le pL1HsSceI.

<sup>c</sup>Étendue minimale et maximale du segment de conversion.

- La conversion prédite par GenBank n'est pas observée dans la lignée GM05849 (variation allélique).
- La conversion observée dans la lignée GM05849 est prédite par GenBank.
- La conversion observée dans la lignée GM05849 n'a pas été prédite par GenBank.

\*Ces bases de l'hétéroduplex ont été réparées par le système de réparation des *mismatches* en faveur de la molécule brisée.

\*\* Des conversions supplémentaires sont observées pour certains événements.

Les positions marquées d'un trait ( \_ ) correspondent à des délétions.

Les lignes verticales délimitent les conversions nécessairement impliquées pour le remplacement du site I-SceI par la portion de 23 pb *NheI-BamHI* (4781 à 4803).

a)

| Position  | LI GenBank <sup>a</sup> | LI cd11 <sup>b</sup> | Étendue conversion min-max <sup>c</sup> |  |
|-----------|-------------------------|----------------------|---|--|
| 4577      | C                       | A                    | 158 - 231                               |  |
| 4588      | C                       | T                    | 80 - 131                                |  |
| 4637      | A                       | G                    | 80 - 131                                |  |
| 4667      | T                       | C                    | 80 - 131                                |  |
| 4672      | A                       | G                    | 80 - 131                                |  |
| 4683      | A                       | C                    | 80 - 131                                |  |
| 4685      | A                       | A                    | 80 - 131                                |  |
| 4688      | G                       | A                    | 80 - 131                                |  |
| 4715      | T                       | C                    | 80 - 131                                |  |
| 4719      | C                       | T                    | 80 - 131                                |  |
| 4740      | C                       | G                    | 80 - 131                                |  |
| 4753      | A                       | A                    | 80 - 131                                |  |
| 4783      | C                       | A                    | 80 - 131                                |  |
| 4795      | A                       | C                    | 80 - 131                                |  |
| 4821      | G                       | T                    | 80 - 131                                |  |
| 4833      | T                       | C                    | 80 - 131                                |  |
| 4860      | T                       | C                    | 80 - 131                                |  |
| 4877      | T                       | T                    | 80 - 131                                |  |
| 4904      | A                       | T                    | 80 - 131                                |  |
| 4924      | A                       | G                    | 80 - 131                                |  |
| 4944      | T                       | C                    | 80 - 131                                |  |
| 4945      | G                       | A                    | 80 - 131                                |  |
| 5040      | G                       | A                    | 80 - 131                                |  |
| 5059      | G                       | A                    | 80 - 131                                |  |
| 5060      | A                       | C                    | 80 - 131                                |  |
| 5069      | A                       | C                    | 80 - 131                                |  |
| 5121      | A                       | T                    | 80 - 131                                |  |
| 5171      | C                       | G                    | 80 - 131                                |  |
| 5174      | T                       | C                    | 80 - 131                                |  |
| 67A10-552 |                         |                      | 158 - 231                               |  |
| 67-F471   |                         |                      | 80 - 131                                |  |
| 67A10-486 |                         |                      | 80 - 131                                |  |
| 67A10-735 |                         | C                    | 80 - 131                                |  |
| 67A10-638 |                         | *                    | 42 - 79                                 |  |
| 67A10-718 |                         |                      | 23 - 66                                 |  |
| 67A10-92  |                         |                      | 50 - 105                                |  |
| 67A10-25  |                         |                      | 94 - 149                                |  |
| 67A10-58  |                         |                      | 141 - 189                               |  |
| 67-F229   |                         |                      | 141 - 189                               |  |
| 67-F493   |                         |                      | 141 - 189                               |  |
| 67-F542   |                         |                      | 141 - 189                               |  |
| 67A10-265 |                         | A                    | 338 - 416                               |  |

b)

| Position                | 4326 | 4334 | 4335 | 4375 | 4384 | 4393 | 4425 | 4449 | 4450 | 4456 | 4482 | 4483 | 4524 | 4525 | 4562 | 4569 | 4588 | 4626 | 4650 | 4683 | 4685 | 4688 | 4715 | 4725 | 4753 | 4783 | 4784 | 4826 | 4861 | 4879 | 4883 | 4896 | 4924 | 4944 | 4994 | 4996 | 5033 | 5060 |     |           |           |
|-------------------------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|-----|-----------|-----------|
| L1 GN05849 <sup>a</sup> | C    | A    | A    | A    | A    | A    | T    | C    | T    | A    | C    | C    | C    | A    | G    | G    | C    | A    | T    | A    | A    | T    | C    | C    | C    | A    | C    | G    | A    | C    | A    | C    | A    | T    | A    | T    | C    | A    | A   |           |           |
| L1 GenBank <sup>a</sup> | G    | A    | A    | G    | C    | A    | G    | T    | C    | T    | A    | A    | C    | C    | G    | A    | C    | A    | T    | A    | A    | T    | T    | G    | C    | C    | C    | G    | G    | C    | C    | T    | A    | T    | C    | C    | C    | A    | A   |           |           |
| L1 cd11 <sup>b</sup>    | G    | A    | A    | G    | A    | G    | G    | T    | C    | A    | A    | A    | A    | T    | G    | A    | A    | T    | G    | A    | G    | T    | C    | T    | C    | A    | A    | T    | C    | G    | A    | C    | C    | G    | C    | T    | A    | C    | C   |           |           |
| 67-R101                 | C    | A    | A    | A    | A    | A    | T    | C    | T    | A    | C    | C    | A    | G    | G    | C    | A    | T    | A    | A    | T    | C    | C    | C    | C    | A    | C    | G    | A    | C    | A    | T    | A    | T    | T    | C    | A    | A    | ... |           |           |
| 67-R71                  |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      | A    | T    | A    | A    | T    | C    | C    | C    | A    | C    | G    |      |      |      |      |      |      |      |      |      |     | 200 - 271 |           |
| 67A10-91, 67-V299       |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |     |           | 88 - 136  |
| 67-F408                 |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |     |           | 111 - 171 |
| 67-F125 et al. (4x)     |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |     |           | 50 - 99   |
| 67-R23                  |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |     |           | 73 - 134  |
| 67A10-679               |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |     |           | 90 - 134  |
| 67A10-602               |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |     |           | 108 - 125 |
| 67-90 et al. (8x)       |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |     |           | 23 - 71   |
| 67-205, 67-F309         |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |     |           | 44 - 107  |
| 67A10-358               |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |     |           | 79 - 124  |
| 67A10-608               |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |     |           | 294-366   |
| 67A10-V304              |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |     |           | 143 - 207 |
| 67A10-365               |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |     |           | 142 - 189 |
| 67-F244                 |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |     |           | 214 - 278 |

(conv. suppl. 4719) \*\*

\*\* (conv. suppl. 4843)

| Étendue conversion min-max <sup>c</sup> |
|---|
| c.o.                                    |
| 200 - 271                               |
| 88 - 136                                |
| 111 - 171                               |
| 50 - 99                                 |
| 73 - 134                                |
| 90 - 134                                |
| 108 - 125                               |
| 23 - 71                                 |
| 44 - 107                                |
| 79 - 124                                |
| 294-366                                 |
| 143 - 207                               |
| 142 - 189                               |
| 214 - 278                               |

c)

| Position  | L1 GenBank <sup>a</sup> | L1 cd11 <sup>b</sup> | Étendue conversion min-max <sup>c</sup> |
|-----------|-------------------------|----------------------|---|
| 4525      | A                       | G                    |   |
| 4569      | G                       | A                    |   |
| 4588      | T                       | T                    |   |
| 4590      | G                       | C                    |   |
| 4638      | G                       | A                    |   |
| 4647      | C                       | T                    |   |
| 4650      | G                       | A                    |   |
| 4658      | G                       | A                    |   |
| 4659      | A                       | A                    |   |
| 4660      | A                       | A                    |   |
| 4662      | C                       | A                    |   |
| 4663      | A                       | A                    |   |
| 4671      | C                       | A                    |   |
| 4679      | T                       | C                    |   |
| 4680      | A                       | G                    |   |
| 4683      | A                       | G                    |   |
| 4685      | A                       | T                    |   |
| 4686      | C                       | T                    |   |
| 4688      | C                       | C                    |   |
| 4696      | C                       | T                    |   |
| 4715      | C                       | T                    |   |
| 4719      | C                       | G                    |   |
| 4721      | C                       | A                    |   |
| 4729      | A                       | C                    |   |
| 4740      | A                       | G                    |   |
| 4753      | C                       | A                    |   |
| 4783      | C                       | A                    |   |
| 4817      | T                       | C                    |   |
| 4833      | T                       | C                    |   |
| 4860      | T                       | C                    |   |
| 4877      | G                       | C                    |   |
| 4904      | A                       | T                    |   |
| 67A10-482 |                         |                      | 215 - 246                               |
| 67-R295   |                         |                      | 68 - 121                                |
| 67A10-V37 |                         |                      | 23 - 62                                 |
| 67A10-164 |                         |                      | 77 - 124                                |

d)

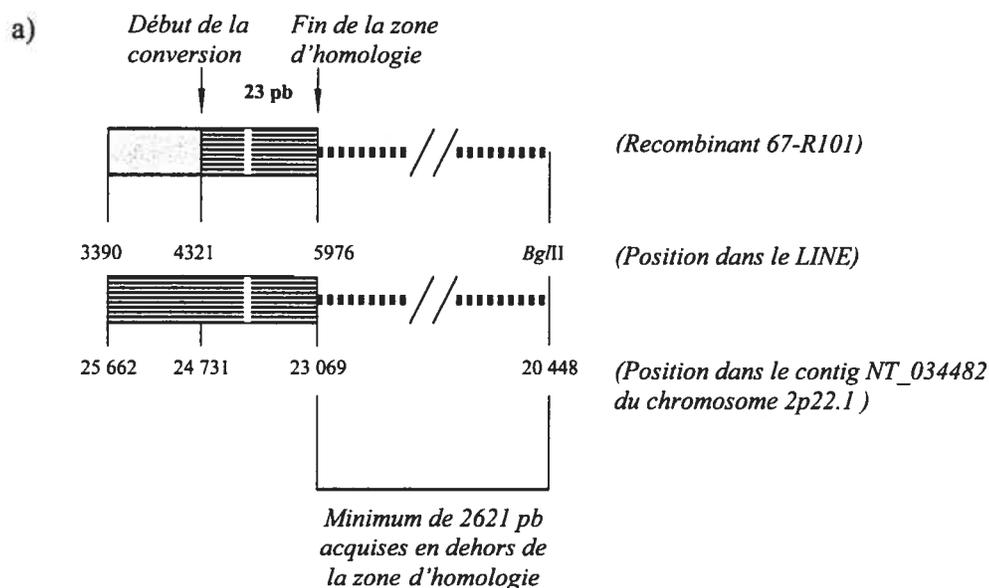
| Position             | Étendue conversion min-max <sup>c</sup> |   |   |   |   |   |   |   |   |   |   |   |   |   |           |
|----------------------|---|---|---|---|---|---|---|---|---|---|---|---|---|---|-----------|
|                      | LI GenBank <sup>d</sup>                 | A | G | T | A | T | C | T | C | T | C | A | G | C | C         |
| LI cd11 <sup>b</sup> | C                                       | T | C | T | C | T | C | T | C | A | G | C | C | C |           |
| 67-F290              |   |   |   |   |   |   |   |   |   |   |   |   |   |   | 181 - 383 |
| 67A10-710            |   |   |   |   |   |   |   |   |   |   |   |   |   |   | 148 - 179 |
| 67-R334              |   |   |   |   |   |   |   |   |   |   |   |   |   |   | 118 - 179 |
| 67-F17 et al. (3x)   |   |   |   |   |   |   |   |   |   |   |   |   |   |   | 88 - 170  |
| 67-F47               |   |   |   |   |   |   |   |   |   |   |   |   |   |   | 78 - 199  |
| 67A10-396            |   |   |   |   |   |   |   |   |   |   |   |   |   |   | 142 - 256 |
| 67A10-614            |   |   |   |   |   |   |   |   |   |   |   |   |   |   | 191 - 343 |
| 67-F325              |   |   |   |   |   |   |   |   |   |   |   |   |   |   | 191 - 343 |
| 67-F480              |   |   |   |   |   |   |   |   |   |   |   |   |   |   | 191 - 343 |
| 67A10-225            |   |   |   |   |   |   |   |   |   |   |   |   |   |   | 191 - 343 |

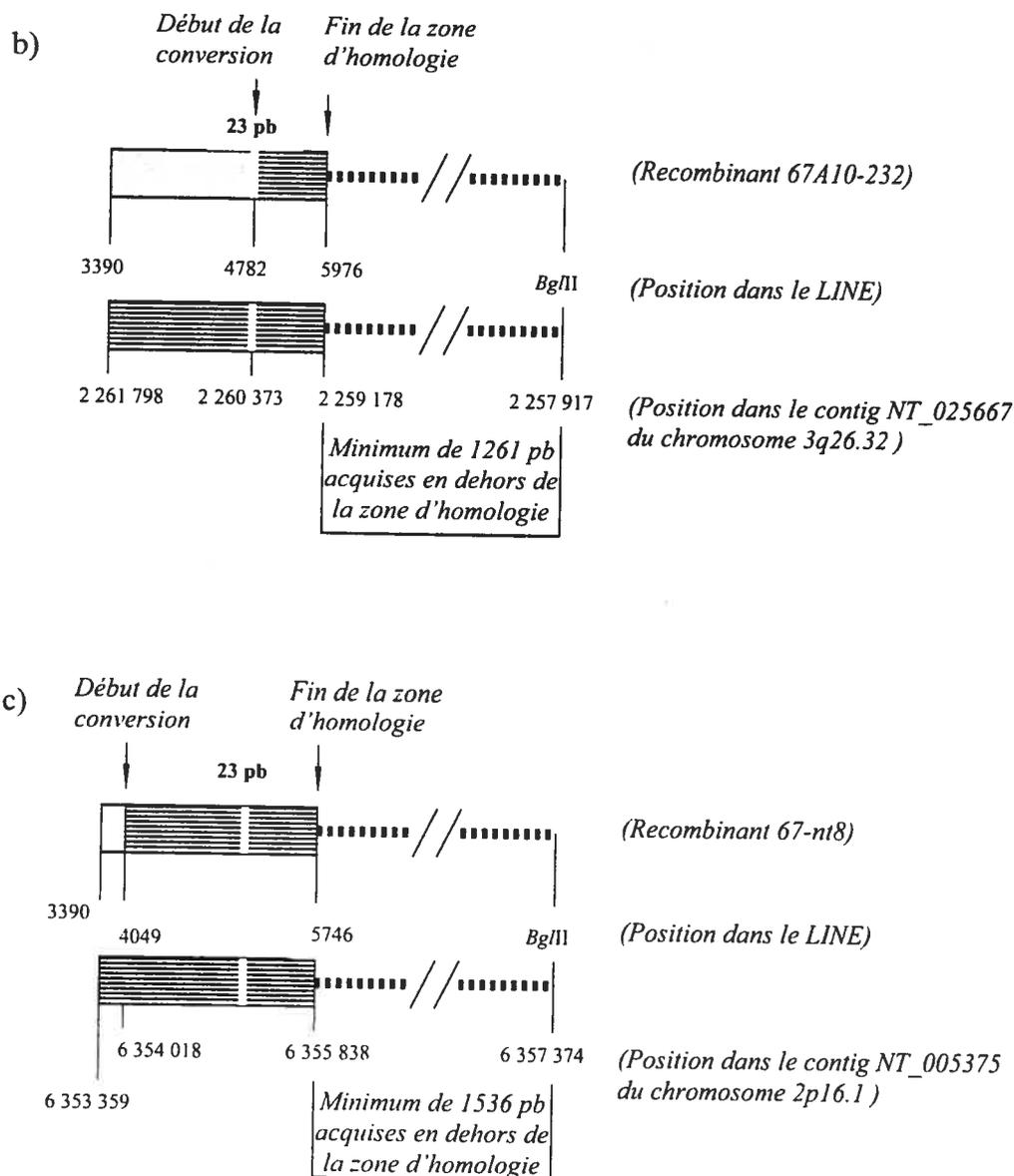
| LI cd11 <sup>b</sup> | Étendue conversion min-max <sup>c</sup> |
|----------------------|---|
| C                    | 191 - 343                               |

### III.6.2.4. Analyse des événements de *crossing-over*

Parmi les 96 événements de RH récupérés, nous avons observé 2 événements de *crossing-over* impliquant l'acquisition de séquences chromosomiques non-LINE. L'événement 67-R101 s'est produit avec le partenaire intrachromosomique 67-B, qui pourrait avoir été impliqué dans près d'un événement sur 3. La conversion a débuté 460 pb en 5' du site de cassure et s'est étendue au-delà de la région d'homologie après la fin du partenaire L1, sur au moins 2,6 kb (figure 24a, pp.109-110). Le second événement de *crossing-over* (67A10-232) s'est produit avec le partenaire interchromosomique 67-H situé sur le locus 3q26.32, qui n'a été relié qu'à cet unique événement. La conversion ne s'étend qu'en 3' du site de cassure et au moins 1260 pb de séquence non-homologue ont été acquises (figure 24b, p.110).

Un troisième événement de *crossing-over* (67-nt8) a été récupéré et analysé, bien qu'il ait été généré spontanément. Puisque générée sans induction de CDB au site I-*SceI*, la conversion a débuté par hasard dans le LINE, à une distance éloignée de 733 pb en 5' du site I-*SceI*. Un minimum de 1536 pb ont été acquises en dehors de la région d'homologie, au locus 2p16.1 du partenaire 67-M (figure 24c, p.110).





**Figure 24 : Structure détaillée des événements de RH par crossing-over**  
a) 67-R101, b) 67-232, c) 67-nt8

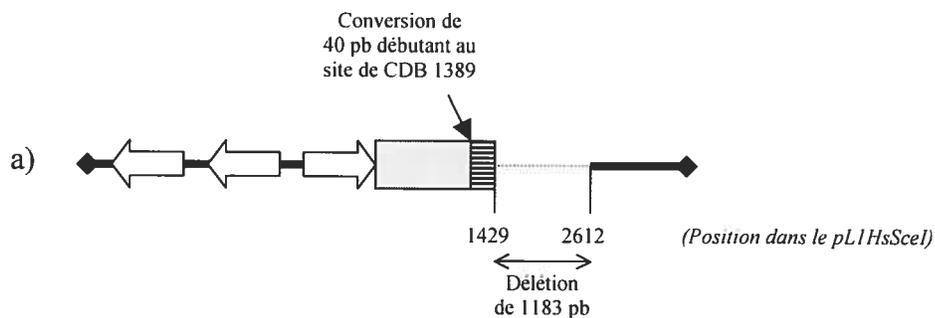
Le début et la fin de la conversion (zones lignées) dans la zone d'homologie sont indiqués par des flèches et sont positionnés dans le LINE et le contig. Les séquences acquises hors de la région d'homologie (pointillés) ont été identifiées jusqu'à la position du site de digestion *BgIII* utilisé pour la récupération des plasmides intégrés.

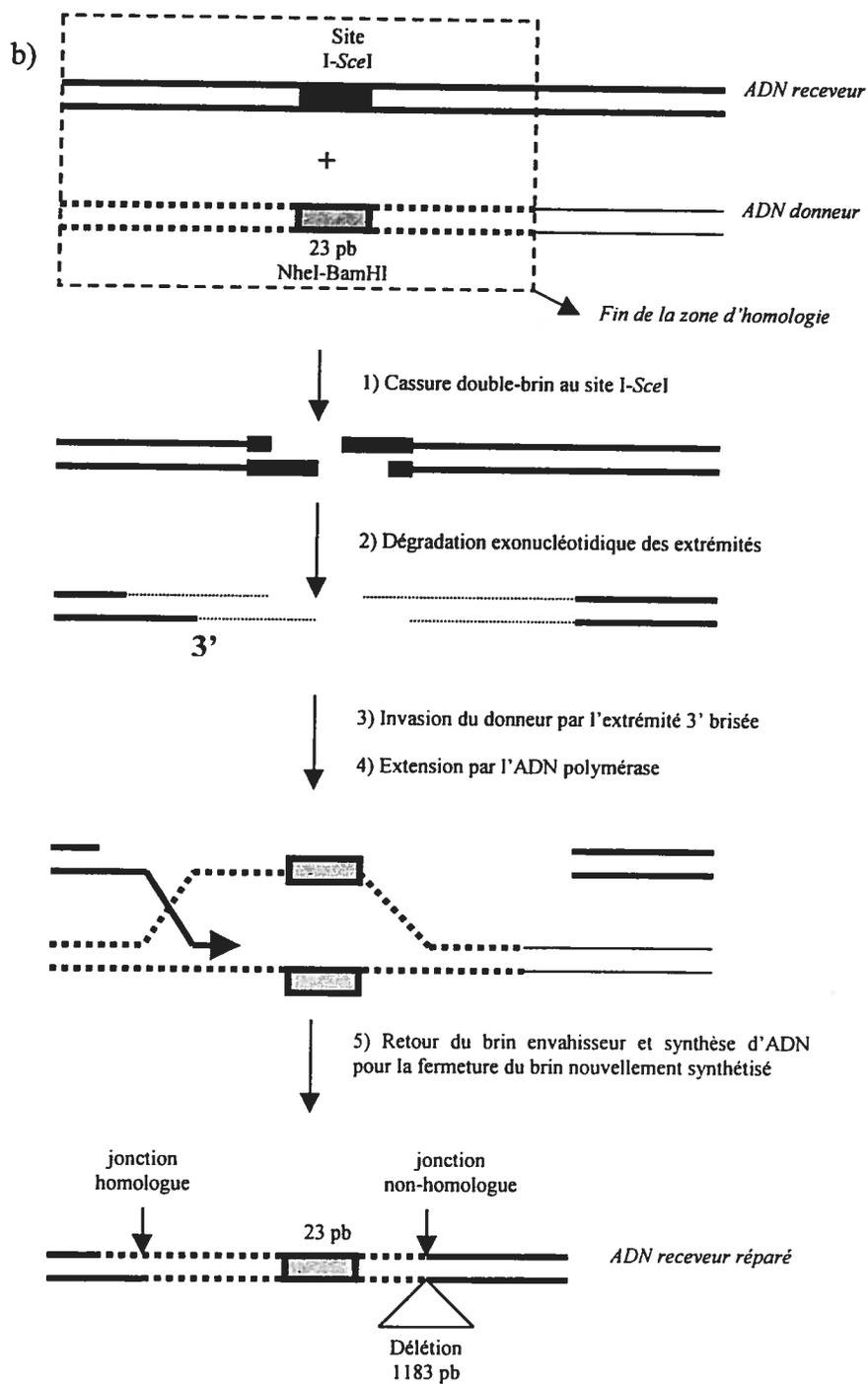
### III.6.2.5. Analyse de l'événement de *one-sided invasion* 67A10-240

Un événement de RH particulier a été récupéré avec le clone 67 (67A10-240). Il implique un segment de conversion de 40 pb à partir du site de cassure, puis une délétion de 1183 pb entraînant la perte de la fin du LINE avant le retour de la séquence plasmidique en 3' (figure 25a, p.111). Cette structure pourrait résulter d'un mécanisme d'invasion par une seule extrémité (*one-sided invasion*) antérieurement décrit par notre laboratoire<sup>119,120</sup> et décrit à la figure 25b (p.112). Il n'est pas exclu que les événements de *crossing-over* décrits à la section précédente puissent en réalité être des événements de *one-sided invasion* dans lesquels des séquences non-homologues seraient acquises et les séquences plasmidiques en 3' seraient perdues avant le retour du brin envahisseur dans le chromosome de l'intégration.

**Figure 25: Description de l'événement de RH 67A10-240**

- Structure du recombinant 67A10-240.
- Mécanisme proposé de RH par *one-sided invasion*.



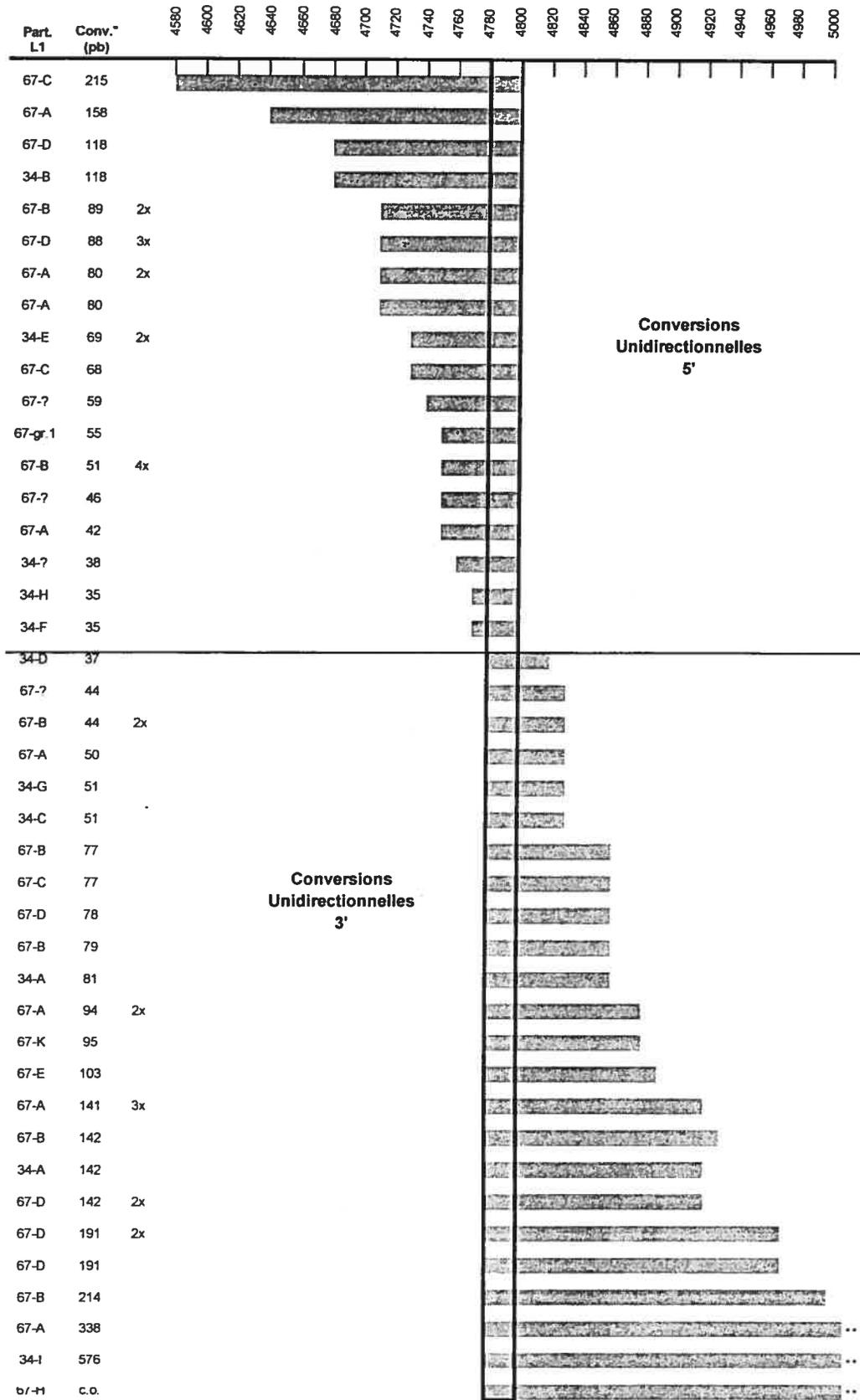


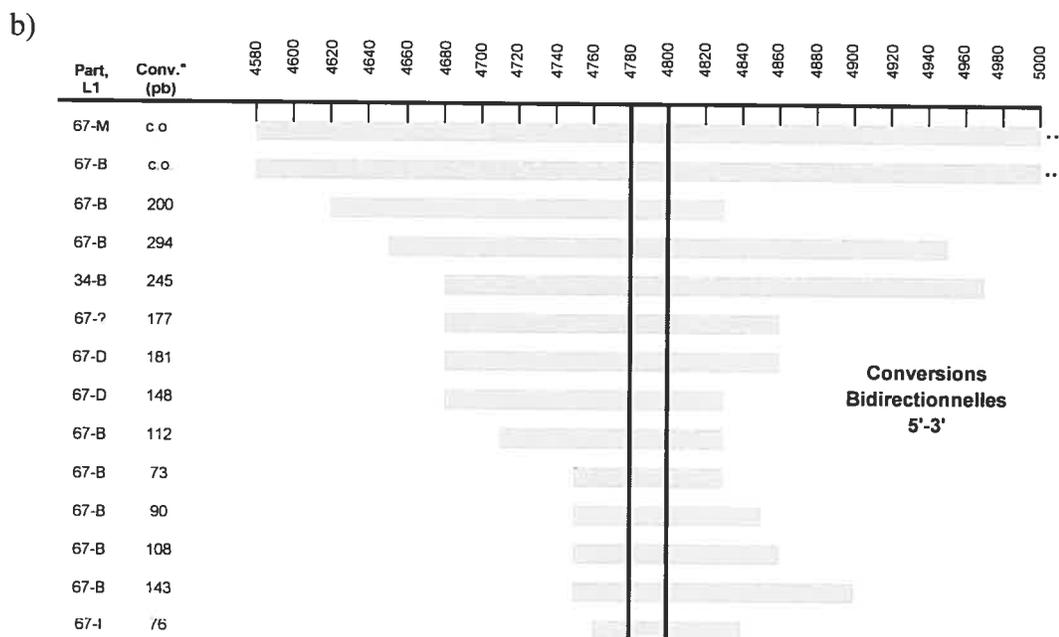
### III.6.3. Analyse des segments de conversion

Afin d'étudier le mécanisme de RH utilisé lors de la réparation d'une CDB par des séquences répétitives ectopiques, nous avons analysé le segment de conversion des événements de RH des clones ATA34 et ATA67 (figure 26, pp.114-115 et tableau XI, p.116). Un total de 43 événements de conversion génique (38%) n'ayant impliqué que le retour des 23 pb endogènes en remplacement du site *I-SceI* n'ont pu être analysés en fonction de leur directionnalité. Tous les autres événements ayant laissé une conversion de directionnalité visible ont été alignés à la figure 26 (pp.114-115) et classés en 3 groupes : unidirectionnalité en 5', unidirectionnalité en 3' et bidirectionnalité.

Quatorze événements (12%) ont laissé une conversion bidirectionnelle de longueur moyenne de 154 pb, 30 événements (27%) ont laissé une conversion unidirectionnelle en 3' de longueur moyenne de 135 pb et 26 événements (23%) ont laissé une conversion unidirectionnelle en 5' de longueur moyenne de 74 pb (tableau XI, p.116)). Globalement, l'ensemble des recombinants ont engendré une conversion moyenne de 79 pb majoritairement unidirectionnelle avec une légère préférence pour le côté 3'. Un mécanisme relié à ces observations sera proposé dans la discussion.

a)





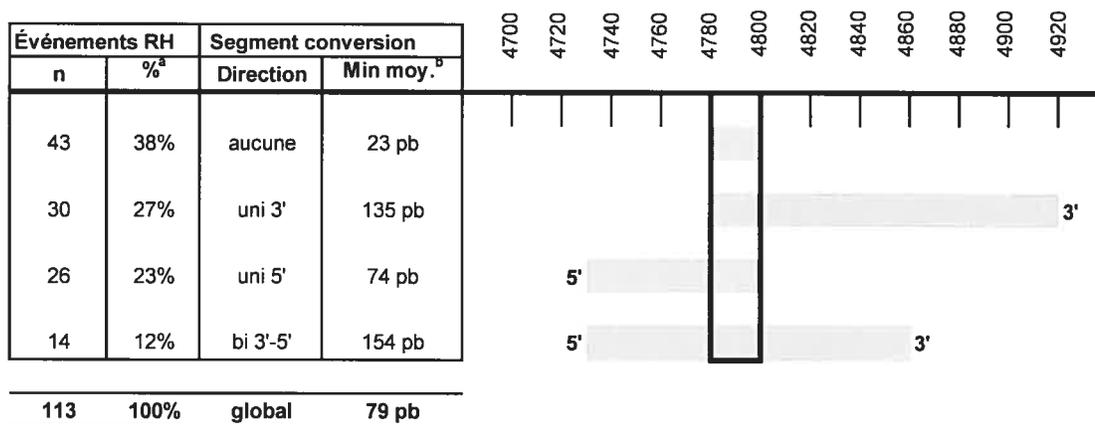
**Figure 26 : Alignement et classement des segments de conversion en fonction de leur longueur et de leur directionnalité**

**a) Événements unidirectionnels**

**b) Événements bidirectionnels**

<sup>a</sup>La longueur du segment de conversion indiquée correspond à la valeur minimale visible dans les recombinants. Les segments de conversion des événements de *crossing-over* sont représentés en partie.

La portion de conversion minimale de 23 pb est encadrée et correspond aux positions 4782 à 4803 de la séquence du L1 cd11. Les 16 événements de conversion génique pour lesquels un partenaire endogène a été identifié mais n'ayant impliqué que le remplacement de la portion 23 pb sans directionnalité visible ne sont pas représentés, de même que les 27 événements n'impliquant que le retour intégral des 23 pb endogènes sans aucune modification.



**Tableau XI : Proportion de l'unidirectionnalité et de la bidirectionnalité chez les événements de conversion génique**

<sup>a</sup>La proportion d'événements unidirectionnels et bidirectionnels est calculée sur le nombre total d'événements de RH ayant un segment de conversion visible attribué à un partenaire particulier. <sup>b</sup>Le calcul de la longueur minimale moyenne du segment de conversion exclue les événements de *crossing-over*.

Une représentation schématique de la moyenne des événements accompagne le tableau. La portion de conversion minimale de 23 pb est encadrée et correspond aux positions 4782 à 4803 de la séquence du L1 cd11.

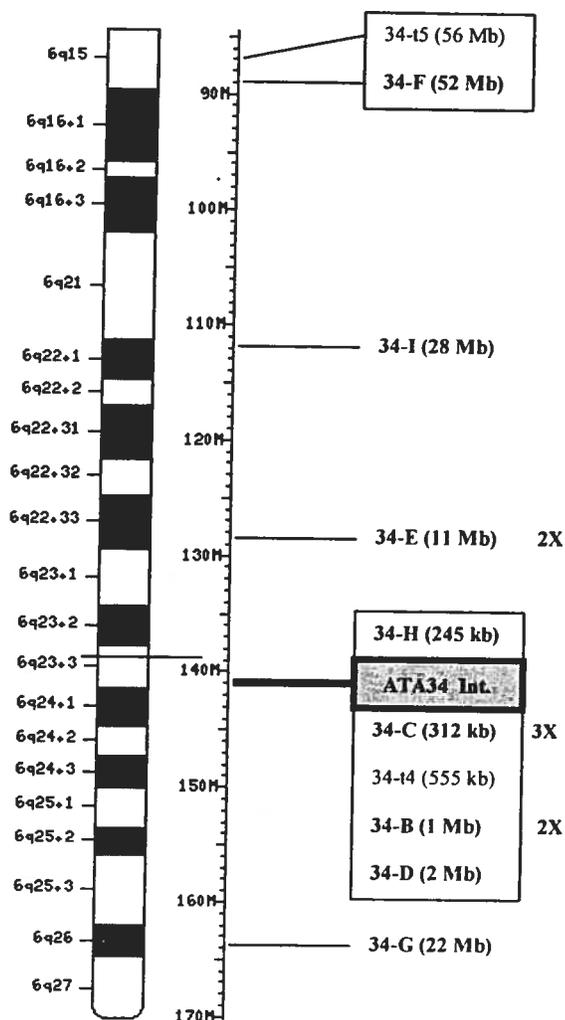
### **III.7. Environnement détaillé des locus d'intégration ATA34 et ATA67**

#### **III.7.1. Distribution des partenaires intrachromosomiques de RH et de NHCR du clone ATA34 sur le chromosome 6q**

Parmi les 9 partenaires de RH identifiés du clone ATA34, 8 sont localisés sur le même bras chromosomique que le locus d'intégration situé en 6q23.3 (figure 27, p.118). Cette portion chromosomique porte également 2 des 6 sites de NHCR. Le plus éloigné des partenaires de RH se situe à 52 Mb (34-F), tout près d'un site de NHCR (34-t5). La plupart des autres sont concentrés à une distance de moins de 2 Mb autour de l'intégration, le plus près étant situé à seulement 245 kb du vecteur intégré (34-H).

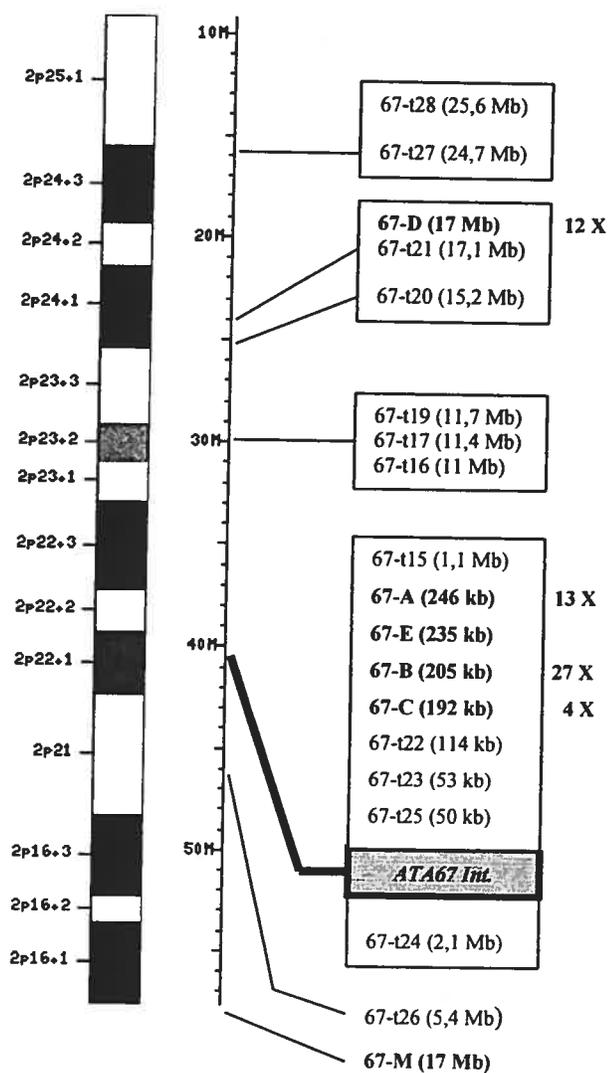
#### **III.7.2. Distribution des partenaires intrachromosomiques de RH et de NHCR du clone ATA67 sur le chromosome 2p**

Sur le bras chromosomique du locus d'intégration du clone ATA67 situé en 2p22.1, nous avons retrouvé 6 partenaires de RH et 13 sites de NHCR (figure 28, p.119). Tout comme pour le clone ATA34, la moitié des partenaires et sites sont retrouvés dans l'environnement immédiat du site d'intégration, à moins de 2 Mb. Les autres sont regroupés sur 3 régions vers la portion télomérique du chromosome. Des sites de regroupement, qui suggèrent une accessibilité préférentielle, sont retrouvés à 11 Mb, 17 Mb et 25 Mb. Le partenaire 67-D, impliqué dans 13 événements de RH, se retrouve au regroupement du 17 Mb. Finalement, un partenaire de RH spontanée (67-M) et un site de translocation (67-t26) sont retrouvés vers la portion centromérique du bras 2p.



**Figure 27: Localisation intrachromosomique des partenaires de RH et des sites de NHCR sur le chromosome 6q (clone ATA34)**

La distance de chaque partenaire de RH (34-A à 34-I) ou site de NHCR (34-t1 à 34-t5) par rapport au locus d'intégration est indiquée entre parenthèses. Si un partenaire est utilisé dans plus d'un événement, la fréquence est indiquée à sa droite.



**Figure 28: Localisation intrachromosomique des partenaires de RH et des sites de NHCR sur le chromosome 2p (clone ATA67)**

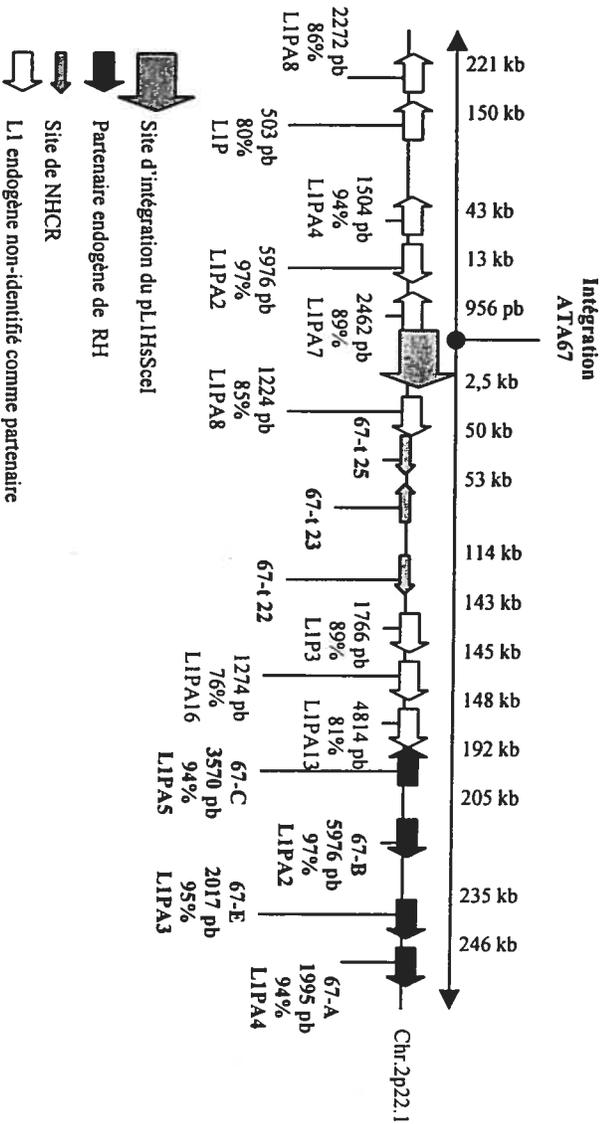
La distance de chaque partenaire de RH (67-A à 67-M) ou site de NHCR (67-t1 à 67-t28) par rapport au locus d'intégration est indiquée entre parenthèses. Si un partenaire est utilisé dans plus d'un événement, la fréquence est indiquée à sa droite.

### III.7.3. Environnement immédiat détaillé du locus ATA34

Puisqu'un nombre élevé de partenaires endogènes de RH ont été localisés très près des locus d'intégration, nous avons entrepris l'analyse détaillée du voisinage immédiat de chacun de ces sites. Sur une étendue de 2 Mb, la région du locus ATA34 regroupe 1 site de NHCR et 4 L1s partenaires de RH, le plus près étant situé à 245 kb de l'intégration (34-H, figure 29, p.121). Cependant, en dedans de 300 kb autour de l'intégration, nous avons observé la présence de 7 autres éléments L1s endogènes possédant également la portion de 23 pb nécessaire à la réparation de la cassure. Contrairement aux partenaires utilisés, la plupart de ces L1s appartenaient à de vieilles sous-familles très peu homologues au cd11. Par contre, un élément de 3694 pb à 93% homologue et un autre de 1280 pb à 97% homologue auraient également eu la possibilité de réparer la cassure. Pour s'assurer que les L1s de familles plus récentes n'étaient pas passés inaperçus dans nos recombinants par un court segment de conversion, nous avons vérifié que chacun des 7 L1s présentaient des conversions distinctives dans la portion du 23 pb réparée, qui aurait permis leur identification immédiate lors de leur utilisation éventuelle.

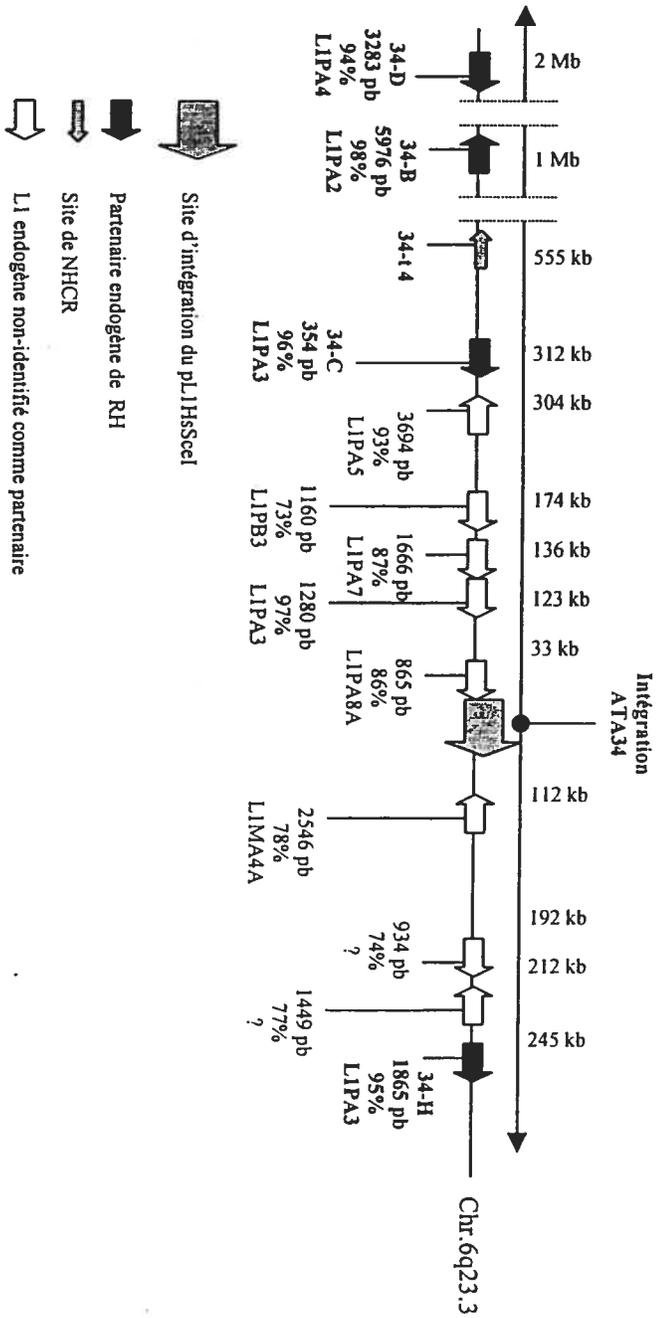
### III.7.4. Environnement immédiat détaillé du locus ATA67

Sur une distance de 250 kb en direction télomérique du locus ATA67, nous avons localisé 3 sites de NHCR et 4 partenaires endogènes de RH qui sont 4 voisins immédiats concentrés sur moins de 50 kb appartenant tous à des sous-familles récentes de L1 (figure 30, p.122). Les 8 autres éléments L1s de la région possédant la portion de 23 pb partagent pour la plupart moins de 90% d'homologie avec le cd11. Par contre, 2 de ces L1s appartiennent aux familles récentes L1PA2 et L1PA4. Ne présentant pas de conversion distinctives dans la portion du 23 pb, un très court segment de conversion aurait masqué leur utilisation. Par contre, puisque ces 2 éléments possèdent des conversions distinctives à partir de 34 pb de la cassure et que la longueur minimale moyenne des segments de conversion observée se situe vers



**Figure 30 : Environnement immédiat détaillé du locus ATA67 sur le chromosome 2p22.1**

Chaque L1 endogène de la région possédant minimalement la portion de 23 pb servant à réparer la cassure est positionné sur le locus avec sa description (taille, % homologie, famille L1). Les flèches indiquent l'orientation des L1s endogènes par rapport au cd11 intégré. Les sites de NHCR sont représentés sur le locus dans l'orientation relative à leur position dans les réarrangements avec le pLIHSSceI.



**Figure 29 : Environnement immédiat détaillé du locus ATA34 sur le chromosome 6q23.3**

Chaque L1 endogène de la région possédant minimalement la portion de 23 pb servant à réparer la cassure est positionné sur le locus avec sa description (taille, % homologie, famille L1). Les flèches indiquent l'orientation des L1s endogènes par rapport au *cd111* intégré. Les sites de NHCR sont représentés sur le locus dans l'orientation relative à leur position dans les réarrangements avec le *pl1HsScel*

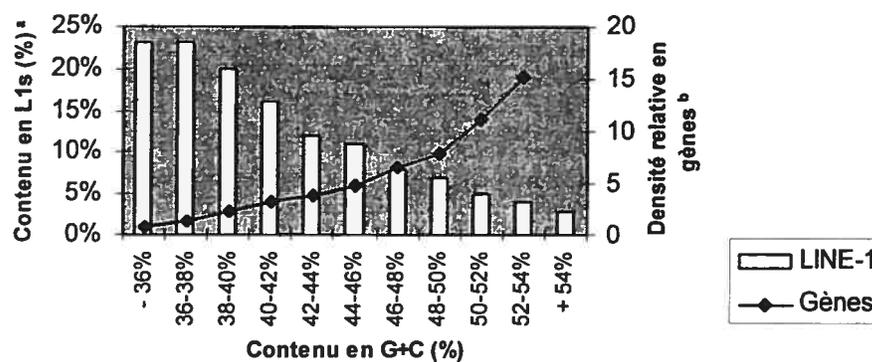
130 pb, l'utilisation de ces 2 L1s aurait probablement été retracée dans les recombinants.

### **III.8. Distribution des gènes et des séquences répétitives sur les chromosomes d'intégration des clones ATA34 et ATA67**

Dans le génome humain, les éléments L1s sont dispersés sur tous les chromosomes. Ils sont toutefois absents des régions centromériques et plus denses dans les bandes G/Q (Giemsa/Quinacrine) riches en A-T, pauvres en gènes et de réplication tardive<sup>2</sup>. Bien qu'en général la densité des L1s soit inversement proportionnelle à celle des gènes (figure 31, p.124), certaines régions du génome sont particulièrement plus denses ou plus pauvres en répétitions<sup>2,208</sup>.

Même si les éléments répétitifs L1s sont généralement absents des régions où se retrouvent les gènes, ils peuvent quand même se retrouver dans des gènes et leurs séquences régulatrices. Ainsi, des événements de RH non-conservatifs impliquant des L1s situés à l'intérieur de gènes peuvent mener à des délétions et des translocations inactivant la fonction d'un gène<sup>53,54</sup>.

Au tableau XII (p.125), nous avons identifié les partenaires L1 endogènes situés dans la séquence d'un ARNm relié à un gène connu ou prédit par GenBank. Sept des 9 partenaires identifiés du clone ATA34 pourraient être situés dans des régions codantes, ainsi que 4 des 10 partenaires identifiés pour le clone ATA67. Puisque la plupart des événements de RH ont impliqué un mécanisme de conversion génique, la fonction de ces gènes ne devrait pas être affectée. Par contre, nous avons récupéré quelques événements de *one-sided invasion* et de *crossing-over*, des mécanismes pouvant impliquer des réarrangements majeurs. L'événement de RH spontané 67-nt8 a impliqué le partenaire endogène 67-M dans un mécanisme de *crossing-over* qui pourrait avoir affecté la fonction d'un gène encore inconnu.



**Figure 31 : Distribution relative des L1s et des gènes dans le génome**

<sup>a</sup>Proportion du génome occupée par des éléments L1s en fonction du contenu en G+C de cette région. <sup>b</sup>Densité relative des gènes en fonction du contenu en G+C du génome (densité relative = proportion de gènes/ proportion du génome).

Les données ont été tirées de la référence 2.

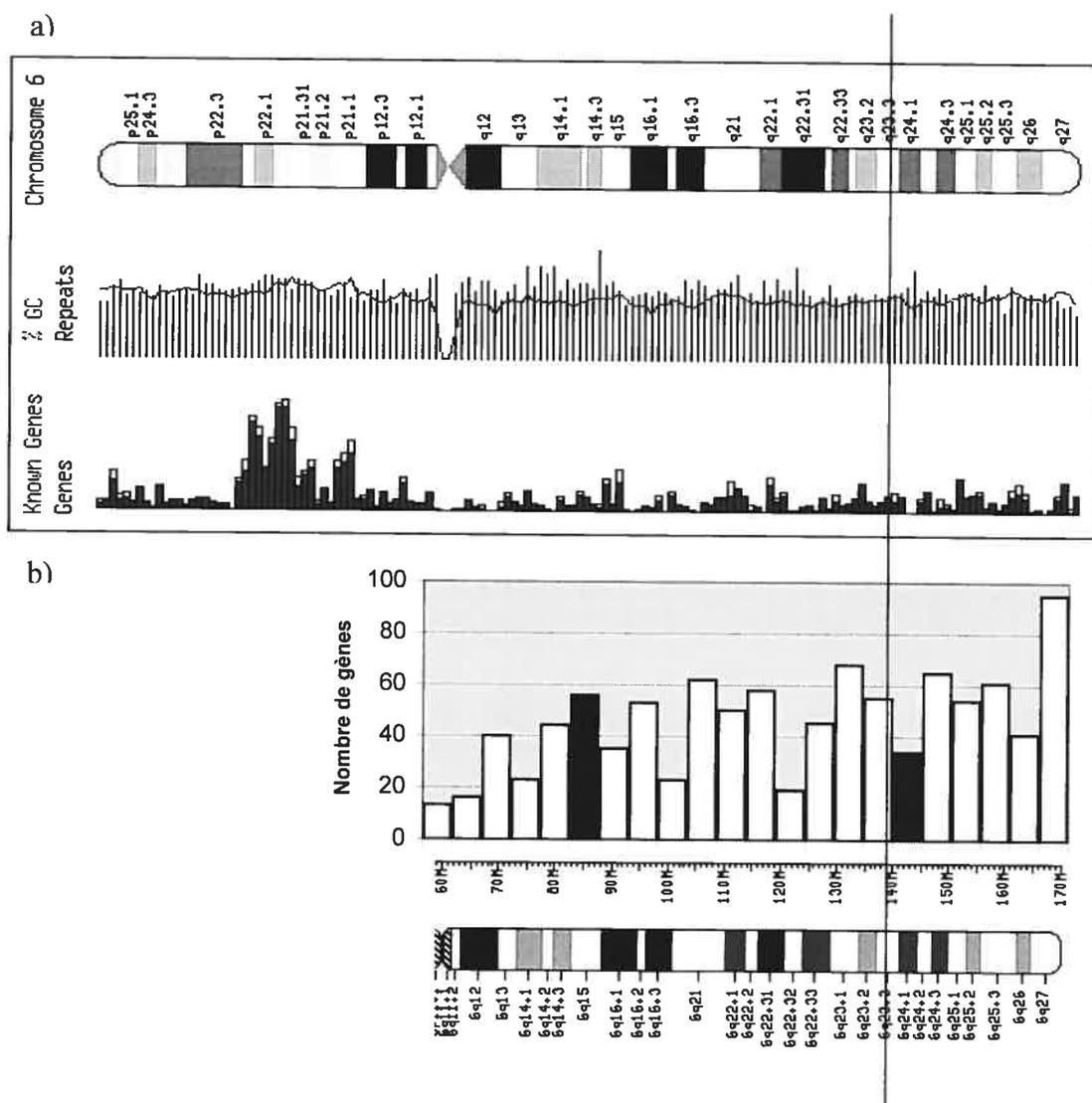
| Partenaire L1 endogène | No accès GenBank | ARNm   |
|------------------------|------------------|--|
| 34-B                   | BQ303919         | M9 spécifique au muscle  |
| 34-C                   | BQ82218          | ARNm inconnu   |
| 34-D                   | A1476439         | NMBR (récepteur neuroméline B)   |
| 34-E                   | Z26653           | LAMA2 (laminine, alpha 2 mérosine)   |
| 34-F                   | AF025654         | RNGTT (ARN guanylyltransférase et 5'P)   |
| 34-I                   | AF194537         | NAG13  |
| 34-G                   | AK057286         | PACRG (parkin co-regulated gene protein)   |
| Int. ATA67*            | M19368           | SLC8A1 (solute carrier family 8 Na <sup>+</sup> /Ca <sup>+</sup> exchanger member 1) |
| 67-C                   | BI049259         | ARNm inconnu   |
| 67-G                   | A1677857         | GTPase régulatrice associée avec pp125 (kinase adhésion focale)                      |
| 67-I                   | AF052166         | PDX1 (complexe pyruvate déshydrogénase, composante lypoyl X)                         |
| 67-M                   | AK055030         | ARNm inconnu   |

**Tableau XII : Localisation des partenaires L1s endogènes dans des séquences codantes**

\* Dans le clone ATA67, le vecteur pL1HsSceI est lui-même intégré dans la séquence d'un gène.

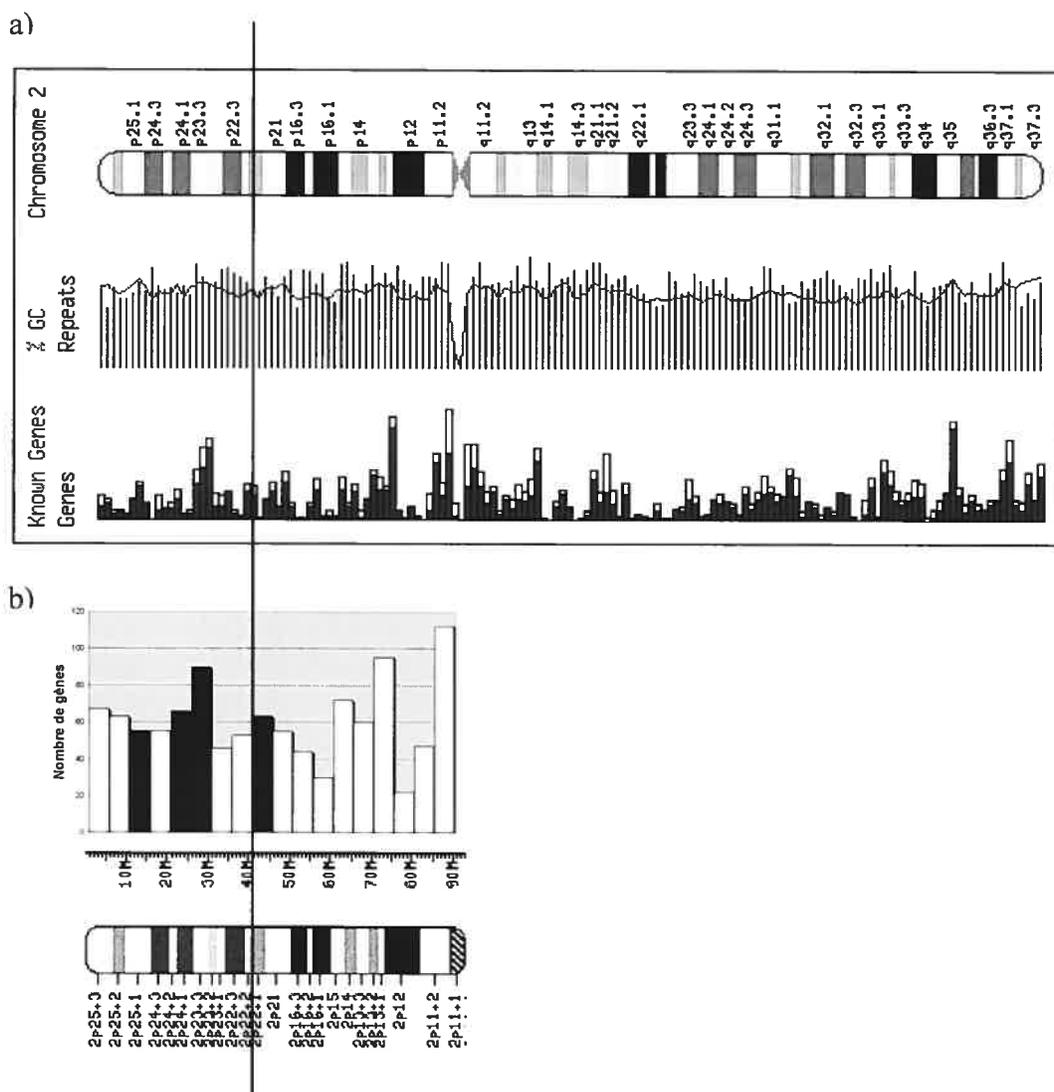
Étant donné la proportion relativement élevée de partenaires L1 endogènes localisés à l'intérieur de séquences codantes, nous avons vérifié si les régions regroupant plusieurs partenaires étaient particulièrement denses en gènes. Il se pourrait qu'une structure de chromatine plus ouverte puisse favoriser l'accessibilité à certaines régions et qu'une région transcriptionnellement active puisse fournir plus facilement la machinerie nécessaire à la recombinaison.

Aux figures 32 (p.127) et 33 (p.128), nous avons illustré pour chacun des chromosomes d'intégration la distribution des gènes, des séquences répétitives et des regroupements de partenaires L1 endogènes. Sur le bras chromosomique 6q, nous n'avons pas observé de densité en gènes plus élevée aux sites de regroupement des partenaires (figure 32, p.127). Par contre, il semble que ces sites correspondent aux régions les plus riches en séquences répétitives. Sur le bras chromosomique 2p, la localisation des sites de regroupement de partenaires ne semble pas être corrélée de façon évidente avec le contenu en gènes ou en séquences répétitives (figure 33, p.128).



**Figure 32 : Distribution des gènes, des séquences répétitives et des regroupements de partenaires du clone ATA34 sur le chromosome 6**

- a) Distribution relative des séquences répétitives (lignes verticales), contenu de l'ADN en G+C (courbe graphique) et répartition des gènes connus (barres noires) et prédits (barres blanches) sur le chromosome 6.
- b) Nombre de gènes retrouvés sur le bras chromosomique 6q, pour chaque segment de 5 Mb. La ligne verticale situe le site d'intégration du pLIHsScel en 6q23.3. Les colonnes noires représentent les régions regroupant plusieurs partenaires L1 endogènes.



**Figure 33 : Distribution des gènes, des séquences répétitives et des regroupements de partenaires du clone ATA67 sur le chromosome 2**

- a) Distribution relative des séquences répétitives (lignes verticales), contenu de l'ADN en G+C (courbe graphique) et répartition des gènes connus (barres noires) et prédits (barres blanches) sur le chromosome 2.
- b) Nombre de gènes retrouvés sur le bras chromosomique 2p, pour chaque segment de 5 Mb. La ligne verticale situe le site d'intégration du pL1HsScel en 2p22.1. Les colonnes noires représentent les régions regroupant plusieurs partenaires L1 endogènes.

## IV. DISCUSSION

### IV.1. Contribution du projet

Suite au séquençage complet du génome humain, l'établissement de son organisation tridimensionnelle est un pas d'importance capitale vers la compréhension du fonctionnement de l'organisme multicellulaire très complexe qu'est l'être humain. À ce jour, certaines règles de base ont pu être établies en ce qui concerne la structure des chromosomes à l'interphase. Cependant, il n'existe toujours pas d'évidence claire que les nombreux et complexes niveaux d'organisation de l'ADN sont conservés et que cette organisation implique bel et bien un rôle fonctionnel important.

Dans le cadre de mon projet de doctorat, nous avons évalué cette possibilité en étudiant les règles et la conservation de l'accessibilité génomique d'un locus donné dans le noyau des cellules humaines à l'interphase. Pour ce faire, nous avons conçu un système innovateur basé sur les interactions homologues entre les séquences répétitives du génome. Par l'utilisation des séquences LINE-1 qui sont dispersées en 500 000 copies sur l'ensemble des chromosomes dans leur disposition naturelle, nous avons pu tracer un portrait très réaliste de l'accessibilité génomique pour deux loci distincts.

Notre modèle nous a également permis d'étudier de façon plus représentative la réparation des CDBs chez les cellules de mammifères, un mécanisme crucial au maintien de l'intégrité génomique. À ce jour, tous les modèles d'étude de la réparation des CDBs par RH étaient basés sur la recombinaison entre 2 séquences artificielles du génome, pour lesquelles l'homologie, la longueur et la proximité étaient des variables contrôlées par l'expérimentateur. De plus, ces systèmes nécessitaient tous la reconstitution d'un marqueur de sélection. Au laboratoire, nous avons été les premiers à étudier la réparation des CDBs par RH avec des séquences répétitives naturelles du génome de souris, au cours de laquelle un libre choix de partenaires homéologues endogènes était offert<sup>125</sup>.

Par mon projet de recherche, nous sommes les premiers à étudier la réparation des CDBs dans les cellules humaines, au cours de laquelle le choix d'un partenaire endogène est libre et très varié en termes de longueur, d'homologie et de proximité. De plus, la méthode de *plasmid rescue* utilisée nous permet de récupérer tous les types d'événements de recombinaison, sans biais pour le mécanisme. Finalement, l'un des aspects les plus innovateurs de notre système d'étude consiste en l'identification et surtout en la localisation des partenaires endogènes impliqués, grâce aux récentes banques de données découlant du séquençage du génome humain.

## IV.2. Nature et fréquence des événements de recombinaison

### IV.2.1. Induction d'une cassure double-brin (CDB)

Dans notre système, l'induction d'une CDB par l'endonucléase I-*SceI* à l'intérieur d'un élément L1 nous a permis de stimuler et de cibler les interactions d'un seul locus avec le reste du génome. Puisqu'une CDB stimule d'un facteur de 100 à 1000 fois les fréquences de recombinaison homologue et illégitime<sup>140,142,143</sup>, ce système a grandement facilité la récupération d'événements qui autrement auraient été faiblement détectables. Lors d'un essai préliminaire de recombinaison spontanée sur 1619 plasmides récupérés, nous n'avons en effet obtenu qu'un seul événement de RH et qu'un seul événement de NHCR. Ainsi, notre fréquence préliminaire de RH spontanée de l'ordre de  $10^{-4}$  est légèrement plus basse que celle obtenue par Meyn<sup>124</sup> avec cette même lignée, soit 5 à  $12 \times 10^{-4}$ . Cependant, le système de recombinaison utilisé dans cette étude impliquait des séquences répétitives homologues à 100% et juxtaposées à un même site chromosomique.

Nos événements spontanés sont quelque peu différents. Résultant d'une perte aléatoire du site I-*SceI*, ils présentent un site de jonction situé 573 pb en 5' du site I-*SceI* (NHCR 67nt-13) ou un segment de conversion débutant à 733 pb en 5' du site I-*SceI* (RH 67nt-8, figure 24c, p.110), contrairement à la plupart des événements

induits par une CDB dont la signature est initiée dans la région immédiate du site I-*Scel*. La recombinaison induite par une CDB pourrait être différente de la recombinaison spontanée pour plusieurs raisons. Alors que des évidences suggèrent que la recombinaison méiotique soit initiée par des CDBs générées par la topoisomérase SPOII<sup>146,209</sup>, les événements moléculaires responsables de l'initiation de la recombinaison spontanée chez les cellules somatiques ne sont pas connus. Il se pourrait que les CDBs soient en cause, mais la localisation de ces cassures, le type d'extrémités qu'elles produisent et le mécanisme par lequel elles sont protégées et traitées pourraient grandement différer d'une coupure induite par l'endonucléase I-*Scel*.

De plus en plus d'évidences suggèrent également que la recombinaison mitotique spontanée puisse être dépendante de la réplication, au cours de laquelle le mécanisme de *crossing-over* serait supprimé en faveur de la conversion génique<sup>142,143,210,211</sup>. Par contre, des expériences de RH spontanée *in vivo* menées au laboratoire avec la technique du *plasmid rescue* sur des souris transgéniques ont démontré que la conversion génique n'était pas le mécanisme le plus utilisé pour la recombinaison entre les séquences répétitives L1. Par ces études, nous avons démontré que dans les cellules somatiques de souris, la RH spontanée mitotique entre un L1 intégré au génome et les L1s endogènes s'effectuait dans plus de 99% des cas par un mécanisme délétère<sup>212</sup>. Puisque ces événements ont tous impliqué des partenaires endogènes différents, il est plus probable qu'ils se soient produits par un mécanisme de *crossing-over* plutôt que par un mécanisme de *single-strand annealing*, qui aurait inévitablement impliqué plusieurs fois un même partenaire intrachromosomique. Ce mécanisme non-conservatif était également très utilisé (~50% des événements) dans les ovules de souris, lors de la RH spontanée entre les répétitions en tandem d'un plasmide intégré ou en voie de l'être<sup>213</sup>. Lors de nos essais préliminaires de RH spontanée en culture cellulaire, il s'avère que le seul événement que nous avons pu récupérer impliquait un mécanisme de *crossing-over*. Beaucoup trop préliminaire pour être indicateur, ce résultat ouvre toutefois la porte à une perspective intéressante sur l'étude de la RH ectopique spontanée entre les séquences répétitives dans un

modèle cellulaire. Bien que très ardue pour cause d'une très faible fréquence spontanée, une telle étude permettrait d'étudier l'importance du contrôle de la RH ectopique sur la stabilité d'un génome truffé d'éléments répétitifs homéologues.

Suite à l'induction d'une CDB, nous n'avons observé un mécanisme de *crossing-over* que dans 2% des événements de RH (tableau II, p.77). Chez des cellules en culture, il se pourrait que le mécanisme de *crossing-over* engendre des réarrangements délétères conduisant à la mort de la cellule. Suite à l'induction des CDBs, seules les cellules viables ont été récupérées, de sorte qu'un biais contre les événements les plus menaçants pour l'intégrité cellulaire ait pu être introduit. Cependant, la lignée cellulaire GM05849 utilisée présente une forte instabilité génomique impliquant un niveau de tolérance supérieur à ce type d'événements délétères. La RH induite par les CDBs a été étudiée par plusieurs groupes et de plusieurs façons. Lorsque la cassure est induite dans l'une de 2 répétitions directes voisines, une RH intrachromosomale est observée dans 30 à 50% des événements de réparation<sup>83,143</sup>. Toutefois, l'augmentation de la divergence entre les 2 répétitions<sup>115,117,169,214</sup> ainsi que leur localisation interchromosomique<sup>142</sup> peut faire diminuer de plusieurs ordres de grandeur la fréquence de réparation par RH. Dans un système cellulaire où une CDB est induite à l'intérieur d'un élément L1 du génome de souris, notre laboratoire a démontré que la fréquence de réparation d'une CDB par RH avec un partenaire endogène homéologue et ectopique se situait à environ 1%<sup>125</sup>, suggérant que la fréquence de RH est modulée par plusieurs facteurs.

#### IV.2.2. Statuts ATM et p53 de la lignée GM05849

Dans un contexte cellulaire instable déficient pour le gène ATM, notre système a permis de générer 3,8% (clone ATA34) et 14,8% (clone ATA67) d'événements de réparation par RH avec des L1s endogènes, la majorité des CDBs ayant plutôt été réparées par un mécanisme de recombinaison non-homologue (tableau II, p.77). Dans la même lignée cellulaire, Meyn avait observé une augmentation de 100 fois de la RH intrachromosomale spontanée entre 2 répétitions directes<sup>204</sup>. Il serait donc

sensé d'expliquer l'élévation marquée du ratio RH:NHEJ dans notre système, en comparaison à celui obtenu chez des cellules de souris plus stables<sup>125</sup>, par la déficience du gène ATM.

Pour établir dans notre système la contribution réelle de la perte de fonction d'ATM sur la fréquence, les mécanismes et les partenaires de RH, nous avons tenté de rétablir dans nos clones la fonction du gène. Cette complémentation avait déjà été réalisée avec succès sur la lignée GM05849 par le laboratoire du Dr Shiloh, par la transfection du vecteur pEBS7YZ5 exprimant le gène ATM<sup>215</sup>. Pour notre complémentation, nous avons dû utiliser une version modifiée de ce vecteur, afin de ne pas entraver notre technique de récupération des plasmides. Malheureusement, cette modification a probablement engendré une perte de stabilité du vecteur, car nous n'avons pas obtenu de clones exprimant l'ADNc d'ATM. La génération éventuelle de clones isogéniques ne différant que par la fonction du gène ATM demeurerait d'une grande utilité pour étudier la contribution et l'impact de la RH entre les séquences répétitives sur l'instabilité des cellules tumorales. Nous pourrions nous demander si dans les mêmes clones, le retour de la fonction d'ATM impliquerait un changement majeur au niveau de la fréquence (plus faible) et du type d'événements de RH spontanée entre les séquences répétitives (plus conservatifs), au niveau du mécanisme de réparation des CDBs, au niveau du choix d'un partenaire homologue (plus restreint) et au niveau de l'accessibilité (moins permissives) d'un locus donné dans le génome.

Puisque la lignée de fibroblastes GM05849 a été immortalisée par l'antigène grand T du virus SV40 (SV40 Ag-T)<sup>216</sup>, la fréquence de RH observée dans ces cellules est également dépendante de son statut P53, une protéine toute aussi cruciale qu'ATM dans le contrôle de la RH. En formant un complexe avec des sites spécifiques de liaison à l'ADN de P53, le SV40 Ag-T inhiberait son activité transcriptionnelle et sa fonction inhibitrice de la RH<sup>133,217</sup>. Ainsi, une hausse du taux de RH de l'ordre de 100 fois a été observée chez la lignée GM00637 immortalisée par SV40 Ag-T, par comparaison avec une lignée où la fonction de P53 était intacte<sup>133</sup>.

### IV.2.3. Différence de fréquence de RH entre les clones

La fréquence de RH observée chez le clone ATA67 surpasse de 4 fois celle observée chez le clone ATA34. Dans le clone ATA67, le pL1HsSceI est intégré à l'intérieur d'un gène probablement transcrit (figure 16, p.69). Il serait légitime de penser que l'ouverture de la chromatine d'une région transcrite puisse augmenter son accessibilité à la machinerie de recombinaison et faciliter l'invasion de l'hétéroduplex par les extrémités brisées. Des évidences directes ont en effet suggéré chez les cellules mammifères que la transcription pouvait stimuler la fréquence de RH extrachromosomale et intrachromosomale spontanée, sans toutefois influencer le mécanisme de recombinaison<sup>185-187</sup>. Par contre, une étude a démontré que la transcription ne semblait pas avoir d'effet stimulateur sur la RH induite par les CDBs<sup>115</sup>, qui est à la base de notre système d'étude.

Pour un nombre équivalent d'événements de réparation, la fréquence de RH du clone ATA67 est nettement et significativement supérieure à celle du clone ATA34 (tableau II, p.77). Cependant, puisque nous avons analysé une quantité supérieure d'événements de RH avec le clone ATA67 pour lequel la fréquence était plus élevée, il nous est difficile de comparer sur la même base le type d'événements obtenus, à moins de générer et d'analyser un nombre supérieur d'événements de réparation avec le clone ATA34, une tâche très ardue compte tenu de sa très faible fréquence de RH. Néanmoins, pour étudier clairement l'effet de l'activité transcriptionnelle, il suffirait de comparer à l'aide de notre modèle les fréquences et les produits de RH pour plusieurs sites génomiques transcrits ou inactifs.

En dehors de l'hypothèse que l'activité transcriptionnelle puisse augmenter l'accessibilité à la machinerie de recombinaison, il se pourrait aussi que la configuration de la chromatine imposée par le stress torsionnel de la transcription devienne en soi une structure plus sensible à l'attaque des nucléases, qui deviendrait en quelque sorte l'événement déclencheur de la recombinaison. Des études sur l'ADN eucaryote ont en effet révélé *in vivo* que l'accessibilité de la chromatine aux différentes nucléases diffèrait grandement entre des sites activement transcrits et des

sites inactifs<sup>189</sup>. Cette hypothèse est appuyée par une augmentation importante de la fréquence de translocations dans le clone ATA67, un mécanisme au cours duquel une seconde CDB doit être induite dans le génome en plus de celle ciblée au site *I-SceI*. Ainsi, les partenaires de réarrangements non-homologues (NHCR) amenés à proximité du locus d'intégration se retrouveraient plus exposés aux nucléases sur place.

Alternativement, l'augmentation concomitante de la fréquence de RH et de NHCR dans le clone ATA67 pourrait s'expliquer par un environnement chromosomique différent pour chacun des clones. Dans le clone ATA67, le locus d'intégration pourrait se situer dans un domaine plus accessible à l'espace interchromatique où sont concentrés les facteurs protéiques nécessaires à la RH et aux NHCR. Inversement, l'intégration du clone ATA34 pourrait être située dans un domaine moins accessible à l'espace interchromatique, forçant la cellule à recoller simplement les extrémités brisées entre elles (NHEJ), sans avoir recours à un partenaire génomique.

### IV.3. Événements de réparation non-homologue

#### IV.3.1. Événements de *non-homologous end-joining* (NHEJ)

La jonction des extrémités brisées (NHEJ) est la voie la plus utilisée par les cellules de mammifères pour réparer une CDB. Bien que des cellules normales puissent refermer 50% des CDBs en recréant un site *I-SceI* fonctionnel<sup>218</sup>, cette refermeture est souvent accompagnée de la délétion et/ou de l'insertion de quelques nucléotides<sup>83,143</sup>, comme c'est le cas dans 80% des événements de NHEJ analysés pour chacun des clones (tableau IV, p.81).

Puisque notre système ne permet la récupération que des événements ayant entraîné la perte du site *I-SceI*, il nous est difficile de nous prononcer sur l'effet possible de l'absence d'ATM sur la fidélité de réparation par NHEJ. Par contre, seulement 15% des recombinants de type NHEJ récupérés présentaient des plus grandes délétions de

plus de 100 pb, un résultat similaire à ceux obtenus chez des cellules génétiquement stables<sup>83,143</sup>. Nos résultats sont donc plus ou moins en accord avec certaines études suggérant que la fidélité du mécanisme de NHEJ soit compromise chez des cellules déficientes pour le gène ATM, en raison d'une perte d'effet protecteur des extrémités brisées<sup>93,94</sup>.

#### IV.3.2. Événements de *non-homologous chromosomal rearrangement* (NHCR)

La fréquence élevée d'événements de translocations et de réarrangements intrachromosomiques non-homologues observée dans notre système dans un contexte cellulaire instable est une caractéristique bien établie des cellules déficientes pour ATM<sup>95,219</sup>. En effet, l'instabilité génomique de ces cellules se reflète par une augmentation marquée des CDBs spontanées et induites par des radiations ionisantes<sup>220</sup>, ainsi que par des aberrations chromosomiques comme des bris de chromosomes, des fragments acentriques et dicentriques, et une fréquence élevée de translocations dans les gènes du système immunitaire<sup>100,203</sup>.

Parmi tous les événements de NHCR (tableau VI, p.85-86), la moitié ont impliqué des sites situés sur le bras chromosomique du locus d'intégration, regroupés à certains endroits préférentiels (figure 28, p.119). Nous pourrions donc présumer que ces réarrangements ne se sont pas produits aléatoirement avec n'importe quelle séquence, mais qu'une organisation de l'ADN en 3 dimensions aurait favorisé l'accessibilité à certains sites, qui pourraient s'avérer plus fragiles aux CDBs. Les translocations représentent également une grande proportion des événements de réparation non-homologues (50%), mais le nombre de sites interchromosomiques analysés n'est pas suffisant pour déceler une préférence pour certains d'entre eux.

Il est important de mentionner qu'étant reconnues instables, les cellules GM05849 utilisées dans ces travaux étaient déjà porteuses d'un certain nombre de réarrangements chromosomiques, avant même que les événements de recombinaison étudiés ne soient générés. Il est donc possible que certains événements de NHCR classés interchromosomiques (translocations) puissent en réalité être des événements de réarrangements intrachromosomiques avec une portion déjà réarrangée. De la même façon, il est possible que l'événement de *crossing-over* 67A10-232 entre le chromosome 2 et le chromosome 3 soit en réalité intrachromosomique, sur un chromosome déjà réarrangé. Puisque le caryotype des clones ATA34 et ATA67 n'a pas été établi, nous ne connaissons pas la structure des chromosomes portant ces sites d'intégration et nous ne pouvons pas être certain que tous les événements interchromosomiques le sont vraiment. Toutefois, les sites interchromosomiques identifiés étant largement distribués, il demeure fort improbable qu'ils soient tous en réalité situés sur un chromosome d'intégration réarrangé.

Dans le cadre de ce projet, l'utilisation de la lignée GM05849 comme modèle cellulaire avait comme principal avantage l'élévation marquée du taux de RH, permettant la génération et l'analyse d'un grand nombre d'événements. Au départ, l'intention était de générer des clones dans la lignée stable GM00637 dans laquelle le gène ATM est fonctionnel, et de comparer les résultats quant à la fréquence de RH, aux mécanismes de recombinaison utilisés et au choix des partenaires endogènes. Cette étude comparative aurait permis de discerner le rôle du gène ATM dans le processus de RH et de mettre en valeur la contribution de la RH entre les séquences répétitives à l'instabilité génomique. Toutefois, l'impossibilité de générer des clones adéquats dans cette lignée cellulaire, ou encore de générer des clones isogéniques par la transfection stable d'un vecteur exprimant le gène ATM dans nos clones ATA34 et ATA67 a rendu impossible cette étude comparative. L'instabilité génomique de la lignée GM05849 est alors devenue un désavantage important affectant les conclusions émanant du côté mécanistique des résultats.

## **IV.4. Événements de RH**

### **IV.4.1. Identification des partenaires L1 endogènes de RH**

L'un des aspects les plus innovateurs de notre modèle consiste en l'identification des partenaires endogènes de RH et surtout en leur localisation dans le génome humain. Par le programme de BLAST du NCBI, nous avons pu identifier avec un bon niveau de certitude le partenaire L1 endogène de 80% des événements de RH du clone ATA34 (tableau VII, p.88-89) mais seulement pour 38% de ceux du clone ATA67 (tableau VIII, p.92-93 et tableau IX p.94-95). Le résultat du BLAST proposait parfois plus d'un partenaire possible pour un même segment de conversion, mais la plupart du temps un seul de ceux-ci se retrouvait sur le bras chromosomique du locus d'intégration et devenait plus probable par sa grande proximité. Notre identification portait donc parfois à interprétation.

Une grande proportion des partenaires du clone ATA67 n'ont pu être identifiés par cette méthode de BLAST pour cause d'un segment de conversion très court ne permettant pas un alignement significatif. Bien que plusieurs de ces événements ne semblent pas avoir subi de conversion s'étendant au-delà de la portion des 23 pb d'origine du L1, il se pourrait que la conversion soit très longue, mais qu'elle demeure invisible à cause d'une très grande homologie du partenaire endogène avec le L1 du vecteur. En effet, si on compare la séquence du L1 cd11 à celle des autres L1s du génome autour du site de CDB, on s'aperçoit qu'une interaction avec un grand nombre de partenaires passerait inaperçue sur un long segment de conversion. Par exemple, une conversion de directionnalité 3' sur 140 pb ne laisserait pas de traces (pas de modifications de bases) pour au moins 100 éléments L1s du génome qui sont homologues à 100% avec le cd11 sur cette portion.

Pour augmenter le taux d'identification des partenaires du clone ATA67, nous avons visuellement comparé le segment de conversion des événements à ceux des

partenaires déjà identifiés (figure 21, pp.99-101). Nous avons alors remarqué de façon évidente que plusieurs de ces orphelins trouvaient leurs parents dans les partenaires déjà identifiés, et même que des partenaires indépendants identifiés par BLAST correspondaient en réalité à un seul et même partenaire. Ainsi, la combinaison de la méthode du BLAST à celle de l'alignement visuel des segments de conversion nous a permis d'augmenter à 64% le taux d'identification des partenaires du clone ATA67. En poussant au maximum notre interprétation, nous aurions même pu attribuer les 24 événements de conversion génique sans différences au partenaire 67-C, puisque la séquence de ce L1 endogène est identique à celle du L1 à réparer sur une grande étendue autour de la CDB. En supposant que c'est réellement le cas, ce partenaire deviendrait le plus utilisé et le plus proximal des 4 partenaires endogènes localisés dans la région immédiate du locus d'intégration (figure 30, p.122).

À première vue, il semblait plutôt étrange que des événements comme le 67-V220 présentant un long segment minimal de conversion (177 pb) et plusieurs bases converties (9 bases modifiées) puissent demeurer sans identification spécifique par la méthode du BLAST. Une étude récente a toutefois démontré que la présence et l'absence des éléments L1s étaient polymorphiques d'un génome à l'autre<sup>221</sup>. Dans la famille récente Ta, 20% des L1s répertoriés n'étaient retrouvés que dans 1 génome sur 3. Il est donc fort possible que des L1s utilisés comme partenaires de RH dans le génome de la lignée GM05849 soient spécifiques à ce génome et que la banque du génome humain ne les ait pas annotés.

Ce polymorphisme pourrait nous faire douter de la validité du résultat de certains BLASTs comme ceux du partenaire 67-B (tableau VIII, pp.92-93) pour lesquels une homologie de 100% n'a pas été rencontrée. Dans ce cas, le véritable partenaire n'est peut-être pas le 67-B, mais un autre L1 non répertorié dans GenBank, qui serait polymorphique et spécifique à la lignée GM05849. Par contre, puisqu'un seul partenaire se démarquait de ces BLASTs et se situait à proximité du site d'intégration, il est beaucoup plus probable qu'il s'agisse du véritable partenaire.

La divergence de 2 à 3% observée entre la séquence de ce L1 dans la lignée et dans GenBank pourrait s'expliquer par une accumulation de mutations ponctuelles aléatoires dans chacun des génomes au cours de leur évolution, les séquences répétitives ne subissant pas de pression sélective<sup>222,223</sup>. Elle pourrait aussi s'expliquer par des événements de RH entre les L1s du génome. La conversion génique entre les séquences répétitives pourrait en effet jouer un rôle significatif dans la sculpture de la diversité génomique<sup>61,62</sup>. Au laboratoire, nous avons démontré que les L1s du génome variaient d'un individu à l'autre, et que leurs interactions homologues par conversion génique seraient impliquées dans leur évolution concertée<sup>63</sup>.

En séquençant le segment de conversion de l'événement 67-R101, nous avons identifié la portion 5' du partenaire 67-C et la portion 3' du partenaire 67-B. Pour impliquer à la fois ces 2 partenaires, l'événement de *crossing-over* analysé aurait nécessité un mécanisme de recombinaison très complexe incluant 2 invasions d'hétéroduplex. Il serait donc beaucoup plus simple et probable que le partenaire 67-B ait antérieurement subi une conversion génique par le partenaire 67-C, produisant un L1 hybride dans la lignée GM05849. Pour prouver ce raisonnement, nous avons tenté d'amplifier par PCR les partenaires 67-B et 67-C du génome de notre lignée cellulaire, dans le but de les séquencer et de démontrer leur conversion. Malheureusement, cette tentative s'est avérée infructueuse.

#### IV.4.2. Mécanisme de RH

Dans la littérature, rares sont les modèles d'étude de la RH ayant permis de récupérer tous les types d'événements de RH et d'en comparer les fréquences, et aucun n'a pu laisser le libre choix d'un partenaire homéologue. Notre étude est donc la première à illustrer clairement le mécanisme de RH suite à une CDB dans les cellules mammifères. Cependant, tel que mentionné précédemment, il se pourrait qu'un certain biais en faveur des mécanismes non-délétères se soit introduit dans notre système lors de la récolte sélective des cellules viables.

Pour chacun des clones, la grande majorité des événements de réparation de CDB par RH ont impliqué un mécanisme de conversion génique alors que seuls 2 événements de *crossing-over* ont été récupérés parmi les 96 événements homologues du clone ATA67 (tableau II, p.77). Outre les mécanismes de conversion génique et de *crossing-over*, le mécanisme homologue de SSA (*single-strand annealing*) observé à une fréquence de 20 à 33% dans d'autres systèmes<sup>83</sup> n'a toutefois pas été observé dans le notre, mis à part l'événement 67-R101 classé comme *crossing-over* intrachromosomique qui aurait tout aussi bien pu être classé comme événement de SSA entre des répétitions directes séparées par 205 kb (figure 24a, p.109).

##### IV.4.2.1. Effet du statut ATM sur la longueur du segment de conversion

D'après le modèle de Szostak<sup>110</sup>, le mécanisme de réparation d'une CDB par RH impliquerait une première étape de dégradation des extrémités brisées. L'étendue de la dégradation des extrémités correspondrait à l'étendue du segment de conversion. En absence de la protéine ATM qui pourrait jouer un certain rôle protecteur aux extrémités brisées, on pourrait s'attendre à observer des segments de conversion particulièrement longs. Pour vérifier cette hypothèse, nous avons calculé la longueur minimale des segments de conversion génique dans les recombinants. Pour l'ensemble des recombinants, nous avons calculé un segment minimal moyen de 79 pb (tableau XI, p.116). Cette valeur étant comparable à celles obtenues dans d'autres

système de RH entre 2 séquences répétitives<sup>111,115,116</sup>, il ne semble pas que la déficience d'ATM dans les cellules GM05849 affecte grandement l'exposition des extrémités brisées aux nucléases.

On remarque que la directionnalité en 3' est un peu plus fréquente que celle en 5' et qu'elle génère des segments en moyenne presque 2 fois plus longs. Un certain biais pourrait s'être introduit contre les événements de longue conversion en 5', qui mèneraient à la perte des séquences d'ampicilline et d'origine de réplication nécessaires à la récupération des plasmides. En réalité, notre système ne permettait de récupérer qu'une fraction des événements de *crossing-over*, c'est-à-dire seuls ceux dirigés en 3'.

#### ***IV.4.2.2. Relation entre le mécanisme et la directionnalité de la conversion***

Toujours selon le modèle de Szostak<sup>110</sup>, chacune des extrémités de la CDB peuvent être dégradées sur une distance variable avant leur invasion dans le partenaire homologue qui servira comme modèle de réparation. Selon le modèle de Stahl<sup>224</sup>, les extrémités ne subissent pas de dégradation extensive mais l'hétéroduplex qui se forme entre les partenaires de RH subit l'action du système de réparation des *mismatches*, habituellement en faveur du partenaire intact. Peu importe le modèle, le segment de conversion résultant sera normalement bidirectionnel.

Chez des cellules mammifères, plusieurs modèles de RH entre des répétitions directes<sup>115,116</sup> et de *gene targeting*<sup>117,118</sup> ont rapporté une proportion plus élevée de conversion bidirectionnelle. À l'exception de la recombinaison allélique qui génère également une majorité de segments bidirectionnels<sup>113,225</sup>, tous les systèmes de RH étudiés chez la levure présentent plutôt une majorité de segments unidirectionnels<sup>111-114</sup>. Certains modèles de recombinaison extrachromosomique chez la levure suggèrent également que l'augmentation de la divergence entre les substrats de recombinaison puisse augmenter la proportion de bidirectionnalité du segment de conversion<sup>111,113</sup>.

À la figure 26 (pp.114-115), nous avons aligné et classé les segments de conversion de chacun de nos recombinants homologues en fonction de leur longueur et de leur directionnalité. Nous avons observé que 50% des segments étaient unidirectionnels en 5' ou en 3', contre seulement 12% de bidirectionnalité (tableau XI, p.116). Pour expliquer la préférence envers une conversion unidirectionnelle, remarquons d'abord que puisque la CDB est induite à l'intérieur d'un site de 18 pb I-SceI qui n'est pas présent dans les L1s endogènes, les extrémités terminales générées ont de courtes régions non-homologues qui pourraient limiter l'invasion à une seule extrémité. Plusieurs études récentes ont démontré que le mécanisme d'invasion par une seule extrémité (*one-sided invasion*)<sup>119,120</sup> était majoritairement utilisé lors de la RH ectopique<sup>142,227,228</sup>. Bien qu'un seul de nos événements porte clairement la signature de ce mécanisme par l'acquisition de séquences non-homologues adjacentes (67A10-240, section III.6.2.5), la prépondérance de segments unidirectionnels pourrait indiquer dans l'ensemble de nos événements l'utilisation préférentielle de ce mécanisme. L'unidirectionnalité pourrait donc refléter l'invasion de l'hétéroduplex par un seul brin, mais pourrait alternativement résulter d'une invasion par les 2 extrémités et de l'action du système de réparation des *mismatches* d'un seul côté de la CDB.

Nous avons par la suite vérifié si le choix de la directionnalité de la conversion était influencé par l'homologie. D'abord, nous n'avons pas observé de relation évidente entre le taux de divergence des partenaires et leur type de directionnalité. Des évidences suggéraient que la longueur de l'homologie puisse stimuler de façon exponentielle la fréquence de RH<sup>168,229,230</sup>. Mais à la lumière de nos résultats, l'utilisation préférentielle d'un partenaire ne semble pas être dépendante de la longueur d'homologie ininterrompue disponible à la CDB. Par exemple, le partenaire le plus populaire (67-B) offre une homologie ininterrompue de 94 pb, comparativement à 210 pb pour le partenaire 67-K qui n'a été utilisé qu'une seule fois (tableau XIII, p.143).

| Partenaire L1   |       | Homologie ininterrompue <sup>a</sup> |                    |                    | Directionnalité <sup>b</sup> |        |        |          |
|-----------------|-------|--------------------------------------|--------------------|--------------------|------------------------------|--------|--------|----------|
| Nom             | Freq. | Longueur<br>(pb)                     | Segment 5'<br>(pb) | Segment 3'<br>(pb) | 23 pb                        | Uni 5' | Uni 3' | BI 5'-3' |
| 67-A            | 13    | 89                                   | 50                 | 39                 | 1                            | 5      | 7      |          |
| 67-B            | 27    | 94                                   | 50                 | 44                 | 8                            | 6      | 5      | 8        |
| 67-C            | 4     | 85                                   | 50                 | 35                 | 1                            | 2      | 1      |          |
| 67-D            | 12    | 166                                  | 88                 | 78                 |                              | 5      | 6      | 1        |
| 67-E            | 1     | 61                                   | 35                 | 26                 |                              |        | 1      |          |
| 67-G            | 1     | 81                                   | 48                 | 33                 | 1                            |        |        |          |
| 67-H            | 1     | 150                                  | 115                | 35                 |                              |        | 1      |          |
| 67-I            | 1     | 75                                   | 37                 | 38                 |                              |        |        | 1        |
| 67-K            | 1     | 210                                  | 115                | 95                 |                              |        | 1      |          |
| 67-M            | 1     | 75                                   | 37                 | 38                 |                              |        |        | 1        |
| <b>moyenne:</b> |       | <b>109</b>                           | <b>63</b>          | <b>46</b>          |                              |        |        |          |
| 34-A            | 3     | 193                                  | 115                | 78                 |                              |        | 3      |          |
| 34-B            | 2     | 145                                  | 75                 | 70                 |                              | 1      |        | 1        |
| 34-C            | 2     | 111                                  | 67                 | 44                 |                              | 1      | 1      |          |
| 34-D            | 1     | 111                                  | 74                 | 37                 |                              |        | 1      |          |
| 34-E            | 2     | 141                                  | 69                 | 72                 |                              | 2      |        |          |
| 34-F            | 1     | 81                                   | 35                 | 46                 |                              | 1      |        |          |
| 34-G            | 1     | 66                                   | 31                 | 35                 |                              |        | 1      |          |
| 34-H            | 1     | 86                                   | 35                 | 51                 |                              | 1      |        |          |
| 34-I            | 1     | 187                                  | 115                | 72                 |                              |        | 1      |          |
| <b>moyenne:</b> |       | <b>125</b>                           | <b>68</b>          | <b>56</b>          |                              |        |        |          |
| <b>moyenne:</b> |       | <b>117</b>                           | <b>66</b>          | <b>51</b>          |                              |        |        |          |

**Tableau XIII : Influence de la longueur d'homologie ininterrompue au site de CDB sur le choix de la directionnalité de la conversion**

<sup>a</sup>La longueur de l'homologie ininterrompue au site de CDB est calculée par la somme des segments parfaitement homologues en 3' et en 5' de la cassure.

<sup>b</sup>La répartition des segments de conversion selon leur directionnalité est indiquée pour chaque partenaire.

En principe, le segment minimal d'homologie parfaite (MEPS) devrait être d'au moins 200 pb pour une recombinaison intrachromosomique<sup>167</sup> et d'au moins 730 pb pour une recombinaison ectopique interchromosomique<sup>231</sup>. Ces conditions sont toutefois très rarement rencontrées dans les événements de RH que nous avons récupérés. Nos résultats démontrent qu'un partenaire intrachromosomique peut effectuer de la RH à haute fréquence avec une homologie ininterrompue de seulement 94 pb (67-B), et que des partenaires interchromosomiques peuvent être impliqués dans des événements en présentant un MEPS de seulement 75 pb (67-I), soit un dixième de l'homologie théoriquement requise. Bien que le MEPS ne semble pas très critique pour l'initiation des événements de conversion génique et de *crossing-over*, il se pourrait qu'il ait une plus grande importance dans l'initiation du mécanisme de SSA, qui n'a pas été observé dans notre système. Bien que le SSA puisse être initié avec aussi peu que 30 pb d'homologie, sa pleine efficacité n'est atteinte qu'avec 200 à 400 pb d'homologie<sup>232</sup>.

Nos résultats démontrent également qu'il n'existe pas de biais de directionnalité en fonction du côté de la CDB offrant la plus longue homologie ininterrompue. Même si la longueur moyenne des segments unidirectionnels en 3' est 2 fois plus grande que celle des segments unidirectionnels en 5' et qu'on note une légère préférence pour la directionnalité 3' (tableau XI, p.116), il demeure que la portion 3' présente en moyenne un plus court segment d'homologie ininterrompue (tableau XIII, p.143). Finalement, comme nous l'avons déjà suggéré<sup>171</sup>, il se pourrait que la restriction de l'homologie puisse favoriser des événements de *one-sided invasion* entre des séquences divergentes, puisque la majorité des conversions sont unidirectionnelles.

#### ***IV.4.2.3. Implication du système de réparation des mismatches (MMR)***

Contrairement au mécanisme de réplication qui a souvent recours au système de réparation des *mismatches* pour corriger ses erreurs, le mécanisme de recombinaison par conversion génique s'effectue surtout par dégradation exonucléotidique du gap et synthèse d'ADN. Bien que la réparation des *mismatches* puisse également être

exercée sur l'hétéroduplex lors de la réparation des CDBs, elle s'effectue normalement de façon continue et fournit rarement des recombinants de type mosaïque<sup>114, 144</sup>. Dans certains événements de RH avec des partenaires fréquents, nous avons observé un segment de conversion hybride (discontinu) entre la séquence du donneur et celle du receveur (événements 67A10-735, 67A10-608, 67-R334 et 67A10-225 à la figure 23, p.104-108). Dans ces événements, une base localisée au tout début du segment de conversion est demeurée inchangée, alors qu'elle aurait dû être modifiée en faveur du partenaire endogène impliqué. Localisées très près du site de cassure et donc probablement incluses dans l'hétéroduplex, ces bases ont pu subir une réparation par le système de réparation des *mismatches*, en faveur de la molécule brisée. Ce genre de discontinuité n'a pas été retrouvé dans les régions distales des segments de conversion, indiquant que la réparation des *mismatches* est limitée à l'hétéroduplex formé au site de CDB et que la synthèse d'ADN est impliquée dans le reste de la conversion. Puisque les recombinants hybrides ainsi retrouvés ne diffèrent que par une seule base du partenaire endogène, la validité de leur identification est très probable. Cependant, nous ne pouvons écarter l'hypothèse que certains recombinants puissent porter un segment de conversion complètement hybride, qui fausserait ou rendrait impossible l'identification du partenaire endogène impliqué.

L'implication des protéines du MMR dans la RH a été rapportée dans les travaux de Elliot et Jasin<sup>176</sup>, qui démontraient que la proportion de segments hybrides augmentait de façon drastique chez des cellules déficientes pour le gène *Msh2*, d'une importance capitale dans le mécanisme de MMR. La protéine P53 aurait également un rôle majeur de reconnaissance spécifique des *mismatches* dans les hétéroduplexs de RH<sup>233</sup>. Dans la lignée GM05849, le mécanisme du MMR semble toutefois fonctionner normalement malgré la liaison de SV40 Ag-T à P53, puisqu'une très faible proportion de nos événements de conversion démontrent une discontinuité.

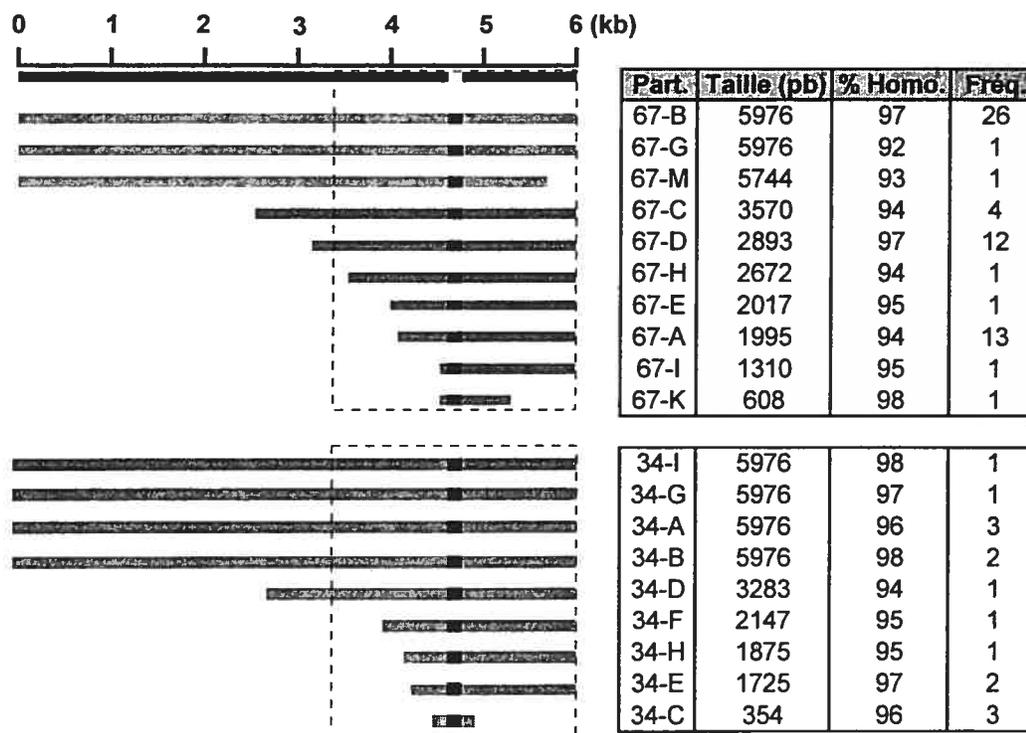
### IV.4.3. Choix du partenaire de RH

#### IV.4.3.1. Effet de la divergence

Pour l'ensemble des partenaires endogènes de RH identifiés avec les clones ATA34 et ATA67, nous avons observé des degrés d'homologie variant de 92 à 98% (tableaux VII pp. 88-89, VIII pp.92-93 et IX pp.94-95). Plus l'homologie est élevée, plus la famille d'un L1 est récente dans l'évolution du génome. Ainsi, les partenaires identifiés appartiennent aux familles récentes L1PA5 à L1PA1, la famille L1PA1 (ou L1Hs) étant la plus jeune et la plus active en rétrotransposition<sup>2</sup>.

Au tableau XIV (p.146), nous avons comparé l'homologie globale de chaque partenaire identifié avec sa fréquence d'utilisation et nous n'avons observé aucune corrélation évidente. Par exemple, le partenaire 67-A qui partage une homologie globale de 94% avec le L1 à réparer a été utilisé dans 13 événements, alors que d'autres partenaires plus homologues comme le 67-K (98%) ou le 67-E (95%) n'ont été utilisés qu'une seule fois. L'homologie partagée autour du site de cassure pourrait avoir une importance supérieure à l'homologie globale dans l'initiation de la recombinaison. Cependant, nous n'avons pas observé de corrélation entre la longueur de l'homologie parfaite autour du site de cassure et la fréquence d'un partenaire (tableau XIV, p.146).

Ces résultats sont consistants avec d'autres travaux de notre laboratoire concernant l'influence de l'homologie sur la fréquence de RH entre les éléments L1s<sup>125,171,201</sup>. Bien que la divergence ne semble pas interférer avec l'efficacité de la RH entre des séquences répétitives ectopiques, elle pourrait toutefois en influencer le mécanisme en faveur d'événements de conversion génique conservatifs pour préserver l'intégrité génomique<sup>171</sup>. La divergence pourrait restreindre la longueur des segments de conversion et empêcher le mécanisme de *crossing-over* qui semble être associé aux longs segments de conversion<sup>234</sup>.



**Tableau XIV : Influence de degré d'homologie et de la longueur sur le choix d'un partenaire L1 endogène de RH**

La région d'homologie partagée entre le L1 du pL1HsScel et les L1s endogènes est encadrée et correspond à la portion cd11B (3390 à 5976).

#### *IV.4.3.2. Effet de la longueur de l'homologie*

Sur les 19 partenaires L1 endogènes identifiés pour l'ensemble des 2 clones, 6 sont des éléments complets et sont impliqués dans 45% des événements. Puisque cette fréquence est de 5 fois supérieure à celle attendue en moyenne pour les L1s appartenant aux familles L1PA1 à L1PA5 (8,5%)<sup>43</sup>, il semble que les L1s de taille complète soient privilégiés dans les interactions homologues. Il serait légitime de penser que la rétrotransposition, un mécanisme impliquant leur transcription, puisse favoriser leur implication dans la RH. Toutefois, un élément L1 complet n'est pas nécessairement actif, car la présence de mutations dans ses cadres de lecture peut empêcher l'expression des protéines nécessaires à sa rétrotransposition<sup>235</sup>. Il semblerait qu'un maximum de 60 éléments L1s soient toujours actifs en rétrotransposition dans le génome humain, et qu'ils appartiendraient à la sous-famille récente Ta<sup>236</sup>. Puisqu'aucun des partenaires de famille récente ne possède les caractéristiques de cette sous-famille et ne partage plus de 98% d'homologie avec un L1 actif consensus, il est très peu probable que leur utilisation privilégiée soit attribuable à leur rétrotransposition.

Alternativement, le fait qu'une grande proportion des partenaires sont de taille quasi complète pourrait s'expliquer par la distribution bimodale des L1s en fonction de leur taille. Une étude récente a en effet démontré que les éléments L1 du génome humain étaient soit très courts (< 1kb), soit très longs (> 5 kb)<sup>237</sup>. Puisque les éléments très courts ne possèdent pas la région de 23 pb servant à réparer la CDB, les éléments très longs deviendraient alors les plus nombreux pour réparer la cassure.

En considérant uniquement la région d'homologie commune avec le L1 du vecteur, on remarque que 78% des événements de RH ont impliqué des partenaires partageant une région d'homologie maximale (tableau XIV, p.146). Cette observation découle bien sûr en partie du fait qu'une grande proportion des partenaires impliqués soient des L1s complets, mais également d'une dépendance exponentielle entre l'étendue de l'homologie et la fréquence de gene targeting<sup>169</sup>.

Il serait difficile de nier qu'en général, le choix d'un partenaire de RH n'implique aucune préférence à l'égard du degré et de la longueur de l'homologie, puisqu'aucun L1 de famille ancestrale n'a été impliqué dans les événements de RH et qu'une préférence marquée semble être accordée aux éléments plus longs. Toutefois, dans notre système où des éléments de toute la gamme de longueur et d'homologie sont disponibles pour réparer la cassure, ces critères ne sont visiblement pas les plus importants dans le choix d'un partenaire.

## **.IV.5. Structure de la chromatine**

### **IV.5.1. Influence de l'activité transcriptionnelle**

Comme mentionné précédemment pour les loci d'intégration, on pourrait poser l'hypothèse que des L1s endogènes situés dans des régions transcriptionnellement actives puissent être plus facilement impliqués dans la RH, considérant la proximité de la machinerie de réparation et l'ouverture de la chromatine. Bien que les régions riches en gènes soient 4 fois plus pauvres en LINEs<sup>2</sup>, qui sont surtout concentrés dans les régions riches en A+T, il n'en demeure pas moins que 60% de tous les partenaires endogènes de RH ont été localisés à l'intérieur de gènes connus ou prédits (tableau XII, p.125).

De plus, parmi les 6 éléments pleine longueur utilisés comme partenaire de RH, 4 sont situés à l'intérieur de séquences de gènes, même si les L1s pleine longueur sont particulièrement sous-représentés dans les régions introniques, en conséquence d'une sélection négative contre leur intégration et leur pouvoir recombogène<sup>39</sup>. Nos résultats suggèrent par conséquent que l'activité transcriptionnelle de la chromatine puisse bel et bien favoriser l'activité de RH entre les séquences répétitives L1s. Toutefois, il serait intéressant de vérifier si les gènes où sont localisés nos partenaires endogènes sont actifs dans la lignée cellulaire GM05849. Pour ce faire, il suffirait de détecter la présence des ARNm par RT-PCR.

### **IV.5.2. Influence de la méthylation**

L'altération de la structure locale de la chromatine par la méthylation pourrait également restreindre l'accessibilité de l'ADN et perturber la RH. Des études ont effectivement démontré que l'hyperméthylation supprimait la RH intrachromosomique ainsi que le *crossing-over* allélique, soit en perturbant le pairage entre 2 séquences homologues avant l'initiation de la recombinaison, soit en empêchant le traitement adéquat des intermédiaires de recombinaison<sup>198</sup>. Il a

également été démontré qu'inversement, l'hypométhylation de l'ADN était accompagnée d'une augmentation de la recombinaison mitotique et des réarrangements chromosomiques, conduisant ultimement à l'instabilité du génome<sup>238</sup>. Il se pourrait que les éléments L1s choisis comme partenaires de RH dans notre modèle soient plus souvent situés dans des régions non-méthylées du génome, et il serait intéressant de le vérifier.

Le génome de la levure, qui est essentiellement constitué de séquences codantes et dépourvu de séquences répétitives ne subit pas de méthylation. Chez le champignon *Ascobolus immersus*, la méthylation est typiquement restreinte aux séquences répétitives de type Mars<sup>239</sup>. Dans le génome des mammifères, la plupart de la méthylation réside également dans les transposons<sup>200</sup>. Dans les cellules différenciées, la méthylation du promoteur des L1s supprimerait d'ailleurs leur activité de rétrotransposition, qui semble être généralement limitée aux cellules non-différenciées comme les cellules germinales<sup>240</sup> et certaines tumeurs<sup>241</sup>. Ces observations nous conduisent à l'hypothèse que la méthylation pourrait stabiliser le génome des eucaryotes supérieurs en supprimant l'activité de rétrotransposition et de RH des séquences répétitives L1s.

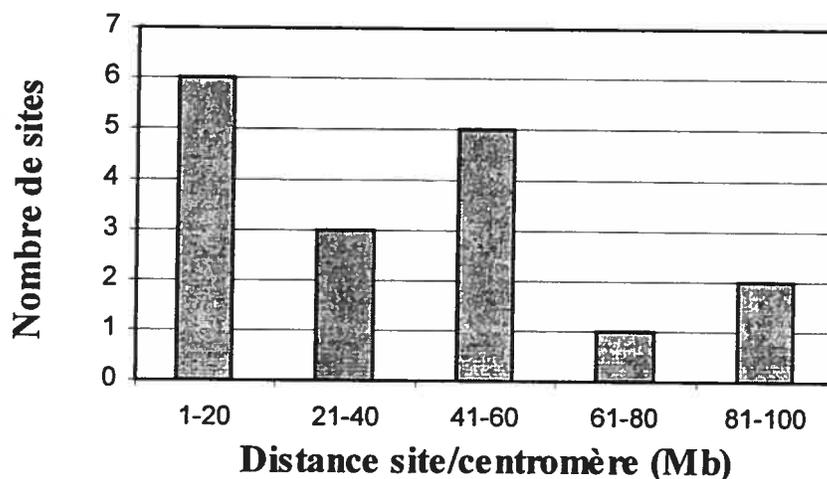
#### IV.5.3. Position télomérique/ centromérique

Tout au long des chromosomes, la structure de la chromatine est influencée par le contenu de l'ADN en gènes et en séquences répétitives. Les L1s sont dispersés uniformément sur les chromosomes, mais sont prédominants dans les bandes G/Q riches en nucléotides A+T et pauvres en gènes, et sont complètement absents du centromère<sup>44</sup>. La position d'un site d'intégration sur le chromosome pourrait donc significativement influencer ses interactions.

Dans chacun des clones, le vecteur s'est intégré aux 2/3 de la portion télomérique des bras chromosomiques 6q et 2p, laissant la liberté à la CDB d'être réparée par interaction avec autant de partenaires de RH ou de NHCR vers le télomère que vers le

centromère (figure 32, p.127 et figure 33, p.128). Cependant, il est frappant de constater que sur le bras court du chromosome 2, 96% des événements ont impliqué un site situé vers le télomère (figure 28, p.119), ce qui suggère une sélection en défaveur de la direction centromérique, qui pourrait être plus délétère par la perte du centromère. La comparaison des distances physiques et génétiques des chromosomes humains suggère en effet une suppression du taux de recombinaison près des centromères, ainsi qu'une augmentation marquée de ce taux dans les 20 à 30 derniers Mb des télomères<sup>2</sup>. Il semblerait également que le taux moyen de recombinaison diminue en fonction de la longueur du bras chromosomique, de façon à augmenter les chances de *crossing-over* sur les plus petits bras lors du brassage des gènes à la méiose<sup>2</sup>. Par contre, 76% des sites de recombinaison interchromosomique identifiés dans notre étude se retrouvent sur des bras longs.

Chez la levure et les mammifères, des études ont démontré que la RH ectopique entre des partenaires situés sur des chromosomes distincts était plus fréquente entre des partenaires localisés à une distance équivalente de leur centromère respectif<sup>142,181</sup>. Pour le clone ATA67, nous avons calculé la distance de chaque site d'interaction chromosomique par rapport au centromère du chromosome où il était localisé (figure 34, p.152). Situé à 55 Mb du centromère du chromosome 2, le site d'intégration du clone ATA67 partage une position équivalente avec 5 des 17 sites d'interaction interchromosomiques, soit 30% des sites. Mais puisqu'une proportion aussi importante de sites sont localisés à moins de 20 Mb du centromère, il ne semble pas y avoir de corrélation évidente à la distance relative du centromère et la préférence d'une interaction.



**Figure 34: Corrélation entre la position centromérique relative des sites d'interaction interchromosomiques et leur fréquence d'interaction**

#### IV.5.4. Sites fragiles

Les sites fragiles sont des régions instables, souvent associées à des bris chromosomiques et des réarrangements à l'origine de cancers. *In vitro*, l'ADN des sites fragiles est hautement recombino-gène<sup>191</sup>. Cette instabilité est induite par des agents qui retardent la réplication ou la réparation de l'ADN. Il a donc été proposé que les sites fragiles puissent être associés aux cassures d'ADN<sup>192</sup>. Selon cette hypothèse, ces sites seraient plus souvent la cible de CDBs et pourraient par conséquent être plus souvent impliqués dans la RH.

Pour vérifier cette hypothèse, nous avons examiné la fragilité des loci d'intégration ainsi que des sites de NHCR et de RH interchromosomiques (tableau XV, p.153). À l'aide de la base de données Mitelman<sup>242</sup>, nous avons calculé le nombre de réarrangements chromosomiques impliquant ces locus dans les néoplasies humaines. Malgré sa localisation dans un gène, le locus d'intégration 2p22 du clone ATA67 n'est associé qu'à un très faible nombre d'aberrations répertoriées. Par contre, le

a)

| NHCR   | Site     | Aberrations* |
|--------|----------|--------------|
| 34-t1  | 20p11.1  | 10           |
| 34-t2  | 1q24.3   | 6            |
| 34-t3  | 3q12.2   | 37           |
| 67-t1  | 8q12.1   | <b>75</b>    |
| 67-t2  | 13q21.33 | 45           |
| 67-t3  | 11q14.2  | <b>138</b>   |
| 67-t4  | 17q21.1  | <b>488</b>   |
| 67-t5  | 9q21.33  | 27           |
| 67-t6  | 9q33.3   | 13           |
| 67-t7  | Yp11.2   | 0            |
| 67-t8  | 7q22.1   | <b>612</b>   |
| 67-t9  | 6q25.1   | <b>124</b>   |
| 67-t10 | 1q31.3   | 24           |
| 67-t11 | 21q22.11 | <b>1456</b>  |
| 67-t12 | 1p12     | <b>67</b>    |
| 67-t13 | 14q21.3  | 7            |
| 67-t14 | 14q23.2  | 6            |
| 67-t18 | 1q25.2   | <b>69</b>    |

b)

| RH   | Site    | Aberrations* |
|------|---------|--------------|
| 34-A | 4q24    | 4            |
| 67-H | 3q26.32 | <b>366</b>   |
| 67-G | 11q22.1 | <b>69</b>    |
| 67-I | 11p13   | <b>171</b>   |
| 67-K | 11p14.1 | 31           |

**Tableau XV : Nombre d'aberrations chromosomiques associées à des néoplasies humaines pour chaque site de translocation (a) et de RH interchromosomique (b)**

Le nombre d'aberrations chromosomiques a été calculé à partir de la banque de données Mitelman (<http://cgap.nci.nih.gov/Chromosomes/Mitelman>). Les loci associés à plus de 50 aberrations sont identifiés en caractère gras. Les loci d'intégration 2p22 et 6q23 ont respectivement été la cible de 4 et de 290 aberrations chromosomiques répertoriées.

locus 6q23 est associé à près de 290 délétions. Il est donc surprenant que la fréquence de RH et de NHCR soit plus faible pour ce clone pour lequel le locus d'intégration semble plus instable et probablement plus affligé de CDBs.

Pour le clone ATA67, 58% des sites d'interaction sont situés dans un locus auquel plus de 50 cas d'aberrations ont été associés. Il pourrait donc y avoir une corrélation entre la fragilité d'un site et son accessibilité génomique. Dans les événements de RH générés avec le clone ATA67, nous avons identifié des partenaires interchromosomiques localisés sur les chromosomes 3q26 (67-H), 11q22 (67-G) et 11p13-14 (67-K et -I) (tableau IX, pp.94-95). Au locus 3q26, 366 cas de cancers sont principalement associés au réarrangement des gènes *EVL1*, *EAP1*, *MDS1* et *BCL6*, alors que le locus 11q22 est associé à 69 réarrangements chromosomiques.

Finalement, le locus 11p13 est associé à 171 cas de cancers impliquant surtout les gènes *RBTN1*, *RBTN2* et *WT1*. Associés à ce site particulièrement fragile, 2 cas de RH délétères entre des éléments L1s ont été rapportés et associés à des pathologies humaines, impliquant respectivement le réarrangement des gènes *PAX6*<sup>55</sup> et *WT1*<sup>56</sup>. Ces gènes ne sont situés qu'à 2,4 Mb du partenaire 67-I lui-même localisé à l'intérieur du gène *PDX1*.

#### **IV.6. Organisation tridimensionnelle du génome**

Pour chacun des clones, la majorité des partenaires de recombinaison ont été localisés sur le même chromosome que le locus d'intégration subissant la CDB (figure 27, p.118 et figure 28, p.119), donc à l'intérieur du territoire occupé par le chromosome. Le territoire chromosomique d'un locus apparaît donc comme étant le compartiment immédiat le plus accessible dans l'organisation tridimensionnelle du génome.

Nos résultats démontrent aussi qu'une CDB peut être réparée à une fréquence significative (20% pour ATA34 et 7% pour ATA67) par recombinaison homologue

avec des séquences situées sur un territoire distinct. Bien que le nombre d'événements de recombinaison interchromosomique récupérés ne soit pas suffisant pour porter des conclusions précises, il semblerait effectivement que l'organisation des territoires dans le noyau permette une accessibilité particulière entre certains territoires. Par exemple, nous avons observé 4 événements de recombinaison du locus ATA67 du chromosome 2 avec des loci situés sur le chromosome 11. Une étude avait antérieurement démontré dans un système *in vivo* que la localisation génomique d'un locus pouvait affecter sa susceptibilité à la recombinaison ectopique interchromosomique<sup>160</sup>. De plus, il existe maintenant quelques évidences de corrélation entre la juxtaposition de certains loci et leur interaction, pouvant conduire à des réarrangements délétères<sup>23, 24, 243</sup>.

À l'intérieur d'un territoire chromosomique, il semblerait que le bras court et le bras long occupent des domaines distincts qui ne se chevauchent pas<sup>28</sup>. Nos résultats sont consistants avec ce second degré d'organisation, puisque tous les partenaires de recombinaison intrachromosomique ont été localisés sur le bras du locus d'intégration (figure 27, p.118 et figure 28, p.119). L'autre bras chromosomique se semble pas être plus accessible que les autres territoires et pourrait même l'être moins que n'importe quel autre, puisque aucun site d'interaction n'y a été retrouvé. Il est pertinent de mentionner qu'à l'interphase, il n'y a pas de pairage préférentiel entre les chromosomes homologues<sup>244</sup> et qu'il pourrait même y avoir une répulsion entre eux<sup>8</sup>. Ainsi, bien qu'en principe les partenaires de recombinaison identifiés se retrouvent également sur l'autre chromosome homologue, il demeure beaucoup plus probable que le partenaire soit situé sur l'homologue contenant le locus d'intégration.

Sur le bras chromosomique de l'intégration, il apparaît évident que tous les sites n'ont pas une accessibilité équivalente. Sur le bras 2p du clone ATA67, plusieurs partenaires apparaissent regroupés à 11 Mb, 17 Mb et 25 Mb du locus d'intégration, suggérant une accessibilité privilégiée dans l'organisation tridimensionnelle du bras chromosomique (figure 28, p.119).

Nos résultats démontrent également qu'un environnement immédiat d'une taille de 1 à 2 Mb fournit à la CDB une grande proportion (56%) des partenaires de recombinaison du clone ATA67 (figure 30, p.122), et que cette région pourrait constituer un domaine subchromosomique dans l'organisation intraterritoriale. Dans ce domaine, les L1s identifiés comme partenaires ne sont cependant pas les seuls L1s possédant les séquences nécessaires à la réparation de la CDB (figure 29, p.121 et figure 30, p.122). Plusieurs de ces éléments ont été localisés beaucoup plus près du site d'intégration, indiquant clairement que le choix d'un partenaire homologue ne s'arrête pas au premier rencontré sur le chromosome linéaire. Bien que la plupart des autres L1s non utilisés appartiennent à des familles ancestrales, certains font partie d'une famille récente et sont même plus proximaux, plus homologues, plus longs et dans la même orientation relative que la plupart des L1s choisis comme partenaires. Ces observations viennent confirmer une fois de plus que l'homologie, la taille et la proximité linéaire ne sont que des critères secondaires dans le choix d'un partenaire de RH.

À l'intérieur des foci de 1 Mb, la chromatine serait repliée sous forme de boucles de chromatine de l'ordre de 30 à 200 kb<sup>7</sup>, dans lesquelles les gènes actifs et inactifs transcriptionnellement auraient des préférences de localisation. Ainsi, les loci d'une boucle ne seraient pas tous accessibles entre eux et certaines interactions seraient privilégiées. Nos résultats sont une fois de plus consistants avec ce dernier niveau d'organisation de la chromatine, puisque des régions spécifiques se révèlent plus accessibles que d'autres à la cassure dans l'environnement immédiat du locus (figure 29, p.121 et figure 30, p.122).

Afin de bien démontrer la proximité physique privilégiée du locus d'intégration avec certains loci du génome situés à l'intérieur du même territoire ou sur un territoire distinct, la meilleure technique serait celle du FISH (*fluorescence in situ hybridization*). Dans les cellules en interphase, nous pourrions démontrer par exemple que le locus 2p22 se retrouve fréquemment à proximité du locus 2p24.1 où sont localisés 2 sites de NHCR et un partenaire de RH impliqués dans 12 événements de réparation. Nous pourrions également démontrer que le territoire du chromosome

2 est situé à proximité de celui du chromosome 11, et que les locus 11p13-14, 11q14 et 11q22 où sont localisés plusieurs partenaires de recombinaison se retrouvent à proximité du locus d'intégration 2p22. Ces loci pourraient se côtoyer près d'un espace interchromatique où la machinerie de réparation serait plus accessible, reflétant la fréquence particulièrement élevée de RH du clone ATA67. En utilisant une sonde localisée dans les séquences plasmidiques du vecteur et une autre localisée dans une portion génomique unique séquencée en 3' des événements de *crossing-over* ou de *one-sided invasion*, nous pourrions démontrer de façon intéressante une proximité tridimensionnelle entre les partenaires de RH impliqués.

Dans notre modèle d'étude, chaque événement de recombinaison a été produit dans une cellule différente. Nous avons démontré que plusieurs sites génomiques pouvaient interagir avec chacun des locus d'intégration, et donc que l'accessibilité génomique possédait un certain dynamisme. Toutefois, puisque des sites d'interaction très privilégiés ont été identifiés, il semblerait que le génome adopte une configuration préférentielle qui pourrait être conservée chez la plupart des cellules à l'interphase. Ce patron d'accessibilité semble être défini par l'organisation tridimensionnelle du génome, pour laquelle une conservation pourrait aussi exister<sup>36-38</sup>. La conservation des différents niveaux d'organisation et du patron d'accessibilité génomique suggère un véritable rôle fonctionnel de cette configuration dans le contrôle des processus cellulaires et le maintien de l'intégrité génomique.

## V. CONCLUSION

Les résultats obtenus dans le cadre de mon projet de doctorat ont des répercussions dans d'importantes sphères de recherche en biologie moléculaire. D'abord, notre étude de l'accessibilité génomique révèle que les interactions entre les différentes régions du génome sont finement contrôlées par les multiples niveaux d'organisation de la chromatine dans le noyau en interphase, et que ce contrôle aurait une importance fonctionnelle dans le maintien de l'intégrité génomique.

Par un modèle d'interactions homologues entre des séquences répétitives naturelles du génome, nous avons démontré pour deux loci distincts que le patron d'accessibilité d'un locus était conservé d'une cellule à l'autre et dicté par l'organisation tridimensionnelle des territoires chromosomiques, plutôt que par la proximité linéaire d'un locus sur le chromosome. L'accessibilité serait également variable en fonction de la position des sites sur le chromosome par rapport au centromère et aux télomères, et de l'état transcriptionnel local de la chromatine. Notre étude des interactions interchromosomiques suggère également qu'une proximité particulière pourrait exister entre les domaines de territoires chromosomiques distincts, et ouvre la voie à l'établissement d'une cartographie tridimensionnelle détaillée de l'accessibilité génomique.

Par ce même système d'étude, nous avons ensuite démontré que la recombinaison ectopique entre les séquences répétitives dispersées était un phénomène relativement fréquent dans le génome humain. Nos résultats démontrent clairement que ces interactions, qui peuvent avoir des conséquences néfastes sur l'intégrité du génome, ne sont pas inhibés par la divergence de séquence. En effet, des partenaires partageant de 92 à 98% d'homologie peuvent interagir par RH, à une fréquence indépendante de leur homologie globale et de la longueur de l'homologie parfaite partagée. La permissivité des interactions homologues ectopiques serait davantage contrôlée par l'organisation tridimensionnelle du génome.

Nous avons également démontré que les interactions homéologues entre les séquences répétitives pouvaient menacer l'intégrité génomique. Puisque l'activité transcriptionnelle de la chromatine semble particulièrement favoriser les interactions homologues, il devient primordial pour la cellule de protéger l'intégrité des séquences codantes en minimisant les risques de réarrangements délétères générés par la recombinaison entre les séquences répétitives qui occupent ces régions. Chez une lignée déficiente dans le contrôle de la RH, nous avons observé de nombreuses interactions impliquant des éléments répétitifs localisés à l'intérieur de gènes, certaines conduisant à des aberrations chromosomiques.

Devant l'exposition constante de notre génome aux dommages à l'ADN et l'implication de la déficience des mécanismes de réparation dans une grande proportion de cancers, notre étude de la réparation des CDBs a également un apport important dans la compréhension de l'instabilité génomique. Nos résultats nous permettent d'abord de suggérer que les sites fragiles du génome, qui sont des régions plus accessibles aux nucléases et donc plus susceptibles aux CDBs, pourraient favoriser l'implication des séquences répétitives locales dans des événements de RH.

Puisque le tiers de notre génome est composé de séquences répétitives dispersées, les CDBs vont très souvent subvenir à l'intérieur de celles-ci. Nous avons donc induit artificiellement une CDB à l'intérieur d'un élément répétitif intégré au génome, pour en étudier le mécanisme de réparation. Dans notre système où tous les types d'événements pouvaient être récupérés, nous avons démontré qu'une CDB induite dans un élément répétitif pouvait être réparée avec un partenaire homologue endogène, par un mécanisme conservatif non-réciproque de conversion génique. L'analyse des segments de conversion générés lors de ces événements nous suggère l'utilisation d'un mécanisme de conversion unidirectionnel impliquant l'invasion de l'hétéroduplex par une seule extrémité. Le mécanisme du *one-sided invasion* pourrait être privilégié lors de la RH entre des séquences divergentes.

Dans une lignée cellulaire déficiente dans le contrôle de la RH, nous avons également observé quelques événements de réparation de CDBs par un mécanisme de type *crossing-over* conduisant à des réarrangements aberrants. Dans un contexte cellulaire instable, ce type d'événements délétères entre les séquences répétitives du génome pourrait menacer significativement l'intégrité génomique et contribuer au processus de tumorigénèse.

Premier en son genre, notre modèle d'étude ne nécessite pas la reconstitution d'un marqueur de sélection, n'induit pas de biais en fonction d'un mécanisme d'interaction particulier et laisse le choix d'une interaction avec une grande variété de partenaires endogènes, qui peuvent être identifiés et localisés grâce à la disponibilité de la séquence complète du génome humain. Pour tous les aspects étudiés, notre modèle reflète beaucoup plus clairement la réalité cellulaire.

## VI. RÉFÉRENCES

1. Watson, J. D. & Crick, F. H. Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. *Nature*. **171**, 737-738 (1953).
2. International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature*. **409**, 860-921 (2001).
3. Rabl, C. Über Zelltheilung. *Morphologisches Jahrbuch*. **10**, 214-330 (1885).
4. Boveri, T. Die BLAST omerenkerne von *Ascaris megalocephala* und die theorie der chromosomenindividualität. *Arch Zellforschung*. **3**, 181-286 (1909).
5. Cremer, T., Baumann, H., Nakanishi, K. & Cremer, C. Correlation between interphase and metaphase chromosome arrangements as studied by laser-UV-microbeam experiments. *Chrom. Today*. **8**, 203-212 (1984).
6. Cremer, T., *et al.* Role of chromosome territories in the functional compartmentalization of the cell nucleus. *Cold Spring Harb. Symp. Quant. Biol.* **58**, 777-792 (1993).
7. Cremer, T. *et al.* Chromosome territories, interchromatin domain compartment, and nuclear matrix: an integrated view of the functional nuclear architecture. *Crit. Rev. Eucaryotic Gene Expression*. **12**, 179-212 (2000).
8. Cremer, T. & Cremer, C. Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat. Rev. Genet.* **2**, 292-301 (2001).
9. Zirbel, R. M., Mathieu, U. R., Kurz, A., Cremer, T. & Lichter, P. Evidence for a nuclear compartment of transcription and splicing located at chromosome domain boundaries. *Chrom. Res.* **1**, 93-106 (1993).
10. Belmont, A. S., Dietzel, S., Nye, A. C., Strukov, Y. G. & Tumbar, T. Large-scale chromatin structure and function. *Curr. Opin. Cell Biol.* **11**, 307-311 (1999).
11. Visser, A. E., Jaunin, F., Fakan, S. & Aten, J. A. High resolution analysis of interphase chromosome domains. *J. Cell Sci.* **113**, 2585-2593 (2000).
12. Sun, H. B., Shen, J. & Yokota, H. Size-dependant positioning of human chromosomes in interphase nuclei. *Biophys. J.* **79**, 184-190 (2000).
13. Croft, J. A. *et al.* Differences in the localization and morphology of chromosomes in the human nucleus. *J. Cell Biol.* **145**, 1119-1131 (1999).

14. Andrulis, E. D., Neiman, A. M., Zappulla, D. C. & Sternglanz, R. Perinuclear localization of chromatin facilitates transcriptional silencing. *Nature*. **394**, 592-595 (1998).
15. Chubb, J. R., Boyle, S., Perry, P. & Bickmore, W. A. Chromatin motion is constrained by association with nuclear compartments in human cells. *Curr. Biol.* **12**, 439-445 (2002).
16. Hsu, T. C. A possible function of constitutive heterochromatin: the bodyguard hypothesis. *Genetics*. **79** (suppl.), 137-150 (1975).
17. Löbrich, M., Cooper, P. K. & Rydberg, B. Non-random distribution of DNA double-strand breaks induced by particle irradiation. *Int. J. Radiat. Biol.* **70**, 493-503 (1996).
18. Folle, G. A., Martinez-Lopez, W., Boccardo, E. & Obe, G. Localization of chromosomal breakpoints: implication of the chromatin structure and nuclear architecture. *Mutat. Res.* **404**, 17-26 (1998).
19. Bornfleth, H., Edelmann, P., Zink, D., Cremer, T. & Cremer, C. Quantitative motion analysis of subchromosomal foci in living cells using four-dimensional microscopy. *Biophys. J.* **77**, 2871-2886 (1999).
20. DeBoni, U. The interphase nucleus as a dynamic structure. *Int. Rev. Cytol.* **150**, 149-171 (1994).
21. Borden, J. & Manuelidis, L. Movement of the X chromosome in epilepsy. *Science*. **242**, 1687-1691 (1988).
22. Vasquez, J., Belmont, A. S. & Sedat, J. W. Multiple regimes of constrained chromosome motion are regulated in the interphase *Drosophila* nucleus. *Curr. Biol.* **11**, 1227-1239 (2001).
23. Kozubek, S. *et al.* Distribution of ABL and BCR genes in cell nuclei of normal and irradiated lymphocytes. *Blood*. **12**, 4537-4545 (1997).
24. Neves, H., Ramos, C., da Silva, M. G., Parreira, A. & Parreira, L. The nuclear topography of ABL, BCR, PML, and RARalpha genes: evidence for gene proximity in specific phases of the cell cycle and stages of hematopoietic differentiation. *Blood*. **94**, 1137-1138 (1999).
25. Tanabe, H., Habermann, F. A., Solovei, I., Cremer, M. & Cremer, T. Non-random radial arrangements of interphase chromosome territories: evolutionary considerations and functional implications. *Mut. Res.* **505**, 37-45 (2000).

26. Skalnikova, M. *et al.* Spatial arrangement of genes, centromeres and chromosomes in human blood cell nuclei and its changes during the cell cycle, differentiation and after irradiation. *Chrom. Res.* **8**, 487-499 (2000).
27. Kozubek, S. *et al.* Spatial distribution of selected genetic loci in nuclei of human leukemia cells after irradiation. *Radiat. Res.* **155**, 311-319 (2001).
28. Dietzel, S. *et al.* Separate and variably shaped chromosome arm domains are disclosed by chromosome arm painting in human cell nuclei. *Chrom. Res.* **6**, 25-33 (1998).
29. Volpi, E. V. *et al.* Large-scale chromatin organization of the major histocompatibility complex and the other regions of human chromosome 6 and its response to interferon in interphase nuclei. *J. Cell. Sci.* **113**, 1565-1576 (2000).
30. Zink, D., Bornfleth, H., Visser, A., Cremer, C. & Cremer, T. Organization of early and late replicating DNA in human chromosome territories. *Exp. Cell Res.* **247**, 176-188 (1999).
31. Jackson, D. A. & Pombo, A. Replicon clusters are stable units of chromosome structure: evidence that nuclear organization contribute to the efficient activation and propagation of S phase in human cells. *J. Cell. Biol.* **140**, 1285-1295 (1998).
32. Sparvoli, E., Levi, M. & Rossi, E. Replicon clusters may form structurally stable complexes of chromatin and chromosomes. *J. Cell Sci.* **107**, 3097-3103 (1994).
33. Mahy, L. N., Perry, P. & Bickmore, W. A. Gene density and transcription influence the localization of chromatin outside of chromosome territories detectable by FISH. *J. Cell Biol.* **159**, 753-763 (2002).
34. Williams, R. R. E., Broad, S., Sheer, D. & Ragoussis, J. Subchromosomal positioning of the epidermal differentiation complex (EDC) in keratinocyte and lymphoblast interphase nuclei. *Exp. Cell Res.* **272**, 163-175 (2002).
35. Münkkel, C. *et al.* Compartmentalization of interphase chromosomes observed in stimulation and experiment. *J. Mol. Biol.* **285**, 1053-1065 (1999).
36. Schmid, M. *et al.* First report on chicken genes and chromosomes 2000. *Cytogenet. Cell Genet.* **90**, 169-218 (2000).
37. Tanabe, H., *et al.* Evolutionary conservation of chromosome territory arrangements in cell nuclei from higher primates. *Proc. Natl. Acad. Sci. USA.* **99**, 4424-4429 (2002).

38. Mahy, N. L., Perry, P. E., Gilchrist, S., Baldock, R. A. & Bickmore, W. A. Spatial organization of active and inactive genes and non-coding DNA within chromosome territories. *J. Cell. Biol.* **157**, 579-589 (2002).
39. Smith, A. F. A. Interspersed repeats and other mementos of transposable elements in mammalian genomes. *Curr. Opin. Gene. Dev.* **9**, 657-663 (1999).
40. Prak, E. L. & Haig, H. K. Jr. Mobile elements and the human genome. *Nature Rev. Genet.* **1**, 134-144 (2000).
41. Feng, Q., Moran, J. V., Kazazian, H. H. & Boeke, J. D. Human L1 retrotransposition encodes a conserved endonuclease required for retrotransposition. *Cell.* **87**, 905-916 (1996).
42. Kazazian, Jr. H. H. An estimated frequency of endogenous insertional mutations in humans. *Nature Genet.* **22**, 130 (1999).
43. Boissinot, S., Entezam, A. & Furano, A. V. Selection against deleterious LINE-1 containing loci in the human lineage. *Mol. Biol. Evol.* **18**, 926-935 (2001).
44. Korenberg, J. R. & Rykowski, M. C. Human genome organisation: alu, lines, and the molecular structure of metaphase chromosome bands. *Cell.* **53**, 391-400 (1988).
45. Bolye, A. L. Ballard, S. G. & Ward, D. C. Differential distribution of long and short interspersed element sequences in the mouse genome: chromosome karyotyping by fluorescence *in situ* hybridization. *Proc. Natl. Acad. Sci. USA.* **87**, 7757-7761 (1990).
46. Narita, N., Nishio, H., Kitoh, Y., Ishikawa, Y. & Minami, R. Insertion of a 5' truncated L1 element into the 3' end of exon 44 of the dystrophin gene resulted in skipping of the exon during splicing in a case of Duchenne muscular dystrophy. *J. Clin. Invest.* **91**, 1862-1867 (1993).
47. Kazazian, H. H. Jr., *et al.* Haemophilia A resulting from de novo insertion of L1 sequences represents a novel mechanism for mutation in man. *Nature.* **332**, 164-166 (1988).
48. Miki, Y., Katagiri, T., Kasumi, F., Yoshimoto, T. & Nakamura, Y. Mutation analysis in the BRCA2 gene in primary breast cancers. *Nat. Genet.* **13**, 245-247 (1996).
49. Miki, Y. *et al.* Disruption of the APC gene by retrotransposal insertion of L1 sequence in a colon cancer. *Cancer Res.* **52**, 643-645 (1992).

50. Moran, J. V., DeBerardinis, R. J., Haig, H. & Kazazian, H. H. Jr. Exon shuffling by L1 retrotransposition. *Science*. **283**, 1530-1534 (1999).
51. Pickeral, O. K., Makalowski, W., Boguski, M. S. & Boeke, J. D. Frequent human genomic DNA transduction driven by LINE-1 retrotransposition. *Genome Res.* **10**, 411-415 (2000).
52. Kidwell, M. G. & Lisch, D. R. Perspective: Transposable elements, parasitic DNA, and genome evolution. *Int. J. Org. Evol.* **55**, 1-24 (2001).
53. Burwinkel, B. & Kilimann, M. Unequal homologous recombination between LINE-1 elements as a mutational mechanism in human genetic disease. *J. Mol. Biol.* **277**, 513-517 (1998).
54. Inoue, H. *et al.* Sequence of the FRA3B common fragile region: Implication for the mechanism of FHIT deletion. *Proc. Natl. Acad. Sci. USA.* **94**, 14584-14589 (1997).
55. Drechsler, M. & Royer-Pokora, B. A LINE element is present at the site of a 300-kb deletion starting in intron 10 of the PAX6 gene in a case of familial aniridia. *Hum. Genet.* **98**, 297-303 (1996).
56. Liu, J. *et al.* LINE-1 element insertion at the t(11:22) translocation breakpoint of a desmoplastic small round cell tumor. *Gene Chr. Cancer.* **18**, 232-239 (1997).
57. Segal, Y. *et al.* LINE-1 elements at the sites of molecular rearrangements in Alport Syndrome-Diffuse Leiomyomatosis. *Am. J. Hum. Genet.* **64**, 62-69 (1999).
58. Öner, C. *et al.* Molecular analysis of the Turkish form of deletion-inversion ( $\delta\beta$ )<sup>o</sup> thalassaemia. *British J. Hematology.* **96**, 229-234 (1997).
59. Fitch, D. H. A. *et al.* Duplication of the  $\gamma$ -globin gene mediated by L1 long interspersed repetitive elements in a early ancestor of simian primates. *Proc. Natl. Acad. Sci. USA.* **88**, 7396-7400 (1991).
60. Schwartz, A. *et al.* Reconstructing hominid Y evolution: X-homologous block, created by X-Y transposition, was disrupted by Yp inversion through LINE-LINE recombination. *Human Mol. Genet.* **7**, 1-11 (1998).
61. Ardlie, K. *et al.* Lower-than expected linkage disequilibrium between tightly linked markers in humans suggests a role for gene conversion. *Am. J. Hum. Genet.* **69**, 582-589 (2001).

62. Frisse, L. *et al.* Gene conversion and different population histories may explain the contrast between polymorphism and linkage disequilibrium levels. *Am. J. Hum. Genet.* **69**, 831-843 (2001).
63. Gauthier, J. Analyse de l'évolution des éléments répétitifs de type LINE-1 chez l'humain. *Mémoire de maîtrise. Université de Montréal* (2001).
64. Roy, A. M. *et al.* Potential gene conversion and source genes for recently integrated Alu elements. *Genome Res.* **10**, 1485-1495 (2000).
65. Lieber, M. R. Pathological and physiological double-strand breaks. *Am. J. Pathology.* **153**, 1323-1332 (1998).
66. Khanna, K. K. & Jackson, S. P. DNA double-strand breaks: signaling, repair and the cancer connection. *Nat. Genet.* **27**, 247-254 (2001).
67. Rogakou, E. P., Pilch, D. R., Orr, A. H., Inova, V. S. & Bonner, W. M. DNA double-strand breaks induced histone H2AX phosphorylation on serine 139. *J. Biol. Chem.* **273**, 5858-5868 (1998).
68. Rogakou, E. P., Boon, C., Redon, C. & Bonner, W. M. Megabase chromatin domains involved in DNA double strand breaks *in vivo*. *J. Cell. Biol.* **146**, 905-916 (1999).
69. Modesti, M. & Kannar, R. DNA repair: spot (lights) on chromatin. *Curr. Biol.* **11**, R229-R232 (2001).
70. Paull, T. T. *et al.* A critical role for histone H2AX in recruitment of repair factors to nuclear foci after DNA damage. *Curr. Biol.* **10**, 886-895 (2000).
71. Shiloh, Y. ATM and ATR: networking cellular responses to DNA damage. *Curr. Opin. Genet. Dev.* **11**, 71-77 (2001).
72. Kim, S. T., Lim, D. S., Canman, C. E. & Kastan, M. B. Substrates specificities and identification of putative substrates of ATM kinase family members. *J. Biol. Chem.* **274**, 37 538-37 543 (1999).
73. Rotman, G. & Shiloh, Y. ATM: from gene to function. *Hum. Mol. Genet.* **7**, 1555-1563 (1998).
74. Chen, G. & Lee, E. Y. P. The product of the ATM gene is a 370-kDa nuclear phosphoprotein. *J. Biol. Chem.* **271**, 33 693-33 697 (1996).
75. Keegan, K. S. *et al.* The Atr and Atm protein kinases associate with different sites along meiotically pairing chromosomes. *Genes Dev.* **10**, 2423-2437 (1996).

76. Canman, C. E. *et al.* Activation of the ATM kinase by ionizing radiation and phosphorylation of p53. *Science*. **281**, 1677-1679 (1998).
77. Smith, G. C. M. *et al.* Purification and DNA binding properties of the ataxia-telangiectasia gene product ATM. *Proc. Natl. Acad. Sci. USA*. **96**, 11 134-11 139 (1999).
78. Suzuki, K., Kodoma, S. & Watanabe, M. Recruitment of ATM protein to double-strand DNA irradiated with ionizing radiation. *J. Biol. Chem.* **274**, 25 571-25 575 (1999).
79. Lim, D. -S. *et al.* Atm binds  $\beta$ -adaplin: a novel explanation for pleiotropic abnormalities in ataxia telangiectasia. *Proc. Natl. Acad. Sci. USA*. **95**, 10 146-10 151 (1998).
80. Xia, S. J. Shammass, M. A. & Shmookler Reis, R. J. Reduced telomere length in ataxia telangiectasia fibroblasts. *Mutat. Res.* **364**, 1-11 (1996).
81. Metcalfe, J. A. *et al.* Accelerated telomere shortening in ataxia telangiectasia. *Nat. Genet.* **13**, 350-353 (1996).
82. Pâques, F. & Haber, J. E. Multiple pathways of recombination induced by double-strand breaks in *Saccharomyces cerevisiae*. *Microbiol. Mol. Biol. Rev.* **63**, 349-404 (1999).
83. Liang, F., Mingguang, H., Romanienko, P. J. & Jasin, M. Homology-directed repeat is a major double-strand break repair pathway in mammalian cells. *Proc. Natl. Acad. Sci. USA*. **95**, 5172-5177 (1998).
84. Nelms, B. E., Maser, R. S., MacKay, J. F., Lagally, M. G. & Petrini, J. H. J. In situ visualization of DNA double-strand break repair in human fibroblasts. *Science*. **280**, 590-592 (1998).
85. Difilippantonio, M. J. *et al.* DNA repair protein Ku80 suppresses chromosomal aberrations and malignant transformation. *Nature*. **404**, 510-514 (1985).
86. Gillert, E. T. *et al.* A DNA damage repair mechanism is involved in the origin of chromosomal translocations t(4 ;11) in primary leukemic cells. *Oncogene*. **18**, 4663-4671 (1999).
87. Dasika, G. K. *et al.* DANN damage-induced cell cycle checkpoints and DNA strand-break repair in development and tumorigenesis. *Oncogene*. **18**, 7883-7899 (1999).
88. Jackson, S. P. Sensing and repairing DNA double-strand breaks. *Carcinogenesis*. **5**, 687-696 (2002).

89. Gao, Y. J. *et al.* Interplay of p53 and DNA repair protein XRCC4 in tumorigenesis genomic stability and development. *Nature*. **404**, 897-900 (2000).
90. Frank, K. M. DNA ligase IV deficiency in mice leads to defective neurogenesis and embryonic lethality via the p53 pathway. *Mol. Cell*. **5**, 993-1002 (2000).
91. Lee, Y., Barnes, D. E., Lindahl, T. & McKinnon, P. J. Defective neurogenesis resulting from DNA ligase IV deficiency requires ATM. *Genes Dev*. **14**, 2576-2580 (2000).
92. Jackson, D. A. & Pombo, A. Replicon clusters are stable units of chromosome structure: evidence that nuclear organization contribute to the efficient activation and propagation of S phase in human cells. *J. Cell. Biol.* **140**, 1285-1295 (1998).
93. North, P., Ganesh, A. & Thacker, J. The rejoining of double-strand breaks in DNA by human cell extracts. *Nucleic Acid Res.* **18**, 6205-6210 (1990).
94. Li, Y. *et al.* Expression of ATM in Ataxia Telangiectasia fibroblasts rescues defects in double-strand break repair in nuclear extracts. *Envir. Mol. Mut.* **37**, 128-140 (2001).
95. Luo, C. *et al.* High frequency and error-prone DNA recombination in Ataxia Telangiectasia cell lines. *J. Biol. Chem.* **271**, 4497-4503 (1996).
96. Thacker, J. The use of integrating DNA vectors to analyse the molecular defects in ionising radiation-sensitive mutants of mammalian cells including ataxia-telangiectasia. *Mutat. Res.* **220**, 187-204 (1989).
97. Powell, S., Whitaker, S., Peacock, J. & McMillan, T. Ataxia telangiectasia: an investigation of the repair defect in the cell line AT5B1VA by plasmid reconstitution. *Mutat. Res.* **294**, 9-20 (1993).
98. Ganesh, A., North, P. & Thacker, J. Repair and misrepair of site-specific DNA double-strand breaks by human cell extracts. *Mutat. Res.* **299**, 251-259 (1993).
99. Dar, M. E., Winters, T. A. & Jorgensen, T. J. Identification of defective illegitimate recombinational repair of oxidatively induced DNA double-strand breaks in ataxia telangiectasia cells. *Mutat. Res.* **384**, 169-179 (1997).
100. Kojis, T. L., Gatti, R. A. & Sparkes, R. S. The cytogenetics of ataxia telangiectasia. *Cancer Genet. Cytogenet.* **56**, 143-156 (1991).

101. Cromie, G. A., Connelley, J. C. & Leach, D. R. F. Recombination at double-strand breaks and DNA ends : conserved mechanisms from phage to humans. *Mol. Cell.* **8**, 1163-1174 (2001).
102. Wood, R. D., Mitchell, M., Sgouros, J. & Lindahl, T. Human DNA repair genes. *Science.* **291**, 1284-1289 (2001).
103. Van Gent, D. C., Hoeijmakers, J. H. J. & Kanaar, R. Chromosomal stability and the DNA double-stranded break connection. *Nat. Genet. Rev.* **2**, 196-206 (2001).
104. Van Dyck, E., Stasiak, A. Z., Stasiak, A. & West, S. C. Binding of double strand-breaks DNA human Rad52 protein. *Nature.* **398**, 728-731 (1999).
105. Chen, J. J., Silver, D., Cantor, S., Livingstone, D. M. & Scully, R. BRAC1, BRCA2 and Rad51 operate in a common DNA damage response pathway. *Cancer Res.* **59**, S1752-S1756 (1999).
106. Moynahan, M. E., Pierce, A. J. & Jasin, M. BRCA2 is required for homology-directed repair of chromosomal breaks. *Mol. Cell.* **7**, 263-272 (2001).
107. Xia, F. *et al.* Deficiency of human BRCA2 leads to impaired homologous recombination but maintains normal nonhomologous end joining. *Proc. Natl. Acad. Sci. USA.* **98**, 8644-8649 (2001).
108. Moynahan, M. E., Chiu, J. W., Koller, B. H. & Jasin, M. BRCA1 controls homology-directed DNA repair. *Mol. Cell.* **4**, 511-518 (1999).
109. Takata, M. *et al.* Chromosome instability and defective recombinational repair in knockout mutants of the five Rad51 paralogs. *Mol. Cell. Biol.* **21**, 2858-2866 (2001).
110. Szostak, J. W., Orr-Weaver, T. L., Rothstein, R. J. & Stahl, F. W. The double-strand break repair model for recombination. *Cell.* **33**, 25-35 (1983).
111. Nelson, H. H., Sweetser, D. B. & Nickoloff, J. A. Effects of terminal nonhomology and homeology on double-strand-induced gene conversion tract directionality. *Mol. Cell. Biol.* **16**, 2951-2957 (1996).
112. Sweetser, D. B., Hough, H., Whelden, J. F., Arbuckle, M. & Nickoloff, J. A. Fine-resolution mapping of spontaneous and double-strand break-induced gene conversion tracts in *Saccharomyces cerevisiae* reveals reversible mitotic conversion polarity. *Mol. Cell. Biol.* **14**, 3863-3875 (1994).

113. Palmer, S., Schildkraut, E., Lazarin, R., Nguyen, J. & Nickoloff, J. A. Gene conversion tracts in *Saccharomyces cerevisiae* can be extremely short and highly directional. *Nucleic Acid Res.* **31**, 1164-1173 (2003).
114. Cho, J. W., Khalsa, G. J. & Nickoloff, J. A. Gene-conversion tract directionality is influenced by the chromosome environment. *Curr. Genet.* **34**, 269-279 (1998).
115. Taghian, D. G. & Nickoloff, J. A. Chromosomal double-strand breaks induce gene conversion at high frequency in mammalian cells. *Mol. Cell. Biol.* **17**, 6386-6393 (1997).
116. Yang, D. & Waldman, A. S. Fine-resolution analysis of products of intrachromosomal homeologous recombination in mammalian cells. *Mol. Cell. Biol.* **17**, 3614-3628 (1997).
117. Elliot, B., Richardson, C., Winderbaum, J., Nickoloff, J. A. & Jasin, M. Gene conversion tracts from double strand break repair in mammalian cells. *Mol. Cell. Biol.* **18**, 93-101 (1998).
118. Li, J. & Baker, M. D. Formation and repair of heteroduplex DNA on both sides of the double-strand break during mammalian gene targeting. *J. Mol. Biol.* **295**, 505-516 (2000).
119. Belmaaza, A., Wallenburg, J. C., Brouillette, S., Gusew, N. & Chartrand, P. Genetic exchange between endogenous and exogenous LINE-1 repetitive elements in mouse cells. *Nucleic Acid Res.* **18**, 6385-6391 (1990).
120. Belmaaza, A. & Chartrand, P. One-sided invasion events in homologous recombination at double-strand breaks. *Mut. Res.* **314**, 199-208 (1994).
121. Lin, F.-L., Sperle, K. & Stenberg, N. Model for homologous recombination during transfer of DNA into mouse L cells; role for DNA ends in the recombination process. *Mol. Cell. Biol.* **4**, 1020-1034 (1984).
122. Karan, P. DNA double-strand break repair in mammalian cells. *Curr. Opin. Gene. Dev.* **10**, 144-150 (2000).
123. Bishop, A. J. R. & Schiestl, R. H. Homologous recombination as a mechanism for genome rearrangements: environmental and genetic effects. *Hum. Mol. Genet.* **9**, 2427-2434 (2000).
124. Meyn, M. S. High spontaneous intrachromosomal recombination rates in ataxia-telangiectasia. *Science.* **260**, 1327-1330 (1993).

125. Tremblay, A., Jasin, M. & Chartrand, P. A double-strand break in a chromosomal LINE element can be repaired by gene conversion with various endogenous LINE elements in mouse cells. *Mol. Cell. Biol.* **20**, 54-60 (2000).
126. Swift, M., Reitnauer, P. J., Morrell, D. & Chase, C. L. Breast and other cancers in families with ataxia-telangiectasia. *N. Engl. Med.* **316**, 1289-1294 (1987).
127. Hecht, F. & Hecht, B. K. Cancer in ataxia-telangiectasia patients. *Cancer Genet. Cytogenet.* **46**, 9-19 (1990).
128. Keldysh, P. L. *et al.* 11q deletions in human colorectal carcinomas: cytogenetics and restriction fragment length polymorphism analysis. *Genes Chr. Cancer.* **6**, 45-50 (1993).
129. Carter, S. L. *et al.* Loss of heterozygosity at 11q22-q23 in breast cancer. *Cancer Res.* **54**, 6270-6274 (1994).
130. Pejovic, T. Genetic changes in ovarian cancer. *Ann. Med.* **27**, 73-78 (1995).
131. Stürzbecher, H.-W., Dondelmann, B., Henning, W., Knippschild, U. & Buchhop, S. P53 is linked directly to homologous recombination processes via RAD51/RecA protein interaction. *EMBO J.* **15**, 1992-2002 (1996).
132. Buchhop, S. *et al.* Interaction of p53 with the human Rad51 protein. *Nucl. Acids Res.* **25**, 3868-3874 (1997).
133. Mekeel, K. L. *et al.* Inactivation of p53 results in high rates of homologous recombination. *Oncogene.* **14**, 1847-1857 (1997).
134. Bertrand, P. *et al.* Increase of spontaneous intrachromosomal homologous recombination in mammalian cells expressing a mutant p53 protein. *Oncogene.* **14**, 1117-1122 (1997).
135. Willers, H., McCarthy, E. E., Alberti, W., Dahm-Daphi, J. & Powell, S. N. Loss of wild-type p53 function is responsible for upregulated homologous recombination in immortal rodent fibroblasts. *Int. J. Radiat. Biol.* **76**, 1055-1062 (2000).
136. Donehower, L. A. *et al.* Mice deficient for p53 are developmentally normal but susceptible to spontaneous tumours. *Nature.* **356**, 215-221 (1992).
137. Game, J. C. DNA double-strand breaks and the RAD50-RAD57 genes in *Saccharomyces*. *Sem. Cancer. Biol.* **4**, 73-83 (1993).

138. Xia, F. *et al.* Deficiency of human BRCA2 leads to impaired homologous recombination but maintains normal non-homologous end-joining. *Proc. Natl. Acad. Sci. USA*. **98**, 8644-8649 (2001).
139. Moynahan, M. E., Chiu, J. W., Koller, B. H. & Jasin, M. Brca1 controls homology-directed DANN repair. *Mol. Cell*. **4**, 511-518 (1999).
140. Jasin, M. Genetic manipulation of genomes with rare-cutting endonucleases. *Trends Genet.* **12**, 224-228 (1996).
141. Rouet, P., Smih, F. & Jasin, M. Expression of a site-specific endonuclease stimulates homologous recombination in mammalian cells. *Proc. Natl. Acad. Sci. USA*. **91**, 6064-6068 (1994).
142. Richardson, C., Moynahan, M. E. & Jasin, M. Double-strand break repair by interchromosomal recombination: suppression of chromosomal translocations. *Gene Dev.* **12**, 3831-3842 (1998).
143. Sargent, R. G., Brenneman, M. A. & Wilson, J. H. Repair of site-specific double strand breaks in a mammalian chromosome by homologous and illegitimate recombination. *Mol. Cell. Biol.* **17**, 267-277 (1997).
144. Donoho, G., Jasin, M. & Berg, P. Analysis of gene targetting and intrachromosomal homologous recombination stimulated by genomic double-strand break in mouse embryonic stem cells. *Mol. Cell. Biol.* **18**, 4070-4078 (1998).
145. Baarends, W. M., Van der Laan, R. & Grootegoed, J. A. DNA repair mechanisms and gametogenesis. *Reproduction*. **121**, 31-39 (2001).
146. Celerin, M., Merino, S. T., Stone, J. E., Menzie, A. M. & Zolan, M. E. Multiples roles of Spo11 in meiotic chromosome behavior. *EMBO J.* **19**, 2739-2750 (2000).
147. Goedecke, W., Eijpe, M., Offenber, H. H., van Aalderen, M. & Heyting, C. Mre11 and Ku70 interact in somatic cells, but are differentially expressed in early meiosis. *Nat. Genet.* **23**, 194-198 (1999).
148. Shulman, M. J., Collins, C., Connor, A., Read, L. R. & Baker, M. D. Interchromosomal recombination is suppressed in mammalian somatic cells. *EMBO J.* **14**, 4102-4107 (1995).
149. Moynahan, M. E. & Jasin, M. Loss of heterozygosity induced by a chromosomal double-strand break. *Proc. Natl. Acad. Sci. USA*. **94**, 8988-8993 (1997).

150. Haber, J. E. A super new twist on the initiation of meiotic recombination. *Cell*. **89**, 163-166 (1997).
151. Fasullo, M. T. & Davis, R. W. Recombination substrates designed to study recombination between unique repetitive sequences in-vivo. *Proc. Natl. Acad. Sci. USA*. **84**, 6215-6219 (1987).
152. Jinks-Robertson, S. & Petes, T. D. Chromosomal translocations generated by high-frequency meiotic recombination between repeated yeast genes. *Genetics*. **114**, 731-752 (1986).
153. Lichten, M., Borts, R. H. & Haber, J. E. Meiotic gene conversion and crossing-over between dispersed homologous sequences occur frequently in *Saccharomyces cerevisiae*. *Genetics*. **115**, 233-246 (1987).
154. Virgin, J. B. *et al.* Crossing-over is rarely associated with mitotic intragenic recombination in *Schizosaccharomyces pombe*. *Genetics*. **157**, 63-77 (2001).
155. Kupiec, M. & Petes, T. D. Allelic and ectopic recombination between Ty elements in yeast. *Genetics*. **119**, 549-559 (1988).
156. Louis, E. J. & Haber, J. E. Mitotic recombination among subtelomeric Y' repeats in *Saccharomyces cerevisiae*. *Genetics*. **124**, 547-559 (1990).
157. Virgin, J. B. & Bailey, J. P. The M26 hotspot of *Schizosaccharomyces pombe* stimulates meiotic ectopic recombination and chromosomal rearrangements. *Genetics*. **149**, 1191-1204 (1998).
158. Wang, P. R., Zhou, R., Zhou, C., Jackson-Cook, C. & Povirk, L. Highly conservative reciprocal translocations formed by apparent joining of exchanged DNA double-strand-break ends. *Proc. Natl. Acad. Sci. USA*. **94**, 12 018-12 023 (1997).
159. Baker, M. D. & Read, L. R. Ectopic recombination within homologous immunoglobulin  $\mu$  gene constant regions in a mouse hybridoma cell line. *Mol. Cell. Biol.* **12**, 4422-4432 (1992).
160. Murti, J. R., Bumbulis, M. & Schimenti, J. C. Gene conversion between unlinked sequences in the germline of mice. *Genetics*. **137**, 837-843 (1994).
161. Richardson, C. & Jasin, M. Frequent chromosomal translocations induced by double-strand breaks. *Nature*. **405**, 697-700 (2000).

162. Shen, P. & Huang, H. V. Homologous recombination in *Escherichia coli*: dependence on substrate length and homology. *Genetics*. **112**, 441-457 (1986).
163. Priebe, S. D., Westmoreland, J., Nilsson-Tillgren, T. & Resnick, M. A. Induction of recombination between homologous and diverged DNAs by double-strand gaps and breaks and role of mismatch repair. *Mol. Cell. Biol.* **14**, 4802-4814 (1994).
164. Harris, S., Rudnicki, K. S. & Haber, J. E. Gene conversions and crossing over during homologous and homeologous ectopic recombination in *Saccharomyces cerevisiae*. *Genetics*. **135**, 5-16 (1993).
165. Chen, W. & Jinks-Robertson, S. The role of the mismatch repair machinery in regulating mitotic and meiotic recombination between diverged sequences in yeast. *Genetics*. **151**, 1299-1313 (1999).
166. Rubnitz, J. & Subramani, S. The minimum amount of homology required for homologous recombination in monkey cells. *Mol. Cell. Biol.* **4**, 2253-2258 (1984).
167. Liskay, R. M., Letsou, A. & Stachelek, J. L. Homology requirement for gene conversion between duplicated chromosomal sequences in mammalian cells. *Genetics*. **115**, 161-167 (1987).
168. Hastly, P., Riviera-Perez, J. & Bradley, A. The length of homology required for gene targeting in embryonic stem cells. *Mol. Cell. Biol.* **11**, 5586-5591 (1991).
169. Deng, C. & Capecchi, M. R. Reexamination of gene targeting frequency as a function of extent of homology between the targeting vector and target locus. *Mol. Cell. Biol.* **12**, 3365-3371 (1992).
170. Waldman, A. S. & Liskay, R. M. Differential effects of base-pair mismatch on intrachromosomal versus extrachromosomal recombination in mouse cells. *Proc. Natl. Acad. Sci. USA*. **84**, 5340-5344 (1987).
171. Belmaaza, A., Milot, E., Villemure, J. & Chartrand, P. Interference or DNA sequence divergence with precise recombinational DNA repair in mammalian cells. *EMBO J.* **13**, 5355-5360 (1994).
172. Evans, E. Alani, E. Roles for mismatch repair factors in regulating genetic recombination. *Mol. Cell. Biol.* **20**, 7839-7844 (2000).
173. Priebe, S. D., Westmoreland, J., Nilsson-Tillgren, T. & Resnick, M. A. Induction of recombination between homologous and diverged DNAs by

- double-strand gaps and breaks and role of mismatch repair. *Mol. Cell. Biol.* **14**, 4802-4814 (1994).
174. Harris, S., Rudnicki, K. S. & Haber, J. E. Gene conversions and crossing over during homologous and homeologous ectopic recombination in *Saccharomyces cerevisiae*. *Genetics*. **135**, 5-16 (1993).
175. Chen, W. & Jinks-Robertson, S. The role of the mismatch repair machinery in regulating mitotic and meiotic recombination between diverged sequences in yeast. *Genetics*. **151**, 1299-1313 (1999).
176. Elliot, B. & Jasin, M. Repair of double-strand breaks by homologous recombination in mismatch repair-defective mammalian cells. *Mol. Cell. Biol.* **21**, 2671-2682 (2001).
177. Negritto, M. T., Wu, X., Kuo, T., Chu, S. & Bailis, A. M. Influence of DNA sequence identity on efficiency of targeted gene replacement. *Mol. Cell. Biol.* **17**, 278-286 (1997).
178. Reitmair, A. H. *et al.* MSH2 deficient mice are viable and susceptible to lymphoid tumours. *Nat. Genet.* **11**, 64-70 (1995).
179. Hunter, N. & Borts, R. H. Mlh1 is unique among mismatch repair proteins in its ability to promote crossing-over during meiosis. *Genes Dev.* **11**, 1573-1582 (1997).
180. Shen, P. & Huang, H. Effect of base pair mismatches on recombination via the RecBCD pathway. *Mol. Gen. Genet.* **18**, 358-360 (1989).
181. Burgess, S. M. & Kleckner, N. Collisions between yeast chromosomal loci in vivo are governed by three layers of organization. *Gene Dev.* **13**, 1871-1883 (1999).
182. Workman, J. L. & Kingston, R. E. Alteration of nucleosome structure as a mechanism of transcriptional regulation. *Annu. Rev. Biochem.* **67**, 545-579 (1998).
183. Keil, R. L. & Roeder, G. S. Cis-acting recombination-stimulating activity in a fragment of the ribosomal DNA of *S. cerevisiae*. *Cell.* **39**, 377-386 (1984).
184. Grimm, C., Schaer, P., Munz, P. & Kohli, J. The strong ADH1 promotor stimulates mitotic and meiotic recombination at the ADE6 gene of *Schizosaccharomyces pombe*. *Mol. Cell. Biol.* **11**, 289-298 (1991).

185. Yancopoulos, G. D. & Alt, F. W. Developmentally controlled and tissue-specific expression of unrearranged VH gene segments. *Cell*. **40**, 271-281 (1985).
186. Nickoloff, J. A. & Reynolds, R. J. Transcription stimulates homologous recombination in mammalian cells. *Mol. Cell. Biol.* **10**, 4337-4845 (1990).
187. Nickoloff, J. A. Transcription enhances intrachromosomal homologous recombination in mammalian cells. *Mol. Cell. Biol.* **12**, 5311-5318 (1992).
188. Thomas, B. J. & Rothstein, R. Elevated recombination rates in transcriptionally active DNA. *Cell*. **56**, 619-630 (1989).
189. Gross, D. S. & Garrard, W. T. Nuclease hypersensitive sites in chromatin. *Annu. Rev. Biochem.* **57**, 159-197 (1988).
190. Boulikas, T. The non-uniform repair of active and inactive chromatin domains. *Int. J. Oncol.* **8**, 65-75 (1996).
191. Glover, T.W. & Stein, C.K. Induction of sister chromatid exchanges at common fragile sites. *Am. J. Hum. Genet.* **41**, 882-890 (1987).
192. Glover, T.W. & Stein, C.K. Chromosome breakage and recombination at fragile sites. *Am. J. Hum. Genet.* **43**, 265-273 (1988).
193. Paldi, A., Gyapay, G. & Jami, J. Imprinted chromosomal regions of the human genome display sex-specific meiotic recombination frequencies. *Curr. Biol.* **5**, 1030-1035 (1995).
194. Ji, W. *et al.* DNA methylation and pericentromeric rearrangements of chromosome 1. *Mutat. Res.* **379**, 33-41 (1997).
195. Miniou, P. M. *et al.* Abnormal methylation pattern in constitutive and facultative (X inactive chromosome) heterochromatin of ICF patients. *Hum. Mol. Genet.* **3**, 2093-2102 (1994).
196. Engler, P. *et al.* A strain-specific modifier on mouse chromosome 4 controls the methylation of independent transgene loci. *Cell*. **65**, 939-947 (1991).
197. Hsieh, C.-L. & Lieber, M. R. CpG methylated minichromosomes become inaccessible for V(D)J recombination after undergoing replication. *EMBO J.* **11**, 315-325 (1992).
198. Maloisel, L. & Rossignol, J. Suppression of crossing-over by DNA methylation in *Ascobolus*. *Gene Dev.* **12**, 1381-1389 (1998).

199. Liang, F. & Jasin, M. Studies of the influence of cytosine methylation on DNA recombination and end-joining in mammalian cells. *J. Biol. Chem.* **270**, 23 838-23 844 (1995).
200. Yoder, J. A., Walsh, C. P. & Bestor, T. B. Cytosine methylation and the ecology of intragenomic parasites. *Trends Genet.* **13**, 335-340 (1997).
201. Richard, M., Belmaaza, A., Gusew, N., Wallenburg, J. C. & Chartrand, P. Integration of a vector containing a repetitive LINE-1 element in the human genome. *Mol. Cell. Biol.* **14**, 6689-6695 (1994).
202. Skowronski, A. F., Fanning, T. G. & Singer, M.F. Unit-length Line-1 transcripts in human teratocarcinoma cells. *Mol. Cell. Biol.* **8**, 1385-1397 (1988).
203. Ziv, Y. *et al.* Cellular and molecular characteristics of an immortalized ataxia-telangiectasia (group AB) cell line. *Cancer Res.* **49**, 2495-2501 (1989).
204. Ausubel, F. M. *et al.* Current protocols in molecular biology. John Wiley & Sons (eds), N.Y. (1992).
205. Zhou, C., Yang, Y. & Jong, A. Y. Mini-prep in ten minutes. *Biotechniques.* **8**, 172-173 (1990).
206. Chabot, C. Étude de l'instabilité génomique médiée par les réarrangements génomiques se produisant suite à l'induction d'une cassure double-brin. *Mémoire de maîtrise . Université de Montréal* (2002).
207. Smit, A. F., Toth, G., Riggs, A. D. & Jurka, J. Ancestral mammalian-wide subfamilies of LINE-1 repetitive sequences. *J. Mol. Biol.* **246**, 401-417 (1995).
208. Korenberg, J. R. & Rykowski, M. C. Human genome organization: Alu, LINEs, and the molecular structure of metaphase chromosome bands. *Cell.* **53**, 391-400 (1988).
209. Wu, T. C. & Lichten, M. Meiosis-induced double-strand break sites determined by yeast chromatin structure. *Science.* **263**, 515-518 (1994).
210. Gloor, G. B. & Lankeau, D. H. Gene conversion in mitotically dividing cells: A view from drosophila. *Trends Genet.* **14**, 43-46 (1998).
211. Rothstein, R., Michel, B. & Gangloff, S. Replication fork pausing and recombination or "gimme a break". *Gene Dev.* **14**, 1-10 (2000).

212. Gusew, N. , Lussier, R., Tremblay, A., Würtele, H. & Chartrand, P. Homologous recombination events between LINE elements obtained in transgenic mice. *Manuscrit en préparation*.
213. Würtele, H., Gusew, N., Lussier, R. & Chartrand, P. Characterization of inter-line recombination in the early embryo. *Manuscrit en préparation*.
214. Te Riele, H., Maandag, E. R. & Berns, A. Highly efficient gene targeting in mouse embryonic stem cells through homologous recombination with isogenic DNA constructs. *Proc. Natl. Acad. Sci. USA.* **89**, 5128-5132 (1992).
215. Ziv, Y. et al. Recombinant ATM protein complements the cellular A-T phenotype. *Oncogene.* **15**, 159-167 (1997).
216. Murnane, J. P. Fuller, L. F. & Painter, R. B. Establishment and characterization of a permanent pSVori-transformed ataxia-telangiectasia cell line. *Exp. Cell Res.* **159**, 119-126 (1985).
217. Wiesmüller, L., Cammenga, J. & Deppert, W. W. In vivo assay of p53 function in homologous recombination between simian virus 40 chromosomes. *J. Virol.* **70**, 737-744 (1996).
218. Lin, Y., Lukacsovich, T. & Waldman, A. S. Multiple pathways for repair of DNA double-strand breaks in mammalian chromosomes. *Mol. Cell. Biol.* **19**, 8353-8360 (1999).
219. Stumm, M. *et al.* High frequency of spontaneous translocations revealed by FISH in cells from patients with the cancer-prone syndromes ataxia telangiectasia and Nijmegen breakage syndrome. *Cytogenet. Cell Genet.* **92**, 186-191 (2001).
220. Takao, N. *et al.* Disruption of ATM in p53-null cells causes multiple functional abnormalities in cellular response to ionizing radiation. *Oncogene.* **18**, 7002-7009 (1999).
221. Myers, J. S. *et al.* A comprehensive analysis of recently integrated human Ta L1 element. *Am. J. Hum. Genet.* **71**, 312-326 (2002).
222. Scott, A. F. *et al.* Origin of the human L1 elements: proposed progenitor genes deduced from a consensus DNA sequence. *Genomic.* **1**, 113-125 (1987).
223. DeBerardinis, R. J. & Kazazian, Jr. H. H. Full-length L1 elements have arisen recently in the same 1-Kb region of the gorilla and human genomes. *J. Mol. Evol.* **47**, 292-301 (1998).

224. Stahl, F. Meiotic recombination in yeast: coronation of the double-strand break repair model. *Cell*. **87**, 965-968 (1996).
225. Nickoloff, J. A., Sweetser, D. B., Clikeman, J. A., Khalsa, G. J. & Wheeler, S. L. Multiple heterologies increase mitotic double-strand break-induced allelic gene conversion tract lengths in yeast. *Genetics*. **153**, 665-679 (1999).
227. Johnson, R. D. & Jasin, M. Sister chromatid gene conversion is a prominent double-strand break repair pathway in mammalian cells. *EMBO J*. **19**, 3398-3407 (2000).
228. Richardson, C. & Jasin, M. Coupled homologous and non-homologous repair of a double-strand break preserves genomic integrity in mammalian cells. *Mol. Cell. Biol.* **20**, 9068-9075 (2000).
229. Waldman, A. S. & Liskay, R. M. Dependence of intrachromosomal recombination in mammalian cells on uninterrupted homology. *Mol. Cell. Biol.* **8**, 5350-5357 (1988).
230. Thomas, K. R. & Capecchi, M. R. Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. *Cell*. **51**, 503-512 (1987).
231. Baker, M. D., Read, L. R., Beatty, B. G. & Ng, P. Requirements for ectopic homologous recombination in mammalian somatic cells. *Mol. Cell. Biol.* **16**, 7122-7132 (1996).
232. Sugawara, N. *et al.* DNA length dependence of the single-strand annealing pathway and the role of *Saccharomyces cerevisiae* RAD59 in double-strand break repair. *Mol. Cell. Biol.* **20**, 5300-5309 (2000).
233. Dudenhöffer, C., Rohaly, G., Will, K., Deppert, W. & Wiesmüller, L. Specific mismatch recognition in heteroduplex intermediates by p53 suggests a role in fidelity control of homologous recombination. *Mol. Cell. Biol.* **18**, 5332-5342 (1998).
234. Aguilera, A. & Klein, H. L. Yeast intrachromosomal recombination: long gene conversion tracts are preferentially associated with reciprocal exchange and require RAD1 and RAD3 gene products. *Genetics*. **123**, 683-694 (1989).
235. Osterag, E. M. & Kazazian, H. H. Biology of mammalian L1 retrotransposons. *Annu. Rev. Genet.* **35**, 501-538 (2001).
236. Boissinot, S., Chevret, P. & Furano, A. V. L1 (LINE-1) retrotransposon evolution and amplification in recent human history. *Mol. Biol. Evol.* **17**, 915-928 (2000).

237. Pavlíček, A., Pačes, J., Zíka, R. & Hejnar, J. Length distribution of long interspersed nuclear elements (LINEs) and processed pseudogenes of human endogenous retroviruses: implications for retrotransposition and pseudogene detection. *Gene*. **300**, 189-194 (2002).
238. Chen, R. Z., Petterson, U., Beard, U., Jackson-Grusby, L. & Jaenish, R. DNA hypomethylation leads to elevated mutation rates. *Nature*. **395**, 89-93 (1998).
239. Goyon, C., Rossignol, J.-L. & Faugeron, G. Native DNA repeats and methylation in *Ascobolus*. *Nucleic Acid Res.* **24**, 3348-3356 (1996).
240. Branciforte, D. & Martin, S. L. Developmental and cell type specificity of a human LINE-1 promoter. *Mol. Cell. Biol.* **10**, 2584-2592 (1994).
241. Brathauer, G. L. & Fanning, T. G. Active LINE-1 retrotransposons in human testicular cancer. *Oncogene*. **7**, 507-510 (1992).
242. Mitelman, F., Johansson, B. & Mertens, F. Mitelman Database of Chromosome Aberrations in Cancer.  
<http://cgap.nci.nih.gov/Chromosomes/Mitelman> (2003).
243. Nikiforova, M. N., Stringer, J.R., Blough, R., Medvedovic, M., Fagin, J. A. & Nikiforova, Y. E. Proximity of chromosomal loci that participate in radiation-induced rearrangements in human cells. *Science*. **290**, 138-141 (2000).
244. Chevret, E., Volpi, E. V. & Sheer, D. Mini review: Form and function in the human interphase chromosome. *Cytogenet. Cell. Genet.* **90**, 13-21 (2000).

