

Université de Montréal

**Rôles des interactions entre loci dans l'organisation spatiale fonctionnelle et  
l'évolution des génomes de mammifères**

Par

**Hugo Würtele**

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

Août 2006

© Hugo Würtele, 2006



QH

506

U54

2007

v. 006

## 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 :**

Rôles des interactions entre loci dans l'organisation spatiale fonctionnelle et l'évolution des génomes de mammifères

**Présentée par :**

Hugo Würtele

a été évaluée par un jury composé des personnes suivantes :

Dr Trang Hoang : président rapporteur  
Dr Pierre Chartrand: directeur de recherche  
Dr Sylvie Mader : membre du jury  
Dr Chantal Autexier : examinateur externe  
Dr Éric Milot : représentant du doyen de la FES



## RÉSUMÉ

Les chromosomes des mammifères sont organisés dans leur séquence mais aussi au niveau de leur positionnement spatial dans le noyau. Bien que les rôles fonctionnels de cette organisation spatiale demeurent peu connus, des travaux récents ont commencé à éclaircir cette question. Ainsi, la fréquence de recombinaison entre loci est liée à leur proximité spatiale et certains loci sont repositionnés dans l'espace nucléaire lors de leur expression. Les travaux présentés dans cette thèse s'inscrivent dans cette lignée d'observations et visent à investiguer comment les possibilités d'interactions entre loci, qui sont modulées par l'organisation spatiale des chromosomes, pourraient influencer les fonctions génomiques. Cette thématique fut explorée à travers trois projets.

Le premier projet analyse la recombinaison entre séquences en proximité linéaire *in vivo* dans l'embryon de souris. Des souris transgéniques furent générées par micro-injection dans des ovules fécondés d'un plasmide contenant deux séquences répétitives LINE-1 (L1) en répétition directe pouvant se recombiner entre elles ou avec des L1 endogènes avant, pendant ou après leur intégration au génome. La configuration des plasmides intégrés fut déterminée. Avant ou pendant l'intégration, les fréquences de recombinaison sont élevées, pouvant impliquer jusqu'à 70% des plasmides, et les délétions homologues sont le type d'événement prédominant. Les plasmides intégrés peuvent aussi se recombiner au fil des générations à plus faible fréquence. Ces résultats suggèrent que la recombinaison entre séquences proximales dans l'embryon a pu influencer l'évolution des L1 et du génome.

Le deuxième projet investigate l'impact de l'environnement spatial d'un locus humain sur le choix de ses partenaires de recombinaison. Notre laboratoire a montré qu'un locus donné utilise certains loci spécifiques à haute fréquence pour se réparer par recombinaison, et ce bien qu'ils ne soient pas les plus rapprochés du locus d'intérêt en terme linéaire. Une explication à ces résultats voudrait que l'organisation spatiale du génome maintienne ces loci préférentiels à proximité du locus d'intérêt. Grâce à la mise au point d'une nouvelle méthode basée sur le Chromosome Conformation Capture (3C), nous démontrons que le locus d'intérêt possède un environnement spatial majoritairement composé de son mégabase linéaire adjacent et de son bras chromosomique. Or, les partenaires de recombinaison préférentiels du locus d'intérêt sont situés dans ces domaines génomiques,

ce qui suggère que l'environnement spatial d'un locus est un facteur déterminant dans le choix de ses partenaires de recombinaison.

Le troisième projet investigate l'impact de l'expression d'un gène sur son environnement spatial. Nous avons réalisé un crible des loci formant l'environnement spatial du gène *HoxB1* de souris avant et après son induction. L'environnement spatial de *HoxB1* peut être divisé en trois parties : un domaine proximal adjacent au gène et couvrant environ 110 kb, un deuxième domaine couvrant 800 kb et un dernier domaine regroupant les interactions distales intra et inter-chromosomiques. De plus, la proportion d'interactions inter-chromosomiques augmente significativement après l'induction du gène, indiquant que l'environnement spatial de *HoxB1* est modifié lors de l'induction de son expression.

Ces travaux supportent l'hypothèse que les chromosomes des mammifères possèdent une organisation spatiale dynamique étroitement liée aux fonctions et à l'évolution du génome.

### **MOTS CLÉS**

Recombinaison homologue et non-homologue, LINE-1, souris transgéniques, cellules embryonnaires, réparation de l'ADN, Chromosome Conformation Capture, organisation nucléaire, repliement de la chromatine, modèle des boucles de 1 mb, *HoxB1*, territoires chromosomiques

## ABSTRACT

Mammalian chromosomes are organized both in their sequence and their spatial positioning in the nucleus. While the functional roles of this spatial organisation remain unclear, recent work has begun to shed light on the issue. Indeed, the frequency of recombination between loci is linked to their spatial proximity and other loci are repositioned in the nuclear space when expressed. The work presented in this thesis also aims to investigate the functional relevance of the spatial organisation of mammalian genomes, with an emphasis on the links between the functions of genomic loci and their possibility of interactions with the genome. We explore this theme through three projects.

The first project analyses *in vivo* recombination between sequences that are in linear proximity in the mouse embryo. Transgenic mice were generated by microinjection of a plasmid containing two LINE-1 (L1) repetitive sequences in direct repeats in fertilized ova. These sequences could recombine with each other and with endogenous L1 before, during and after their integration to the genome. We determined the configuration of the integrated plasmids. Before or during integration, recombination frequencies are high, involving up to 70% of plasmids with homologous deletions being the most frequent outcome. Integrated plasmids can also be modified by recombination from generation to generation at reduced frequencies. These results suggest that recombination between proximal sequences in the embryo could have had an impact on the evolution of L1 and of the genome.

The second project investigates whether the spatial environment of a human locus influences the choice of its recombination partners. Our lab showed that a genomic locus utilizes specific loci at high frequencies to repair itself by recombination, even if these loci are not the closest to the locus of interest in linear terms. These observations could be explained if the preferred loci are in closer spatial proximity to the locus of interest than other under-utilized loci. Using a novel method based on Chromosome Conformation Capture (3C), we show that the locus of interest possesses a spatial environment mostly made of its adjacent megabase and its chromosomal arm. Interestingly, the preferential recombination partners of this locus are located within these domains, suggesting that the spatial environment of a locus is an important factor in the choice of its recombination partners.

Finally, a third project investigates the impact of the expression of a gene on its spatial environment. We have screened the loci forming the spatial environment of the mouse HoxB1 gene before and after its induction. HoxB1's spatial environment can be divided in three parts: a first part is represented by a proximal domain covering approximately 110 kb around the gene, a second part extends to a domain of 800 kb and a third part consists of distal intra- and interchromosomal interactions. The proportion of inter-chromosomal interactions increases after the gene's induction, indicating that HoxB1's spatial environment is modified during its induction.

This work supports the hypothesis that mammalian chromosomes possess a dynamic spatial organisation that is directly linked to the functions and evolution of the genome.

### **KEYWORDS**

Homologous and non-homologous recombination, LINE-1, transgenic mouse, embryonic cells, DNA repair, Chromosome Conformation Capture, nuclear organization, chromatin folding, megabase loop model, HoxB1, Chromosomal territories

**TABLE DES MATIÈRES**

|   |            |
|---|------------|
| <b>IDENTIFICATION DU JURY .....</b>   | <b>II</b>  |
| <b>RÉSUMÉ.....</b>  | <b>III</b> |
| <b>MOTS CLÉS .....</b>  | <b>IV</b>  |
| <b>ABSTRACT .....</b>   | <b>V</b>   |
| <b>KEYWORDS .....</b>   | <b>VI</b>  |
| <b>TABLE DES MATIÈRES .....</b>   | <b>VII</b> |
| <b>LISTE DES FIGURES .....</b>  | <b>XII</b> |
| <b>LISTE DES TABLEAUX.....</b>  | <b>XIV</b> |
| <b>LISTE DES SIGLES ET ABBRÉVIATIONS.....</b>   | <b>XV</b>  |
| <b>REMERCIEMENTS .....</b>  | <b>XIX</b> |
| <b>INTRODUCTION.....</b>  | <b>1</b>   |
| <b>CONTEXTE ET PROJETS DE RECHERCHE.....</b>  | <b>1</b>   |
| <b>PREMIÈRE PARTIE : RÔLES DES INTERACTIONS ENTRE LOCI DANS LE REMODELAGE DU GÉNOME PAR RECOMBINAISON GÉNÉTIQUE .....</b> | <b>2</b>   |
| <b>Recombinaison génétique : mécanismes et rôles dans le remodelage des génomes de mammifères .....</b>                   | <b>3</b>   |
| Deux types de mécanismes de recombinaison distincts : homologues et illégitimes   | 4          |
| Reconnaissance des CDB et signalisation initiale .....  | 5          |
| Recombinaison non-homologue : mécanismes et facteurs protéiques impliqués.....  | 7          |
| Recombinaison homologue : mécanismes et facteurs protéiques impliqués .....   | 9          |
| Implication de la recombinaison dans le remodelage et l'évolution des génomes de mammifères.....                          | 14         |
| Système d'étude : Recombinaison dans l'embryon de souris et son impact sur l'évolution des génomes de mammifères.....     | 17         |
| Système d'étude : Les séquences répétitives L1 .....  | 18         |

|  |           |
|--|-----------|
| <b>DEUXIÈME ET TROISIÈME PARTIES : RÔLES DES INTERACTIONS ENTRE LOCI DANS LA DÉFINITION DES ENVIRONNEMENTS SPATIAUX GÉNOMIQUES ET L'ORGANISATION FONCTIONNELLE DU GÉNOME .....</b>   | <b>20</b> |
| <b>L'ADN existe sous forme de chromatine compacte dans le noyau .....</b>  | <b>21</b> |
| Rôles des histones dans les mécanismes de compaction de la chromatine.....   | 22        |
| Modulation du degré de compaction de la chromatine par les modifications post-traductionnelles aux histones et les facteurs de remodelage de la chromatine .....   | 24        |
| La compaction d'une fibre de chromatine dépend de son contenu en gène.....   | 25        |
| La localisation des modifications post-traductionnelles aux histones et des facteurs de remodelage lie le contenu en gène et le repliement supérieur de la chromatine .  | 26        |
| <b>Organisation supérieure de la chromatine et compartimentalisation fonctionnelle du noyau.....</b>   | <b>27</b> |
| Les chromosomes occupent des territoires distincts dans l'espace nucléaire .....   | 28        |
| Les territoires chromosomiques peuvent être séparés en plusieurs sous-domaines dans l'espace nucléaire.....  | 31        |
| Les limites à la diffusion de la chromatine dans l'espace nucléaire permettent une certaine stabilité dans l'organisation nucléaire .....  | 32        |
| Positionnement des domaines chromosomiques entre eux et par rapport aux structures nucléaires .....  | 33        |
| Les interactions spécifiques entre loci existent et facilitent leur recombinaison.....   | 35        |
| Système d'étude : lien entre l'environnement spatial d'un locus et le choix d'un partenaire de recombinaison .....   | 37        |
| Positionnement spécifique de régions génomiques à la périphérie du noyau.....  | 37        |
| Positionnement de loci en lien avec les structures nucléaires.....   | 39        |
| Positionnement de loci spécifiques par rapport à leur territoire chromosomique et lien avec la transcription .....   | 40        |
| Usines de transcription.....   | 42        |
| Contacts entre séquences régulatrices et gènes.....  | 43        |
| Système d'étude : Modifications de l'environnement spatial génomique du gène HoxB1 lors de l'induction de son expression dans le contexte de la différenciation de cellules embryonnaires de souris induite par l'acide rétinoïque ..... | 46        |
| La biologie des gènes Hox et leur induction par l'acide rétinoïque.....  | 46        |

|  |    |
|--|----|
| <b>MATÉRIEL ET MÉTHODE</b> .....   | 50 |
| <b>1. RECOMBINAISON ENTRE SÉQUENCES EN RÉPÉTITION DIRECTE DANS L'EMBRYON DE SOURIS ET SON IMPACT SUR L'ÉVOLUTION DES GÉNOMES DE MAMMIFÈRES</b> .....   | 50 |
| <b>2. LIEN ENTRE L'ENVIRONNEMENT SPATIAL D'UN LOCUS ET LE CHOIX DE SES PARTENAIRES DE RECOMBINAISON</b> .....  | 52 |
| <b>3. MODIFICATIONS DE L'ENVIRONNEMENT SPATIAL GÉNOMIQUE DU GÈNE HoxB1 LORS DE L'INDUCTION DE SON EXPRESSION DANS LE CONTEXTE DE LA DIFFÉRENCIATION DE CELLULES EMBRYONNAIRES DE SOURIS INDUITE PAR L'ACIDE RÉTINOÏQUE</b> ..... | 54 |
| <b>RÉSULTATS</b> .....   | 60 |
| <b>1. RECOMBINAISON ENTRE SÉQUENCES EN RÉPÉTITION DIRECTE DANS L'EMBRYON DE SOURIS ET SON IMPACT SUR L'ÉVOLUTION DES GÉNOMES DE MAMMIFÈRES</b> .....   | 60 |
| Objectifs .....  | 60 |
| Design expérimental.....   | 60 |
| La recombinaison homologe extra-chromosomique est extrêmement efficace dans l'embryon précoce de souris .....  | 64 |
| Caractérisation des réarrangements de plasmides.....   | 70 |
| Implication des L1 endogènes dans certains événements de recombinaison homologues dans l'embryon de souris.....  | 72 |
| Modification des plasmides intégrés par recombinaison homologe au fil des générations de souris transgéniques.....   | 72 |
| <b>2. LIEN ENTRE L'ENVIRONNEMENT SPATIAL D'UN LOCUS ET LE CHOIX DE SES PARTENAIRES DE RECOMBINAISON</b> .....  | 75 |
| Objectifs .....  | 75 |
| Design expérimental.....   | 77 |
| Spécificité des produits de REM-3C.....  | 79 |
| Distribution des partenaires de fixation intra-chromosomiques .....  | 85 |
| Distribution des partenaires de REM-3C inter-chromosomiques.....   | 88 |
| L'accessibilité spatiale privilégiée au Mb génomique adjacent est une caractéristique commune à plusieurs loci .....   | 88 |
| <b>3. MODIFICATIONS DE L'ENVIRONNEMENT SPATIAL GÉNOMIQUE DU GÈNE HoxB1 LORS DE L'INDUCTION DE SON EXPRESSION DANS LE CONTEXTE DE LA DIFFÉRENCIATION DE CELLULES EMBRYONNAIRES DE SOURIS INDUITE PAR L'ACIDE RÉTINOÏQUE</b> ..... | 91 |

|   |            |
|---|------------|
| Objectifs .....   | 91         |
| Design expérimental.....  | 92         |
| Récupération à haute fréquence des loci proximaux par inverse-PCR 3C .....  | 96         |
| L'environnement nucléaire du gène HoxB1 peut être séparé en domaines génomiques distincts d'interactions spatiales .....                            | 102        |
| L'environnement génomique spatial du gène HoxB1 change durant le traitement des cellules ES de souris à l'acide rétinoïque.....                     | 105        |
| Distribution génomique des partenaires de fixation inter et intra-chromosomiques distaux .....  | 110        |
| <b>DISCUSSION .....</b>   | <b>114</b> |
| <b>1. RECOMBINAISON ENTRE SÉQUENCES EN RÉPÉTITION DIRECTE DANS L'EMBRYON DE SOURIS ET SON IMPACT SUR L'ÉVOLUTION DES GÉNOMES DE MAMMIFÈRES.....</b> | <b>114</b> |
| Description de l'essai .....  | 114        |
| Fréquences de recombinaison homologue observées .....   | 116        |
| Fréquences des différents produits de recombinaison .....   | 119        |
| Mécanismes de recombinaison .....   | 121        |
| Recombinaison avec L1 endogènes .....   | 123        |
| Modification de la structure intégrée au fil des générations et impact sur l'évolution des génomes de mammifères .....                              | 123        |
| Conclusions.....  | 124        |
| <b>2. LIEN ENTRE L'ENVIRONNEMENT SPATIAL D'UN LOCUS ET LE CHOIX DE SES PARTENAIRES DE RECOMBINAISON .....</b>                                       | <b>125</b> |
| La méthodologie du REM-3C.....  | 125        |
| Choix du système d'étude .....  | 126        |
| Localisation des partenaires de REM-3C.....   | 127        |
| Modèles d'organisation de la chromatine en domaines spatiaux.....   | 128        |
| Partenaires distaux et inter-chromosomiques.....  | 129        |
| Impact de l'environnement spatial d'un locus sur ses choix des partenaires de recombinaison .....   | 130        |
| Usages et limitations de la méthodologie du REM-3C.....   | 132        |
| Conclusions.....  | 133        |



|  |            |
|--|------------|
| <b>3. MODIFICATIONS DE L'ENVIRONNEMENT SPATIAL GÉNOMIQUE DU GÈNE HOXB1 LORS DE L'INDUCTION DE SON EXPRESSION DANS LE CONTEXTE DE LA DIFFÉRENCIATION DE CELLULES EMBRYONNAIRES DE SOURIS INDUITE PAR L'ACIDE RÉTINOÏQUE .....</b> | <b>134</b> |
| Développement de la méthodologie de l'inverse-PCR 3C .....   | 134        |
| Choix du système d'étude .....   | 135        |
| Description de l'environnement spatial de base du gène HoxB1 .....   | 136        |
| Modifications de l'environnement spatial du gène HoxB1 suite au traitement des cellules ES à l'acide rétinoïque .....  | 139        |
| Validations, interprétations et limitations des méthodologies de 3C ouvertes développées .....   | 144        |
| <b>CONCLUSIONS GÉNÉRALES.....</b>  | <b>149</b> |
| <b>RÉFÉRENCES.....</b>   | <b>153</b> |

## LISTE DES FIGURES

|   |    |
|---|----|
| Figure 1 : La recombinaison non-homologue ou illégitime.....  | 8  |
| Figure 2 : Conséquences de la recombinaison homologue entre séquences non-alléliques sur l'organisation linéaire du génome .....  | 10 |
| Figure 3 : Le <i>single-strand annealing</i> .....  | 11 |
| Figure 4 : La recombinaison homologue .....   | 13 |
| Figure 5 : Le modèle de recombinaison induit par une brisure simple-brin ou <i>nick</i> tel que proposé par Meselson et Radding (Meselson et Radding 1975).....           | 15 |
| Figure 6 : Le modèle de recombinaison par <i>one-sided invasion</i> .....   | 16 |
| Figure 7 : Structure des éléments répétitifs LINE-1 .....   | 19 |
| Figure 8 : Plusieurs niveaux hiérarchiques de repliement supérieur de la fibre de chromatine .....  | 23 |
| Figure 9 : Organisation du noyau et modèle des territoires chromosomiques.....  | 31 |
| Figure 10 : Le modèle d'organisation de la chromatine en boucles de 1 Mb .....  | 32 |
| Figure 11 : Influence de l'organisation spatiale du génome sur les événements de recombinaison .....  | 36 |
| Figure 12 : Impact des contacts entre séquences spécifiques sur l'organisation des chromosomes et la régulation de l'activité transcriptionnelle.....                     | 45 |
| Figure 13 : Organisation des gènes Hox, du complexe HoxB et du gène HoxB1 .....   | 48 |
| Figure 14 : Plasmide utilisé dans l'étude et stratégie d'analyse .....  | 61 |
| Figure 15 : Distribution de fréquences des différentes configurations de plasmides présentes dans les structures intégrées pour les fondateurs et leurs descendances..... | 65 |
| Figure 16 : Configuration agglomérée des plasmides intégrés chez les fondateurs .....   | 69 |
| Figure 17 : Réarrangements d'intérêt particulier .....  | 71 |
| Figure 18 : La méthode de REM-3C.....   | 78 |
| Figure 19 : Un "code-barre" de l'environnement spatial du locus chr2: 40872835 dans la lignée cellulaire IMR90.....   | 80 |
| Figure 20 : Distribution des partenaires de REM-3C intra-chromosomiques.....  | 84 |
| Figure 21 : La méthodologie de l'inter-L1 3C .....  | 89 |
| Figure 22 : Design expérimental de l'inverse-PCR 3C réalisé sur le gène HoxB1 de souris .....   | 94 |

|   |     |
|---|-----|
| Figure 23 : Analyse par RT-PCR de l'expression des gènes HoxB1 et Oct4 durant le protocole de différenciation des cellules OS25 de souris par l'acide rétinoïque .....              | 95  |
| Figure 24 : Fréquence de clonage des fragments HindIII génomiques proximaux qui entourent le locus HoxB1 .....  | 99  |
| Figure 25 : Analyse par 3C classique des fragments HindIII génomiques proximaux.....  | 101 |
| Figure 26 : Proportion cumulative des partenaires de fixation récupérés par inverse-PCR 3C à différentes distances du locus HoxB1 .....   | 105 |
| Figure 27 : Distribution des partenaires de fixation clonés du gène HoxB1 dans différents intervalles génomiques.....   | 106 |
| Figure 28 : Contrôle d'efficacité du traitement 3C .....  | 109 |
| Figure 29 : Analyse du cycle cellulaire des populations de cellules souches embryonnaires traitées à l'acide rétinoïque par cytométrie en flux avec le marqueur Hoechst 33342 ..... | 110 |
| Figure 30 : Distribution des proportions des clones de partenaires de fixation dans les bandes intra-chromosomiques.....  | 111 |
| Figure 31 : Distribution des partenaires inter-chromosomiques de fixation récupérés par Inverse-PCR 3C sur le locus du gène HoxB1 .....   | 113 |

**LISTE DES TABLEAUX**

|   |    |
|---|----|
| Tableau 1 : Inverse-PCR 3C réalisés sur le locus du gène HoxB1 .....  | 56 |
| Tableau 2 : Séquences des amorces utilisées dans l'étude de 3C réalisée sur le locus HoxB1 .....                                    | 58 |
| Tableau 3 : Fréquences des produits de recombinaison chez les souris fondatrices.....   | 62 |
| Tableau 4 : Nombre de copies de transgène.....  | 67 |
| Tableau 5 : Comparaison statistique des distributions de fréquences des différents membres des lignées de souris transgéniques..... | 73 |
| Tableau 6 : Intervalles de confiance autour des distributions de fréquences des souris 4400 et 4701 .....                           | 75 |
| Tableau 7 : Partenaires REM-3C .....  | 83 |
| Tableau 8 : Partenaires potentiels de REM-3C situés dans une fenêtre génomique de 2 Mb autour du locus de départ .....              | 87 |
| Tableau 9 : Inter-L1 3C sur les cellules ATA67 .....  | 90 |

## LISTE DES SIGLES ET ABBRÉVIATIONS

|                 |   |
|-----------------|---|
| %               | Pourcentage                                     |
| °C              | Degré(s) celcius                                |
| 53BP1           | p53 binding protein 1                           |
| 3C              | Chromosome Conformation Capture                 |
| ABL             | Oncogène Abelson                                |
| ADN             | Acide désoxyribonucléique                       |
| ADNc            | ADN complémentaire                              |
| ARN             | Acide ribonucléique                             |
| ARNm            | ARN messenger                                   |
| A/T             | Adénine/thymidine                               |
| ATM             | Ataxia-telangiectasia mutated                   |
| ATP             | Adenosine triphosphate                          |
| ATR             | ATM and Rad3 related                            |
| BASC            | BRCA-associated complex                         |
| BCR             | Breakpoint cluster region protein               |
| BLM             | Protéine du syndrome de Bloom                   |
| BRCA1,2         | Breast cancer susceptibility gene 1, 2          |
| BRCT            | BRCA1 C-Terminal domain                         |
| CDB             | Cassure double-brin                             |
| CDK             | Cyclin dependant kinase                         |
| CHO             | Chinese hamster ovary                           |
| CHK1,2          | Checkpoint kinase 1, 2                          |
| chr             | Chromosome                                      |
| CMH             | Complexe majeur d'histocompatibilité            |
| CO <sub>2</sub> | Dioxyde de carbone                              |
| CT-IC           | Chromosome territory-interchromatin compartment |
| DMEM            | Dulbecco's minimal essential medium             |
| DNA-PK          | DNA-dependant protein kinase                    |
| DNAse           | Désoxyribonuclease                              |
| DR1,2,5         | Direct repeat 1,2,5                             |

|           |  |
|-----------|--|
| DSB       | Double-strand break  |
| E.coli    | Escherichia coli   |
| EN        | Endonucléase   |
| ERCC1/XPF | excision repair cross-complementing rodent repair deficiency, complementation group 1/Xeroderma pigmentosum, complementation group F |
| ES        | Embryonic Stem   |
| ESA1      | Essential <i>SAS</i> family acetyltransferase 1  |
| EZH2      | Enhancer of zeste, drosophila homolog 2  |
| FBS       | Fetal bovine serum   |
| FGF       | Fibroblast growth factor   |
| FISH      | Fluorescence in situ hybridization   |
| FRAP      | Fluorescence recovery after photobleaching   |
| GAPDH     | Glyceraldehyde-3-phosphate dehydrogenase   |
| G/C       | Guanine/cytosine   |
| GCN5      | General control of amino acid synthesis protein 5-like 2   |
| HDAC      | Histone désacétylase   |
| HP1       | Heterochromatin protein 1  |
| Hbb-1     | Hemoglobin beta 1  |
| HS        | Site hypersensible à la DNase  |
| INO80     | Inositol-requiring protein 80  |
| J-1, 1, 3 | Jour -1, 1, 3  |
| kb        | Kilopaire(s) de base   |
| L1        | LINE-1   |
| L1Hs      | LINE-1 human specific  |
| LCR       | Locus control region   |
| LIF       | Leukemia inhibitory factor   |
| LINE      | Long interspersed nuclear element  |
| LTR       | Long terminal repeat   |
| M         | Molaire  |
| MAR (s)   | Matrix attachment region (s)   |
| Mb        | Mégapaire(s) de base   |

|         |   |
|---------|---|
| MDM2    | Mouse double minute 2                         |
| min     | Minute(s)                                     |
| mL      | Millilitres(s)                                |
| mM      | Millimolaire                                  |
| MRN     | Mre11/Rad50/Nbs1                              |
| MSH     | MutS homologous gene                          |
| MTG     | Monothioglycerol                              |
| NBS1    | Nijmegen breakage syndrome                    |
| NCBI    | National center for biology information       |
| N-CoR   | Nuclear receptor co-repressor                 |
| NHEJ    | Non-homologous end-joining                    |
| NIGMS   | National institute of general medical science |
| nm      | Nanomètre (s)                                 |
| NuA4    | Nucleosomal acetyltransferase of H4           |
| NuRD    | Nucleosome remodeling and deacetylase         |
| NURF    | Nucleosome remodeling factor                  |
| OCT1, 4 | Octamer binding transcription factor 1, 4     |
| pb      | Paire(s) de base                              |
| PBS     | Phosphate buffered saline                     |
| PCR     | Polymerase chain reaction                     |
| PML     | Promyelocytic leukemia                        |
| Pol     | Polymérase                                    |
| RAR     | Retinoic acid receptor alpha                  |
| RARE    | Retinoic acid response element                |
| Rb      | Retinoblastoma                                |
| REM-3C  | Repetitive-element-mediated 3C                |
| RET     | Receptor tyrosine kinase proto-oncogene       |
| RFC     | Replication factor C                          |
| RNAse   | Ribonucléase                                  |
| RPA     | Replication protein A                         |
| RSC     | <i>Remodel the structure of chromatin</i>     |
| RT-PCR  | Réverse transcriptase PCR                     |
| RXR     | Retinoid x receptor                           |

|            |   |
|------------|---|
| s          | Secondes(s)   |
| SC35       | Spliceosome component 35 kDa  |
| Set1, 7    | Su(var) 3-9 (suppressor of position effect variegation), E(z) (enhancer of zeste), and Trx (trithorax) 1, 7 |
| SIN3       | Switch independent 3  |
| SINE       | Short interspersed nuclear element  |
| SIR2, 3, 4 | Silent information regulator 2, 3, 4  |
| SMRT       | Silencing mediator of retinoic acid and thyroid hormone receptor  |
| snRNA      | Small nuclear RNA   |
| snRNP      | Small nuclear ribonucleoprotein   |
| SSA        | Single-strand annealing   |
| Suv39h     | suppressor of variegation 3-9 homolog 1   |
| SV40       | Simian virus 40   |
| SWI/SNF    | Switch/sucrose non-fermenting   |
| SWR1       | Swi2/Snf2 related protein 1   |
| TFIIB      | Transcription factor IIB  |
| TIF1       | Transcriptional Intermediary Factor 1   |
| TK         | Thymidine kinase  |
| TSA        | Trichostatin-A  |
| U          | Unité enzymatique   |
| UCSC       | University of California, Santa Cruz  |
| µg         | Microgramme(s)  |
| µL         | Microlitre(s)   |
| UTR        | Untranslated region   |
| VP16       | Viral Protein 16  |
| WRN        | Werner syndrome protein   |
| X-gal      | 5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside   |
| XRCC2,3,4  | X-ray cross-complementing gene 2,3,4  |



## **REMERCIEMENTS**

Le chemin qui mène aux études graduées et à la complétion de celles-ci est tortueux et parsemé de hasards. Il faut donc savoir tirer profit de chaque rencontre et de chaque expérience réalisée en cours de route. La première de ces rencontres heureuses fut celle de mon directeur de recherche, le Dr Pierre Chartrand. Je tiens à lui rendre hommage et à le remercier, car il fut un mentor idéal qui m'a permis d'appriivoiser et de comprendre le fonctionnement du monde de la recherche. Je lui suis reconnaissant de m'avoir donné la possibilité d'explorer à fond mes idées les plus audacieuses, d'avoir été disponible pour discuter des projets et des résultats expérimentaux et de m'avoir enseigné à interpréter ces derniers avec rigueur. Je tiens aussi à remercier les membres passés du laboratoire, en commençant par Roxane Lussier, qui m'a appris les rudiments techniques du métier de chercheur et avec qui j'ai collaboré sur plusieurs projets. Merci aussi à Nadine Gusew pour sa participation à certains projets sur lesquels j'ai travaillé. Il me serait impossible de passer sous silence les innombrables discussions que j'ai eues avec Kevin Little, sur la vie autant que sur la recherche, discussions qui ont mené à l'élaboration d'un nombre incalculable de projets scientifiques aussi saugrenus qu'inspirants. Merci aux filles du labo, Catherine Chabot, Jacinte Gauthier et Hélène d'Anjou, avec qui j'ai partagé l'intensité des cours de maîtrise et de la vie de laboratoire. Je remercie particulièrement Jacinte, sur qui j'ai aussi eu la chance de pouvoir compter en tant qu'assistante de recherche, pour son travail remarquable et son enthousiasme contagieux. Merci également à Edlie St-Hilaire, avec qui j'ai eu la chance de collaborer lors des dernières années de mon doctorat : son éthique de travail impeccable et son investissement dans mes projets de doctorat ont été essentiels à la réussite de ceux-ci. Je tiens aussi à remercier le Dr Guy Sauvageau et son équipe, qui ont eu l'amabilité de m'accueillir comme un des leurs dans leur laboratoire durant mes dernières années de doctorat et qui m'ont fait profiter de leur expertise scientifique et technique. Merci aussi aux membres de l'Institut du cancer de Montréal et de l'Institut de recherche en immunologie et en cancérologie qui ont participé d'une façon ou d'une autre à ma formation et à mon travail de recherche. Je remercie d'autre part le Fonds de la recherche en santé du Québec, le programme de biologie moléculaire et la faculté des études supérieures de l'Université de Montréal et la fondation J. A. DeSève pour leur soutien financier. Finalement, je voudrais exprimer ma gratitude envers ma famille pour leur soutien moral, financier et personnel. Merci donc à mes parents, Diane et Paul et à mes

sœurs Elsa et Heidi, pour tout ce qu'ils ont fait et continuent de faire pour moi. Je tiens aussi à remercier tout particulièrement ma conjointe Myriam pour avoir partagé avec moi les moments difficiles et heureux, pour avoir su faire face aux nombreux sacrifices demandés par la complétion de mon doctorat, et surtout pour m'avoir donné la chance participer au plus merveilleux des projets, celui d'être père d'un petit garçon, mon petit Félix.

## INTRODUCTION

### Contexte et projets de recherche

Le noyau est une organelle spécifique aux eucaryotes. Ses fonctions sont cruciales pour la survie des cellules : régulation de l'expression des gènes, transcription et exportation des ARN messagers, réplication et réparation de l'ADN et même possiblement traduction (Cremer et Cremer 2001; Iborra et al. 2001; Lamond et Earnshaw 1998). Toutes ces fonctions doivent être coordonnées dans l'espace nucléaire malgré la complexité des génomes de mammifères et la concentration élevée de facteurs protéiques et de sites de liaison présents. Pour mener à bien toutes ses tâches, le génome est fonctionnellement organisé et compartimentalisé dans sa séquence autant que dans l'espace nucléaire de façon à regrouper certains loci et facteurs protéiques dans l'espace et le temps. Cette organisation réduit la dépendance vis-à-vis des processus de diffusion aléatoire en favorisant certains contacts privilégiés entre molécules. En bref, diviser pour régner. Étant donné que certaines régions génomiques possèdent des adresses spatio-temporelles préférentielles dans le noyau, les loci génomiques possèdent un environnement spatial dynamique relativement défini composé d'un nombre restreint d'autres loci génomiques.

Le thème central de cette thèse est le suivant : les interactions et la proximité spatiale entre loci génomiques sont liées aux fonctions du génome et à son organisation. Nous explorerons cette thématique à travers trois projets de recherche. Un premier projet concerne l'étude des activités de recombinaison à l'œuvre *in vivo* dans l'embryon de souris. Ce projet explore comment les interactions entre régions d'ADN situées en grande proximité les unes des autres peuvent conduire à des événements de recombinaison homologue, et comment ces événements peuvent modifier l'organisation linéaire du génome de façon héritable. Un deuxième projet analyse l'influence de l'environnement spatial d'un locus donné sur le choix de ses partenaires de recombinaison. Les résultats obtenus suggèrent que les loci qui sont fréquemment à proximité spatiale d'un locus donné et qui forment son environnement spatial sont favorisés pour interagir par recombinaison avec ce dernier. Finalement, un troisième projet investigate l'environnement spatial du gène *HoxB1* pendant l'induction de son expression dans le contexte de la différenciation de cellules souches embryonnaires (ES) de souris induite par l'acide rétinoïque. Il illustre

comment l'environnement spatial d'un locus peut être modifié en lien avec son activité transcriptionnelle, et comment ce dynamisme modifie les possibilités d'interactions entre ce locus et le reste du génome.

### **Première partie : Rôles des interactions entre loci dans le remodelage du génome par recombinaison génétique**

Suite au séquençage des génomes de la souris et de l'humain, il est devenu évident que le génome est hautement organisé au niveau de sa séquence linéaire (Lander et al. 2001; Waterston et al. 2002). Plusieurs gènes co-régulés tels que les gènes Hox, les gènes d'immunoglobuline, ou les gènes de globine sont en effet organisés en complexes, alors que des gènes ayant des fonctions ou des patrons d'expression similaires (tels les gènes ubiquitaires ou *housekeeping genes*) sont souvent regroupés dans le génome (Lercher et al. 2002). Des regroupements de gènes faiblement exprimés dans certains tissus spécifiques (Reymond et al. 2002) ou faisant parties de voies de signalisations spécifiques ont aussi été observés dans les génomes de plusieurs organismes procaryotes et eucaryotes (Cohen et al. 2000; Lee et Sonnhammer 2003). Ces regroupements de gènes suggèrent ainsi l'existence d'environnements spécifiques au niveau de la séquence même du génome. Bien qu'étant de nature modulaire, les gènes semblent donc être évolutivement maintenus dans des environnements spécifiques formés par d'autres loci génomiques.

L'observation de ce type d'organisation renforce une vision intuitive du génome en tant qu'entité statique : celui-ci étant la source de ce que nous sommes, il se doit d'être un gardien fiable, constant et stable de l'identité des organismes vivants. Or, malgré l'importance que peut avoir leur organisation linéaire, les génomes sont remarquablement dynamiques et plusieurs mécanismes existent pour les remodeler et en modifier l'organisation. Ceci est en effet suggéré par de nombreuses observations, allant des insertions de rétrotransposons causant diverses maladies ou modifiant extensivement les génomes des mammifère ou de plantes (Kalendar et al. 2000; Kazazian 1998; Kazazian et Moran 1998) jusqu'à la démonstration de l'existence de variations dans le nombre de copies de vastes régions génomiques chez la souris et l'humain (Iafrate et al. 2004; Li et al. 2004). Les cellules peuvent donc non seulement survivre à d'importants remodelages de

leur matériel génétique, mais ces remodelages sont probablement des vecteurs puissants de l'évolution des génomes lorsqu'ils donnent naissance à des modifications héritables.

Parmi les mécanismes de remodelage du génome, la recombinaison génétique est particulièrement importante. Ces mécanismes impliquent par définition des interactions physiques entre séquences d'ADN. Les régions génomiques qui sont en proximité linéaire, comme les séquences en répétition directe par exemple, ont ainsi l'opportunité d'interagir à haute fréquence entre elles. Ces interactions entre séquences qui font partie d'un micro-environnement génomique particulier pourraient dans ce contexte être évolutivement utiles pour homogénéiser les séquences régulatrices ou codantes et ainsi participer à la définition fonctionnelle de cet environnement. En effet, l'activité de ces mécanismes de recombinaison dans certains types cellulaires comme les cellules embryonnaires et germinales peut conduire à des modifications héritables des chromosomes et ainsi participer à leur évolution. Par contre, peu d'études ont à ce jour investigué la recombinaison se produisant *in vivo* dans ces types cellulaires. Une première partie de cette thèse décrit le développement d'une nouvelle méthodologie qui permet de caractériser la fréquence des divers mécanismes de recombinaison à l'œuvre *in vivo* dans l'embryon de souris et de démontrer qu'ils peuvent conduire à des modifications héritables du génome. Les prochaines sections d'introduction décrivent donc les mécanismes de recombinaison génétique et les facteurs impliqués dans leurs fonctions.

### **Recombinaison génétique : mécanismes et rôles dans le remodelage des génomes de mammifères**

La recombinaison est impliquée dans plusieurs événements cellulaires importants tels que la méiose, le remodelage des locus d'immunoglobuline et d'anticorps (recombinaison V(D)J et *class-switch recombination*), l'hypermutableté somatique de ces loci et la réparation de l'ADN. Elle peut aussi participer à l'évolution du génome si elle se produit dans les cellules germinales ou embryonnaires. Étant donné qu'elle est stimulée par les dommages à l'ADN tels que les cassures double-brin (CDB), la recombinaison est généralement perçue comme étant un mécanisme de réparation du matériel génétique (Liang et al. 1998; Sargent et al. 1997). La plupart des modèles qui décrivent les mécanismes de recombinaison illégitime et homologue ont donc pour point de départ

l'introduction d'une CDB (Valerie et Povirk 2003), mais d'autres modèles décrivent aussi la recombinaison induite par une brisure simple-brin ou *nick* dans l'ADN (Lee et al. 2004; Meselson et Radding 1975). Notons que dans le cas de la recombinaison spontanée, la recombinaison pourrait provenir de l'introduction d'une brisure simple ou double-brin dans l'ADN ou même être initiée avant toute brisure. Les mécanismes impliqués dans ces événements ne sont donc pas toujours clairs (Malone et al. 1990).

Les CDB sont fréquentes dans la vie cellulaire : un fibroblaste humain peut en encourir environ 50 par cycle cellulaire, en particulier à cause de problèmes lors de la réplication (Vilenchik et Knudson 2003). Certaines enzymes telles que les topoisomérases peuvent aussi, pour relaxer le stress torsionnel de l'ADN, introduire des brisures dans l'ADN (Ikeda 1994). D'autre part, des agents extérieurs tels que les radiations ionisantes et certains agents chimiques peuvent générer des CDB. Finalement, la recombinaison observée dans le système immunitaire et lors de la méiose est induite par des enzymes spécifiques qui introduisent des CDB dans un contexte régulé (Dudley et al. 2005; Keeney 2001).

Les CDBs non régulées peuvent être particulièrement nocives pour la cellule. D'ailleurs, la présence d'une seule de celles-ci peut conduire à l'arrêt du cycle cellulaire dans des cellules en culture (Huang et al. 1996). En effet, après la reconnaissance du dommage, un arrêt du cycle est normalement signalé pour permettre à la cellule de réparer son ADN. Si cette réparation est impossible, la cellule se dirigera généralement vers l'apoptose (Rich et al. 2000). Puisqu'une réparation erronée des dommages à l'ADN peut conduire à des mutations, un contrôle efficace des mécanismes de réparation de l'ADN est crucial.

### Deux types de mécanismes de recombinaison distincts : homologues et illégitimes

La recombinaison génétique peut être divisée en deux types principaux : homologue et illégitime ou non-homologue (Valerie et Povirk 2003). La recombinaison homologue fait usage de l'homologie entre séquences pour remodeler l'ADN. Ce type de mécanisme est prépondérant en phase S et G2, alors les chromosomes sont dupliqués, en particulier à cause de la régulation transcriptionnelle des gènes des facteurs de recombinaison impliqués (Takata et al. 1998). La recombinaison illégitime, aussi appelée *non-homologous end-joining* (NHEJ) ou recombinaison non-homologue, permet quant à elle de ligaturer des

extrémités libres d'ADN sans avoir recours à l'homologie (Pastwa et Blasiak 2003). Ce type de mécanisme est actif pendant tout le cycle cellulaire, mais le manque de compétition de la part de la recombinaison homologue en G1 rend la recombinaison illégitime particulièrement prépondérante durant cette phase du cycle (Takata et al. 1998). Le NHEJ est généralement moins fidèle que la recombinaison homologue, puisque la ligation des extrémités est souvent accompagnée de délétions et de réarrangements de celles-ci, présumément pour mettre à jour des régions de micro-homologie facilitant la réparation (Nicolas et Young 1994; Roth et Wilson 1986). La recombinaison illégitime semble être le mécanisme le plus fréquemment utilisé pour réparer les dommages à l'ADN chez les mammifères. En effet, certaines études rapportent que les brisures sont réparées par ces mécanismes dix fois plus fréquemment que par la recombinaison homologue (Sargent et al. 1997). Par contre, d'autres études ont rapporté que la recombinaison homologue pouvait dans certains cas représenter jusqu'à 50% des événements de réparation (Liang et al. 1998).

#### Reconnaissance des CDB et signalisation initiale

La réparation d'une lésion par recombinaison requiert d'abord sa détection. Les PI3-kinases ATM, ATR et DNA-PK sont centrales pour la détection des CDB chez les mammifères et pour la signalisation qui y est associée, et ce de façon plus ou moins redondante pour ce qui est de ATR et ATM (Durocher et Jackson 2001). La façon dont ces protéines reconnaissent les dommages n'est pas claire, mais pourrait entre autre être liée à la détection de changements dans la structure supérieure de la chromatine (Bakkenist et Kastan 2003). Elles peuvent aussi se lier directement aux extrémités d'ADN et font partie d'un complexe de reconnaissance des dommages appelé BASC (*BRCA-associated complex*), qui contient aussi plusieurs protéines de réparation (Wang et al. 2000). Ces interactions leur permettent donc d'être parmi les premiers effecteurs de la cascade de signalisation couplée à la réponse aux dommages à l'ADN.

Les kinases ATM et ATR sont impliquées dans le contrôle du cycle cellulaire et dans les arrêts induits par les dommages à l'ADN. Entre autres, elles régulent l'activité de p53 en la phosphorylant et en phosphorylant son inhibiteur, MDM2. p53 est une protéine centrale dans la réponse aux dommages et lorsqu'activée, elle peut en réguler un grand nombre d'aspects. Par exemple, p53 interagit avec des protéines telles que Rad51, BLM, WRN et

BRCA2 qui sont importantes pour la réparation. Elle peut aussi se lier directement à divers intermédiaires de recombinaison tels que les jonctions de Holliday et influencer leur résolution (Valerie et Povirk 2003). Finalement, p53 induit la transcription de p21, un inhibiteur de CDK (kinase dépendante des cyclines ou *cyclin dependent kinase*) qui permet un arrêt du cycle cellulaire en G1/S. ATM et ATR phosphorylent aussi les protéines CHK1 et 2, qui sont impliquées dans les arrêts en G1/S et en G2. D'autre part, ATM et ATR phosphorylent aussi des facteurs de réparation de l'ADN tels que NBS1, BRCA1, l'hélicase BLM, la protéine RAD51 et c-Abl, qui peut réguler Rad51 par phosphorylation (Valerie et Povirk 2003).

Un des premiers événements se produisant au site de CDB est la phosphorylation de l'histone  $\gamma$ -H2AX par ATR et ATM (Burma et al. 2001). Cette phosphorylation couvre un domaine d'environ 1 Mb autour de la lésion chez les mammifères (Rogakou et al. 1998) et facilite le recrutement de plusieurs protéines impliquées dans la réponse aux CDBs. Parmi celles-ci, mentionnons des facteurs de remodelage de la chromatine tels que NuA4, INO80, SWR1 (Downs et al. 2004; Morrison et al. 2004; van Attikum et al. 2004) et plusieurs protéines contenant des motifs BRCT, tels que BRCA1 et Nbs1 (Celeste et al. 2003; Kobayashi et al. 2002).

Le complexe MRN (Mre11/Rad50/Nbs1) est aussi important pour la réponse initiale aux CDB et est rapidement recruté aux sites de dommages. Il interagit avec les extrémités d'ADN et les maintient à proximité les unes des autres, ce qui facilite le NHEJ (Huang et Dynan 2002). Mre11 est une nucléase qui est nécessaire pour la dégradation des extrémités d'ADN précédant la recombinaison homologe ou illégitime, comme nous le verrons plus bas (Paull et Gellert 2000). La protéine Nbs1 possède un domaine BRCT qui interagit avec l'histone H2AX phosphorylée de façon à stabiliser le complexe MRN au site de dommage (Kobayashi et al. 2002). D'autre part, le complexe MRN est capable d'ouvrir les duplexes d'ADN et pourrait ainsi faciliter des étapes subséquentes de recombinaison (Paull et Gellert 1999).

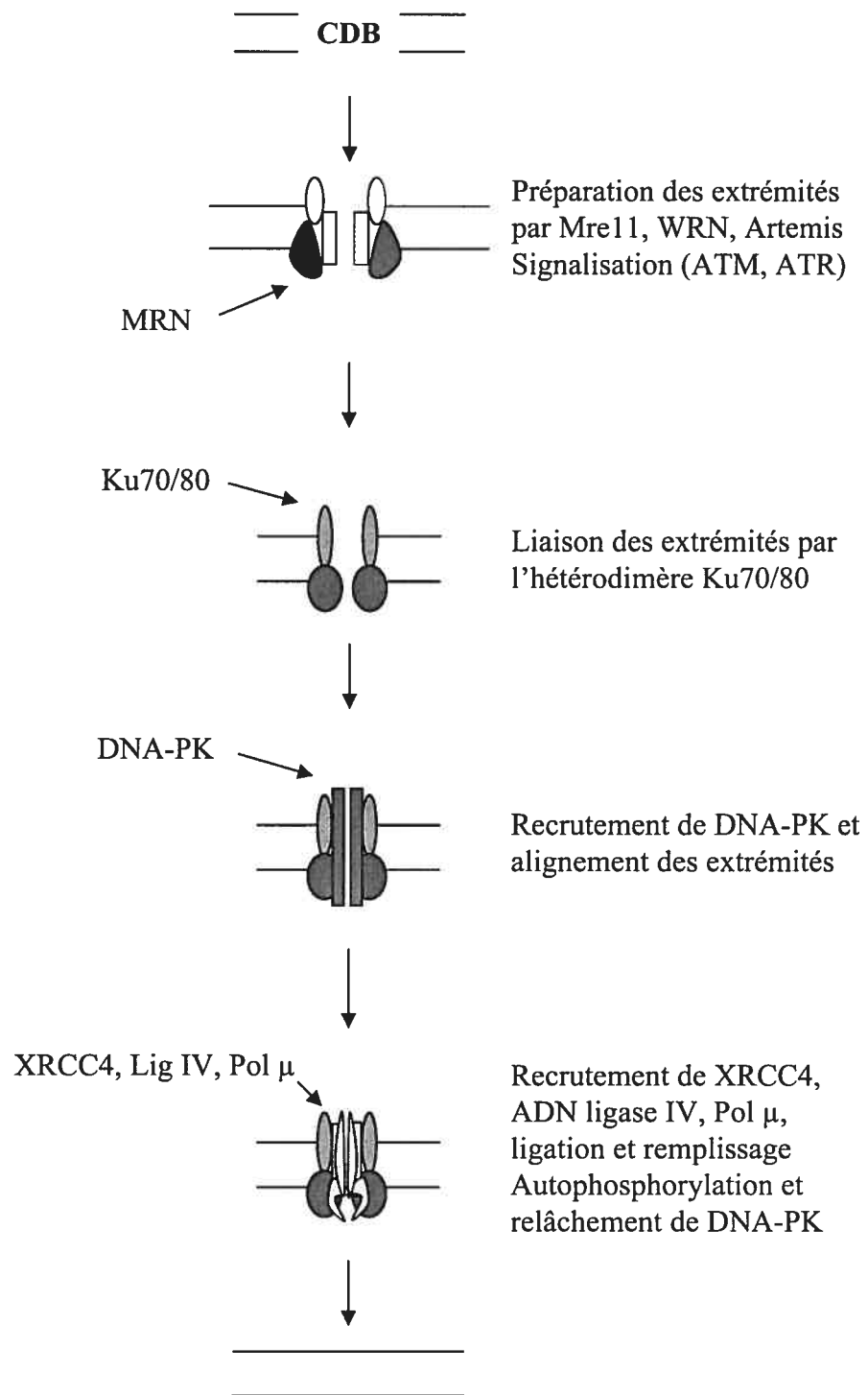
Finalement, en plus de la phosphorylation de l'histone  $\gamma$ -H2AX, d'autres modifications aux histones ont un rôle à jouer dans la réponse aux dommages à l'ADN (Wurtele et Verreault 2006). L'acétylation de la lysine 56 et la méthylation de la lysine 79 de H3 sont d'un intérêt



particulier (Giannattasio et al. 2005; Masumoto et al. 2005). Ces modifications sont présentes dans les histones nouvellement synthétisées et ont des rôles importants dans la réparation de l'ADN puisque des levures incapables de modifier ces résidus sont sensibles aux dommages à l'ADN lors de la réplication. Le rôle de l'acétylation en lysine 56 de H3 n'est pour le moment pas clair, mais pourrait agir au niveau du recrutement de facteurs aux sites de dommage situés près des fourches de réplication. Les lysines 79 méthylées sont reconnues par des protéines contenant des domaines tudor, dont Rad9 ou 53BP1 qui sont impliquées dans les arrêts du cycle cellulaire en G1 et en S (Huyen et al. 2004).

### Recombinaison non-homologue : mécanismes et facteurs protéiques impliqués

Après la reconnaissance du dommage et la signalisation initiale, la cellule doit choisir le mode de réparation de la lésion. Considérons premièrement la recombinaison non-homologue. Ce type de mécanisme permet la ligation d'extrémités libres d'ADN (Figure 1) mais introduit fréquemment des mutations au site de cassure puisque les extrémités de la CDB sont souvent dégradées de quelques paires de bases durant les étapes initiales de réparation (Nicolas et Young 1994). L'objectif de ces dégradations semble être la mise à jour de régions de micro-homologie entre les extrémités, qui peuvent faciliter leur appariement et leur ligation (Roth et Wilson 1986). Les dégradations autour du site de cassure peuvent cependant être plus extensives et conduire à des délétions de plusieurs kb (Wurtele et al. 2003). Les protéines Mre11 et WRN, ainsi qu'un complexe formé des protéines DNA-PK et Artemis, participent à cette préparation des extrémités (Cooper et al. 2000; Huang et al. 2000; Ma et al. 2002; Paull et Gellert 2000). Cependant, Mre11 est une exonucléase 3'-5' et cette directionnalité est incompatible avec les intermédiaires de recombinaison connus. La protéine Artemis possède une directionnalité convenable (5'-3'), mais sa délétion ne conduit qu'à une modeste réduction de l'efficacité de recombinaison. L'identité des nucléases effectrices impliquées dans la résection des extrémités est donc peu claire en ce moment. Notons que la signalisation des dommages peut être liée à ces dégradations par le recrutement de facteurs liant les extrémités d'ADN: l'initiation de la signalisation et la dégradation des extrémités pourraient donc être concomitantes (Lisby et al. 2004). En phase S et G2, une compétition existe pour les étapes subséquentes de réparation entre les facteurs de recombinaison homologue et non-homologue : ces étapes préparatoires sont en effet partagées au moins en partie entre les deux mécanismes.

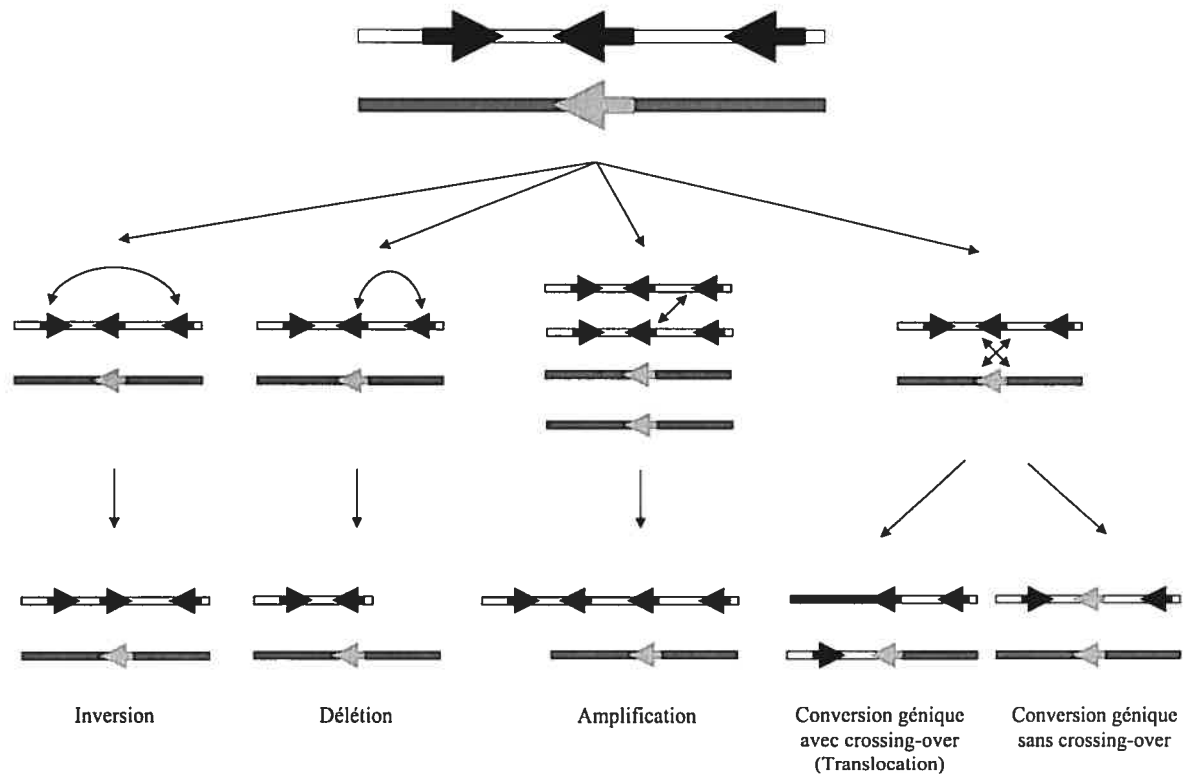


**Figure 1 : La recombinaison non-homologue ou illégitime**

Après la préparation des régions situées en bordure du dommage, l'hétérodimère Ku70/80 se lie aux extrémités d'ADN (Walker et al. 2001) et recrute la PI3-kinase DNA-PK, qui peut, entre autres, s'autophosphoryler et phosphoryler les protéines Ku. Son autophosphorylation résulte en son éventuel déplacement du site de brisure (Pastwa et Blasiak 2003). Les protéines Ku et DNA-PK sont importantes pour maintenir les extrémités d'ADN à proximité les unes des autres, de par leurs interactions Ku-Ku et DNA-PK-DNA-PK (Cary et al. 1997; DeFazio et al. 2002). La ligation est réalisée par les protéines XRCC4 et ADN ligase IV alors que d'autres protéines finalisent la réparation, telle que la polymérase  $\mu$  qui peut combler les régions simple brin par exemple (Pastwa et Blasiak 2003).

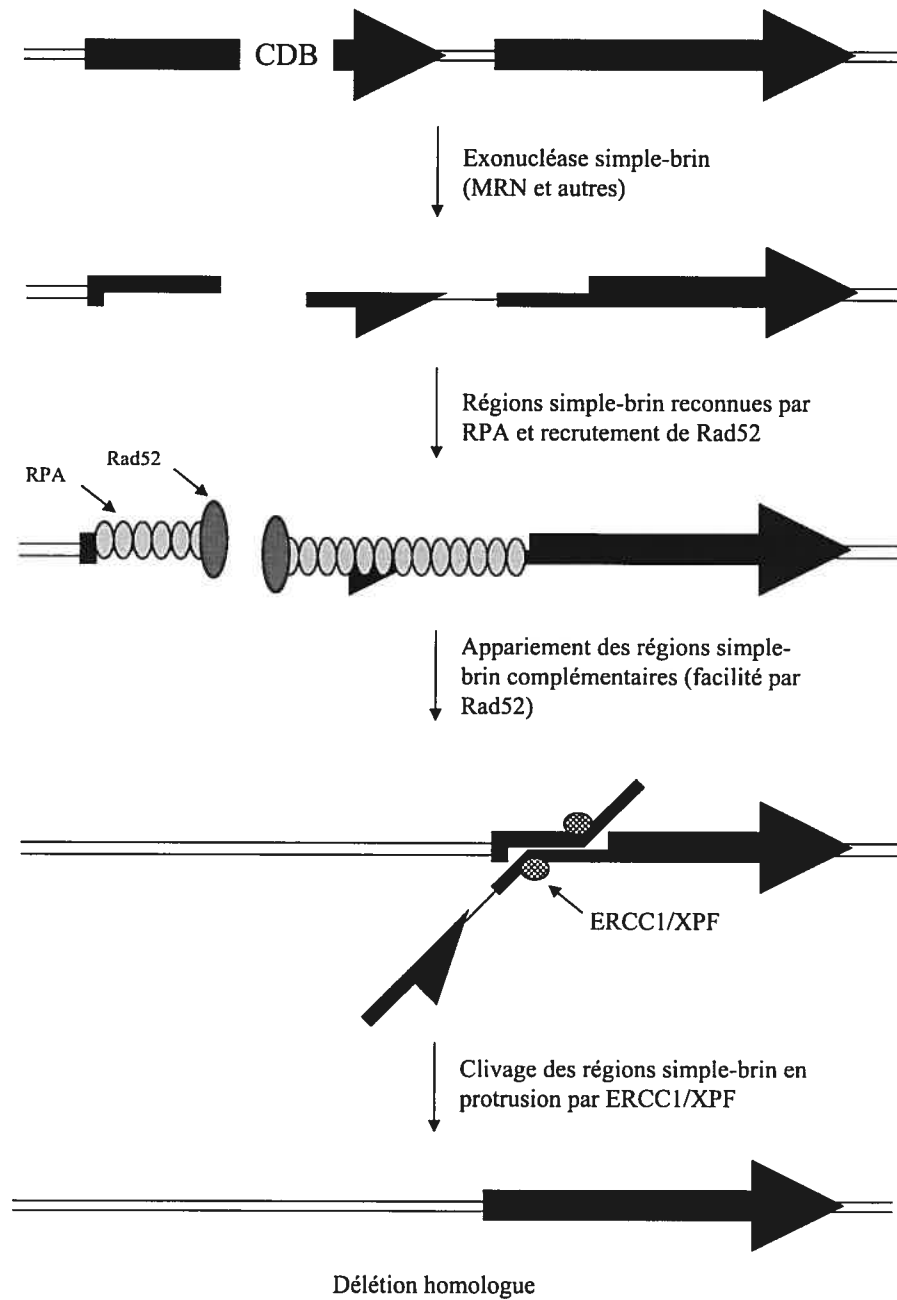
### Recombinaison homologue : mécanismes et facteurs protéiques impliqués

La recombinaison homologue fait usage de l'homologie entre séquences et peut donner naissance à une foule de produits, tels que des amplifications, délétions, inversions ou translocations, selon la configuration des séquences impliquées (Figure 2). Après la préparation des extrémités, une région plus ou moins grande d'ADN est convertie sous forme simple-brin, avec l'aide probable du complexe MRN et de d'autres activités nucléases moins connues (Trujillo et al. 1998). Plusieurs mécanismes peuvent à ce moment entrer en jeu. Considérons premièrement le *single-strand annealing* ou SSA (Figure 3). Ce mécanisme agit surtout sur les séquences en répétition directe, mais peut aussi agir si deux brisures double-brin sont présentes dans des séquences homologues plus distantes ou même situées sur des molécules d'ADN différentes (Lin et al. 1984; Lin et al. 1990a; Lin et al. 1990b). La zone de dégradation simple brin est premièrement reconnue par RPA (*replication protein A*) qui coopère avec Rad52 (Shinohara et al. 1998). Rad52 facilite l'appariement entre séquences homologues : les séquences homologues en répétition directe « glissent » donc les unes sur les autres jusqu'à leur appariement (Van Dyck et al. 2001). Les régions non-appariées sont clivées par l'endonucléase ERCC1/XPF (Sargent et al. 2000). Les régions simple-brin restantes sont comblées par des polymérases inconnues. Ce mécanisme résulte en une délétion des séquences situées entre les répétitions, mais peut conduire à d'autres produits (translocation par exemple) selon qu'il se produit entre des régions d'un même chromosome ou entre chromosomes différents.



**Figure 2 : Conséquences de la recombinaison homologue entre séquences non-alléliques sur l'organisation linéaire du génome**

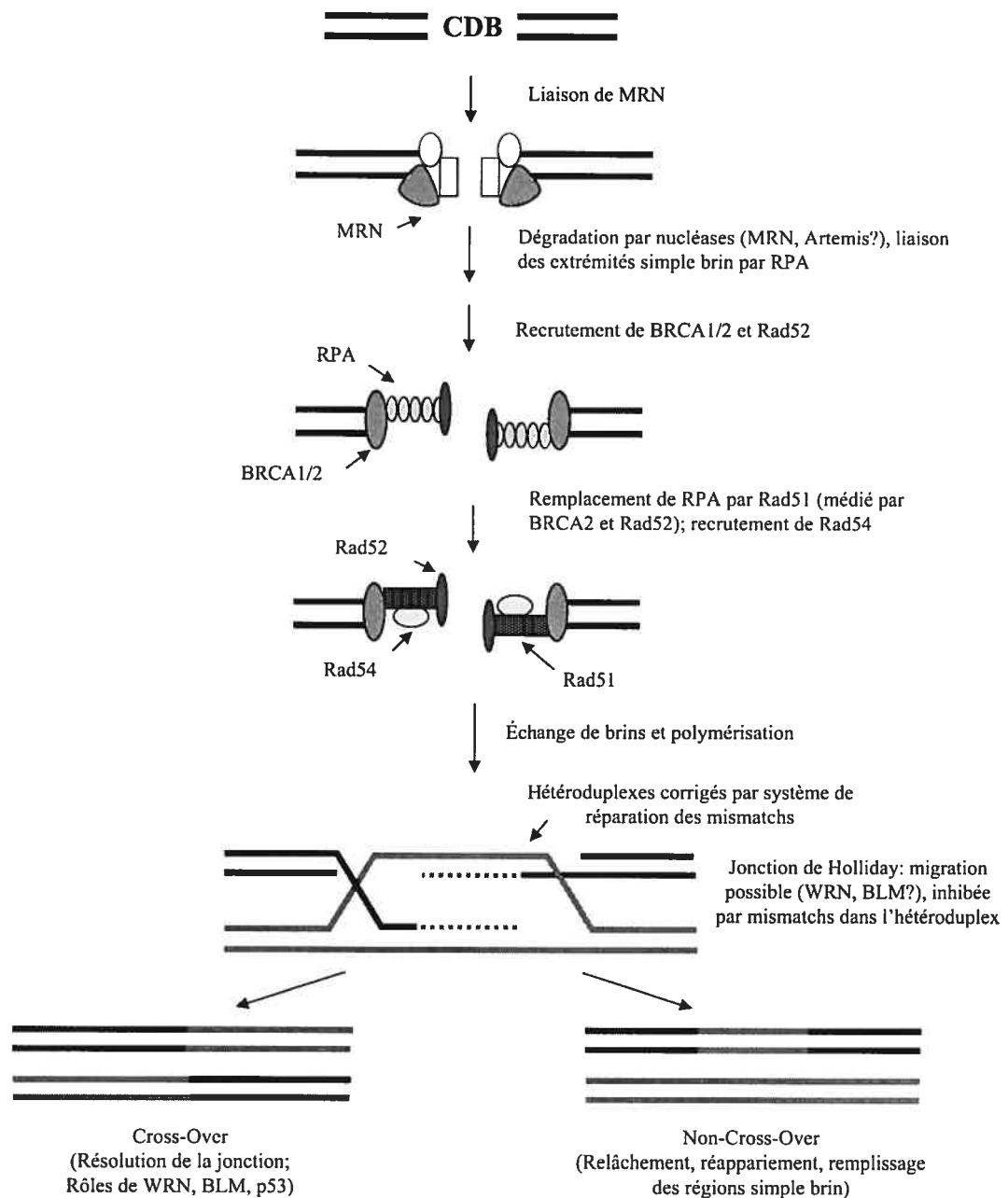
Les flèches représentent des séquences homologues. Les amplifications sont possibles après une étape de réplication du génome (grâce à la présence de la chromatide sœur).



### Figure 3 : Le *single-strand annealing*

Les deux flèches représentent des séquences homologues en répétition directe. Rad52 se lie sous forme d'heptamère aux extrémités libres d'ADN, mais est représenté par un seul ovale pour plus de clarté. La finalisation de la réparation implique un remplissage simple-brin par des facteurs inconnus.

Les mécanismes de « *strand invasion* » ou d'invasion de duplex d'ADN peuvent avoir comme point de départ une CDB. Plusieurs modèles furent proposés, mais ceux-ci sont probablement médiés par les mêmes facteurs protéiques (Paques et Haber 1999)(Figure 4). Le modèle classique d'invasion à partir d'une brisure double-brin fut décrit à partir d'observations réalisées chez la levure (Szostak et al. 1983). Une des premières protéines qui interagit avec les régions simple brin d'ADN est RPA, qui peut aussi participer à la signalisation du dommage (Wold 1997). BRCA2 et Rad52 sont ensuite recrutés et facilitent le remplacement de RPA par Rad51 et sa déposition sur l'ADN (Chen et al. 1999; New et al. 1998; Sung 1997; Yang et al. 2002). Rad51 peut alors catalyser les réaction d'échange de brin entre le duplexe d'ADN brisé et une séquence « donneuse » intacte (Baumann et al. 1996; Sugiyama et al. 1997; Sung 1994). Les protéines Rad52 et Rad54 facilitent l'invasion entamée par Rad51 (Dronkert et al. 2000; Shinohara et Ogawa 1998). Après l'invasion, une structure appelée « jonction de Holliday » est formée : la migration des branches de cette structure peut conduire à une région plus ou moins longue de conversion génique. De plus, il peut se produire une polymérisation d'ADN à partir des brins d'ADN envahisseurs : l'étendue de cette polymérisation détermine aussi la zone de conversion génique résultante. Les hélicases WRN et BLM interagissent avec les jonctions de Holliday et pourrait être impliquées dans leur migration et/ou dans leur résolution (Mohaghegh et al. 2001; Saintigny et al. 2002). Finalement, la résolution de la jonction de Holliday par l'action de nucléases ou par un relâchement suivi d'un réappariement peut conduire à des produits de conversion génique avec ou sans crossing-over. La région hétéroduplexe créée, dans laquelle chaque brin provient d'une molécule d'ADN distincte, peut d'autre part être un substrat pour les mécanismes de correction des mésappariements par les protéines de la famille MSH (Alani et al. 1994). Les produits résultant d'une telle correction d'hétéroduplexes posséderont nécessairement une région de conversion génique en faveur d'un ou l'autre des brins formant l'hétéroduplexe : l'existence de ces mécanismes avait d'ailleurs été postulée par le modèle original de réparation des CDB (Modrich 1991). Les étapes de polymérisation sont probablement médiées par la polymérase  $\eta$  chez les mammifères (McIlwraith et al. 2005).



#### Figure 4 : La recombinaison homologue

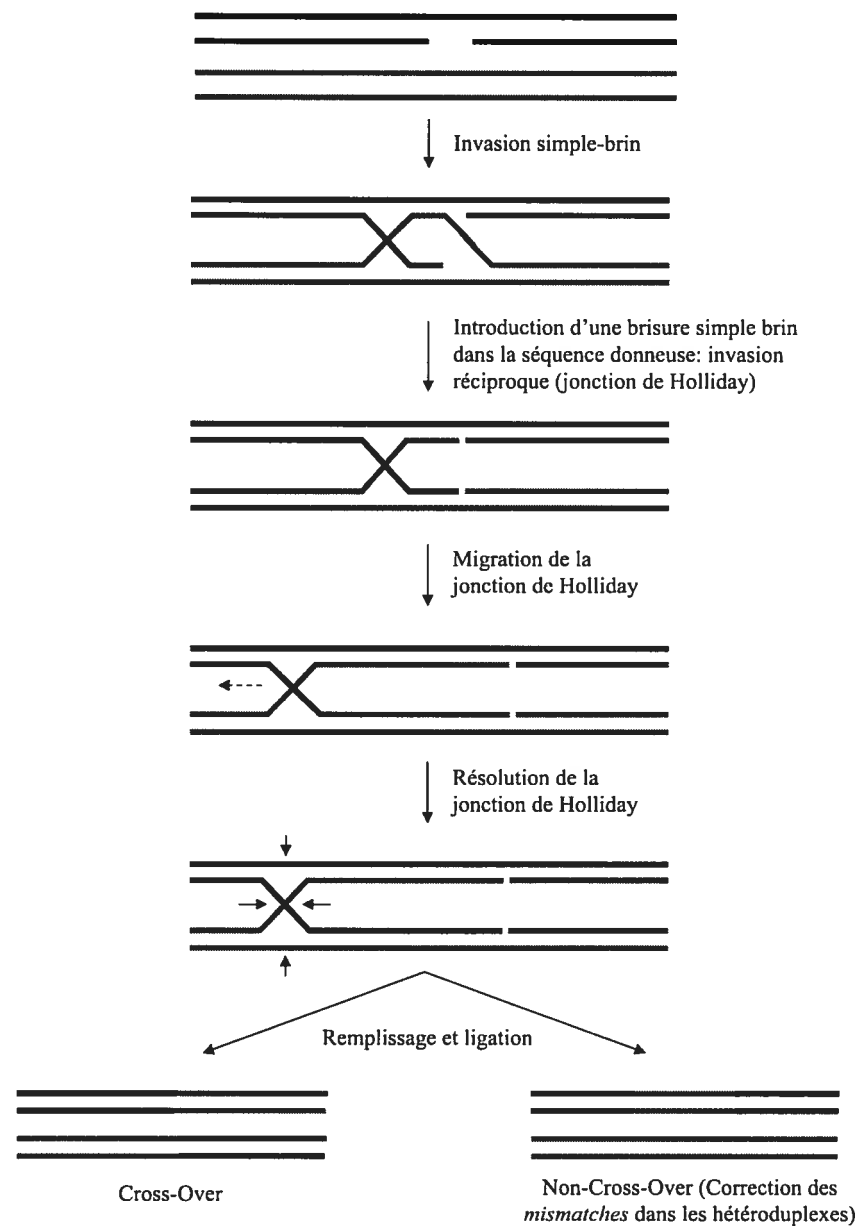
Le mécanisme d'invasion par les deux côtés de la CDB est illustré (Valerie et Povirk 2003); plusieurs autres modèles mécanistiques ont été proposés (voir plus bas).

D'autres modèles furent proposés pour décrire les mécanismes de recombinaison homologue par invasion de duplex d'ADN. Le modèle de Meselson-Radding ou modèle « Avimore » postule que la recombinaison est initiée par une brisure simple-brin : la résolution de l'échange de brins peut ainsi mener à des événements de conversion génique avec ou sans crossing-over (Figure 5) (Lee et al. 2004; Meselson et Radding 1975). Finalement, le mécanisme de *one-sided invasion* suppose pour sa part qu'une seule extrémité de la région endommagée participe à l'invasion du duplexe donneur (Figure 6) (Belmaaza et Chartrand 1994). Ce mécanisme permet de polymériser des régions génomiques qui dépassent la région d'homologie, et des mécanismes conceptuellement similaires permettent même de dupliquer des chromosomes entiers chez la levure (*break-induced replication*) (Malkova et al. 1996). Ce dernier mécanisme peut donc transférer des séquences d'un endroit à l'autre dans le génome.

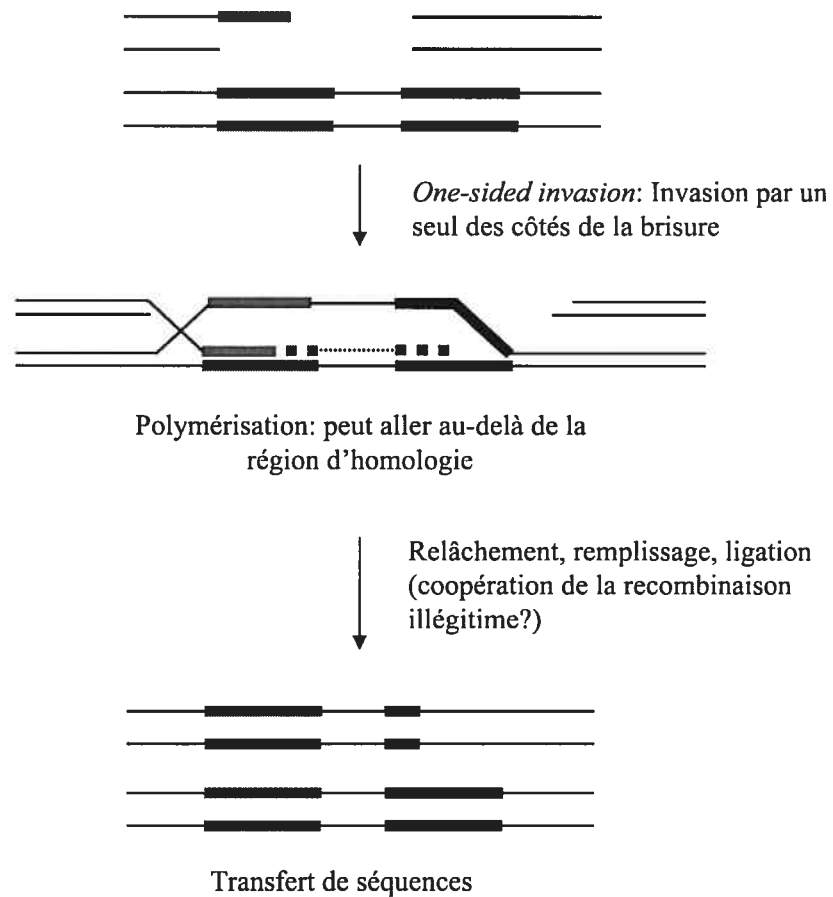
#### Implication de la recombinaison dans le remodelage et l'évolution des génomes de mammifères

Les événements de recombinaison non-homologue peuvent générer des mutations locales et des produits nocifs pour l'intégrité du génome comme des translocations ou des délétions extensives. La recombinaison homologue est donc généralement considérée comme un mécanisme de réparation plus fidèle que le NHEJ. Étant donné qu'elle se produit principalement en phase S ou G2, les séquences donneuses sont le plus souvent situées sur la chromatide sœur, ce qui limite les réarrangements (Johnson et Jasin 2000; Takata et al. 1998). Cependant, la recombinaison entre substrats non-alléliques est possible, et peut donner naissance à des translocations, inversions et délétions (voir Figure 2) (Elliott et Jasin 2002). De plus, des événements de « one-sided invasion » pourraient propager des séquences codantes ou régulatrices à travers le génome. Le génome étant parsemé de nombreuses séquences répétitives comme les LINE et les SINE (Long et Short Interspersed Nuclear Element, respectivement), des événements de recombinaison homologue non-alléliques les impliquant pourraient avoir d'importantes conséquences pour le génome. Dans les cellules embryonnaires ou germinales, ces événements peuvent d'ailleurs conduire à des modifications héréditaires ayant un impact sur l'évolution du génome.





**Figure 5 : Le modèle de recombinaison induit par une brisure simple-brin ou *nick* tel que proposé par Meselson et Radding (Meselson et Radding 1975).**



**Figure 6 : Le modèle de recombinaison par *one-sided invasion***

Le mécanisme de *one-sided invasion* permet le transfert de séquences qui flanquent la région d'homologie (Belmaaza et Chartrand 1994). Il faut noter que la résolution de la jonction de Holliday créé après l'invasion ou le relâchement du brin invasif suivi de son appariement avec l'autre côté de la CDB permet aussi théoriquement la création de produits de conversion génique avec ou sans crossing-over (non-illustré).

Dans les cellules souches embryonnaires en culture, la conversion génique sans-crossing-over est le résultat le plus fréquent des événements de réparation des CDB lorsque les séquences sont situées sur des chromosomes différents (Richardson et al. 1998). Par contre, lorsqu'une brisure double-brin est induite dans une séquence en répétition directe, la situation est moins claire. Il fut rapporté que les produits résultant de la réparation d'une brisure double-brin dans une répétition directe étaient majoritairement des délétions dans un système de cellule souche embryonnaire (Liang et al. 1998). Au contraire, dans des

cellules d'ovaires de hamster chinois (*Chinese hamster ovaries cells* ou *CHO*), ce type de recombinaison résulte surtout en des événements de conversion génique sans crossing-over (Bill et Nickoloff 2001). Cette étude suggère aussi que la recombinaison spontanée donne un spectre de produit de recombinaison différent de celle induite par les CDB, avec une prépondérance de délétions lors de la recombinaison spontanée. Contrastant avec ces derniers résultats, un autre groupe a démontré exactement le contraire dans les cellules L de souris (Liskay et al. 1984). Les mécanismes et les conséquences des événements de recombinaison spontanée entre séquences en répétition directe chez les mammifères sont donc peu clairs pour le moment.

Tel qu'indiqué ci-haut, la recombinaison homologue entre séquences non-alléliques peut modifier le génome dans un système de cellules embryonnaires en culture. Certaines évidences plus « archéologiques » suggèrent que des événements de recombinaison homologue non-allélique peuvent se produire *in vivo* dans les cellules germinales ou embryonnaires. Par exemple, les séquences répétitives LINE-1 (L1) sont plus homologues entre elles à l'intérieur d'une espèce qu'entre différentes espèces (Saxton et Martin 1998; Verneau et al. 1998), ce qui suggère que des mécanismes de conversion génique entre L1 sont à l'origine de cette homogénéisation. La recombinaison homologue entre L1 a aussi été impliquée dans la création de nouveaux éléments formés par la combinaison d'éléments existants (Saxton et Martin 1998) et dans la duplication du gène de la  $\gamma$ -globine qui se serait produite dans un ancien ancêtre des primates (Fitch et al. 1991). Finalement, des événements ancestraux de conversion génique sans crossing-over entre des gènes de  $\gamma$ -globine non-alléliques (Powers et Smithies 1986) et des réarrangements homologues intra-chromosomiques entre le gène BRCA1 et un pseudogène rapproché (Puget et al. 2002) furent retracés. Dans tous ces cas, cependant, le type cellulaire et le moment précis auxquels ces événements se sont produits sont peu clairs.

#### Système d'étude : Recombinaison dans l'embryon de souris et son impact sur l'évolution des génomes de mammifères

La plupart des études concernant les mécanismes de recombinaison et leurs prévalences ont été réalisées en cellules en culture en utilisant des systèmes d'induction de CDB. Peu de systèmes expérimentaux permettant de caractériser les événements de recombinaison

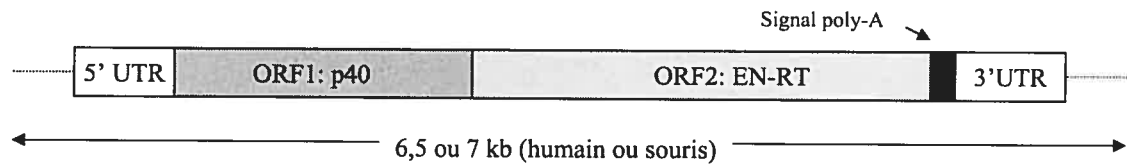
spontanée dans un système *in vivo* ont été rapportés. D'autre part, la nature précise des événements conduisant à la recombinaison spontanée n'étant pas connue (CDB, *nick* ou autre), il n'est pas clair à quel point les études concernant la réparation des cassures double-brin en cellules en culture sont représentatives des événements se produisant spontanément *in vivo*.

Jusqu'à maintenant, il fut démontré que la conversion génique sans crossing-over spontanée était fréquente dans les spermatozoïdes de souris (Murti et al. 1992). D'autre part, les événements de single-strand annealing entre plasmides linéarisés dans la région d'homologie peuvent se produire dans l'ovule et l'embryon pré-implantation de souris (Fiorenza et al. 2001; Pieper et al. 1992; Shimoda et al. 1991). Finalement, l'existence d'intégration de transgènes en tandem chez les souris peut être expliquée par des activités de recombinaison homologue dans l'embryon (Folger et al. 1982; Palmiter et Brinster 1986).

Aucune caractérisation systématique des mécanismes de recombinaison à l'oeuvre dans l'ovule fécondé et l'embryon de souris n'a été rapportée. Une première partie de cette thèse concerne donc l'étude des événements de recombinaison se produisant dans des ovules fécondés ou des embryons de souris par une nouvelle stratégie basée sur la micro-injection d'un substrat de recombinaison dans un ovule fécondé et sa caractérisation subséquente dans la souris transgénique résultante. Le substrat de recombinaison utilisé dans notre étude contient deux séquences répétitives L1 en répétition directe qui peuvent recombiner entre eux ou avec les L1 endogènes de souris. Donc, bien qu'ils ne soient pas le focus principal des projets décrits dans cette thèse, il convient de décrire brièvement la biologie des L1, puisque nous les utiliserons au niveau technique.

### Système d'étude : Les séquences répétitives L1

Les séquences répétitives L1 sont des rétrotransposons autonomes qui représentent environ 20% des génomes de souris et humain (Furano 2000; Ostertag et Kazazian 2001). Les L1 complets de souris mesurent environ 7 kb contre 6,5 kb chez l'humain, et possèdent deux cadres de lecture qui encodent une protéine qui lie et transporte l'ARN (p40) et une autre ayant des activités reverse-transcriptase et endonucléase (Figure 7).



**Figure 7 : Structure des éléments répétitifs LINE-1**

Les LINE-1 (Long Interspersed Nuclear Element ou L1) contiennent deux cadres de lecture (ORF1 et ORF2). La protéine p40 peut lier et servir au transport cytoplasmique de l'ARN. La protéine EN-RT est une endonucléase et réverse-transcriptase. Les séquences L1 possèdent un promoteur interne pour la polymérase III.

Les L1 se propagent par rétrotransposition : ils sont transcrits en un ARNm qui est traduit en réverse-transcriptase et endonucléase, puis cet ARNm est réverse-transcrit et intégré au génome. Les modèles de rétrotransposition des L1 postulent que l'endonucléase crée une brisure simple-brin dans une séquence génomique riche en A/T, qui peut alors s'apparier avec la région 3' de l'ARNm. Cette région sert alors d'amorce pour la réverse-transcription. Étant donné la faible processivité de leur réverse-transcriptase, les L1 sont souvent tronqués en 5'. Une minorité des L1 sont pleine longueur et actifs : on estime qu'il existe environ 3000 L1 actifs chez la souris, contre seulement 60 chez l'humain (Kazazian et Goodier 2002).

Les L1 sont distribués dans tout le génome, mais sont plus fréquents dans les régions riches en A/T et pauvres en gènes. Bien que l'intégration des L1 soit aléatoire, ils semblent avoir été retirés des régions riches en G/C et en gènes au cours de l'évolution, potentiellement à cause d'interférences transcriptionnelles (Pavlicek et al. 2001). Il existe d'ailleurs plusieurs cas d'intégration de L1 dans des séquences codantes qui sont associés à des pathologies (Ostertag et Kazazian 2001). Les L1 n'ont pas de fonctions connues, mais leur rétrotransposition peut copier et déplacer des séquences génomiques non-L1 flanquantes, ce qui suggère qu'ils ont pu participer à l'évolution du génome entre autre par le transfert d'exons ou de séquences régulatrices (*exon shuffling*) (Moran et al. 1999; Morrish et al. 2002). D'autres séquences répétitives utilisent aussi la machinerie des L1 pour se propager, tels que les SINEs (séquences Alu chez l'humain par exemple) et les pseudogènes (Dewannieux et al. 2003; Esnault et al. 2000; Wei et al. 2001).

Les L1 accumulent un grand nombre de mutations au fil des années (Hardies et al. 1986; Lander et al. 2001). Malgré tout, les séquences L1 d'une espèce sont en moyenne homologues à environ 95% entre elles. Les mutations subies par les L1 peuvent être utilisées pour les classer en plusieurs familles. L'existence de ces familles suggère d'ailleurs des mécanismes d'homogénéisation inter-alléliques, présumément par conversion génique inter-L1 et/ou par rétrotransposition active d'un nombre limité de L1 à un moment donné dans le parcours évolutif d'une espèce (Saxton et Martin 1998; Tremblay et al. 2000; Verneau et al. 1998).

### **Deuxième et troisième parties : Rôles des interactions entre loci dans la définition des environnements spatiaux génomiques et l'organisation fonctionnelle du génome**

Les interactions entre loci proximaux peuvent donc conduire à des modifications de la séquence linéaire d'une molécule d'ADN et ainsi participer à l'évolution des environnements génomiques décrits précédemment. Au-delà de ce niveau linéaire d'organisation de sa séquence, l'ADN est aussi spatialement organisé dans le noyau. Cette organisation spatiale conduit à une certaine spécificité dans les interactions entre loci génomiques, ce qui définit des environnements spatiaux correspondant aux loci auxquels un locus donné a préférentiellement accès et avec lesquels il peut interagir plus fréquemment.

Une étude récente de notre laboratoire a démontré que les partenaires de recombinaison d'un locus humain donné ne sont pas aléatoirement distribués dans le génome et que certains partenaires préférentiels existent (D'Anjou et al. 2004). Dans une deuxième partie de cette thèse, nous décrivons l'environnement spatial génomique de ce locus, de façon à vérifier si la composition de cet environnement est liée à son choix de partenaires de recombinaison.

La raison première expliquant la spécificité des interactions entre loci est probablement l'organisation fonctionnelle du noyau. La transcription est une des fonctions nucléaires qui est liée à l'organisation spatiale de l'ADN. Dans une troisième partie de cette thèse, nous caractérisons l'environnement spatial du locus HoxB1 durant l'induction de son expression. Nous analyserons comment l'environnement spatial d'un locus donné peut être modifié par

des processus nucléaires comme la transcription, et comment ce dynamisme influence les relations entre le locus exprimé et le reste du génome.

Étant donné les limitations des méthodes existantes, nous avons développé de nouvelles méthodologies, basées sur le Chromosome Conformation Capture (3C) originalement décrit par Dekker et ses collaborateurs (Dekker et al. 2002), qui nous ont permis de déterminer quels sont les loci qui forment l'environnement spatial d'un locus donné sans présumer de leur identité au préalable comme c'est le cas pour l'approche 3C. Cette thèse décrit donc aussi le développement de ces nouvelles méthodologies originales. Les prochaines sections d'introduction expliqueront comment le génome est organisé dans le noyau, comment cette organisation peut contribuer à déterminer l'environnement spatial spécifique d'un locus et en quoi cette organisation peut être liée aux différentes fonctions du génome.

### **L'ADN existe sous forme de chromatine compacte dans le noyau**

L'organisation spatiale de l'ADN débute par sa compaction sous forme de chromatine. L'ADN est en effet enroulé et compacté via ses interactions avec les histones pour former une structure nucléo-protéique appelée nucléosome. Les histones forment un octamère composé de deux exemplaires de chacune des histones « centrales » (*core histones*) H2A, H2B, H3 et H4 ou certains variants (Margueron et al. 2005; Pusarla et Bhargava 2005) qui enroule 1,75 tours d'ADN comptant pour 147 paires de bases (pb) (Luger et al. 1997). Le nucléosome est donc le premier niveau d'organisation spatiale de la chromatine. Lorsque la chromatine est décondensée, elle ressemble à un « collier de perle » (*beads on a string*) : dans cette conformation, il y a peu de contacts entre les nucléosomes et ceux-ci sont séparés par une région d'ADN appelée ADN de liaison ou « *linker DNA* ». Par contre, en conditions physiologiques les queues N-terminales des histones centrales contactent l'ADN de liaison et peuvent influencer l'empaquetage de l'ADN par les nucléosomes en collaboration avec les histones de liaison. Une fibre de chromatine peut ainsi se compacter *in vitro* en fibre de 30 nm de diamètre en présence de concentration de sel physiologique, alors que la forme en « collier de perle » n'est observée que lorsque les concentrations en sel sont basses (Thoma et Koller 1977; Thoma et al. 1979). Des fibres de plus grand diamètre qui sont dues à des interactions intra ou inter-fibres de chromatine sont aussi observées *in vitro* à haute concentration de sel. Ces observations *in vitro* furent confirmées

*in vivo* par la visualisation de la chromatine dans des noyaux à l'interphase (Belmont et Bruce 1994). Des fibres de chromatine d'un diamètre de 30 nm furent observées dans le noyau, suivies de niveaux supérieurs de compaction formant des fibres de 60-80 nm et de 100-130 nm nommées « fibre chromonème ». Par contre, la conformation en « collier de perle » de la chromatine ne fut pas observée directement *in vivo*, ce qui suggère que la majorité des fonctions de la chromatine ont pour substrat premier une fibre compactée à divers degrés (Figure 8).

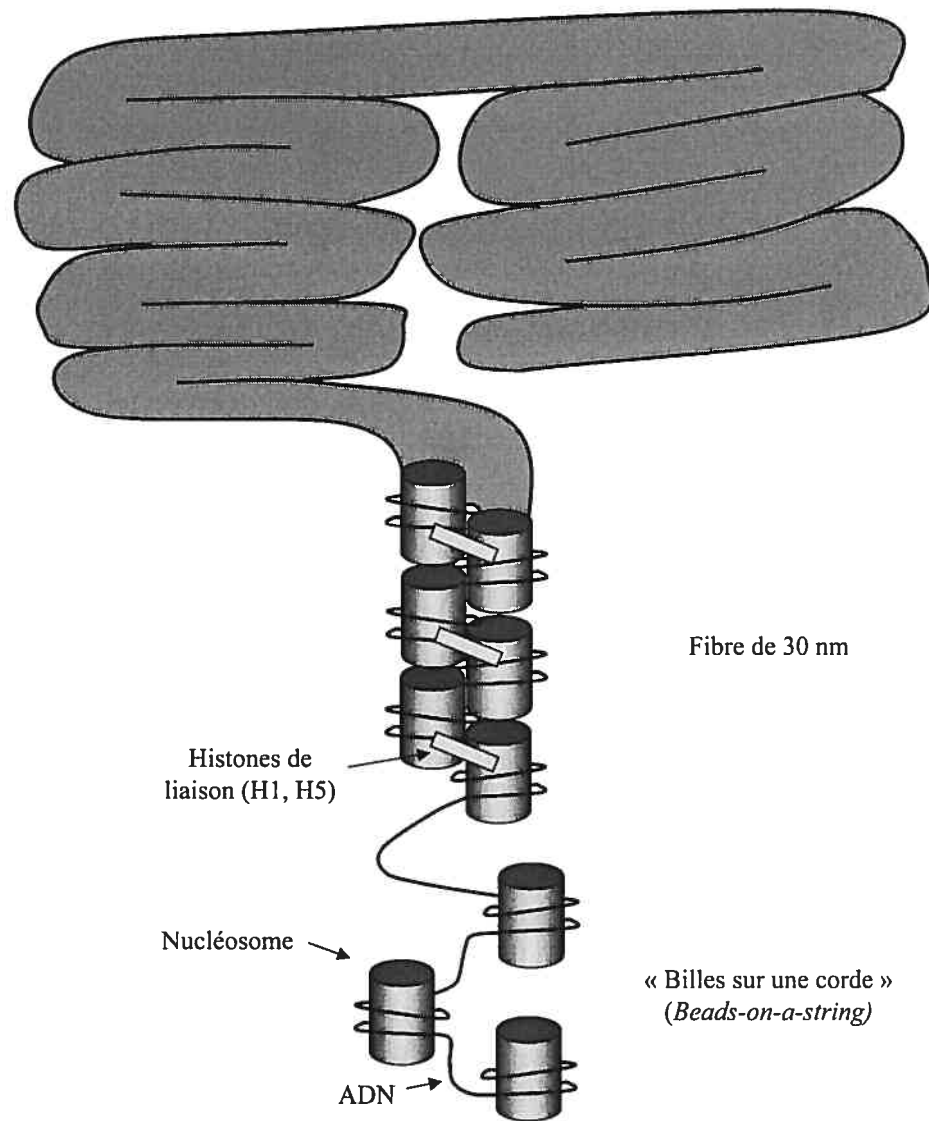
### Rôles des histones dans les mécanismes de compaction de la chromatine

Les mécanismes de repliement supérieur de la fibre de chromatine dépendent en grande partie des histones et des interactions qu'elles ont entre elles et avec l'ADN. L'histone de liaison H1 (et son variant H5 dans certains tissus) influence les angles d'entrée et de sortie de l'ADN du nucléosome (Grigoryev 2001) et est impliquée dans l'empaquetage serré des nucléosomes puisque leur perte résulte en une impossibilité de formation de fibre de 30 nm *in vitro* (Hizume et al. 2005; Leuba et al. 1998a; Leuba et al. 1998b). L'action des histones de liaison sur l'angle d'entrée et de sortie de l'ADN du nucléosome peut donc présumément conduire à divers degrés de compaction (Bednar et al. 1998; Woodcock et al. 1993). Les histones de liaison peuvent aussi s'oligomériser de façon dépendante de l'ADN via leur domaine globulaire et maintenir plusieurs molécules ou régions d'ADN à proximité les unes des autres, ce qui pourrait avoir un impact sur la formation des fibres « chromonèmes » (Thomas et al. 1992).

Les histones centrales sont évidemment essentielles au repliement supérieur de la fibre de chromatine. Une fibre de chromatine contenant des histones centrales H2A/B et H3/4 trypsinisées pour enlever les queues N-terminales ne peuvent former des structures d'ordre supérieur (Allan et al. 1982; Fletcher et Hansen 1995; Moore et Ausio 1997; Tse et Hansen 1997). Des évidences cristallographiques suggèrent aussi que les domaines N-terminaux de l'histone H4 d'un nucléosome interagissent avec les histones H2A/B d'un autre nucléosome pour compacter un rang de nucléosome (Luger et al. 1997). D'autre part, certains variants d'histones peuvent avoir un impact sur le repliement supérieur de la chromatine. Par exemple, un ensemble de nucléosome contenant le variant H2AZ ne peut se condenser en forme supérieures au-delà de la forme de 30nm (Fan et al. 2002).



Fibre chromonème (60 à 130 nm): repliement d'ordre supérieur de la chromatine et contacts inter-fibres



**Figure 8 : Plusieurs niveaux hiérarchiques de repliement supérieur de la fibre de chromatine**

Le repliement de la fibre de chromatine permet sa compaction et les associations entre fibres de chromatine. Le modèle est adapté des figures de Horn et Peterson et de Hansen (Hansen 2002; Horn et Peterson 2002).

## Modulation du degré de compaction de la chromatine par les modifications post-traductionnelles aux histones et les facteurs de remodelage de la chromatine

Une fibre de chromatine possède donc plusieurs niveaux de repliement supérieur, et cette organisation le long de la fibre dépend en grande partie des histones. Or, ce type d'organisation se doit d'être dynamique pour pouvoir répondre aux besoins cellulaires changeants. Pour modifier la compaction locale de la chromatine, la cellule peut opérer une multitude de modifications post-traductionnelles aux histones, telles que l'acétylation, la phosphorylation et la méthylation (Eberharter et Becker 2002). Ces modifications forment le « code histone » qui permet de former des structures permissives ou répressives aux fonctions de l'ADN (Jenuwein et Allis 2001; Strahl et Allis 2000). Plusieurs de ces modifications se produisent sur les queues N-terminales des histones centrales car des modifications dans la charge de ces résidus peuvent influencer le repliement de la chromatine en modifiant l'affinité des histones pour l'ADN ou pour d'autres histones. D'autres modèles ont aussi été proposés dans lesquels les interactions entre les histones centrales sont modifiées par des interactions des résidus d'histone modifiés avec des protéines non-histones, ce qui pourrait influencer la structure de la chromatine (Tse et Hansen 1997).

Plusieurs modifications aux histones ont un impact sur le repliement supérieur de la chromatine. L'acétylation des histones centrales conduit en effet à une structure moins compacte de la chromatine *in vitro* (Wang et al. 2001) alors qu'un traitement de cellules humaines par l'inhibiteur de désacétylases trichostatin-A (TSA) résulte en une décondensation marquée de leur chromatine (Toth et al. 2004). La formation du repliement supérieur de la chromatine *in vitro* est aussi inversement corrélée avec la proportion des sites acétylables d'histones qui sont acétylés (Tse et al. 1998). La phosphorylation des histones H1 est aussi impliquée dans le repliement supérieur de la chromatine car elle conduit au détachement de H1 de la chromatine (Dou et al. 1999). De plus, l'effet de l'acétylation des histones centrales est plus important en absence d'histone H1 (Garcia-Ramirez et al. 1995). La méthylation des histones est elle aussi impliquée dans le repliement supérieur de la chromatine, puisqu'elle permet le recrutement de protéines qui affectent ce repliement (voir plus bas) (Grigoryev 2001).

Le recrutement de facteur de remodelage de la chromatine ATP-dépendant tels INO80, NURF, SWI/SNF ou RSC par des facteurs séquence-spécifiques ou par certaines modifications aux histones peut aussi permettre de rendre la chromatine moins compacte localement (Johnson et al. 2005; Narlikar et al. 2002). Par exemple, des mutations dans les histones centrales résultant en une perte de compaction de la chromatine peuvent compenser la perte d'action de SWI/SNF chez la levure, ce qui suggère qu'un des rôles de ce facteur est de modifier le degré de compaction de chromatine (Horn et al. 2002; Wechsler et al. 1997). L'expression de certains gènes lors de la mitose dépend aussi de SWI/SNF, ce qui suggère aussi que ce facteur est nécessaire pour contrer la condensation de la chromatine en mitose (Krebs et al. 2000). La cellule possède donc plusieurs façons de moduler l'organisation de la chromatine qui dépendent des histones, en particulier son degré de compaction.

#### La compaction d'une fibre de chromatine dépend de son contenu en gène

Fonctionnellement, l'existence de régions génomiques de compaction variable est particulièrement liée à leur contenu en gène. En effet, la distance physique entre des signaux de FISH (*fluorescence in situ hybridization* ou hybridation par fluorescence *in situ*) séparés d'une distance génomique équivalente est plus faible dans les bandes chromosomiques pauvres en gènes que dans les bandes riches en gènes, ce qui suggère que la chromatine des régions pauvres en gène est plus compacte (Yokota et al. 1997). Une étude à grande échelle de la répartition des zones de chromatine ouverte et compacte a d'ailleurs confirmé ces observations en montrant par centrifugation analytique que les zones de chromatine ouverte correspondent généralement aux régions possédant une haute densité de gènes (Gilbert et al. 2004).

Étant donné que l'expression des gènes dépend du contexte cellulaire, le degré de condensation de la chromatine doit être modulé de façon dynamique. Par exemple, le ciblage d'un activateur transcriptionnel VP16 à une région contenant plusieurs sites de liaison en tandem cause une nette décondensation de la chromatine, présumément à cause d'un relâchement de la forme « chromonème » de 100 nm (Tumbar et al. 1999). Dans un contexte plus naturel, l'induction de l'expression des gènes du complexe HoxB murin

conduit elle aussi à une importante décondensation de la chromatine (Chambeyron et Bickmore 2004).

### La localisation des modifications post-traductionnelles aux histones et des facteurs de remodelage lie le contenu en gène et le repliement supérieur de la chromatine

La compaction de la chromatine est liée aux modifications post-traductionnelles des histones et à la présence de gènes. Or, l'abondance et la localisation de modifications post-traductionnelles aux histones sont directement liées à l'activité génique. La liaison de facteurs de transcription aux gènes actifs peut en effet conduire au recrutement d'activités de modification des histones, telles que les histones acetyl-transférases Gcn5 et Esa1 (Robert et al. 2004) ou l'histone méthyltransférase Set1 (Ng et al. 2003). Chez la levure, l'acétylation des queues N-terminales des histones H3 (lysines 9 et 14 en particulier) et H4 (lysines 5, 8, 12, 16) est en effet augmentée dans les régions promotrices et des codons de départ des gènes et est corrélée avec le taux de transcription des gènes (Pokholok et al. 2005).

Les régions pauvres en gènes présentent pour leur part une chromatine plus compacte, l'hétérochromatine (Grigoryev 2001), qui peut être séparée en deux catégories : facultative et constitutive (Craig 2005). L'hétérochromatine facultative est assemblée *de novo* alors que l'hétérochromatine constitutive est permanente. Le chromosome X inactif des mammifères est un exemple d'hétérochromatine facultative alors que la chromatine péri-centromérique et les régions proches des télomères sont des exemples d'hétérochromatine constitutive. L'établissement de l'hétérochromatine implique généralement des HDAC (histones désacétylases) telles que les complexes SIN3 et NuRD qui sont entre autre recrutés à certains gènes régulés différemment selon le cycle cellulaire (Robert et al. 2004). Les désacétylases peuvent être recrutées par des facteurs de répression transcriptionnelle ou interagir avec l'ADN méthylé (fréquent dans les régions génomiques inactives en transcription) (Wade 2001). Cette interaction n'est pas à sens unique, puisque la méthylation répressive des histone peut aussi conduire à la méthylation de l'ADN, de façon à inactiver certains loci de façon plus permanente (Lehnertz et al. 2003).

D'autre part, la désacétylation des histones par les HDAC peut s'accompagner du recrutement d'histone méthyltransférases qui méthylent entre autres l'histone H3 en lysine 9 et 27 et H4 en lysine 20 (Craig 2005; Sims et al. 2003). Ces histones méthylées en lysine peuvent par la suite interagir avec les protéines du groupe polycomb ou HP1 (Lachner et al. 2001). Ces protéines ont la propriété de se multimériser (Orlando 2003) : les loci contenant des complexes HP1-histones méthylées peuvent ainsi recruter d'autres loci semblables pour former des structures compactes (Grigoryev 2001). Ceci explique aussi le fait que les régions d'hétérochromatine puissent s'étendre le long des chromosomes et soient capables de recruter des loci distaux. D'autres protéines ayant des propriétés semblables aux HP1, telles que les protéines SIR2, 3, et 4, peuvent aussi se lier aux histones et interagir entre elles (Hecht et al. 1995). Ces protéines sont par exemple recrutées aux télomères et se lient aux régions N-terminales des histones H3 et H4 (Guarente 1999). Ces interactions entre régions de chromatine permettent donc à divers degrés le repliement supérieur de la chromatine.

Quel est l'effet des modifications à la compaction de la chromatine sur l'accessibilité au génome? Une chromatine moins compacte est plus accessible aux facteurs protéiques : par exemple, la sensibilité à la DNase d'un fragment de chromatine est modulée par son état d'acétylation (Hebbes et al. 1994). L'acétylation des histones H3 et H4 facilite aussi l'accès de facteurs de transcription à leur site de reconnaissance (Lee et al. 1993; Vettese-Dadey et al. 1996) et conduit à une augmentation de la dispersion de molécules de dextran de grande taille dans le noyau (Gorisch et al. 2005). Ces observations suggèrent donc que l'accessibilité des protéines au matériel génétique est modulée entre autre par le degré d'acétylation et par le degré de compaction et de condensation de la chromatine.

### **Organisation supérieure de la chromatine et compartimentalisation fonctionnelle du noyau**

Une fibre de chromatine possède donc diverses régions de compaction liées aux fonctions nucléaires. Ces niveaux de compaction dépendent d'interactions entre nucléosomes et entre régions de fibre de chromatine. Au-delà de ces niveaux de compaction, le repliement supérieur de la chromatine forme des domaines chromosomiques fonctionnellement définis qui regroupent des régions génomiques à certains endroits du noyau à des moments définis

de la vie cellulaire. L'organisation fonctionnelle de ces domaines dans l'espace nucléaire est le sujet des prochaines sections d'introduction.

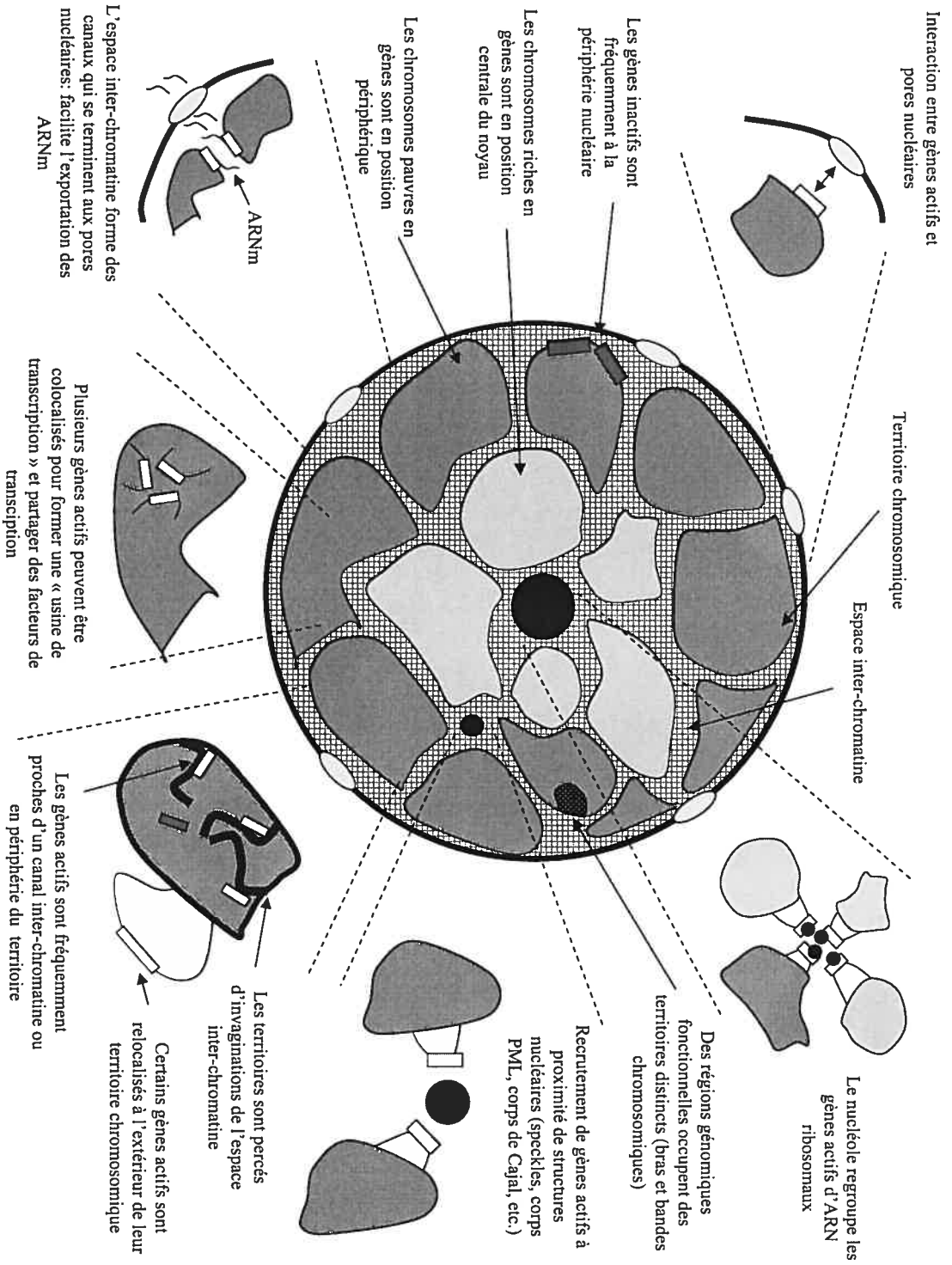
### Les chromosomes occupent des territoires distincts dans l'espace nucléaire

Les chromosomes sont le premier niveau fonctionnel d'organisation spatiale du génome dans le noyau. Les chromosomes occupent des territoires distincts qui se mêlent peu les uns aux autres dans l'espace nucléaire (Cremer et Cremer 2001; Visser et al. 2000; Zink et al. 1998). Cette observation a mené au modèle du « *chromosome territory-interchromatin compartment* (CT-IC) » voulant que des masses chromosomiques relativement compactes définissent un espace inter-chromosomique servant de canaux d'accès empruntés par divers facteurs protéiques. Ces canaux seraient relativement libres d'encombrement dû à l'ADN et relieraient les pores nucléaires à l'espace inter-chromatine. Selon ce modèle, les gènes actifs devraient avoir un accès privilégié à l'espace inter-chromatine, de façon à pouvoir recruter les facteurs de transcription et à faciliter l'exportation des ARNm produits. Cet espace contiendrait en outre d'autres structures nucléaires telles que les « *speckles* » SC35 et les corps PML (Figure 9).

Ce modèle présente l'avantage de définir une solution aux questions soulevées par le fait que malgré une concentration en ADN et en facteurs protéique élevée, des molécules de taille relativement grande et des ARN messagers peuvent se déplacer rapidement d'un bout à l'autre du noyau (Phair et Misteli 2000; Politz et al. 1999). De plus, il fut démontré que les ARNm se répandent dans le noyau en suivant des trajectoires lobulaires qui évitent les zones marquées intensément au Hoescht, un agent qui colore l'ADN (Zachar et al. 1993). Des observations similaires ont aussi été réalisées dans d'autres systèmes : par exemple, des molécules de vimentines peuvent se polymériser en formant un réseau réticulé qui contourne les territoires chromosomiques humains (Bridger et al. 1998).

Certaines observations réalisées en microscopie électronique suggèrent par contre que les zones inter-territoriales sont plus ou moins remplies de chromatine (Dehghani et al. 2005) et que les canaux inter-territoriaux représenteraient des zones de densité moindre de chromatine organisée en lattice ou « mailles de filet ». De façon compatible avec cette idée, les territoires chromosomiques ne sont pas totalement imperméables aux facteurs

protéiques, mais sont probablement percés d'invaginations ou de canaux provenant du domaine inter-chromosomique qui les rendent accessibles aux divers facteurs protéiques (Visser et al. 2000). En effet, les sites de transcription ne sont pas restreints aux canaux inter-chromatine mais sont aussi retrouvés à l'intérieur des territoires chromosomiques, ce qui suggère que la frontière entre territoire chromosomique et espace inter-chromatine n'est pas strictement définie et que des facteurs protéiques peuvent pénétrer dans les territoires (Mahy et al. 2002a; Mahy et al. 2002b; Sadoni et Zink 2004; Verschure et al. 1999). Cependant, comme nous le verrons plus loin, certains loci sont effectivement localisés préférentiellement en périphérie des territoires chromosomiques de façon dépendante de la transcription. Ceci suggère donc que bien que les territoires chromosomiques ne soient pas imperméables aux divers facteurs protéiques, la proximité entre un locus et un canal intra ou inter-territorial pourrait faciliter l'accessibilité des facteurs à leurs sites d'activité.



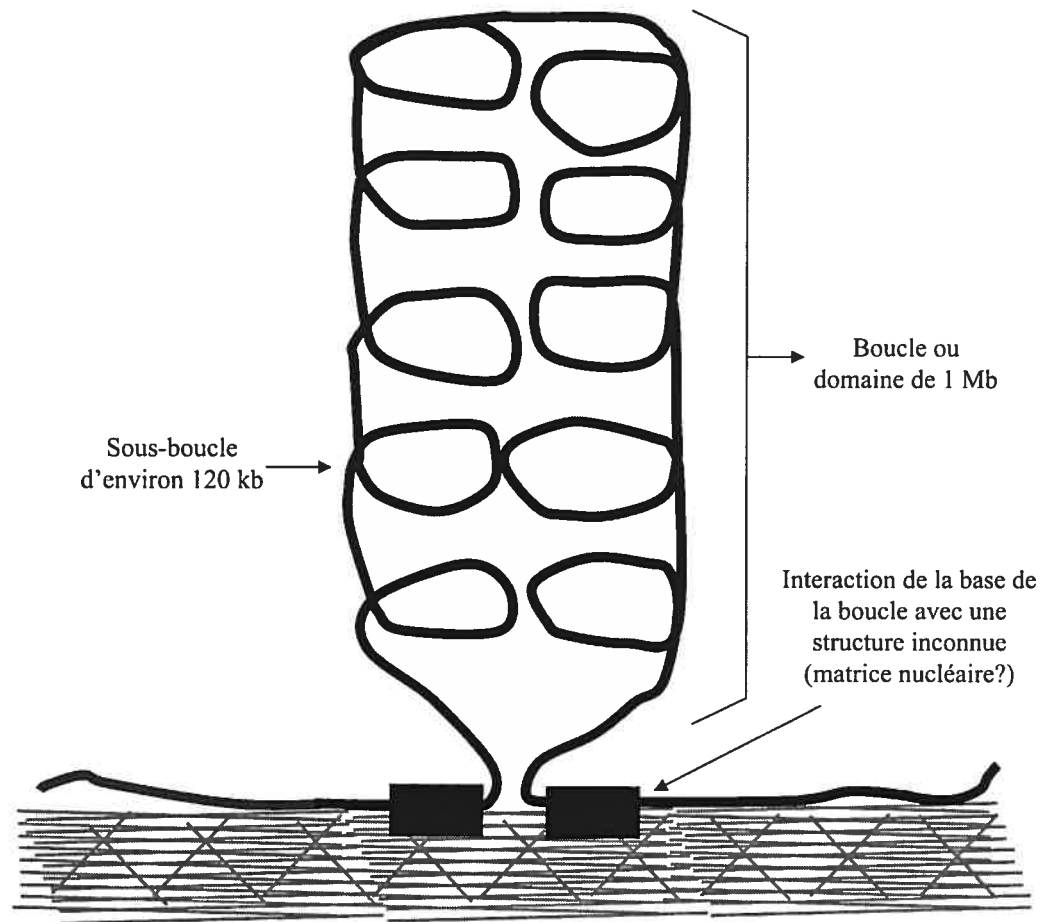


**Figure 9 : Organisation du noyau et modèle des territoires chromosomiques**

Les territoires chromosomiques gris pâle représentent des chromosomes riches en gènes alors que les gris foncé sont pauvres en gène. Les rectangles blancs représentent des gènes actifs. Les rectangles gris foncé représentent des gènes inactifs.

Les territoires chromosomiques peuvent être séparés en plusieurs sous-domaines dans l'espace nucléaire

À l'instar des chromosomes entiers, plusieurs sous-unités fonctionnelles des chromosomes occupent aussi des territoires restreints dans le noyau (Figure 9). Des études de FISH ont montré que les bras chromosomiques et que certaines bandes des chromosomes humains occupent des territoires distincts et n'entrent en contact entre eux que sur une surface limitée (Dietzel et al. 1998), ce qui suggère que des domaines chromosomiques définis entre autre par la compaction de leur chromatine peuvent être compartimentalisés dans le noyau. À plus petite échelle, la comparaison entre la distance génomique séparant des sondes de FISH et la distance physique entre les signaux résultant a mené à la proposition d'un modèle dans lequel la chromatine serait organisée sous forme de boucles de chromatine d'une taille de l'ordre de 1 Mb possédant une structure interne aléatoire et qui seraient disposées de façon plus ou moins aléatoire le long d'une structure pour le moment mal définie (Sachs et al. 1995; Yokota et al. 1995). Il existe en effet une relation biphasique marquée entre la distance génomique et la distance physique entre les sondes de FISH, qui est caractérisée par deux phases linéaires possédant un point de cassure à environ 1 Mb. À cette distance génomique, le taux d'augmentation des distances entre les signaux de FISH diminue brusquement. Ceci suggère qu'après ce seuil la séparation génomique n'est pas le seul facteur entrant en jeu dans la détermination de la distance spatiale entre deux loci. De plus, pour certaines régions génomiques, la distance physique entre les signaux de FISH entre dans une phase de diminution après une phase initiale d'augmentation selon la séparation génomique, ce qui pourrait s'expliquer par la formation de boucles de chromatine. Des modélisations ont par la suite suggéré que l'intérieur des domaines de 1 Mb serait constitué de sous-boucles d'une taille de l'ordre de 100 kb (Munkel et al. 1999) (Figure 10).



**Figure 10 : Le modèle d'organisation de la chromatine en boucles de 1 Mb**

Les limites à la diffusion de la chromatine dans l'espace nucléaire permettent une certaine stabilité dans l'organisation nucléaire

Le fait que des territoires et domaines chromosomiques puissent être observés par FISH suggère que la diffusion de la chromatine est limitée dans l'espace nucléaire (Gasser 2002). En effet, bien que d'importants mouvements de chromatine furent observés dans certaines situations chez la drosophile (Csink et Henikoff 1998), la chromatine est généralement peu mobile dans le noyau. Chez la levure et la drosophile, les mouvements de chromatine sont généralement confinés dans une région restreinte du noyau (Marshall et al. 1997). Chez les mammifères, des expériences de FRAP (*Fluorescence Recovery After Photobleaching*) réalisées sur les cellules HeLa ou Swiss 3T3 ont montré qu'une zone de chromatine « blanchie » peut persister pendant plus d'une heure, suggérant que cette zone et celles qui l'entourent sont plus ou moins immobiles (Abney et al. 1997). D'autre part, un marquage

de plusieurs centromères de cellules HeLa a démontré qu'ils étaient pratiquement immobiles sur des périodes de plus de 2 heures (Shelby et al. 1996).

Les limites à la diffusion de l'ADN dans le noyau ne sont pas simplement dues à l'encombrement stérique entre molécules, puisque plusieurs molécules telles que les ARNm diffusent de façon relativement libre dans l'espace nucléaire. Ainsi, l'association de la chromatine avec certains compartiments ou structures nucléaires peut en limiter les mouvements. Les loci associés à certaines structures nucléaires comme le nucléole ou la périphérie sont en effet nettement moins mobiles que d'autres loci (Chubb et al. 2002). Les pores nucléaires et d'autres facteurs de la membrane nucléaire interagissent aussi directement avec certains segments de chromatine et pourraient potentiellement en limiter les mouvements (Brown et al. 1997; Casolari et al. 2004).

Plusieurs études suggèrent d'autre part l'existence d'une « matrice nucléaire » semblable au cytosquelette et qui aurait pour fonction de maintenir l'organisation nucléaire (Berezney et Coffey 1974). Cette matrice serait formée de protéines telles que la topoisomérase II (Berrios et al. 1985), et servirait de point d'ancrage à des boucles de chromatine (Earnshaw et Heck 1985). Ces points d'ancrage sont appelés MAR (*matrix attachment region*), et sont des séquences d'environ 300 pb peu conservées et riches en A/T (Boulikas 1995). L'existence même de cette matrice est un sujet de controverse (Pederson 1998). En effet, les procédures utilisées pour son isolation pourraient causer une précipitation des protéines nucléaires, résultant en une apparence filamenteuse artéfactuelle. De plus, l'existence de la matrice n'a pu être confirmée *in vivo* par microscopie. Enfin, il fut montré par FRAP que la topoisomérase II, une des prétendue composante majeure de la matrice nucléaire, était très mobile dans le noyau (Christensen et al. 2002), ce qui met en doute la possibilité qu'elle fasse partie d'un squelette nucléaire stable. L'influence de la matrice nucléaire sur l'organisation du noyau n'est donc pas claire pour le moment.

#### Positionnement des domaines chromosomiques entre eux et par rapport aux structures nucléaires

La diffusion limitée de la chromatine dans le noyau implique une certaine stabilité dans l'arrangement des domaines chromosomiques. D'ailleurs, plusieurs études ont observé que

l'arrangement des territoires chromosomiques entre eux est effectivement maintenu de façon stable durant l'interphase (Gerlich et al. 2003; Walter et al. 2003), ce qui permet des interactions préférentielles entre territoires et domaines chromosomiques particuliers. Par contre, ces associations ne semblent pas strictement conservées dans les cellules individuelles d'une population (Cornforth et al. 2002; Lesko et al. 1995). À cause de ceci, les modèles d'organisation du génome dans le noyau sont généralement probabilistes : les loci génomiques ne sont pas aléatoirement distribués dans l'espace nucléaire, mais il existe un certain degré d'incertitude quant à leur position exacte par rapport à l'espace nucléaire et au reste du génome (Parada et al. 2003). Un grand nombre d'études concernant le positionnement radial (c'est-à-dire en relation avec le centre ou la périphérie) des chromosomes dans l'espace nucléaire ont ainsi observé des divergences significatives par rapport à un ordre aléatoire (Cremer et Cremer 2001).

Le contenu en gène est impliqué dans la détermination du positionnement radial des chromosomes dans l'espace nucléaire par des mécanismes mal connus (Figure 9). Par exemple, les chromosomes 18 et 19 ont une taille similaire mais un contenu en gène différent : le 18 est plus pauvre en gène que le 19, et ce dernier est situé en position plus centrale dans le noyau (Croft et al. 1999). Ce type d'arrangement a d'ailleurs été conservé par l'évolution puisque le positionnement radial des chromosomes 18 et 19 de primates est similaire à celui des chromosomes humains (Tanabe et al. 2002a; Tanabe et al. 2002b). Finalement, une étude de tous les chromosomes humains a démontré que leur contenu moyen en gène était corrélé avec leur positionnement radial dans le noyau, les chromosomes riches en gènes étant en position plus centrale que les chromosomes pauvres en gènes (Boyle et al. 2001).

Le contenu en gène peut donc potentiellement expliquer en partie le fait que certains arrangements spatiaux de chromosomes soient statistiquement plus fréquents que d'autres. Les chromosomes 1, 16, 17, 19 et 22 humains sont souvent dans le centre du noyau, ce qui résulte en leur fréquente proximité (Cornforth et al. 2002). Des arrangements spécifiques des chromosomes X, Y et 8, 11 furent aussi observés dans les cellules quiescentes humaines (Nagele et al. 1999). D'autre part, les chromosomes 9 et 22 sont aussi associés dans le noyau et sont impliqués dans des translocations récurrentes (BCR-ABL) (Kozubek et al. 1999). Ce lien entre translocations et positionnement des chromosomes est un thème

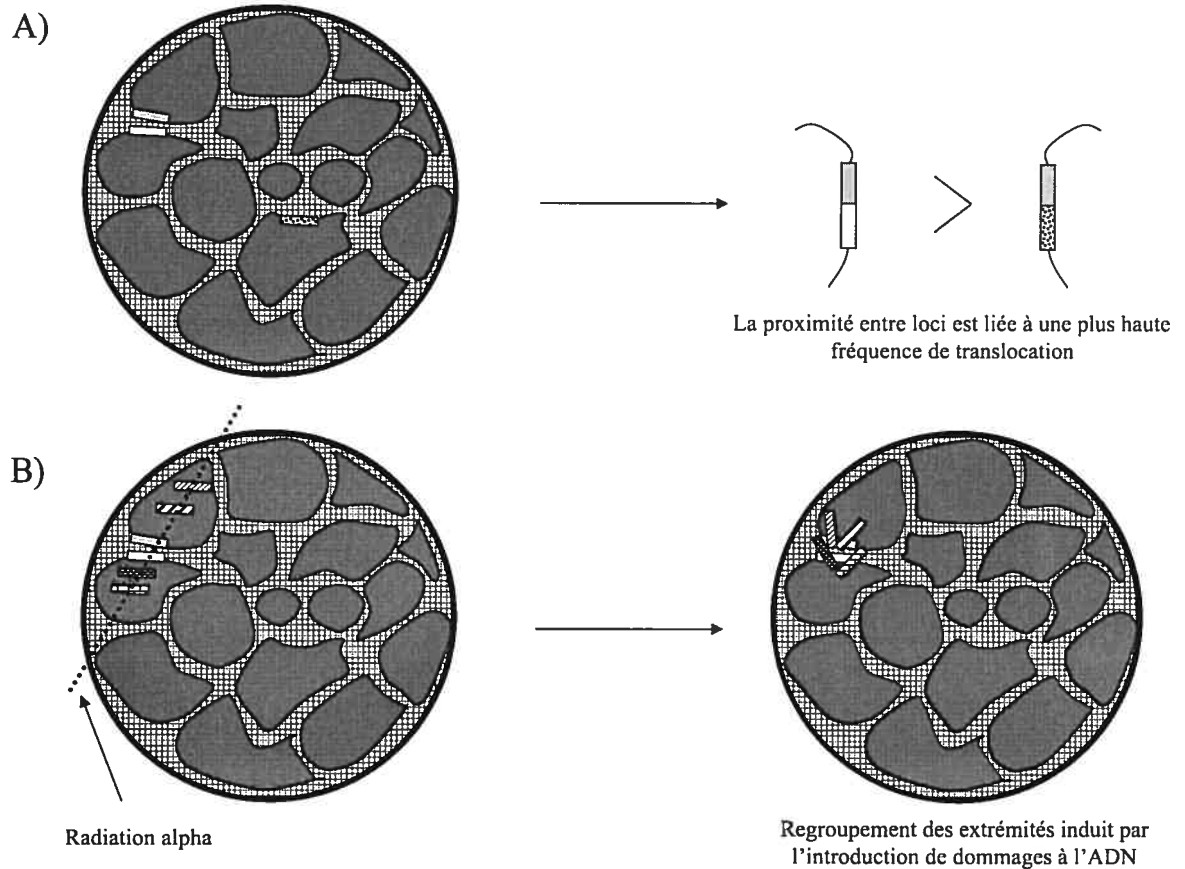
récurrent : par exemple, plusieurs associations spécifiques entre territoires chromosomiques existent chez la souris, sont tissus-spécifiques et sont liées à des translocations fréquentes entre les chromosomes en question (Parada et al. 2004).

Si de telles interactions préférentielles entre territoires chromosomiques sont conservées dans une population cellulaire et sont fonctionnellement importantes, il serait logique qu'elles soient transmises aux cellules filles lors de la division cellulaire. Or, la stabilité de ces arrangements dépend de la stringence avec laquelle ils sont évalués. En effet, certaines études ont montré que l'organisation relative des chromosomes était grossièrement transmise aux cellules filles après la mitose (Essers et al. 2005; Gerlich et al. 2003; Nagele et al. 1999; Sun et Yokota 1999). À l'inverse, d'autres études ont conclu que l'arrangement des chromosomes n'était pas strictement conservé après une division (Thomson et al. 2004; Walter et al. 2003). Ces différences sont surtout liées à la définition « conservatrice » de la transmission des interactions entre chromosomes utilisée par certains auteurs. En effet, même dans les études rapportant l'absence d'une stricte conservation dans l'organisation des chromosomes après une mitose, l'organisation globale des chromosomes apparaissait tout de même conservée jusqu'à un certain point dans les cellules-filles. Ces interprétations sont compatibles avec une vision probabiliste de l'organisation nucléaire du génome et suggèrent que certaines informations positionnelles sont transmises lors de la division cellulaire.

#### Les interactions spécifiques entre loci existent et facilitent leur recombinaison

L'existence des territoires chromosomiques et sous-chromosomiques et le positionnement non-aléatoire des chromosomes dans l'espace nucléaire impliquent que certains loci génomiques sont préférentiellement à proximité les uns des autres dans le noyau. La recombinaison génétique semble être particulièrement influencée par la proximité spatiale entre loci (Figure 11A). Ainsi, les gènes BCR et ABL situés sur les chromosomes 9 et 22 participent à des translocations dans la leucémie myéloïde chronique et sont fréquemment à proximité l'un de l'autre dans l'espace nucléaire (Kozubek et al. 1997; Lukasova et al. 1997). Les gènes RET et H4 sont aussi impliqués dans des translocations induites par les radiations dans les cancers de la thyroïde. Bien que ces gènes soient séparés de 30 Mb, ils sont fréquemment juxtaposés dans les cellules de la thyroïde (Nikiforova et al. 2000).

Finalement, les gènes MYC, BCL et des immunoglobulines sont colocalisés dans l'espace nucléaire des cellules B et sont impliqués dans des translocations dans les lymphomes des cellules B (Roix et al. 2003).



**Figure 11 : Influence de l'organisation spatiale du génome sur les événements de recombinaison**

A) Les loci à proximité les uns des autres sont plus fréquemment impliqués dans des événements de recombinaison causant des translocations. B) Dans certains cas, l'induction de dommages à l'ADN qui est à la source des événements de recombinaison causant les translocations peut modifier l'organisation et l'environnement spatial des loci impliqués.

Par contre, d'autres mécanismes peuvent entrer en jeu dans le choix d'un partenaire de recombinaison. Par exemple, si plusieurs brisures se produisent en suivant une ligne dans le noyau (en irradiant des cellules par des radiations alpha, par exemple), les extrémités libres d'ADN générées peuvent se regrouper dans l'espace nucléaire pour être réparées, ce qui

peut conduire à des appariements illégitimes (Aten et al. 2004). De plus, chez la levure, plusieurs extrémités d'ADN peuvent être recrutées vers un regroupement de facteurs de réparation (Lisby et al. 2003). Dans ces cas, la proximité entre loci pourrait être induite par le dommage à l'ADN (Figure 11B). Par contre, le recrutement de loci vers un regroupement de facteurs donné pourrait aussi être influencé par une proximité spatiale préalable entre ces loci.

#### Système d'étude : lien entre l'environnement spatial d'un locus et le choix d'un partenaire de recombinaison

Une deuxième partie expérimentale de cette thèse concerne l'étude de l'environnement spatial d'un locus humain et de son influence sur le choix des partenaires de recombinaison de ce locus. Une étude réalisée dans notre laboratoire a récemment démontré que l'accessibilité d'un locus à d'autres loci génomiques pour participer à des événements de recombinaison dans des cellules humaines était restreinte et spécifique (D'Anjou et al. 2004). Ainsi, une brisure double-brin ciblée dans une séquence répétitive L1 est réparée par recombinaison en utilisant d'autres L1 du génome comme séquences donneuses, mais ceux-ci ne sont pas distribués aléatoirement dans le génome. La plupart des partenaires de recombinaison préférentiels du locus d'intérêt sont ainsi situés à l'intérieur d'un Mb du site de brisure ou dans le même bras chromosomique. Nous avons donc caractérisé l'environnement spatial de ce locus en définissant quels sont les loci génomiques qui sont fréquemment à proximité spatiale de celui-ci, de façon à vérifier si cet environnement spatial pouvait avoir un impact sur le choix de ses partenaires de recombinaison.

#### Positionnement spécifique de régions génomiques à la périphérie du noyau

Certains loci sont donc positionnés les uns par rapport aux autres de façon spécifique dans l'espace nucléaire et l'expression génique semble être une fois de plus un déterminant important de ce positionnement. En effet, en plus d'être lié au positionnement de chromosomes entiers, le contenu en gène est impliqué dans le positionnement radial de régions de chromatine plus courtes (Figure 9). Les bandes de chromatine riches en gènes sont ainsi généralement en position centrale du noyau, alors que les bandes pauvres en gènes sont souvent en périphérie (Ferreira et al. 1997; Sadoni et al. 1999). Plusieurs régions

d'ADN peu actives transcriptionnellement, telles les régions sub-téломériques et péricentromériques, ont donc tendance à se retrouver à la périphérie chez plusieurs organismes comme le blé et la drosophile, ce qui donne naissance à la « conformation de Rabl », dans laquelle le centromère et les télomères d'un chromosome donné sont situés à des pôles opposés du noyau. Cette conformation n'est par contre pas strictement observée chez les mammifère : bien qu'il existe une tendance vers une localisation périphérique des centromères chez l'humain (Weierich et al. 2003), certains d'entre eux sont retrouvés autour du nucléole (Carvalho et al. 2001) ou à des positions internes du noyau (Weimer et al. 1992). Pour leur part, les télomères sont souvent en périphérie du noyau chez la souris mais au centre chez l'humain (Weierich et al. 2003).

Ces observations amènent à se poser la question suivante : le positionnement radial des régions de chromatine dans le noyau est-il une cause ou un effet de l'activité génique? Une réponse partielle à cette question est apportée par le fait que l'activité transcriptionnelle de certains loci peut être régulée par leur recrutement vers la périphérie, et plus généralement, vers des régions nucléaires riches en hétérochromatine (Figure 9). En effet, le recrutement artificiel de loci vers la périphérie, en créant une protéine de fusion possédant un site de liaison spécifique à l'ADN couplé à une protéine de la membrane nucléaire, résulte en leur inactivation (Andrulis et al. 1998). Plusieurs gènes régulés dans les cellules lymphoïdes sont aussi recrutés à l'hétérochromatine péricentrique avec une fréquence inversement proportionnelle à leur activité transcriptionnelle (Brown et al. 1997). D'autre part, l'insertion d'un fragment d'hétérochromatine dans un gène actif conduit à son inactivation par recrutement du locus actif vers une région d'hétérochromatine chez la drosophile (Csink et Henikoff 1996; Dernburg et al. 1996). Finalement, le gène de la collagénase humaine est relâché de la périphérie lors de son activation dans les cellules en voie de sénescence (Imai et al. 1997) alors que le locus inactif des immunoglobulines est à la périphérie dans les cellules pro-T, mais est relocalisé en position plus centrale lorsqu'activé dans les cellules pro-B (Kosak et al. 2002). Ces quelques exemples suggèrent que le recrutement actif de loci vers des zones d'hétérochromatine telle que la périphérie nucléaire ou les régions périnucléolaires inhibe la transcription alors que l'activité génique pourrait conduire à un relâchement des loci de ces régions nucléaires et à un retour à un positionnement central « par défaut ». Selon ce modèle, le positionnement central de



régions de chromatine serait ainsi une conséquence indirecte de leur non-recrutement vers des zones d'hétérochromatine.

Ces observations pointent vers un compartiment périnucléaire réfractaire à la transcription. Or, dans certains cas un positionnement à la membrane nucléaire semble faciliter la transcription (Figure 9). En effet, des régions hypersensibles à la DNase se retrouvent parfois à la périphérie dans certaines lignées tumorales (Park et De Boni 1998) alors que l'expression de certaines séquences est corrélée avec leur migration vers la périphérie (Park et De Boni 1996). Finalement, chez la levure certains loci sont associés aux pores nucléaires de façon dépendante de leur expression, ce qui pourrait faciliter l'exportation de leurs ARNm (Casolari et al. 2004). Les règles régissant le positionnement radial de la chromatine ne sont donc pas absolues et peuvent varier en fonction du locus étudié.

#### Positionnement de loci en lien avec les structures nucléaires

Plusieurs exemples montrent que d'autres compartiments nucléaires sont spécifiquement associés à certaines régions de chromatine (Figure 9). Le nucléole est une structure assemblée autour des gènes d'ARN ribosomiaux (Scheer et Hock 1999) qui sont présents sur plusieurs chromosomes (13, 14, 15, 21 et 22 chez l'humain). Le nucléole regroupe les gènes d'ARN ribosomiaux en un seul endroit qui permet la transcription par la polymérase I et la maturation des ARN ribosomiaux. Certains centromères sont aussi recrutés au nucléole (Carvalho et al. 2001) alors que d'autres loci situés près du nucléole ont une mobilité réduite, ce qui suggère que cette structure peut leur servir de point d'ancrage à certaines régions de chromatine (Chubb et al. 2002).

Les « speckles » contiennent des ribonucléoprotéines nucléaires (snRNP) impliquées dans l'épissage et la maturation des ARNm et sont présents à environ 20 à 40 copies par noyau. Ils sont probablement des réservoirs de facteurs d'épissage et de maturation des ARNm (Misteli et al. 1997). Ces structures recrutent des gènes impliqués dans la myogénèse et les gènes du collagène type 1 alpha 1 et 2 lors de leur activation (Moen et al. 2004; Shopland et al. 2003), ce qui pourrait faciliter la maturation de leurs ARNm (Smith et al. 1999). Les bandes riches en gènes sont de plus fréquemment à proximité de ces structures, ce qui suggère qu'elles peuvent faciliter l'expression génique (Shopland et al. 2003).

Les corps de Cajal sont des structures mobiles présentes à environ 1 à 5 copies par noyaux qui sont souvent associées avec le nucléole ou avec des régions de chromatine. Ils contiennent des ARN polymérases, des facteurs de transcription et des protéines impliquées dans l'épissage et la maturation des ARNm (Gall 2000) et sont potentiellement des sites de pré-assemblage de complexes de transcription (Gall 2001). Ces corps sont associés à la chromatine de façon ATP et transcription-dépendante, en particulier avec les gènes d'histones et de certains snRNA (Jacobs et al. 1999; Platani et al. 2002; Shopland et al. 2001).

Les corps PML (promyelocytic leukemia) sont au nombre d'environ 10 par noyaux (Dellaire et Bazett-Jones 2004). En plus de la protéine PML, environ 40 autres protéines s'y retrouvent à un moment ou un autre. Ceci suggère qu'ils sont impliqués dans de nombreuses fonctions nucléaires, allant de la réparation de l'ADN à la régulation transcriptionnelle, en passant par l'apoptose. Plusieurs régions génomiques transcriptionnellement actives sont fréquemment localisées à proximité de ces structures, incluant les gènes d'histones (Wang et al. 2004), du CMH (Shiels et al. 2001) et de la protéine p53 (Sun et al. 2003).

Notons en terminant qu'en plus d'être à proximité de certains facteurs spécifiques à une structure, le fait de repositionner certains loci dans l'espace inter-chromatine (puisque ces structures sont toutes dans ce compartiment) pourrait dans certains cas avoir un effet sur leurs capacités transcriptionnelles en facilitant l'accès à d'autres facteurs qui ne sont pas forcément liés aux structures nucléaires en question.

#### Positionnement de loci spécifiques par rapport à leur territoire chromosomique et lien avec la transcription

Les premiers modèles des territoires chromosomiques suggéraient que les loci actifs en transcription seraient localisés à la périphérie des territoires, en contact avec l'espace inter-chromatine, ce qui pourrait faciliter l'exportation des ARNm et les contacts de ces loci avec les facteurs de transcription (Cremer et Cremer 2001; Zirbel et al. 1993). Cependant, la transcription peut se produire à l'intérieur des territoires, ce qui suggère que ces derniers ont probablement la consistance d'une éponge traversée par des canaux d'espace inter-

chromatine (Verschure et al. 1999). Ces observations pourraient suggérer que le positionnement d'un locus par rapport à son territoire chromosomique est sans importance. Or, bien que certains gènes actifs puissent être distribués à l'intérieur de leur territoires chromosomiques, les régions riches en gènes ont une nette tendance à être en périphérie ou même à l'extérieur de leurs territoires (Mahy et al. 2002a) et ce positionnement est souvent conservé entre la souris et l'humain (Mahy et al. 2002b) (Figure 9).

Ces tendances ont été confirmées par des observations suggérant que l'activation de la transcription est parfois liée à un repositionnement des loci hors de leur territoire chromosomique (Figure 9). Par exemple, des régions de chromatine du locus du complexe majeur d'histocompatibilité sont repositionnées hors de leur territoire chromosomique lors de leur induction par l'interféron gamma (Volpi et al. 2000). Le « *epidermal differentiation complex* » est lui aussi repositionné à des positions périphériques ou externes par rapport à son territoire chromosomique dans les kératinocytes, en corrélation avec son activité transcriptionnelle (Williams et al. 2002). Le locus de la beta-globine de souris et de l'humain est lui aussi relocalisé hors de son territoire avant son activation (Ragoczy et al. 2003). Finalement, le locus du gène HoxB1 est extrudé de son territoire chromosomique en réponse à l'induction de son expression lors de la différenciation de cellules souches embryonnaires (ES) de souris induite par l'acide rétinoïque (Chambeyron et Bickmore 2004).

Les mécanismes menant à ce type de décondensation ne sont pas totalement compris à ce jour. Certaines expériences réalisées avec les domaines d'activation des protéines p53 et pRb ont montré que ces événements de décondensation massive pouvaient dépendre d'une augmentation de l'acétylation des histones. Par contre, le recrutement du récepteur aux estrogènes ou de la protéine BRCA1 à des sites de liaison répétés en tandem n'augmente pas l'acétylation des histones de la région malgré une importante décondensation (Carpenter et al. 2004; Carpenter et Belmont 2004; Nye et al. 2002; Ye et al. 2001). De plus, un traitement au TSA de cellules ES ne récapitule pas la décondensation observée pour le gène HoxB1 en réponse au traitement à l'acide rétinoïque, bien qu'il induise une hyperacétylation massive à ce locus (Chambeyron et Bickmore 2004). Certains facteurs pourraient donc avoir une activité particulière leur permettant de décondenser la chromatine à grande échelle. À cet effet, une protéine appelée COBRA1 peut interagir avec BRCA1 et

est nécessaire pour induire la décondensation associée au recrutement de ce facteur (Ye et al. 2001).

D'autres événements cellulaires peuvent aussi affecter le positionnement de loci par rapport à leur territoire. Par exemple, la différenciation cellulaire peut rendre les territoires chromosomiques diffus, de telle sorte que plusieurs loci sont relocalisés en périphérie de leurs territoires sans égard à leur statut transcriptionnel (Stadler et al. 2004). Il convient donc de s'interroger quant à la relation causale existant entre expression génique et repositionnement dynamique hors du territoire chromosomique par décondensation des fibres de chromatine. D'une part, l'extrusion du locus de la beta-globine précède l'activation du gène (Ragoczy et al. 2003) et fut proposée comme faisant partie d'un processus de préparation à la transcription. Par contre, le remplacement du LCR du complexe de la beta-globine par un élément répressif permet le recrutement de ce locus vers l'hétérochromatine par un mécanisme dépendant de l'extrusion du locus hors de son territoire (Ragoczy et al. 2003). Les mécanismes de repositionnement à grande échelle de la chromatine peuvent donc avoir des effets positifs et négatifs sur la transcription.

### Usines de transcription

Le fait que certains loci soient relocalisés en lien avec leur expression amène l'observateur à se poser la question suivante : où vont-ils? D'une part, une relocalisation peut mener les loci vers l'espace inter-chromatique ou en proximité de structures nucléaires. Ils peuvent aussi être recrutés vers des régions d'hétérochromatine pour être réprimés. Par contre, d'autres modèles proposent que les loci soient relocalisés de façon à partager des machineries transcriptionnelles. En effet, la complexité des machineries transcriptionnelles suggère qu'elles sont probablement difficiles à assembler et à maintenir. Il pourrait donc être plus simple dans certains cas de mobiliser des régions de chromatine vers des machineries fixes plutôt que l'inverse (Cook 1999). Le modèle des « usines de transcription » postule ainsi que plusieurs régions codantes soient regroupées autour d'une même machinerie de transcription (Figure 9).

Le système le plus souvent cité en support d'un tel modèle est le nucléole. En effet, plusieurs polymérases sont regroupées et s'affairent à la transcription d'un grand nombre de

gènes ribosomiaux situés sur plusieurs chromosomes. La transcription par les polymérase II et III semble elle aussi être compatible avec le modèle des usines de transcription. En effet, la transcription est organisée sous forme de foci contenant plusieurs polymérase actives dans les cellules humaines. Chez les cellules HeLa, on dénombre environ 2000 de ces foci contenant chacun environ 20 polymérase (Iborra et al. 1996). Le modèle des usines de transcription suggère donc qu'il y aurait environ 20 à 30 unités de transcription et leurs polymérase associées par usine (Jackson et al. 1998).

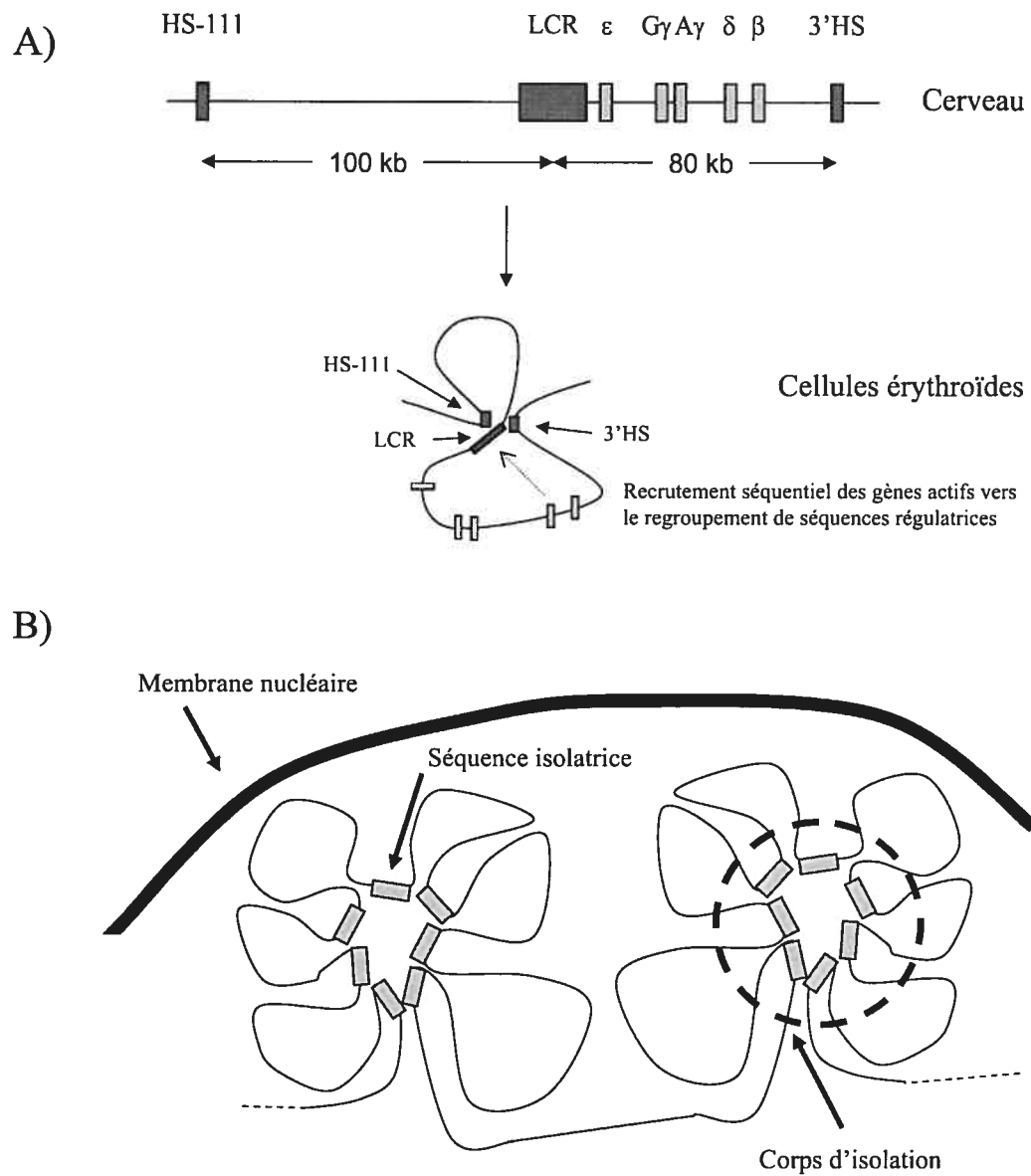
De récentes observations ont plus directement mesuré le recrutement de gènes spécifiques vers des usines de transcription. Les gènes du complexe de l'alpha-globine de souris sont recrutés vers un regroupement de gènes ubiquitaires ou « *housekeeping genes* » situé à proximité lorsqu'ils sont activés (Zhou et al. 2006). D'autre part, certains gènes tels Hbb-1 et Eraf, qui sont séparés par de grandes distances génomiques, sont recrutés à la même machinerie de transcription de façon dépendante de la transcription (Osborne et al. 2004). Cette dernière étude a de plus démontré que les gènes étaient recrutés vers des « usines » déjà assemblées, ce qui suggère que les gènes ne sont pas actifs avant le recrutement, et que ce recrutement est à la base de leur expression.

#### Contacts entre séquences régulatrices et gènes

La régulation des gènes peut aussi avoir une influence sur l'environnement spatial d'un locus. Les gènes peuvent posséder des séquences régulatrices proximales et distales. Pour expliquer l'action des séquences régulatrices distales, il fut suggéré que celles-ci pourraient dans certains cas être amenées à proximité du gène à réguler par un repliement de la chromatine. Des études menées sur le complexe de la  $\beta$ -globine ont permis de confirmer l'existence de ces mécanismes. Dans les cellules érythroïdes, des régions régulatrices faisant partie du « *locus control region* » du complexe de la  $\beta$ -globine de souris se regroupent dans l'espace nucléaire, alors que les séquences intervenantes sont extrudées pour former des boucles. Ces séquences régulatrices sont distantes de plusieurs kb entre elles et sont séparées des gènes actifs du complexe de la  $\beta$ -globine d'environ 40 à 60 kb. Lors de leur expression, les gènes sont tour à tour recrutés vers ces séquences régulatrices (de Laat et Grosveld 2003; Drissen et al. 2004; Palstra et al. 2003; Patrinos et al. 2004; Tolhuis et al. 2002). D'autre part, la configuration de ces loci est linéaire dans le cerveau

(qui n'exprime pas les gènes de globine), ce qui suggère que l'état en forme de boucles de chromatine maintient cette région sous forme « active ». L'environnement spatial et la proximité entre loci spécifiques ont donc des conséquences fonctionnelles dans ce cas (Figure 12).

D'autres contacts entre loci peuvent avoir des conséquences différentes sur la transcription et l'organisation de la chromatine. Les séquences isolatrices possèdent à cet effet plusieurs propriétés intéressantes (Capelson et Corces 2004; Labrador et Corces 2002). Ces séquences bloquent la progression de l'hétérochromatine ou contrent l'effet d'un « *enhancer* » sur un gène : elles définissent donc des domaines de chromatine indépendants. Par exemple, les bandes chromosomiques sont flanquées par des séquences isolatrices chez la drosophile. Au nombre de 500 environ, elle sont regroupées à 25 sites nucléaires appelés « corps d'isolation » qui sont à la périphérie du noyau (Figure 12) (Gerasimova et al. 2000). Ces regroupements d'éléments isolateurs donnent lieu à des boucles de chromatine, et il fut démontré que des changements transcriptionnels peuvent conduire à des réorganisations de ces boucles.



**Figure 12 : Impact des contacts entre séquences spécifiques sur l'organisation des chromosomes et la régulation de l'activité transcriptionnelle.**

A) Repliement dynamique du locus de la globine de souris. LCR : *Locus Control Region*. HS : Site hypersensible à la DNase. Adapté du modèle de Patrinos et collaborateurs (Patrinos et al. 2004). B) Organisation spatiale des séquences isolatrices et formation de « corps d'isolation » à la périphérie nucléaire. Des boucles de chromatine sont créées par les interactions entre séquences isolatrices. Adapté du modèle de Labrador et Corces (Labrador et Corces 2002).

## Système d'étude : Modifications de l'environnement spatial génomique du gène HoxB1 lors de l'induction de son expression dans le contexte de la différenciation de cellules embryonnaires de souris induite par l'acide rétinoïque

Une partie de cette thèse décrit l'étude de l'environnement spatial génomique du gène HoxB1 de souris lors de l'induction de son activation, dans le contexte de la différenciation de cellules embryonnaires souches de souris induite par l'acide rétinoïque. Nous avons vu que l'activité transcriptionnelle et la différenciation peuvent modifier l'organisation spatiale de la chromatine. Nous avons donc voulu vérifier comment l'activité transcriptionnelle d'un gène modèle (HoxB1) pouvait modifier les interactions et les relations existant entre ce gène et le reste du génome. Le gène HoxB1 est relocalisé à l'extérieur de son territoire chromosomique lors de son activation par l'acide rétinoïque, ce qui fait que son environnement spatial est modifié. Notre objectif était de décrire cet environnement dynamique.

Cette description est axée sur le développement d'une nouvelle méthodologie de *Chromosome Conformation Capture* ouverte. Nous ne nous intéresserons donc pas directement à la biologie des gènes Hox en tant que telle. Ce gène fut utilisé en tant que système modèle d'induction de gène qui résulte en un changement d'environnement spatial et en tant que « preuve de principe » de la méthodologie en question. Néanmoins, il convient ici de faire une brève revue de la biologie de cette classe de gènes.

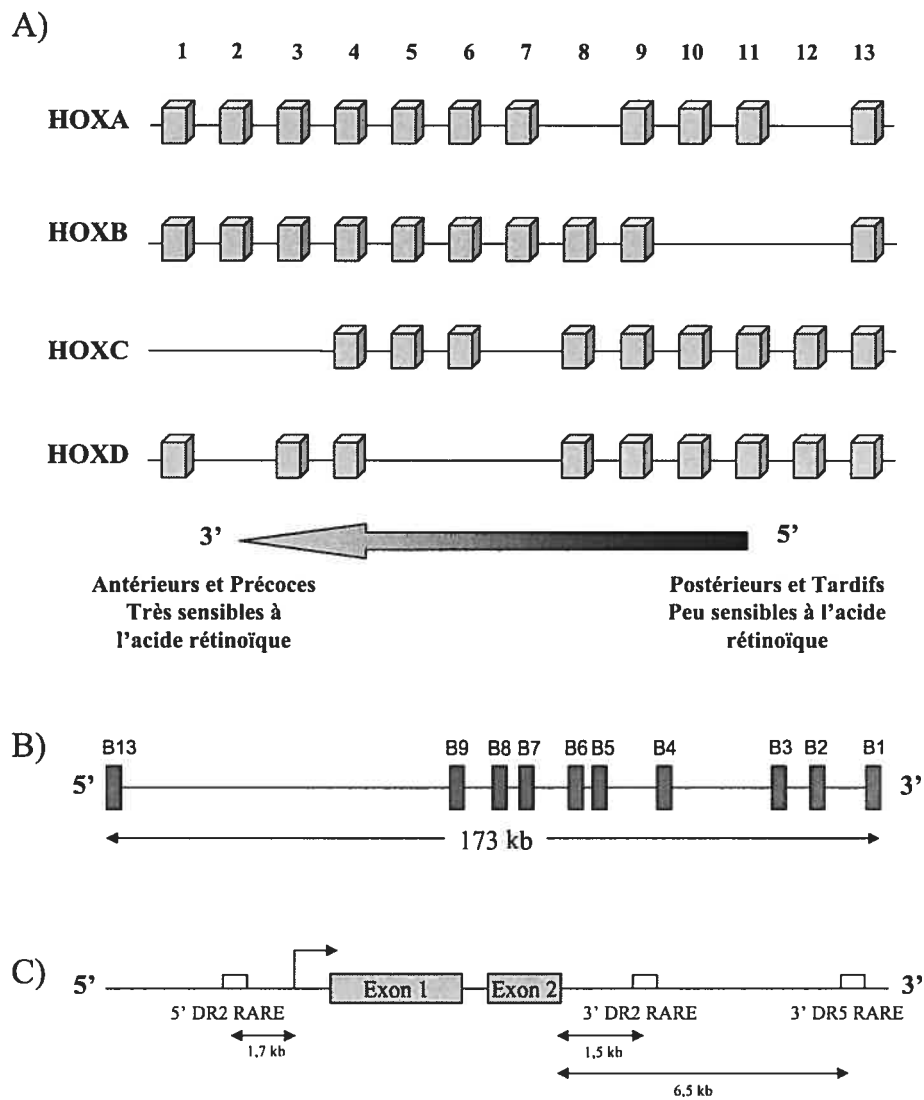
### La biologie des gènes Hox et leur induction par l'acide rétinoïque

Les gènes Hox sont une famille de gènes qui encodent des facteurs de transcription possédant un domaine « homeobox » capable de lier l'ADN (Maconochie et al. 1996; Marshall et al. 1996). Ils sont présents en quatre complexes distincts chez les vertébrés (HoxA, HoxB, HoxC, HoxD) et sont impliqués dans l'établissement de patrons antéro-postérieurs durant le développement embryonnaire. Les complexes comptent plusieurs gènes organisés de façon co-linéaire dans le génome par rapport à leur domaine spatio-temporel d'expression dans l'embryon. Ces gènes forment 13 groupe paralogues : par exemple, HoxA1, B1, D1 sont paralogues et sont similairement situés par rapport aux autres gènes de leurs complexes respectifs. Les gènes Hox sont différenciellement sensibles



à l'acide rétinoïque selon leur position dans le complexe. En effet, les gènes situés en 3' des complexes sont exprimés tôt dans l'embryogénèse et dans des tissus plus antérieurs que les 5', et leur sensibilité à l'acide rétinoïque est co-linéaire avec leur position génomique, les gènes en 3' étant plus sensibles que les gènes en 5'. L'organisation de la sensibilité à l'acide rétinoïque le long des complexes de gènes Hox est d'autre part conservée dans les cellules en cultures (Figure 13A). À cause de la colinéarité spatio-temporelle régissant leur expression, les mutations dans les gènes Hox conduisent à des phénotypes « homéotiques » caractérisés par des déplacements ou des duplications de structures anatomiques le long des axes antéro-postérieurs embryonnaires.

L'acide rétinoïque est donc central dans la régulation des gènes Hox et agit comme ligand pour des récepteurs spécifiques, appelés RAR ou RXR (Chambon 1996; Marshall et al. 1996). Ces protéines font partie de la famille des récepteurs nucléaires qui comprend entre autre le récepteur à l'hormone thyroïdienne, aux oestrogènes et aux gluco-corticoïdes. Les divers récepteurs à l'acide rétinoïque ont différentes spécificités de substrats. En effet, les RAR ont pour ligands naturels l'acide rétinoïque 9-cis et all-trans, alors que les récepteurs RXR ne lient que l'acide rétinoïque 9-cis. Ces récepteurs se lient à des séquences appelées RARE ou « *retinoic acid response element* », qui sont composées de deux séquences en répétition directe, ou DR, dont la séquence consensus est PuG(GT)TCA(X)<sub>n</sub>PuG(GT)TCA (Chambon 1996; Marshall et al. 1996). Le nombre variable de nucléotides entre les deux répétitions dicte quels dimères de récepteurs pourront se lier et quel en sera l'effet sur la transcription. En effet, RAR et RXR doivent agir en hétéro-dimères pour lier chacun un demi-site consensus, mais RXR peut aussi s'homo-dimériser ou interagir avec d'autres récepteurs nucléaires. Les éléments DR2 et DR5 (espace de 2 ou 5 nucléotides entre les répétitions) sont préférentiellement liés par les RAR/RXR et favorisent la transcription alors que d'autres conformations comme DR1 favorisent le dimère RXR/RXR et sont associées à la répression. Cette liaison n'est pas nécessairement dépendante du ligand (Weston et al. 2003).



**Figure 13 : Organisation des gènes Hox, du complexe HoxB et du gène HoxB1**

A) Organisation génomique comparative des complexes de gènes Hox de souris. Adapté de (Maconochie et al. 1996). B) Organisation génomique du complexe HoxB situé sur le chromosome 11 de souris. C) Organisation génomique du gène HoxB1 de souris. La figure n'est pas à l'échelle.

Les récepteurs nucléaires recrutent plusieurs co-facteurs agissant entre autres sur la machinerie transcriptionnelle de base ou sur la structure de la chromatine. Les interactions entre les récepteurs à l'acide rétinoïque et divers co-facteurs peuvent dépendre de la liaison au ligand. Par exemple, les récepteurs peuvent recruter des répresseurs tels que N-CoR et

SMRT de façon indépendante du ligand (Chen et Evans 1995; Horlein et al. 1995). Suite à un changement conformationnel induit par la liaison du ligand, la spécificité du domaine d'activation est modifiée pour recruter des activateurs de la transcription tels que p300/CBP (une histone acétyltransférase), TFIIB (machinerie transcriptionnelle de base) et TIF1 (Le Douarin et al. 1995).

L'organisation du complexe HoxB de souris est illustrée à la Figure 13B. Ce complexe contient 10 gènes (HoxB1-9, 13) et est situé sur le chromosome 11. Chez le poulet, les gènes HoxB1-5 sont sensibles à l'acide rétinoïque alors que les gènes HoxB6-13 sont régulés par le FGF (Fibroblast Growth Factor) (Bel-Vialar et al. 2002). La biologie du gène HoxB1 est d'un plus grand intérêt pour cette thèse (Figure 13C). HoxB1 possède deux RARE en 3' du gène : un DR2 impliqué dans l'expression dans le tube neural, le mésoderme et les membres antérieurs, et un DR5 important pour l'expression dans l'endoderme des viscères, le mésoderme, les tissus extra-embryonnaires, les poumons et les membres antérieurs (Huang et al. 2002). Il existe aussi un RARE et un « enhancer » situés en 5' du gène qui sont impliqués dans l'expression dans le rhombomère 4 et son exclusion des rhombomères 3 et 5 (Marshall et al. 1996).

La chromatine du gène HoxB1 est remodelée de façon concomitante avec son expression (Chambeyron et Bickmore 2004). En effet, lors de son activation par un traitement des cellules à l'acide rétinoïque, les histones du complexe HoxB sont modifiées par l'acétylation en lysine 9 de l'histone H3 et la di-méthylation de la lysine 4, qui sont des modifications correspondant à une activation transcriptionnelle. Cependant, même un gène inactif du complexe (HoxB9) est lui-aussi modifié par acétylation lors du traitement à l'acide rétinoïque. Il semble donc que les modifications aux histones précèdent l'activation des gènes individuels du complexe. Le gène HoxB1 est repositionné hors de son territoire chromosomique de façon concomitante avec son expression, et les gènes HoxB1 et B9 sont visiblement séparés à ce moment par FISH. Ceci suggère que le repliement d'ordre supérieur de la chromatine du complexe HoxB, et l'environnement spatial du gène HoxB1 en particulier, sont modifiés de façon concomitante avec son expression.

## MATÉRIEL ET MÉTHODE

### 1. Recombinaison entre séquences en répétition directe dans l'embryon de souris et son impact sur l'évolution des génomes de mammifères

**Plasmide pXH:** Le plasmide pXH est dérivé du pASB-HindIII (Belmaaza et al. 1990). pASB-HindIII contient le fragment KpnI/EaeI de 3 kb situé à l'extrémité 3' du L1 de souris L1MdA2 cloné par Loeb et collaborateurs (Loeb et al. 1986). pASB-HindIII fut digéré par HindIII, les extrémités simple-brin furent remplies avec la Klenow (New England Biolabs) et un oligonucléotide de 8 pb contenant un site de restriction XhoI y fut ligaturé pour générer pASB-XhoI. Ce plasmide fut digéré par AflIII et XmnI tandis que pASB-HindIII fut digéré par HincII et XmnI. Les fragments contenant les séquences L1 de chacun de ces plasmides furent ligaturés ensemble pour former le plasmide pXH, qui contient deux fragments de L1 en répétition directe. Ces deux L1 sont donc identiques excepté pour les sites HindIII et XhoI.

**Souris transgéniques:** Avant la micro-injection, le plasmide pXH fut linéarisé avec NarI (Amersham-Pharmacia Inc.). L'ADN linéarisé fut séparé par électrophorèse sur gel d'agarose, excisé et purifié à l'aide d'une colonne de fibre de verre. Les plasmides linéaires furent micro-injectés dans le pronucléus mâle d'ovules de souris B6C 3F1 fertilisés en utilisant les procédures standard (Gordon et Ruddle 1983). La présence du transgène dans les souris résultantes fut détectée par buvardage de Southern sur l'ADN génomique de queue digéré par PstI (New England Biolabs) selon les méthodes standards. La sonde qui fut utilisée correspond à la partie plasmidique du pXH.

**Détermination du nombre de copies de transgène :** Lorsque possible (souris 4400, 4701, 4431), le nombre de copies de transgène fut estimé en divisant le nombre de clones bactériens analysés par plasmid rescue par le nombre de clones représentant un réarrangement unique du plasmide. Nous avons assumé dans ces cas que le réarrangement en question était présent à raison d'une seule copie dans l'intégration en tandem de la souris concerné (en raison de sa structure moléculaire unique). Le nombre de copies des autres souris fut déterminé par PCR quantitatif en utilisant le SYBR Green PCR master mix (Applied Biosystems Inc.) et un appareil ABI 7700 équipé du logiciel Sequence Detector

v1.7. Une paire d'amorces fut utilisée pour amplifier un fragment de 54 bp du gène amplicilline du plasmide (5' AACCCACTCGTGACCCAACTGAT 3' et 5' AAACGCTGGTGAAAGTAAAAGATGCTGAA 3'). Une deuxième paire d'amorces fut utilisée pour amplifier un fragment de 56 pb du gène Rb de souris (5' AGCAGCCTCAGCCTTCCATACTCAGA 3' et 5' ACAGGGCAAGGGAGGTAGATTTCAATG 3'). Une courbe standard fut générée pour chaque paire d'amorces. Ces courbes ont été générées à partir d'amplification en triplicata sur des dilutions d'ADN génomique de queue de la souris 4469 (les dilutions utilisées pour générer les courbes standards couvraient 4 log). Environ 5 ng d'ADN génomique de queue des souris sélectionnées fut soumis à des amplifications en triplicata pour évaluer leur nombre de copies de transgène intégrées. Le nombre de copies fut estimé en comparant les résultats obtenus avec la paire d'amorce amplicilline pour les souris 4400, 4701 et 4431 avec les résultats obtenus pour les autres souris. Ces résultats furent normalisés pour la quantité d'ADN initiale utilisée pour chaque souris en utilisant les résultats de l'amplification Rb. La moyenne des trois valeurs de nombre de copies obtenues (par comparaison avec les valeurs des souris 4400, 4701 et 4431) est rapportée. Pour ce qui est des souris dont le nombre de copies ne fut pas évalué par PCR quantitatif, une comparaison de l'intensité de signal par buvardage de Southern entre les membres évalués et non-évalués d'une même lignée nous a permis d'estimer leur similarité en ce qui a trait au nombre de copies de transgène. La comparaison d'intensité de signal fut réalisée à l'aide du logiciel ImageJ v1.25S de Wayne Rasband. La sonde de buvardage de Southern utilisée correspondait à la partie plasmidique du plasmide pXH. Le buvardage de Southern fut réalisé sur de l'ADN génomique de queue de souris digéré par PstI/SacI, qui excise le fragment plasmidique du pXH.

**Plasmid rescue et transformation bactérienne:** L'ADN génomique de queue de souris fut extrait selon les méthodes standards et fut digéré par PstI (New England Biolabs). L'ADN digéré fut nettoyé par extraction au phénol : chloroforme, puis 5 µg de cet ADN fut ligaturé avec 1600U de T4 DNA ligase (New England Biolabs) dans 200 µl. Les produits de ligation furent électroporés en utilisant les bactéries électrocompétentes Rec A- DH10B (Invitrogen Inc.) selon le protocole du fabricant.

**Extraction d'ADN plasmidique et analyse:** L'ADN plasmidique fut extrait en utilisant une modification de la miniprep par lyse alcaline (Zhou et al. 1990). L'ADN fut analysé par analyse de restriction en utilisant les enzymes appropriées (New England Biolabs) et par séquençage. Les analyses de séquences furent réalisées avec NCBI Blast ([www.ncbi.nlm.nih.gov/BLAST](http://www.ncbi.nlm.nih.gov/BLAST)) et le UCSC genome browser ([www.genome.cse.ucsc.edu](http://www.genome.cse.ucsc.edu)).

## **2. Lien entre l'environnement spatial d'un locus et le choix de ses partenaires de recombinaison**

**Culture cellulaire :** Les lignées cellulaires IMR90 (fibroblastes fœtaux humain non-immortalisés) et ATA67 (clone cellulaire dérivé de la lignée de fibroblastes humains déficients en ATM GM005849) furent utilisées. La lignée cellulaire GM005849 fut obtenue du NIGMS Human Mutant Cell Repository (Coriell Institute for Medical Research). La lignée IMR90 fut obtenue d'ATCC (American Type Culture Collection). Les cellules furent maintenues dans du DMEM (Dulbecco modified Eagle medium) enrichi de L-glutamine, de sérum de veau fœtal (FBS) à 10% du volume total (Wisent) et de 50ug/ml de gentamycine et penstrep (Wisent) et cultivées à 37°C avec 5% CO<sub>2</sub>.

**REM-3C:** Une modification de la procédure de 3C publiée fut suivie, exceptée pour l'étape de PCR (Tolhuis et al. 2002). 10<sup>7</sup> cellules en croissance exponentielle furent rincées avec une solution tampon phosphate (PBS ou *Phosphate Buffered Saline*), récoltées par trypsinisation et traitées avec du DMEM/10% FBS supplémenté de 2% formaldéhyde pour 10 minutes à température de la pièce. Le formaldéhyde fut inactivé avec de la glycine 1M. Les noyaux furent extraits en présence d'inhibiteurs de protéase et traités avec 600U d'enzyme de restriction BglII ou EcoRI pour 16 heures. La chromatine digérée fut ensuite ligaturée avec la T4 ADN ligase en conditions diluées (2,5ng/μl). Les liens chimiques introduits par le formaldéhyde furent renversés en incubant à 65 °C avec de la protéinase K pour la nuit (*overnight*). L'ADN fut ensuite extrait et soumis au PCR.

**REM-PCR et inter-L1 PCR :** 200 ng d'ADN fut utilisé pour chaque PCR. Les amplifications furent réalisées en utilisant le Expand High-Fidelity PCR system (Roche). Pour l'étape d'élongation du REM-3C, les conditions de cyclage thermique étaient : 15s à

94°C, 30s à 65°C, 2 :30 min à 68°C; 70 cycles. L'amorce utilisée pour cette étape était 408F38 (5'-GGGGACCAACAAAAAAACTT-3'). Cette amorce est complémentaire à un locus situé sur le chromosome 2 humain, en position 40872835 (UCSC genome browser version may 2004). Pour l'étape d'amplification du REM-3C, les conditions étaient : 15s à 94°C, 30s à 55°C, 2:30 min à 68°C pour 10 cycles suivi de 15s à 94°C, 30s à 65°C, 2 :30 min + 5s/cycle à 68°C pour 30 cycles. 1/50 de la réaction d'élongation de 70 cycles fut utilisé dans cette étape d'amplification. Les amorces utilisées étaient : 408F94 (5'-TGTCTTTTATAGGGAGGCATCTTTATAC-3') et l'amorce complémentaire à la séquence consensus L1 HDF5461 (5'-CCATTTGACCCAGCCATCCCATTACT-3'). Ces conditions d'amplification furent aussi utilisées pour le inter-L1 PCR, mais dans ce cas seule l'amorce HDF5461 fut utilisée.

**Clonage des produits de PCR et préparation plasmidique :** Les produits de REM-3C furent séparés par électrophorèse sur gel d'agarose de 1%. Le gel fut découpé en 5 fragments égaux et l'ADN fut extrait en utilisant le Sephaglas Bandprep kit (Amersham Pharmacia Biotech). L'ADN de chaque fragment fut cloné de façon indépendante en utilisant les kits Qiagen A-addition et PCR cloning (Qiagen) pour minimiser le biais dans l'efficacité de clonage dû à la taille des fragments de PCR. Les ligations provenant d'un PCR donné furent ensuite regroupées et transformées dans des bactéries DH5 $\alpha$  électrocompétentes (Invitrogen). Pour la méthode REM-3C, les clones positifs furent identifiés par hybridation *in situ* d'une sonde correspondant à l'ADN génomique situé entre l'amorce 408F94 et le site BglII utilisé dans l'expérience sur le chromosome 2 humain, tel que décrit (Buluwela et al. 1989). Pour la méthodologie inter-L1 3C, la sélection bleu/blanc utilisant le X-gal fut utilisée pour identifier les clones contenant une insertion. L'ADN plasmidique fut extrait en utilisant une modification de la méthode de la lyse alcaline (Zhou et al. 1990).

**Analyses de séquences:** Les séquences furent analysées avec le UCSC Human Genome Browser (version May 2004) ([www.genome.ucsc.edu](http://www.genome.ucsc.edu)).

### 3. Modifications de l'environnement spatial génomique du gène HoxB1 lors de l'induction de son expression dans le contexte de la différenciation de cellules embryonnaires de souris induite par l'acide rétinoïque

**Cellules souches embryonnaires de souris et différenciation :** Les cellules souches embryonnaires de souris OS25 furent gracieusement fournies par le laboratoire du Dr Wendy Bickmore. Les cellules furent maintenues sur des pétris recouverts de gélatine 0,1% dans du DMEM supplémenté de 10% FCS, d'acides aminés non-essentiels, de 0,3 mg/ml de L-glutamine, de 0,15mM A-Monothioglycérol (MTG), de 1000U/ml de LIF murin (Chemicon international ESGRO) et de 100 ug/ml G418 et furent conservées dans une atmosphère de 5% CO<sub>2</sub> à 37°C. Le milieu de culture des cellules souches embryonnaires non-différenciées fut changé quotidiennement. Ces cellules furent étiquetées « jour -1 ». La différenciation à l'aide de l'acide rétinoïque fut réalisée tel que décrit précédemment (Chambeyron et Bickmore 2004). Brièvement, 5X10<sup>5</sup> cellules furent étalées sur un pétri de culture cellulaire sans gélatine en utilisant le même milieu que précédemment mais sans LIF et sans G418 pour 1 journée (jour 0); 5X10<sup>-6</sup> M d'acide rétinoïque fut ensuite ajouté à ce milieu pour 4 jours (jour 1 à 4). Après 2 jours de traitement à l'acide rétinoïque, le milieu fut supplémenté de 2,5 µM de ganciclovir pour 2 jours (jusqu'au jour 4 du protocole). Ensuite, le milieu fut supplémenté uniquement avec du ganciclovir (sans acide rétinoïque) pour 3 jours (jour 5 à 7). Le milieu fut changé à tous les deux jours du protocole de différenciation.

**Analyse par RT-PCR:** L'ARN total fut extrait en utilisant le réactif Trizol (Invitrogen). Entre 500 ng et 5 ug d'ARN total fut traité avec 1U de DNase I sans RNase (Invitrogen). Les cDNA furent générés en utilisant 200U de M-MLV Reverse transcriptase (Invitrogen) et 5 ng/ul d'hexanucléotide aléatoire (Invitrogen). Les PCR furent réalisés en utilisant 1/20 de la préparation d'ADNc par PCR. Séquences des amorces: Oct4: 5'-GGCGTTCTCTTTGGAAAGGTGTTTC-3' et 5'-CTCGAACCACATCCTTCTCT-3'; HoxB1: 5'-CCATATCCTCCGCCGAG-3' et 5'-CGGACTGGTCAGAGGCATC-3'; β-actine: 5'-GGTCAGAAGGACTCCTATGTGG-3' et 5'-TCTCAGCTGTGGTGGTGAAG-3'. Les PCR semi-quantitatifs furent réalisés en utilisant 1U de Taq polymérase par réaction et les conditions de cyclage suivantes : 2 min à 94°C; 15s à 94°C, 30s à 60°C, 45s à 72°C pour 25 cycles (β-actin) ou 35 cycles (HoxB1 et Oct4); 5 min à 72°C. Les amplifications



par PCR quantitatif furent réalisées en duplicata, en utilisant un appareil Stratagene Mx3000P, le logiciel MxPro (Stratagene) et le Quantitect SYBR green PCR kit (Qiagen). Chaque point du graphique présenté en Figure 23 représente la moyenne des résultats obtenus à partir de quatre à huit traitements indépendants à l'acide rétinoïque. Les conditions de cyclage thermique utilisées furent similaires à celles de l'approche semi-quantitative, à l'exception du nombre total de cycles, qui fut fixé à 40. Des courbes standards furent réalisées pour chaque paire d'amorces en duplicata en utilisant des dilutions 1 :10, 1 :100 et 1 :1000 d'une préparation de cDNA du jour -1.

**Chromosome Conformation Capture (3C):** Une modification de la procédure de 3C standard fut utilisée, excepté pour l'étape du PCR (Tolhuis et al. 2002). Entre  $5 \times 10^5$  et  $5 \times 10^6$  cellules environ furent rincées avec du PBS et traitées avec du DMEM/10% FBS supplémenté de 2% de formaldéhyde pour 10 minutes à température de la pièce. Le formaldéhyde fut ensuite inactivé en ajoutant de la glycine 1M, et les cellules furent récupérées avec un grattoir à cellules. Les cellules furent pipettées pour obtenir une suspension de cellules homogène. Les noyaux furent extraits en présence d'inhibiteurs de protéases et traités pour la nuit (*overnight*) avec 600 U d'enzyme de restriction HindIII (New England Biolabs). La chromatine digérée fut ligaturée en utilisant la T4 ADN ligase (New England Biolabs) en conditions diluées (2,5 ng/ul). Les liens chimiques du formaldéhyde furent renversés par une incubation pour la nuit (*overnight*) à 65 °C avec de la protéinase K. L'ADN fut ensuite extrait et soumis au PCR.

**Inverse-PCR:** L'ADN traité par 3C fut digéré avec l'enzyme de restriction MseI (New England Biolabs). L'ADN digéré fut ensuite nettoyé par extraction au phénol/chloroforme et ligaturé dans des conditions diluées (2,5 ng/ul) en utilisant la T4 ADN ligase (New England Biolabs). Les inverse-PCR sur le locus HoxB1 furent réalisées en utilisant le Expand High Fidelity PCR system (Roche) et 50 à 200 ng d'ADN traité par le protocole 3C, digéré par MseI et ligaturé. Un premier PCR fut réalisé en utilisant les amorces 387F (5'- ATGGGGGTTCTGGGATAAGTAA-3') et 317R (5'- GGGACAAAAGGATGACTAGGAAGGGAGAG-3'). 1/50 de ce premier PCR fut ensuite utilisé pour une deuxième amplification par PCR en utilisant les amorces 548F (5'- TTGGGAGGGGAGTAAAAGTCTT-3') et 269R (5'- CAGGGTGTGAGCCAGGGGTCTAAAATC-3') pour obtenir suffisamment de produits

de PCR pour les étapes de clonage subséquentes. Les conditions de PCR furent les suivantes : 2 min à 94°C; 15s à 94°C, 30s à 60°C, 3 min à 68°C pour 10 cycles; 15s à 94°C, 30s à 60°C, 3 min + 5s/cycle d'extension automatique à 68°C pour 20 cycles; 5 min à 68°C. Le Tableau 1 décrit les différents inverse-PCRs qui furent clonés et séquencés. Dans certains cas, plusieurs inverse-PCRs indépendants furent regroupés avant le clonage pour s'assurer d'une grande diversité de produits analysés.

| Jour du traitement | Noms des ADN <sup>a</sup> | Noms des PCR | # de PCR regroupés <sup>b</sup> | # de clones <sup>c</sup> |
|--------------------|---------------------------|--------------|---------------------------------|--------------------------|
| J-1                | B1                        | A            | 1                               | 94                       |
| J-1                | B1                        | B            | 1                               | 86                       |
| J-1                | B1                        | C            | 1                               | 60                       |
| J-1                | B1                        | D            | 1                               | 77                       |
| J-1                | B1                        | E            | 1                               | 77                       |
| J-1                | B1                        | F            | 1                               | 23                       |
| J-1                | B1                        | 1-9          | 9                               | 70                       |
| J-1                | B1                        | 10-18        | 9                               | 66                       |
| J-1                | F1                        | H            | 1                               | 58                       |
| J-1                | F1                        | N            | 1                               | 60                       |
| J-1                | F1                        | 1-9          | 9                               | 72                       |
| J-1                | F1                        | 10-18        | 9                               | 82                       |
| J-1                | F1                        | 19-27        | 9                               | 78                       |
| J1                 | F2                        | I            | 1                               | 78                       |
| J1                 | F2                        | J            | 1                               | 88                       |
| J1                 | F2                        | 1-10         | 10                              | 77                       |
| J1                 | F2                        | 11-20        | 10                              | 80                       |
| J1                 | F2                        | 21-30        | 10                              | 77                       |
| J1                 | F3                        | 1-7          | 7                               | 87                       |
| J1                 | F3                        | 8-14         | 7                               | 82                       |
| J1                 | F3                        | 15-21        | 7                               | 83                       |
| J3                 | E1                        | 1-10         | 10                              | 68                       |
| J3                 | E1                        | 11-20        | 10                              | 77                       |
| J3                 | E1                        | 21-28        | 8                               | 77                       |
| J3                 | E1                        | 29-36        | 8                               | 64                       |
| J3                 | E1                        | 37-44        | 8                               | 71                       |
| J3                 | E1                        | 45-52        | 8                               | 57                       |
| J3                 | G1                        | 1-9          | 9                               | 146                      |
| J3                 | G2                        | 1-9          | 9                               | 165                      |

**Tableau 1 : Inverse-PCR 3C réalisés sur le locus du gène HoxB1**

<sup>a</sup>: Chaque préparation d'ADN provient de cellules traitées indépendamment

<sup>b</sup>: Nombre d'inverse-PCR indépendants regroupés avant l'étape de clonage

<sup>c</sup>: Nombre de clones qui furent séquencés et localisés dans les banques de données du génome

**Clonage des produits de PCR :** Les produits de PCR furent séparés par électrophorèse sur gel d'agarose. L'ADN fut extrait des fragments de gel en utilisant le Sephaglas Bandprep kit (Amersham Pharmacia Biotech) et clonés en utilisant le Qiagen A-addition et PCR cloning kit (Qiagen). Les produits de ligation furent ensuite transformés dans des bactéries DH5 $\alpha$  électrocompétentes (Invitrogen). La sélection bleu/blanc avec le X-gal fut utilisée pour identifier les clones contenant une insertion. L'ADN plasmidique fut extrait ou le produit cloné fut amplifié en réalisant un PCR directement sur les colonies bactériennes positives.

**Analyse de séquences:** Les séquences furent analysées en utilisant le UCSC genome browser ([www.genome.ucsc.edu](http://www.genome.ucsc.edu); assemblage du génome de souris March 2005) et le Ensembl Karyoview ([www.ensembl.org](http://www.ensembl.org)).

**Analyse de 3C:** Huit fragments génomiques proximaux furent choisis pour les analyses de 3C. Ces fragments correspondent aux loci situés aux positions -143122, -66631, -30994, -17379, 5631, 12515, 36214 et 53223 de la Figure 24. Des ADN contrôles contenant un fragment du locus HoxB1 ligaturé aux fragments génomiques d'intérêt furent générés par PCR, en utilisant une amorce complémentaire à la région située en amont du site HindIII du locus HoxB1 utilisé pour l'inverse-PCR 3C (548F) et une amorce spécifique aux loci d'intérêt (voir Tableau 2). Le nombre de copies de chacune des préparations d'ADN contrôles fut déterminé par PCR quantitatif Taqman en utilisant des amorces et une sonde situées dans le fragment HoxB1 commun à toutes les préparations (amorces AvHIII et 1A et sonde 1S). Une courbe standard (5 log) fut générée pour cette paire d'amorces et cette sonde pour permettre cette quantification. Les préparations contrôles quantifiées furent diluées en série (4 log) et utilisées pour générer des courbes standard normalisées avec les paires d'amorces correspondantes et qui permettent d'amplifier les produits de ligation entre le locus HoxB1 et les loci d'intérêt par PCR quantitatif Taqman (amorce 1A, sonde 1S et amorces 2B à 10B du Tableau 2). Pour simuler la complexité des véritables préparations d'ADN 3C, 100 ng d'ADN génomique de souris fut ajouté à chacune des réactions des courbes standard. Ces courbes nous permettent de corriger pour les légères différences d'amplification entre les paires d'amorces et de normaliser les signaux des amplifications réalisées sur les préparations d'ADN 3C. L'abondance des différents produits de ligation d'intérêt fut déterminée par PCR quantitatif Taqman sur les

préparations d'ADN 3C des jours -1 et 3, en rapportant les signaux obtenus pour chaque paire d'amorces à la courbe standard normalisée correspondante. Afin de tenir compte des différences dans la quantité et la qualité des ADN 3C utilisés pour ces amplifications, les résultats furent normalisés à l'interne en calculant le ratio du signal obtenu pour une paire d'amorces donnée par rapport au signal obtenu pour la paire d'amorces amplifiant le produit de ligation entre HoxB1 et le locus 5631, et ce pour chaque amplification réalisée sur un ADN 3C donné. Chaque point du graphique de la Figure 25 représente la moyenne de trois à six signaux relatifs. Les amplifications furent réalisées en utilisant le Applied Biosystems Inc. Taqman master mix, un appareil Stratagene Mx3000P et le logiciel MxPro ou un appareil Applied Biosystems Inc. 7900HT et le logiciel SDS 2.2.2. Les conditions des PCR quantitatifs étaient: 10 min à 94°C; 15s à 94°C, 1 min à 60°C pour 50 cycles.

| Nom    | Locus           | Séquence (5'-3')           |
|--------|-----------------|----------------------------|
| 1S     | Locus de départ | TTAGCCCATTGGCCTGGGAGAGAT   |
| 548F   | Locus de départ | TTGGGAGGGGAGTAAAAGTCTT     |
| AvHIII | Locus de départ | AAGGGGCTAGGGAGTGGGGGTG     |
| 1A     | Locus de départ | GCTGGGACTGCCAAACTCTG       |
| 2B     | -17379          | AGAGCAGACTTAGCCCAAGATTCC   |
| 3B     | -30994          | TCTCAGCTCAGGCCACTTCCTGTA   |
| 5B     | -66631          | CTGTATGAGGAGGTGGTGAGACAA   |
| 6B     | -143122         | TCCTAAGGATTCTTCACCATAGACTC |
| 7B     | 5631            | TCAGAGATGCTTGCTCTGTCCTCT   |
| 8B     | 12515           | TGAAGGAATTCACCACTCTTCTGT   |
| 9B     | 36214           | GACTGAAACACTGGCCCAAG       |
| 10B    | 53223           | AGTCTCGAGAAAGGACACATTGGC   |

**Tableau 2 : Séquences des amorces utilisées dans l'étude de 3C réalisée sur le locus HoxB1**

**Contrôle d'efficacité de 3C :** Le locus du gène  $\beta$ -actine fut choisi pour réaliser le contrôle d'efficacité de 3C (voir la Figure 28 et le texte pour les détails). La fréquence de fixation du fragment HindIII génomique contenant le gène  $\beta$ -actine au fragment adjacent situé immédiatement en amont fut évaluée. Premièrement, un PCR de 25 cycles fut réalisé sur les préparations d'ADN 3C en utilisant les amorces ActNestA : 5'-ACTTAGGTGTACCTGTGTGTGCCT-3' et ActAdj: 5'-AGCAGTGGTTTCTATTGGCTGTCG-3'. Ce premier PCR fut réalisé avec le Expand High Fidelity PCR system de Roche Diagnostics. Les conditions PCR étaient : 2 min à

94°C; 20s à 94°C, 30s à 60°C, 1 min à 72°C pour 25 cycles. Un PCR quantitatif SYBR green fut réalisé sur 1/50 de cette première amplification en utilisant l'amorce interne ActA: 5'-CTTCTGACCTAGAACTCTTGATCCC-3' et l'amorce ActAdj dans les mêmes conditions PCR, à l'exception du nombre de cycles qui fut fixé à 35. Une courbe standard fut réalisée pour le PCR quantitatif par SYBR green en utilisant des dilutions sériées sur 4 log d'un premier PCR tel que décrit plus haut et les amorces ActA et ActAdj. Pour comparer le nombre de copies du fragment HindIII génomique contenant le gène  $\beta$ -actine dans les différentes préparations d'ADN 3C (et ainsi normaliser pour la quantité d'ADN présente dans chaque échantillon), une courbe standard fut réalisée en utilisant des dilutions sériées (4 log) d'ADN génomique de souris et les amorces ActA, ActB: 5'-CCCTCTACACACTCAGAATTCATC-3' et la sonde Taqman ActS: 5'-CGGGGAACGGTACCACACTCAGTTT-3'. Les conditions de ce PCR quantitatif Taqman étaient les suivantes: 94°C; 15s à 94°C, 1 min à 60°C pour 50 cycles. Ce PCR amplifie un fragment situé à l'intérieur du fragment génomique HindIII contenant le gène  $\beta$ -actine. Le même PCR fut réalisé sur chacun des ADN 3C et le signal fut comparé à la courbe standard pour obtenir le nombre de copies de ce fragment dans chacune des préparations d'ADN 3C. Les quantifications furent réalisées au moins en triplicata avec le Applied Biosystem Inc. Taqman master mix ou le Qiagen Quantitect SYBR green master mix et un appareil à PCR en temps réel Stratagene Mx3000P et le logiciel MxPro.

**Analyse du cycle cellulaire par cytométrie en flux avec le marqueur Hoechst 33342 :**

Les cellules furent incubées une heure dans un milieu de marquage préchauffé à 37°C (HBSS+FCS2%+10mM HEPES pH 7.4) contenant 10  $\mu$ g/mL de Hoechst 33342 (Molecular Probe, Invitrogen) et furent ensuite rincées dans le même milieu froid sans Hoechst 33342. Entre 139000 et 196000 événements par condition furent recueillis en utilisant un cytomètre en BD LSR II (BD Biosciences, San Jose, CA). Les résultats furent analysés avec le logiciel FCS Express (DeNovo Software, Ontario, Canada).

## RÉSULTATS

### 1. Recombinaison entre séquences en répétition directe dans l'embryon de souris et son impact sur l'évolution des génomes de mammifères

#### Objectifs

Peu d'études ont jusqu'à présent caractérisé la recombinaison homologe se produisant *in vivo*. Cette première partie concerne le développement d'une méthodologie permettant la caractérisation des activités de recombinaison à l'oeuvre dans l'embryon de souris. Les objectifs spécifiques de cette partie étaient de :

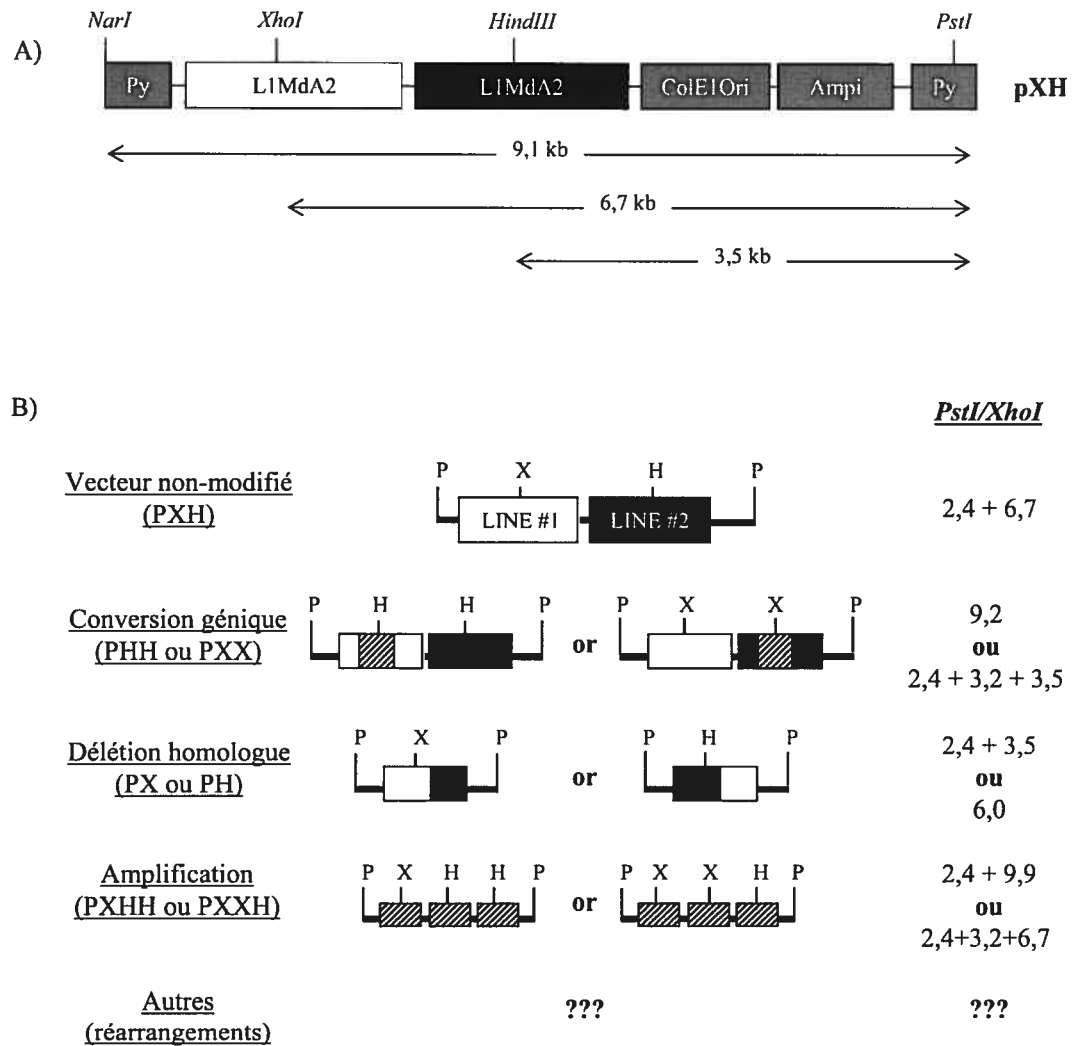
-Développer un essai expérimental permettant d'investiguer la prévalence des différents mécanismes de recombinaison génétique à l'oeuvre *in vivo* dans l'embryon de souris

-Caractériser les événements de recombinaison intra ou inter-moléculaires impliquant un plasmide contenant deux séquences LINE-1 en répétition directe et se produisant avant, pendant ou après son intégration au génome d'embryons de souris

**NOTE :** Cette étude a fait l'objet d'une publication (Wurtele et al. 2005).

#### Design expérimental

Pour atteindre les objectifs décrits plus haut, nous avons généré des souris transgéniques par micro-injection du plasmide pXH dans des ovules fécondés de souris (Figure 14A). Ce plasmide contient deux fragments de LINE-1 (L1) presque identiques provenant de la famille L1Mda2 arrangés en répétition directe. La seule différence entre ces deux L1 est une modification d'un site de restriction HindIII endogène (position 1316 du fragment) en un site XhoI dans une des deux séquences (voir Matériel et méthode). Les deux fragments de L1 du pXH sont des substrats pour des événements de recombinaison homologe avant, durant et après l'intégration du plasmide au génome : ils peuvent en effet se recombiner entre eux ou avec les L1 endogènes.



**Figure 14 : Plasmide utilisé dans l'étude et stratégie d'analyse**

A) Plasmide pXH. Le plasmide fut linéarisé par *NarI* avant la microinjection dans les ovules fécondés. Les séquences L1 du pXH font partie de la famille L1MdA2 (voir Matériel et méthode). Py: Séquences de l'origine de réplication du virus du polyome murin. ColE1Ori: Origine de replication bactérienne du plasmide, Amp<sup>r</sup>: Gène de résistance à l'ampicilline. B) Stratégie d'analyse des plasmides recombinants: Les plasmides récupérés furent digérés par *PstI/XhoI*. Le site *PstI* utilisé pour la procédure de *plasmid rescue* est dupliqué dans cette figure pour plus de clarté. Les produits de recombinaison homologue attendus et les patrons de restriction *PstI/XhoI* correspondants sont indiqués (la taille des produits de digestion attendus est indiquée en kb).

Quinze souris transgéniques fondatrices furent générées par micro-injection dans des ovules fécondés de souris du plasmide pXH linéarisé par l'enzyme de restriction NarI (Tableau 3). 5 de ces souris furent choisies pour générer des lignées transgéniques (4400, 4402, 4469, 4472 et 4475; Figure 15). Toutes les souris fondatrices furent analysées par buvardage de Southern sur de l'ADN génomique de queue en utilisant plusieurs digestions de restriction (résultats non-montrés). Cette analyse a indiqué que toutes les souris fondatrices avaient intégré des plasmides qui avaient probablement été impliqués dans des événements de recombinaison homologue et/ou illégitime. Cette analyse nous a aussi suggéré que le nombre de copies de transgènes intégrées dans les différentes souris était très variable.

| Souris | Nombre de copies <sup>a</sup> | Colonies repiquées <sup>b</sup> | PXH   | PXX   | PHH   | PX    | PH    | PXHH | PXXH | Autres |
|--------|-------------------------------|---------------------------------|-------|-------|-------|-------|-------|------|------|--------|
| 4400   | Moyen                         | 347                             | 39,5% | 42,7% | 0,0%  | 0,3%  | 0,3%  | 0,0% | 7,8% | 9,5%   |
| 4401   | Bas                           | 235                             | 42,1% | 14,0% | 0,0%  | 43,0% | 0,4%  | 0,0% | 0,0% | 0,4%   |
| 4405   | Bas                           | 111                             | 45,0% | 53,2% | 0,0%  | 0,0%  | 1,8%  | 0,0% | 0,0% | 0,0%   |
| 4408   | Moyen                         | 129                             | 59,7% | 2,3%  | 9,3%  | 11,6% | 14,0% | 1,6% | 0,0% | 1,6%   |
| 4413   | Bas                           | 138                             | 58,0% | 0,0%  | 19,6% | 7,2%  | 15,2% | 0,0% | 0,0% | 0,0%   |
| 4419   | Moyen                         | 139                             | 64,7% | 3,6%  | 12,2% | 15,1% | 0,7%  | 0,0% | 0,0% | 2,9%   |
| 4431   | Bas                           | 183                             | 69,4% | 0,0%  | 0,0%  | 1,1%  | 9,8%  | 0,0% | 0,0% | 19,7%  |
| 4438   | Bas                           | 169                             | 79,9% | 0,0%  | 0,0%  | 1,2%  | 18,9% | 0,0% | 0,0% | 0,0%   |
| 4454   | Haut                          | 188                             | 45,7% | 3,2%  | 8,0%  | 17,6% | 20,7% | 3,2% | 1,6% | 0,0%   |
| 4465   | Haut                          | 143                             | 63,6% | 0,7%  | 7,7%  | 20,3% | 6,3%  | 0,7% | 0,0% | 0,7%   |
| 4469   | Haut                          | 534                             | 31,5% | 1,3%  | 3,7%  | 21,0% | 38,2% | 2,2% | 0,6% | 1,5%   |
| 4472   | Haut                          | 140                             | 40,7% | 6,4%  | 1,4%  | 31,4% | 17,1% | 0,0% | 0,0% | 2,9%   |
| 4475   | Moyen                         | 88                              | 83,0% | 0,0%  | 3,4%  | 8,0%  | 1,1%  | 3,4% | 0,0% | 1,1%   |
| 4479   | Moyen                         | 138                             | 92,8% | 0,0%  | 0,7%  | 1,4%  | 2,9%  | 2,2% | 0,0% | 0,0%   |
| 4486   | Haut                          | 126                             | 86,5% | 3,2%  | 0,0%  | 7,1%  | 0,8%  | 0,0% | 0,0% | 2,4%   |

**Tableau 3 : Fréquences des produits de recombinaison chez les souris fondatrices**

<sup>a</sup>: Catégories de nombre de copies provenant du Tableau 4

<sup>b</sup>: Nombre de plasmides analysés

Pour investiguer la structure des plasmides intégrés dans chaque souris, nous avons utilisé la méthode de *plasmid rescue* sur l'ADN génomique de queue de souris. L'objectif de cette expérience était de déterminer comment les plasmides intégrés avaient été remodelés lors de leur intégration au génome des embryons de souris. En effet, il est connu que la recombinaison extra-chromosomique (se produisant avant ou pendant l'intégration des plasmides) est beaucoup plus fréquente que la recombinaison chromosomique (se produisant après l'intégration des plasmides) : il était donc probable qu'une fois intégrés, les plasmides soient relativement stables, et que les modifications observées dans leur



structure soient représentatives des événements de recombinaison auxquels ils avaient participé avant ou pendant leur intégration. Comme nous le verrons plus loin, cette présomption était généralement fondée, bien que dans certains cas les plasmides puissent être modifiés par recombinaison après leur intégration au fil des générations de souris. D'autre part, étant donné que l'intégration se produit tôt dans l'embryogénèse (voir plus bas), les modifications subies par les plasmides lors de leur intégration reflètent les activités de recombinaison à l'œuvre *in vivo* dans l'embryon précoce ou dans l'ovule fécondé.

L'ADN génomique des différentes souris fut donc digéré par l'enzyme de restriction PstI, qui clive le plasmide pXH à un seul endroit. Lorsque coupés par cette enzyme, les sites d'intégration de plasmides en répétition directe génèrent des plasmides de taille unitaire. L'ADN digéré fut ensuite ligaturé dans des conditions diluées et utilisé pour transformer des bactéries Rec A-. Les colonies résistantes à l'ampicilline furent repiquées, et leur ADN plasmidique fut digéré par PstI et XhoI : cette analyse de restriction nous a permis de classer les plasmides récupérés selon le type de produit de recombinaison qu'ils représentent (Figure 14B). Il est à noter que nous ne pouvons nous prononcer que sur le produit final des différents événements de recombinaison, et non sur le mécanisme précis qui fut impliqué dans leur génération.

Les produits de recombinaison furent catégorisés en 4 classes (Figure 14) : conversion génique sans crossing-over, délétion homologue (auquel cas un seul fragment de L1 demeure dans le plasmide, et les séquences situées entre les deux répétitions directes originales sont supprimées), amplification (auquel cas le plasmide résultant contient 3 fragments de L1) et d'autres réarrangements (cette catégorie regroupe tous les plasmides ne pouvant être assignés à aucune autre catégorie et contient entre autres les réarrangements illégitimes). Finalement, une dernière classe de plasmides récupérés est représentée par les plasmides qui ne furent pas modifiés de façon détectable par des événements de recombinaison.

Nous nous sommes assurés du fait que les plasmides récupérés par *plasmid rescue* n'étaient pas modifiés de façon importante dans les bactéries utilisées pour leur clonage, et qu'ils étaient donc représentatifs des structures d'intégration de plasmides en tandem présentes dans les souris. Nous avons donc testé la stabilité du plasmide pXH dans les bactéries

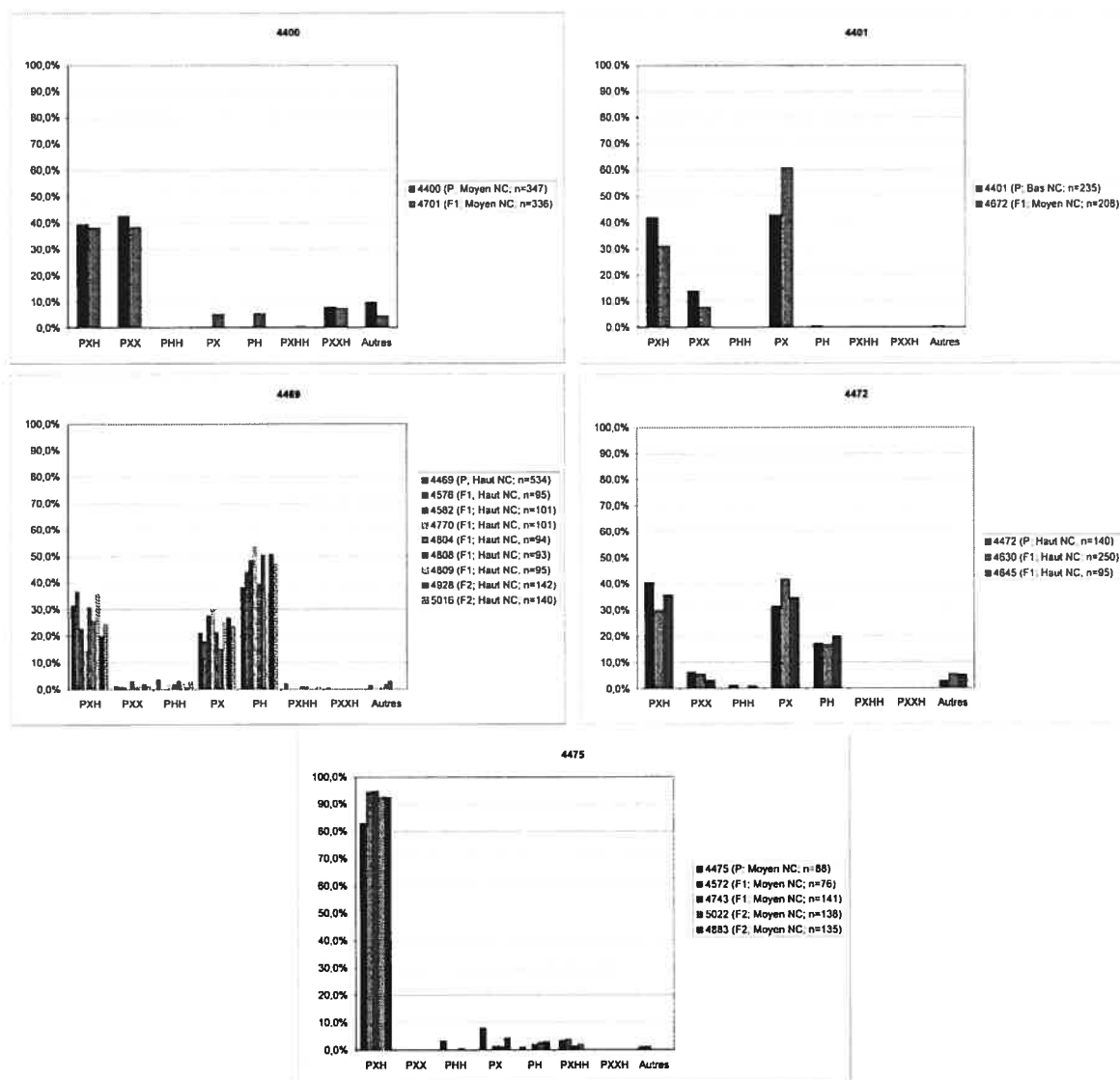
DH10B utilisées pour le *plasmid rescue*. Le plasmide circulaire fut transformé dans ces cellules et 168 colonies bactériennes furent repiquées et analysées : aucun événement de recombinaison ne fut détecté. Pour reproduire le plus fidèlement possible la procédure de *plasmid rescue*, le plasmide linéarisé par PstI fut mélangé avec de l'ADN génomique de souris en quantités équivalant à 15 copies de plasmides par génome. Cette préparation fut ligaturée et utilisée pour transformer des bactéries. 45 colonies furent repiquées et analysées : aucune molécule recombinante ne fut découverte. Finalement, le plasmide linéarisé par PstI fut utilisé seul pour transformer des bactéries DH10B. Une seule colonie sur 72 analysées était représentative d'un événement de délétion homologue (PH). Le bruit de fond artefactuel de recombinaison de notre méthodologie de *plasmid rescue* est donc bas, probablement de l'ordre de 1% ou moins.

### **La recombinaison homologue extra-chromosomique est extrêmement efficace dans l'embryon précoce de souris**

Le Tableau 3 et la Figure 15 présentent les fréquences relatives des différentes configurations de plasmide pour chaque souris. Les fréquences des différents types d'événements de recombinaison sont très variables entre les fondateurs. Cependant, la distribution des fréquences est semblable pour les différents individus d'une même lignée transgénique, ce qui suggère que les souris n'étaient généralement pas des mosaïques et que l'intégration du plasmide s'est produite tôt dans l'embryogénèse. Le fait que les descendants d'une lignée soient semblables entre eux suggère aussi que la plupart des événements de recombinaison se sont produits avant ou pendant l'intégration des plasmides, et qu'une fois intégrés, les plasmides deviennent beaucoup plus stables.

La distribution des types de produits de recombinaison dans les souris fondatrices, et le fait que ces distributions semblent relativement stables une fois les plasmides intégrés nous permet d'évaluer la prévalence des différents types d'événements de recombinaison se produisant dans l'embryon de souris au moment de l'intégration des plasmides. Les résultats présentés au Tableau 3 montrent que toutes les souris fondatrices ont intégré des plasmides non-modifiés (pXH). Les fondateurs ont aussi tous intégré des plasmides ayant participé à des événements de recombinaison : les fréquences de recombinaison homologue semblent particulièrement élevées, approchant 70% dans certains cas. 14 des 15 et 13 des

15 fondateurs ont intégré des plasmides impliqués dans des événements de délétion homologue (PX, PH) et dans des événements de conversion génique sans crossing-over (PXX, PHH), respectivement. Finalement, 8 fondateurs ont intégré des plasmides impliqués dans des amplifications et 10 fondateurs ont d'autres réarrangements dans leurs structures intégrées.



**Figure 15 : Distribution de fréquences des différentes configurations de plasmides présentes dans les structures intégrées pour les fondateurs et leurs descendances**

Les configurations de plasmides sont expliquées en Figure 14b. Bas, Moyen ou Haut NC : Catégories de nombre de copies provenant du Tableau 4. n : Nombre de plasmides analysés.

Le nombre de copies de transgènes intégrés fut évalué pour chaque souris fondatrice et pour un certain nombre de descendants et est présenté au Tableau 4. Ce nombre de copies représente notre échantillonnage des plasmides recombinants présents dans la cellule au moment de l'intégration. Lorsque possible, le nombre de copies fut estimé en utilisant un réarrangement présumément présent à une seule copie dans l'intégration en tandem comme valeur-étalon. En divisant le nombre de clones d'un réarrangement unique donné par le nombre de clones analysés par plasmid rescue pour la souris correspondante, il est possible d'obtenir une estimation du nombre de copies intégrées. Cette procédure fut utilisée pour les souris 4400, 4701 et 4431 (voir Figure 17D et E). Par exemple, dans la souris 4431, un réarrangement particulier correspondant à la jonction transgène-génome représente environ 20% ou 1/5 des plasmides récupérés (Tableau 3). Étant donné que cette jonction devrait être présente à raison d'une seule copie dans l'intégration de plasmides en tandem, le nombre total de plasmides intégrés pour cette souris devrait être de cinq par génome. Les évaluations du nombre de copies de transgènes des souris 4400, 4701 et 4431 furent ensuite utilisées comme valeurs-étalon pour la détermination du nombre de copies des autres souris par PCR quantitatif (voir Matériel et méthode). Les résultats de l'estimation du nombre de copies furent utilisés pour catégoriser les souris en nombre de copies faible (<10), moyen (10-100) et haut (>100). Ces estimations sont une indication de l'abondance relative du transgène entre les différentes souris et ne doivent pas être interprétées comme des valeurs exactes.

| Souris | Génération | Nombre de copies moyen <sup>a</sup> | Catégorie <sup>b</sup> |
|--------|------------|-------------------------------------|------------------------|
| 4400   | P          | <i>10</i>                           | Moyen                  |
| 4701   | F1         | <i>22</i>                           | Moyen                  |
| 4401   | P          | 4                                   | Bas                    |
| 4672   | F1         | <i>25</i>                           | Moyen                  |
| 4405   | P          | 4                                   | Bas                    |
| 4408   | P          | <i>16</i>                           | Moyen                  |
| 4413   | P          | 6                                   | Bas                    |
| 4419   | P          | <i>63</i>                           | Moyen                  |
| 4431   | P          | 5                                   | Bas                    |
| 4438   | P          | 6                                   | Bas                    |
| 4454   | P          | <i>152</i>                          | Haut                   |
| 4465   | P          | <i>294</i>                          | Haut                   |
| 4469   | P          | <i>235</i>                          | Haut                   |
| 4576   | F1         | <i>SB</i>                           | Haut                   |
| 4582   | F1         | <i>SB</i>                           | Haut                   |
| 4770   | F1         | <i>213</i>                          | Haut                   |
| 4804   | F1         | <i>SB</i>                           | Haut                   |
| 4808   | F1         | <i>SB</i>                           | Haut                   |
| 4809   | F1         | <i>SB</i>                           | Haut                   |
| 4928   | F2         | <i>469</i>                          | Haut                   |
| 5016   | F2         | <i>395</i>                          | Haut                   |
| 4472   | P          | <i>773</i>                          | Haut                   |
| 4630   | F1         | <i>427</i>                          | Haut                   |
| 4645   | F1         | <i>1457</i>                         | Haut                   |
| 4475   | P          | 33                                  | Moyen                  |
| 4572   | F1         | 43                                  | Moyen                  |
| 4743   | F1         | 36                                  | Moyen                  |
| 5022   | F2         | 76                                  | Moyen                  |
| 4883   | F2         | 47                                  | Moyen                  |
| 4479   | P          | 90                                  | Moyen                  |
| 4486   | P          | 117                                 | Haut                   |

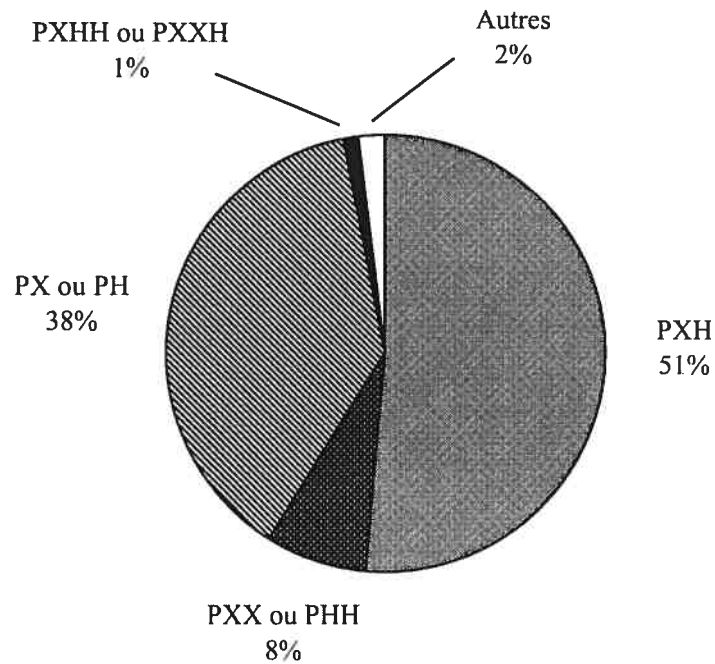
**Tableau 4 : Nombre de copies de transgène**

<sup>a</sup>: Lorsque le nombre de copies est en italique, il fut déterminé en divisant le nombre de plasmides analysés par le nombre de réarrangements récupérés. Ces valeurs furent ensuite utilisées comme valeur-étalon pour l'estimation du nombre de copies des autres souris par PCR quantitatif ou par buvardage de Southern (voir Matériel et méthode). La moyenne des calculs réalisés en utilisant les trois valeurs-étalon est présentée. SB: Le nombre de copies de ces souris est semblable à celui des autres membres de la lignée par buvardage de Southern.

<sup>b</sup>: Les souris furent catégorisées sur la base de leur nombre de copies de transgène de la façon suivante: Bas (moins de 10 copies), Moyen (entre 10 et 100 copies), Haut (plus de 100 copies).

Nous avons appliqué le test statistique du Chi carré aux distributions des souris fondatrices : cette analyse révèle que les structures intégrées sont statistiquement différentes les unes des autres ( $p < 0,0001$ ; Test Chi carré de Pearson avec approximation de Monte Carlo). Il existe donc un certain degré de variabilité dans la composition des intégrations en tandem. Il est donc possible que les embryons se soient comportés différemment les uns des autres au niveau des événements de recombinaison. Par contre, il est aussi possible que cette variabilité provienne du fait que le nombre de copies est très différent entre les souris. Ainsi, les intégrations à bas nombre de copies pourraient ne pas refléter la population de plasmides recombinants présente dans la cellule au moment de l'intégration simplement parce que le nombre limité de plasmides intégrés résulte en un faible échantillonnage statistique de cette population initiale. Cependant, lorsque l'analyse est reprise en utilisant seulement les souris fondatrices à moyen et haut nombre de copies (4419, 4454, 4465, 4469, 4472, 4475, 4479 et 4486), il demeure évident que certains embryons furent beaucoup plus actifs en recombinaison que d'autres. Par exemple, les souris 4475, 4479 et 4486 possèdent relativement peu de plasmides recombinants intégrés (entre 83% et 92,8% sont non-recombinants), alors que les souris 4469 et 4472 possèdent seulement 32% et 41% de non-recombinants respectivement. Nous en concluons donc que les embryons des différentes souris possédaient des populations de plasmides de compositions différentes au moment de l'intégration de ceux-ci.

Pour obtenir une vue d'ensemble des activités de recombinaison à l'œuvre dans l'embryon au moment de l'intégration des plasmides, nous avons multiplié le nombre de copies moyen trouvé au Tableau 4 par la fréquence des différents types de produits de recombinaison pour chaque fondateur. Les résultats de ces multiplications pour chaque souris furent ensuite additionnés par catégorie de produits (Figure 16).



**Figure 16 : Configuration agglomérée des plasmides intégrés chez les fondateurs**

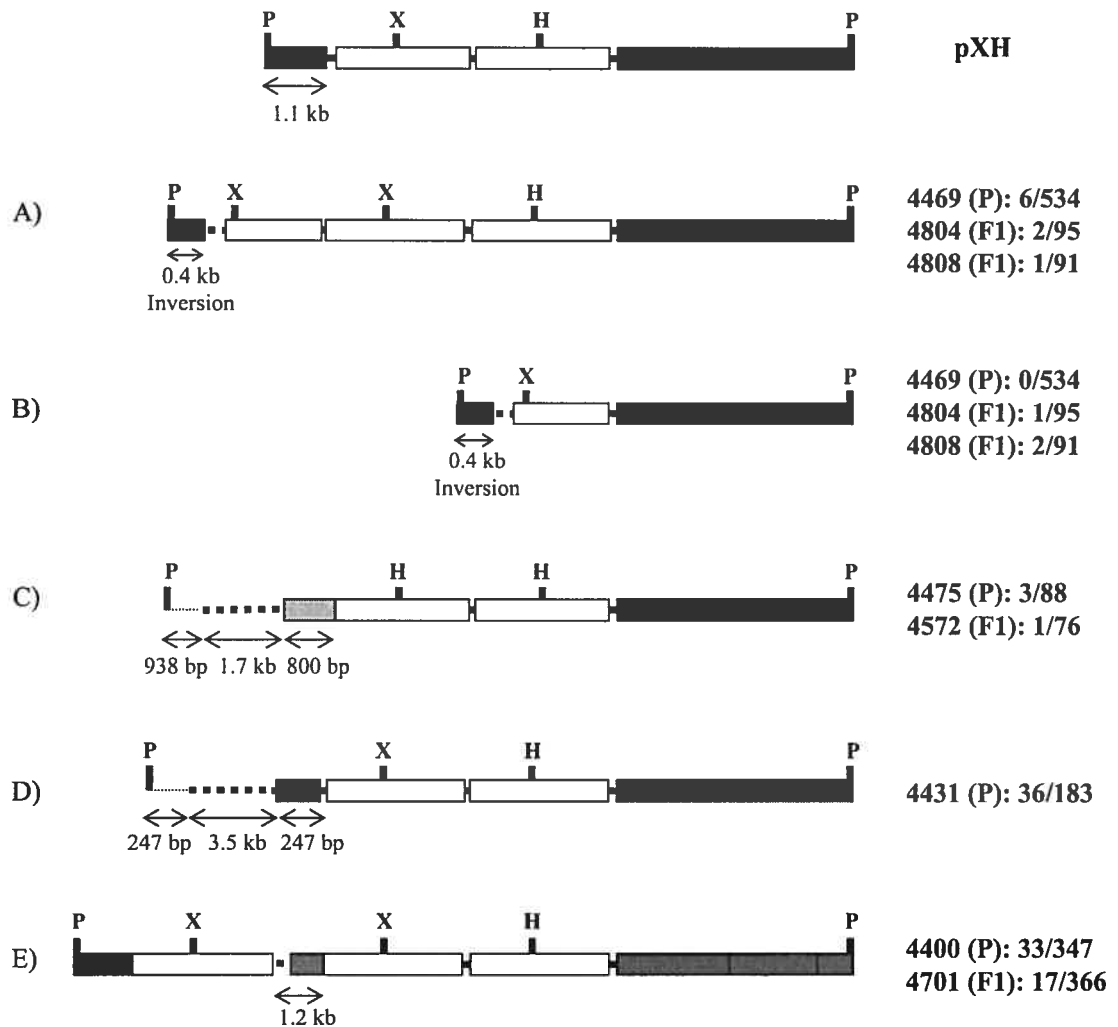
Le nombre de copies moyen de chaque souris (Tableau 4) fut multiplié par la fréquence des différentes configurations de plasmides retrouvée dans cette souris. Les résultats de toutes les souris furent ensuite additionnés par catégorie pour obtenir une image globale des activités de recombinaison agissant sur les plasmides avant ou pendant leur intégration dans l'embryon. Les catégories de produits de recombinaison sont décrites en Figure 14.

Environ 50% de tous les plasmides intégrés ont été impliqués dans des événements de recombinaison homologue alors que seuls 2% furent impliqués dans des événements de recombinaison non-homologue. Les événements de délétions homologues sont 5 fois plus fréquents que les événements de conversion génique sans crossing-over (38% vs 8%) dans notre essai. De façon concordante avec ces observations, le Tableau 3 montre que seules deux souris fondatrices (4400 et 4405) ont une plus grande incidence de conversion génique sans crossing-over que de délétions homologues alors que 11 d'entre elles sont dans la situation inverse. Finalement, deux souris fondatrices ont des fréquences égales de délétions homologues et de conversion génique sans crossing-over. La distribution des événements parmi les différentes souris est donc similaire à celle observée chez la souris « moyenne » représentée en Figure 16.

## Caractérisation des réarrangements de plasmides

Des plasmides réarrangés ne présentant pas une des structures attendues pour un événement de recombinaison homologue (classifiés comme « Autres » dans le Tableau 3 et la Figure 14) furent retrouvés pour la plupart des fondateurs. Cinq réarrangements sont d'un intérêt particulier et sont présentés en Figure 17. Les plasmides réarrangés obtenus à partir des souris 4400, 4701 et 4431 furent utilisés comme valeur-étalon lors de la détermination du nombre de copies des autres souris (Figure 17D et E). Les réarrangements récupérés à partir de la souris 4469 et de deux de ses descendants sont présentés en Figure 17A et B. Le réarrangement en Figure 17A fut retrouvé dans les trois souris. Il contient 3 fragments de L1, mais celui de gauche est incomplet. Une partie des séquences plasmidiques a subi une inversion (entre le site de restriction PstI et le fragment de L1 le plus à gauche). Le réarrangement de la Figure 17B semble être un descendant du réarrangement en A : il possède la même inversion caractéristique mais ne possède que le fragment de L1 incomplet à gauche du réarrangement en A. Ce réarrangement ne fut pas retrouvé à partir de 4469, mais fut retrouvé dans les descendants F1 de cette souris fondatrice (4804 et 4808). La relation la plus probable entre ces différents réarrangements est que le réarrangement en A ait donné naissance à celui en B en participant à un événement de délétion homologue impliquant le fragment de L1 le plus à droite et le fragment incomplet le plus à gauche, résultant en une délétion du fragment de L1 central. Le fait que les deux réarrangements (A et B) existent dans les souris F1 suggère qu'il existait plus d'une copie du réarrangement A dans la souris fondatrice, et que ce réarrangement a été amplifié avant ou pendant l'intégration du plasmide. Ceci suggère donc que les copies intégrées peuvent être modifiées par recombinaison homologue au fil des générations.





### Figure 17 : Réarrangements d'intérêt particulier

Les lignes pointillées épaisses indiquent des régions non-séquencées. Les lignes pointillées minces indiquent des séquences d'ADN génomique de souris. P: Site de restriction PstI. X: Site de restriction XhoI. H: Site de restriction HindIII. Pour faciliter la comparaison, le plasmide pXH non-modifié est illustré. La fréquence de récupération par *plasmid rescue* de ces réarrangements dans les différentes souris est indiquée. Pour le réarrangement présenté en c), la boîte gris pâle représente des séquences L1 endogènes.

### **Implication des L1 endogènes dans certains événements de recombinaison homologues dans l'embryon de souris**

Le réarrangement provenant de la souris 4475 (qui a aussi été récupéré à partir de la souris 4572, son descendant F1) comprend une jonction homologue entre le fragment L1 de gauche du pXH et un L1 endogène (Figure 17C). La séquence d'ADN à proximité du site de *plasmid rescue* PstI fut déterminée et comparée aux bases de données du génome. Le site d'intégration de ce réarrangement est situé sur le chromosome 8 et le L1 endogène impliqué à ce site est 97,5% homologue au fragment de L1 du pXH. Il est intéressant de noter que la séquence L1 situées en aval de la jonction homologue (à l'intérieur du L1 original du pXH) ne possède aucune modification par rapport à la séquence attendue, à l'exception de la modification du site XhoI en site HindIII. Le L1 endogène possède un site HindIII, mais sa séquence présente 16 bases de différence en comparaison du L1 du pXH (entre la jonction homologue et le site HindIII). Il est donc peu probable que le site HindIII retrouvé dans ce réarrangement provienne du L1 endogène. Le scénario le plus plausible voudrait donc que le plasmide en Figure 17C ait été impliqué dans un événement de recombinaison intra ou inter-plasmide pXH avant de recombiner avec le L1 endogène. Le mécanisme permettant d'expliquer le plus simplement cet événement est le *one-sided invasion*, comme nous en discuterons plus loin.

### **Modification des plasmides intégrés par recombinaison homologue au fil des générations de souris transgéniques**

Nous avons vérifié si les plasmides intégrés pouvaient être impliqués dans des événements de recombinaison homologue ou illégitime dans les cellules germinales des fondateurs ou dans les embryons des descendants. De tels événements donneraient naissance à des souris descendantes possédant des variations dans leurs distributions de fréquence de configuration de plasmides intégrés. Les distributions de fréquence des divers types de recombinant des fondateurs furent comparées avec les distributions des descendants en utilisant le test Chi carré avec la correction de Bonferonni (niveau de confiance de 95%). Ce test a montré que 4 des 5 lignées transgéniques montraient des différences significatives dans la composition des intégrations de plasmides en tandem entre les fondateurs et leur descendance (Tableau 5).

| Fondateur | P vs F          | F vs F       |
|-----------|-----------------|--------------|
| 4400      | $p < 0,0001$    | NA           |
| 4672      | $p < 8,000E-04$ | NA           |
| 4469      | $p = 0,0043$    | $p = 0,0158$ |
| 4472      | $p = 0,2613$    | ND           |
| 4475      | $p = 0,0054$    | $p = 0,4586$ |

**Tableau 5 : Comparaison statistique des distributions de fréquences des différents membres des lignées de souris transgéniques**

L'approximation de Monte Carlo du test Exact fut utilisée au lieu du test de Chi-carré de Pearson. P vs F : Comparaison de tous les membres de la lignée, incluant le fondateur. F vs F : Comparaison des descendants entre eux. NA : Ne s'applique pas. ND : Non-déterminé.

Une des raisons expliquant les différences entre les membres d'une lignée transgénique pourrait être la présence de plusieurs sites d'intégration dans le fondateur, qui seraient ségrégués différenciellement dans les descendants. Cette situation pourrait avoir eu lieu dans la lignée provenant du fondateur 4475. En effet, dans cette lignée les descendants sont similaires entre eux mais sont légèrement différents du fondateur. Dans la lignée 4401, il y a une importante augmentation (7 fois) dans le nombre de copies du descendant F1 par rapport au fondateur (Tableau 4). Ceci pourrait être dû à du mosaïcisme dans le fondateur résultant en la présence de différents sites d'intégration en tandem dans une fraction des cellules ou des tissus. Alternativement, la structure intégrée en tandem a pu être amplifiée durant le développement embryonnaire du descendant ou dans les cellules germinales du fondateur par des erreurs de réplication ou par un échange inégal avec la chromatide sœur. Dans le cas de la lignée provenant du fondateur 4469, une analyse par Southern avec une sonde complémentaire aux séquences plasmidiques du pXH a indiqué qu'il y avait une différence dans le patron de bandes obtenues pour le descendant 4770 par rapport aux autres membres de cette lignée (résultats non-montrés). Une comparaison statistique des distributions de fréquences du fondateur et du descendant 4770 montre d'ailleurs qu'elles sont significativement différentes ( $p = 2,000E-04$ ). Ceci suggère que cet individu possède une structure intégrée remodelée et qu'il est probablement responsable des différences statistiques existant entre les membres de cette lignée.

Nous avons reconstruit statistiquement la structure intégrée en tandem pour la souris 4400 et son descendant 4701 en générant des intervalles de confiance à 95% autour des valeurs des distributions de fréquence (Tableau 6). Cette reconstruction nous a indiqué que les délétions homologues (PX ou PH) étaient statistiquement absentes de la structure intégrée du fondateur 4400 (test chi carré de Pearson avec correction de Bonferroni), alors que le descendant F1 4701 présente une copie intégrée de chaque type de délétion. En effet, en multipliant le nombre de copies des souris par les proportions indiquées au Tableau 6, nous pouvons conclure à l'absence statistique de délétions dans 4400 et à leur présence dans la souris 4701. Ceci pourrait être expliqué si la souris 4400 était une mosaïque et que 4701 avait hérité d'une structure intégrée différente de celle qui est présente dans la queue du fondateur : c'est-à-dire que l'intégration en tandem présente dans la queue du fondateur était différente de celle présente dans ses cellules germinales. Cependant, dans ces deux souris, un réarrangement unique fut récupéré (Figure 17E). Ceci démontre que la structure intégrée présente dans la queue du descendant était de la même origine que celle présente dans la queue du fondateur. Étant donné que les deux souris ont une structure intégrée de la même origine, l'explication la plus probable de la genèse de ces deux délétions homologues est que des événements de recombinaison homologue se sont produits entre les plasmides intégrés dans les cellules germinales du fondateur ou dans l'embryon du descendant F1. Ces résultats suggèrent donc que les L1 intégrés peuvent être impliqués dans des événements de recombinaison homologue, et que ces modifications peuvent mener à des changements héréditaires du génome.

|          | PXHP  |       | PXXP  |       | PPHP |     | PXP  |      | PHP  |      | PXHHP |      | PXXHP |       | Autres |       |
|----------|-------|-------|-------|-------|------|-----|------|------|------|------|-------|------|-------|-------|--------|-------|
|          | Inf   | Sup   | Inf   | Sup   | Inf  | Sup | Inf  | Sup  | Inf  | Sup  | Inf   | Sup  | Inf   | Sup   | Inf    | Sup   |
| 4400 (P) | 32,3% | 46,7% | 35,4% | 49,9% | NA   | NA  | 0,0% | 1,1% | 0,0% | 1,1% | NA    | NA   | 3,8%  | 11,7% | 5,2%   | 13,8% |
| 4701(F1) | 31,3% | 45,2% | 31,6% | 45,5% | NA   | NA  | 2,0% | 8,4% | 2,2% | 8,7% | 0,0%  | 1,6% | 3,6%  | 11,1% | 1,6%   | 7,7%  |

**Tableau 6 : Intervalles de confiance autour des distributions de fréquences des souris 4400 et 4701**

Le test utilisé est le Chi carré de Pearson avec correction de Bonferonni pour comparaisons multiples. L'intervalle de confiance est de 95%. Inf : Limite inférieure de l'intervalle. Sup : Limite supérieure de l'intervalle. Le nombre de copies pour ce test statistique fut fixé à 15 (moyenne entre les nombres de copies déterminés pour chacune des souris au Tableau 4) : il y a donc absence statistique des configurations PXP et PHP du fondateur, alors que ces configurations sont présentes dans le descendant.

## **2. Lien entre l'environnement spatial d'un locus et le choix de ses partenaires de recombinaison**

### **Objectifs**

La recombinaison entre séquences homologues L1 intégrées qui sont en proximité linéaire les unes des autres peut conduire à des modifications héréditaires du génome, et semble être particulièrement active dans l'embryon de souris. Ces observations suggèrent que les éléments L1 chromosomiques du génome pourraient être impliqués dans des événements de recombinaison dans l'embryon et/ou dans les cellules germinales, et que ces interactions pourraient avoir des conséquences évolutives importantes sur l'organisation du génome. Cependant, le système que nous avons utilisé était artificiel, car il juxtaposait les séquences homologues d'intérêt. Or, comme nous l'avons vu, la proximité spatiale entre loci semble pouvoir favoriser leurs interactions par recombinaison (Kozubek et al. 1999; Kozubek et al. 1997; Nikiforova et al. 2000; Parada et al. 2004) : la présence de répétitions directes représente donc certainement un facteur facilitant pour ces événements.

Dans un contexte plus naturel, il est probable que l'accessibilité préférentielle d'un locus à certaines régions du génome, à cause de l'organisation spatiale de la chromatine de ce locus, puisse influencer le choix de ses partenaires de recombinaison. Notre laboratoire a

récemment publié une étude démontrant que le choix des partenaires de recombinaison d'un locus particulier n'était pas aléatoire chez l'humain, et qu'un locus donné possédait certains partenaires de recombinaison préférentiels (D'Anjou et al. 2004). En effet, une brisure double-brin introduite dans un L1 marqué peut théoriquement être réparée par recombinaison homologue avec une panoplie d'autres séquences L1 éparpillées à travers le génome. Malgré ceci, certains L1 spécifiques furent préférentiellement utilisés pour réparer le L1 en question. La plupart de ces séquences étaient situés à l'intérieur d'une distance génomique de 1 Mb du L1 endommagé. Les séquences plus distantes étaient généralement situées dans le même bras chromosomique que le L1 brisé, alors que l'autre bras chromosomique ne semblait pas plus accessible que d'autres régions inter-chromosomiques.

Il était donc possible que ce choix de partenaires de recombinaison reflète une accessibilité spatiale plus aisée du locus endommagé à ces régions génomiques particulières. Selon cette hypothèse, les loci les plus fréquemment à proximité spatiale du locus brisé et formant son environnement spatial immédiat devraient être favorisés lors des événements de recombinaison. Pour vérifier cette hypothèse, nous avons décidé d'investiguer l'influence de l'environnement spatial du locus utilisé dans l'étude citée ci-haut (D'Anjou et al. 2004) sur le choix de ses partenaires de recombinaison. Les objectifs de cette partie étaient les suivants :

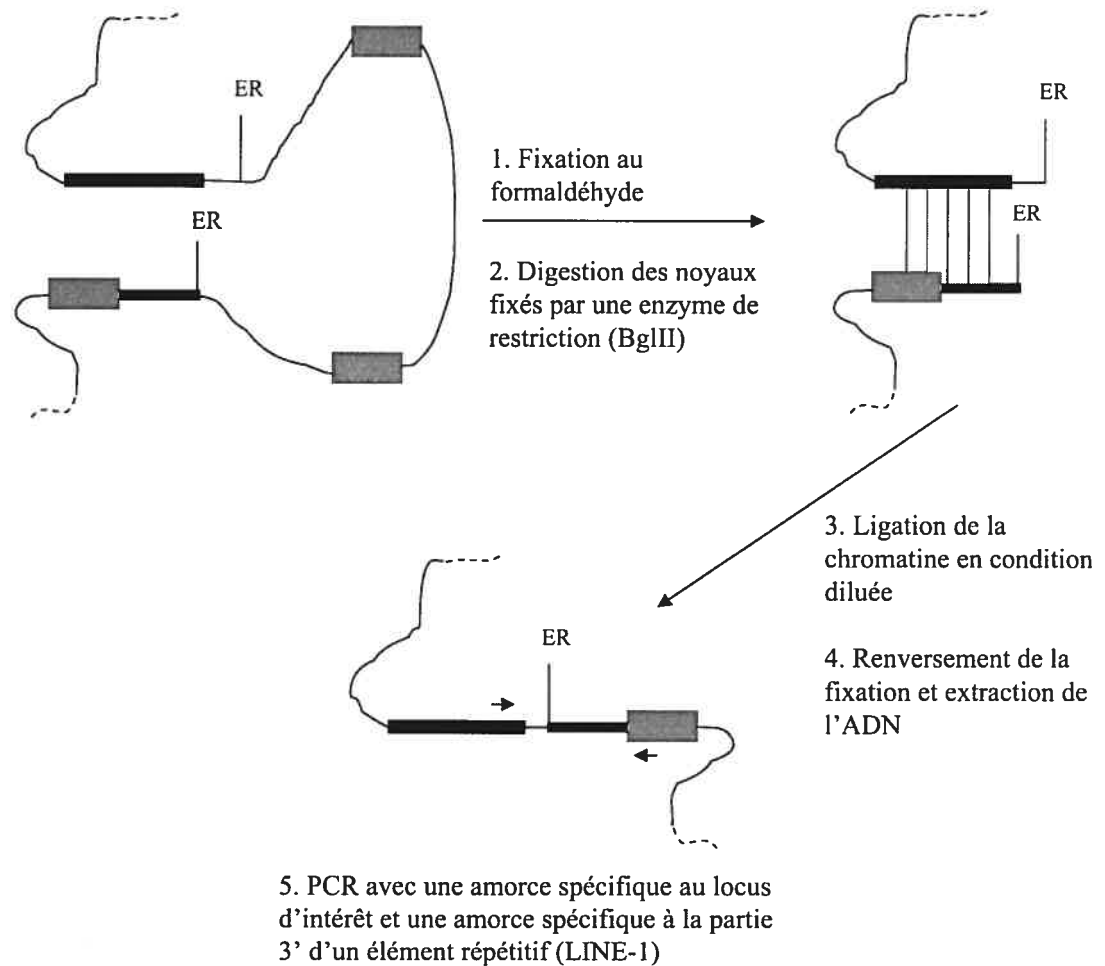
-Développer une méthode basée sur la méthodologie du *Chromosome Conformation Capture* (3C) (Dekker et al. 2002) qui permettrait de décrire l'environnement spatial génomique d'un locus donné sans a priori, en identifiant les loci qui sont le plus fréquemment à proximité spatiale du locus d'intérêt dans les noyaux de cellules humaines

-Utiliser cette méthodologie pour décrire l'environnement spatial génomique d'un des locus caractérisé par un essai de recombinaison entre séquences L1 (situé sur le chromosome 2p22.1 humain) (D'Anjou et al. 2004), de façon à pouvoir investiguer le lien entre la fréquence à laquelle certaines régions génomiques sont à proximité spatiale d'un locus donné et leur accessibilité à ce locus lors d'événements de recombinaison

## Design expérimental

Le but de cette étude était d'identifier quels loci sont fréquemment à proximité d'un locus génomique d'intérêt dans l'espace nucléaire de cellules humaines en croissance. Pour atteindre ce but, nous avons développé une modification de la méthodologie du 3C (Dekker et al. 2002). Cette modification fut baptisée *Repetitive-Element-Mediated Chromosome Conformation Capture* (REM-3C; Figure 18). Dans cette méthode, des cellules sont traitées avec une solution de formaldéhyde, ce qui fixe chimiquement les loci génomiques qui sont à proximité spatiale les uns des autres dans l'espace nucléaire. Les noyaux fixés sont digérés avec une enzyme de restriction appropriée puis sont traités par une ligase en conditions diluées. Cette dilution favorise les ligations entre molécules fixées par le formaldéhyde plutôt que les ligations inter-moléculaires aléatoires. La fréquence de ligation de deux loci est donc proportionnelle à la fréquence à laquelle ils sont à proximité l'un de l'autre dans l'espace nucléaire. Finalement, la fixation est renversée, l'ADN est extrait et les produits de ligation sont analysés par PCR.

Dans la procédure 3C originale, l'abondance de certains produits de ligation spécifiques est déterminée par PCR quantitatif en utilisant des amorces complémentaires aux loci d'intérêt. Nous avons voulu rendre cette étape de PCR plus ouverte, de façon à pouvoir déterminer quels loci sont à proximité d'un locus d'intérêt, mais sans présumer de leur identité. La méthodologie du REM-3C permet ce type d'analyse : l'étape de PCR utilise à cet effet une amorce spécifique au locus d'intérêt et une amorce complémentaire à la séquence consensus d'une séquence répétitive. Étant donné que les séquences répétitives sont distribuées dans le génome, une grande proportion des produits de ligation générés par la procédure de 3C devrait pouvoir être amplifiée par cette méthode. Les produits de PCR sont ensuite clonés et séquencés, puis leur localisation génomique est déterminée par comparaison avec les banques de données du génome. Un locus cloné fréquemment à partir de plusieurs expériences de REM-3C est donc fréquemment à proximité spatiale du locus d'intérêt.



### Figure 18 : La méthode de REM-3C

La méthode est basée sur la technologie du 3C développée à l'origine par Dekker et collaborateurs (Dekker et al. 2002). Boîtes grises : séquences répétitives (dans notre cas, séquences répétitives LINE-1). Boîtes noires : loci qui sont à proximité spatiale les uns des autres. ER : Site de restriction (BglII dans notre cas)

Un des locus analysé dans l'étude sur l'accessibilité des sites de recombinaison à un L1 marqué fut choisi pour l'étude par REM-3C de son environnement spatial (D'Anjou et al. 2004). Il s'agit d'un locus situé en position 40872835 du chromosome 2p22.1 humain. Deux lignées cellulaires différentes furent utilisées : un clone (ATA67) dérivé de la lignée de fibroblastes humains déficients en ATM GM005849 et la lignée de fibroblastes de poumons fœtaux humain non-immortalisés IMR90. La lignée ATA67 fut utilisée dans l'étude de recombinaison : à cet effet, un des deux allèles de la région investiguée en



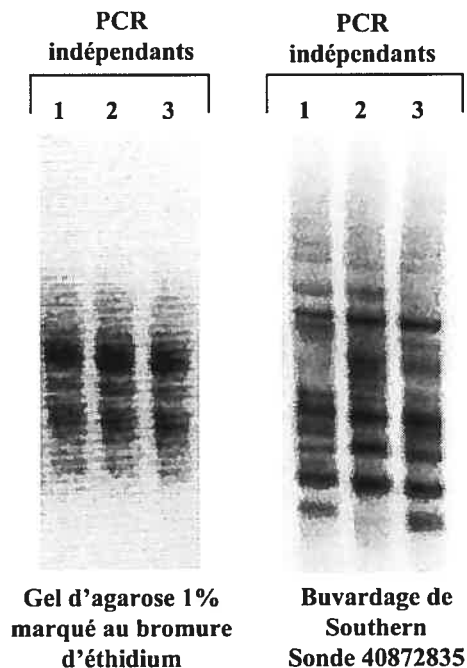
2p22.1 fut modifié par l'intégration d'un plasmide contenant un site de restriction unique permettant l'introduction d'une brisure double-brin. Les deux allèles du site choisi peuvent donc être distingués dans cette lignée. Par sa part, la lignée IMR90 fut utilisée pour pouvoir comparer l'organisation du locus d'intérêt dans une situation « normale » avec une situation « pathologique » représentée par les cellules déficientes en ATM.

Les expériences de REM-3C furent réalisées comme suit. Les cellules en croissance furent récoltées et traitées au formaldéhyde selon le protocole publié par Tolhuis et collaborateurs (Tolhuis et al. 2002). Les noyaux furent digérés par l'enzyme de restriction BglII et la chromatine fut ligaturée. L'étape de PCR fut réalisée tel que décrit dans la section Matériel et méthode. Brièvement, une amorce qui est située à environ 350 pb en amont du site BglII du locus en 2p22.1 fut utilisée pour réaliser 70 cycles d'élongation. Un aliquot de cette réaction d'élongation fut ensuite utilisé pour un PCR en utilisant une amorce interne située à environ 300 pb du site BglII et une amorce spécifique à la séquence consensus des éléments L1. Cette stratégie de PCR fut baptisée REM-PCR et est nécessaire à l'obtention de produits PCR spécifiques en utilisant une amorce spécifique à un locus donné et une amorce complémentaire à un élément répétitif. En effet, ce type de PCR résulte habituellement en une amplification peu productive de la part de l'amorce spécifique au locus d'intérêt étant donné que l'amorce L1 génère d'elle-même plusieurs produits à partir de sites génomiques dans lesquels deux séquences L1 sont en orientations opposées. Ces produits sont nombreux et prennent le dessus sur ceux provenant de l'utilisation des deux amorces. Le fait de réaliser une amplification linéaire à partir du locus d'intérêt (l'étape d'élongation) augmente la quantité de matrice d'intérêt par rapport à l'ADN génomique responsable du bruit de fond problématique. Ceci permet donc l'amplification de produits provenant de l'utilisation des amorces L1 et 2p22.1. Des méthodes similaires ont d'ailleurs déjà été utilisées pour amplifier les séquences génomiques inconnues flanquant des sites d'insertion rétrovirale ou de transgènes (Minami et al. 1995; Nelson et al. 1991; Sonza et al. 1996). Une fois cette amplification réalisée, les produits obtenus peuvent être clonés.

### **Spécificité des produits de REM-3C**

Les résultats d'une expérience de REM-3C typique sont présentés en Figure 19. Comme nous pouvions nous y attendre, les produits de REM-PCR sont nombreux et forment une

trainée sur un gel d'agarose : plusieurs produits de ligation distincts peuvent donc être amplifiés par cette méthode. Pour vérifier lesquels de ces produits sont d'authentiques produits REM-PCR (par rapport à des produits inter-L1 qui doivent être considérés comme du bruit de fond), un buvardage de Southern fut réalisé sur trois REM-3C indépendants (réalisés sur le même ADN de départ). La sonde utilisée est un fragment d'ADN génomique correspondant aux 300 pb situés entre l'amorce interne du chromosome 2 : 40872835 et le site BglII utilisé. Ce fragment doit donc par définition se retrouver dans tous les produits REM-3C spécifiques. Les résultats de l'expérience en Figure 19 démontrent que plusieurs produits de ligation spécifiques sont effectivement amplifiés dans plusieurs PCR indépendants, ce qui donne naissance à un « code-barre » de l'environnement spatial de ce locus. Étant donné que seulement deux produits de ligations peuvent être générés à partir d'une cellule donnée (une pour chaque allèle du locus en 2p22.1), ce résultat suggère fortement que certains produits de ligation particuliers furent générés dans plusieurs cellules de la population étudiée.



**Figure 19 : Un «code-barre» de l'environnement spatial du locus chr2: 40872835 dans la lignée cellulaire IMR90.**

Chaque puit représente un PCR indépendant réalisé à partir d'un même échantillon d'ADN traité par 3C. Plusieurs produits spécifiques sont obtenus reproductiblement à partir d'un même ADN traité par 3C.

Pour déterminer l'identité des produits de ligation obtenus par REM-3C, les produits de PCR furent clonés et séquencés. Étant donné que les produits inter-L1 sont beaucoup plus abondants que les produits REM-PCR spécifiques, les colonies bactériennes positives furent identifiées par « transfert de colonies bactériennes » ou *colony lift* suivi d'une hybridation *in situ* avec la même sonde que celle utilisée à la Figure 19. Les clones furent ensuite séquencés : les clones de REM-3C consistent en une première région provenant du chromosome 2 : 40872835 jusqu'au site BglII, suivie de séquences provenant d'une autre région génomique non-contigüe. Pour la lignée ATA67, les produits originant des deux allèles purent être distingués à cause des différences de séquences introduites par l'insertion d'un plasmide dans cette région. La localisation des partenaires de ligation récupérés est présentée au Tableau 7 et à la Figure 20.

Il est à noter que certains partenaires particuliers furent clonés plusieurs fois à partir d'un même PCR, mais qu'ils ne furent jamais reclonés par la suite à partir d'autres PCR réalisés sur le même échantillon d'ADN traité par 3C. Ceci suggère que le REM-PCR peut parfois mener à une sur-représentation fortuite de certains événements à l'intérieur d'une réaction donnée. C'est donc pour prévenir ce problème potentiel d'interprétation qu'un événement donné ne fut comptabilisé qu'une seule fois par PCR cloné. Par exemple, un partenaire comptabilisé cinq fois au Tableau 7 fut en fait récupéré dans cinq PCR indépendants; par contre, il a pu être récupéré plusieurs fois dans chacun de ces PCR. Étant donné que les événements récupérés à partir de PCR indépendants proviennent nécessairement de cellules distinctes, la récupération d'un même événement à partir de plusieurs PCR indépendants démontre que le locus en question fut à proximité du locus de départ dans plusieurs cellules indépendantes de la population cellulaire étudiée.

| Position                                  | Bande   | Distance (Mb) <sup>a</sup> | PCR IMR90 | PCR ATA67n | PCR ATA67m | Taille PCR (bp) <sup>b</sup> | Amorce <sup>c</sup> | Rec ATA67 (D'anjou et al. 2004) |
|---|---------|----------------------------|-----------|------------|------------|------------------------------|---------------------|---------------------------------|
| <b>Chromosome 2 (Intra-chromosomique)</b> |         |                            |           |            |            |                              |                     |                                 |
| 10027523                                  | 2p25.1  | 30,85                      |           | 1          |            | 582                          | 23/26               |                                 |
| 11828583                                  | 2p25.1  | 29,04                      |           | 1          |            | 1183                         | 25/26               |                                 |
| 14672835                                  | 2p24.3  | 26,20                      |           |            |            |                              |                     | 1 (67-t28)                      |
| 15572835                                  | 2p24.3  | 25,30                      |           |            |            |                              |                     | 1 (67-t27)                      |
| 18121659                                  | 2p24.2  | 22,75                      |           | 1          |            | 866                          | 24/26               |                                 |
| 19059605                                  | 2p24.2  | 21,81                      |           | 1          |            | 850                          | 24/26               |                                 |
| 23272835                                  | 2p24.1  | 17,60                      |           |            |            |                              |                     | 1 (67-t21)                      |
| 25172835                                  | 2p23.3  | 15,70                      |           |            |            |                              |                     | 1 (67-t20)                      |
| 28672835                                  | 2p23.2  | 12,20                      |           |            |            |                              |                     | 1 (67-t19)                      |
| 28972835                                  | 2p23.2  | 11,90                      |           |            |            |                              |                     | 1 (67-t17)                      |
| 29272637                                  | 2p23.2  | 11,60                      |           |            |            |                              |                     | 12 (67-D)                       |
| 29372835                                  | 2p23.2  | 11,50                      |           |            |            |                              |                     | 1 (67-t16)                      |
| 39936915                                  | 2p22.1  | 0,94                       | 5         | 4          | 1          | 813                          | 26/26               |                                 |
| 40075863                                  | 2p22.1  | 0,80                       | 3         |            | 2          | 1063                         | 23/26               |                                 |
| 40159049                                  | 2p22.1  | 0,71                       |           |            |            |                              |                     | 13 (67-A)                       |
| 40169994                                  | 2p22.1  | 0,70                       |           |            |            |                              |                     | 1 (67-E)                        |
| 40199469                                  | 2p22.1  | 0,67                       |           |            |            |                              |                     | 27 (67-B)                       |
| 40205376                                  | 2p22.1  | 0,67                       | 3         | 1          |            | 861                          | 22/26               |                                 |
| 40209246                                  | 2p22.1  | 0,66                       |           |            |            |                              |                     | 4 (67-C)                        |
| 40219835                                  | 2p22.1  | 0,65                       |           |            |            |                              |                     | 1 (67-t22)                      |
| 40280835                                  | 2p22.1  | 0,59                       |           |            |            |                              |                     | 1 (67-t23)                      |
| 40283835                                  | 2p22.1  | 0,59                       |           |            |            |                              |                     | 1 (67-t24)                      |
| 40403114                                  | 2p22.1  | 0,47                       | 3         | 1          |            | 1124                         | 24/26               |                                 |
| 40872835                                  | 2p22.1  | <i>(Locus de départ)</i>   |           |            |            |                              |                     |                                 |
| 41413565                                  | 2p22.1  | 0,54                       |           |            |            | 1171                         | 26/26               |                                 |
| 41742111                                  | 2p21    | 0,87                       | 3         | 2          |            | 665                          | 26/26               |                                 |
| 41894780                                  | 2p21    | 1,02                       | 1         | 1          |            | 570                          | 17/26               |                                 |
| 45712835                                  | 2p21    | 4,84                       |           |            |            |                              |                     | 1 (67-t26)                      |
| 48815177                                  | 2p16.3  | 7,94                       | 1         |            |            | 477                          | 22/26               |                                 |
| 49514242                                  | 2p16.3  | 8,64                       | 1         |            |            | 996                          | 25/26               |                                 |
| 52374128                                  | 2p16.3  | 11,50                      | 1         |            |            | 978                          | 26/26               |                                 |
| 52413953                                  | 2p16.3  | 11,54                      |           | 1          |            | 668                          | 26/26               |                                 |
| 54490069                                  | 2p16.2  | 13,62                      | 1         |            |            | 466                          | 21/26               |                                 |
| 62888313                                  | 2p15    | 22,02                      |           | 1          |            | 894                          | 26/26               |                                 |
| 63715570                                  | 2p15    | 22,84                      |           |            |            |                              |                     | 1 (67-M)                        |
| 66486209                                  | 2p14    | 25,61                      | 1         |            |            | 1693                         | 26/26               |                                 |
| 79866900                                  | 2p12    | 38,99                      | 1         |            |            | 933                          | 25/26               |                                 |
| 96528546                                  | 2q11.2  | 55,66                      | 1         |            |            | 580                          | 26/26               |                                 |
| 159121119                                 | 2q24.1  | 118,25                     | 1         |            |            | 1277                         | 26/26               |                                 |
| 161763925                                 | 2q24.2  | 120,89                     | 1         |            |            | 1158                         | 26/26               |                                 |
| <b>Inter-Chromosomique</b>                |         |                            |           |            |            |                              |                     |                                 |
| chr4: 74596698                            | 4q13.3  | -                          |           | 1          |            | 897                          | 25/26               |                                 |
| chr5: 57500786                            | 5q11.2  | -                          |           | 1          |            | 1557                         | 25/26               |                                 |
| chr6: 62575157                            | 6q11.1  | -                          |           | 1          |            | 953                          | 22/26               |                                 |
| chr6: 80349723                            | 6q14.1  | -                          | 1         |            |            | 965                          | 26/26               |                                 |
| chr7: 91113943                            | 7q21.2  | -                          |           |            | 1          | 1068                         | 26/26               |                                 |
| chr7: 111652604                           | 7q31.1  | -                          |           | 1          |            | 869                          | 26/26               |                                 |
| chr7: 115648342                           | 7q31.2  | -                          |           | 1          |            | 888                          | 26/26               |                                 |
| chr8: 57890675                            | 8q12.1  | -                          | 1         |            |            | 863                          | 26/26               |                                 |
| chr9: 98733129                            | 9q22.33 | -                          |           | 1          |            | 889                          | 25/26               |                                 |
| chr13: 31051340                           | 13q12.3 | -                          |           | 1          |            | 1115                         | 25/26               |                                 |
| chr14: 69841764                           | 14q24.2 | -                          |           |            | 1          | 1158                         | 24/26               |                                 |
| chrX: 131805807                           | Xq26.2  | -                          | 1         |            |            | 892                          | 22/26               |                                 |

### Tableau 7 : Partenaires REM-3C

La numérotation des positions commence au télomère du bras p du chromosome.

Rangées grises: Partenaires préférentiels qui furent récupérés dans plusieurs REM-PCR indépendants.

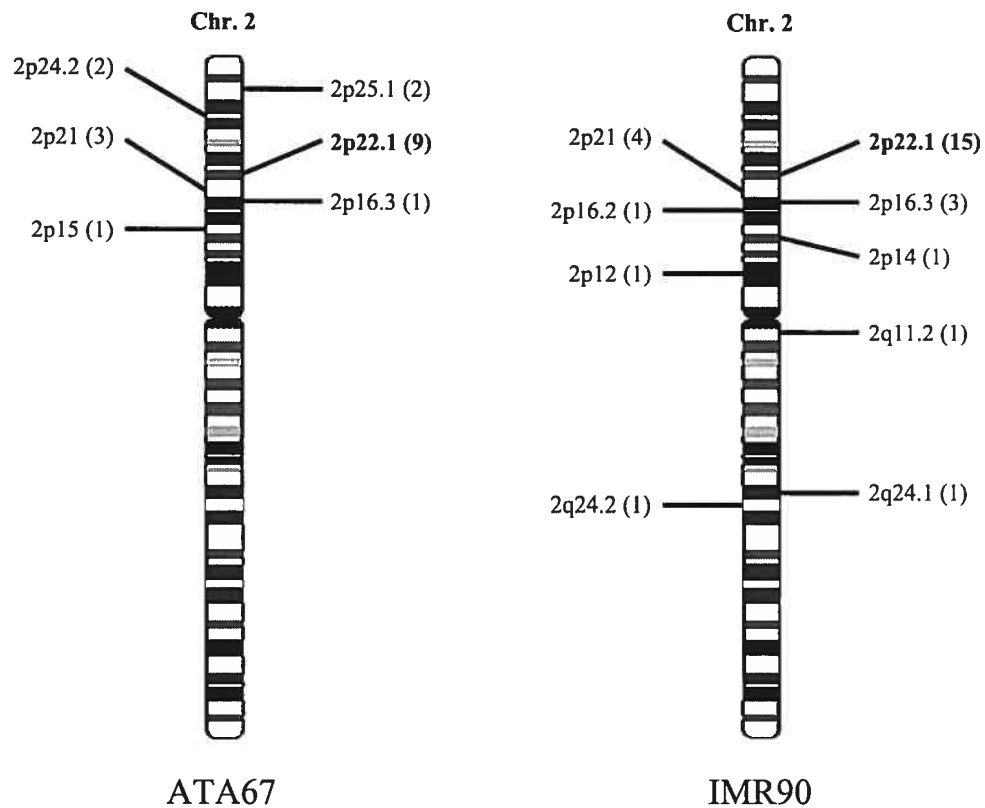
PCR IMR90, PCR ATA67n, PCR ATA67m: Nombre de REM-PCR indépendants à partir desquels un partenaire donné a été cloné. Le partenaire fut retrouvé ligaturé aux allèles de la lignée IMR90, à l'allèle naturel de ATA67 et à l'allèle modifié de ATA67, respectivement.

<sup>a</sup>: Distance génomique entre un partenaire et le locus de départ

<sup>b</sup>: Taille des produits de REM-PCR clonés

<sup>c</sup>: Nombre de bases complémentaires entre l'amorce L1 utilisée dans le REM-PCR et la séquence génomique du partenaire correspondant.

Rec ATA67 (D'Anjou et al. 2004) : Nom donné dans l'article de D'Anjou et al. et nombre d'événements de recombinaison entre ce locus génomique et le locus de départ dans la lignée ATA67. Les recombinants contenant un t (par exemple, 67-t26) sont des translocations provenant de recombinaison illégitime alors que les recombinants identifiés par une lettre (par exemple 67-E) sont des recombinants homologues.



**Figure 20 : Distribution des partenaires de REM-3C intra-chromosomiques.**

Les nombres entre parenthèse indiquent le nombre d'événements indépendants retrouvés dans une bande chromosomique particulière. La bande 2p22.1 (en caractères gras) contient le locus de départ.

De façon concordante avec les résultats présentés en Figure 19, plusieurs partenaires distincts furent récupérés par la méthodologie de REM-3C. Les partenaires intra-chromosomiques (46 événements indépendants) furent récupérés plus fréquemment que les partenaires inter-chromosomiques, tel qu'attendu selon les modèles de territoires chromosomiques (Cremer et Cremer 2001). D'autre part, il existe une nette préférence pour le bras chromosomique du locus de départ : seuls 3 événements ont impliqué l'autre bras. Ces résultats sont compatibles avec des données de FISH montrant que les bras chromosomiques occupent des territoires distincts dans l'espace nucléaire (Dietzel et al. 1998), et avec les résultats de recombinaison de d'Anjou et collaborateurs (D'Anjou et al. 2004). Les résultats montrent aussi que 31 événements sur 46 proviennent d'événements de ligation avec des loci situés à moins de 1 Mb de chaque côté du locus de départ. Les partenaires préférentiels (récupérés à partir de plus d'un PCR indépendant) sont d'ailleurs

tous situés à l'intérieur de ce domaine de 2 Mb. Les positions des partenaires de recombinaison identifiés par l'étude de D'Anjou et collaborateurs sont aussi indiquées : un grand nombre de loci ayant participé à des événements de recombinaison sont dans la région génomique couvrant les positions 40403114 à 39936915, qui comprend entre autres la majorité des partenaires préférentiels. Ceci suggère que l'accessibilité à cette région a effectivement pu influencer le choix des partenaires de recombinaison du locus investigué. De plus, la majorité des partenaires de recombinaison sont situés vers le télomère du chromosome, ce qui est aussi le cas des partenaires de fixation. Il convient aussi de noter que peu d'événements de REM-3C furent récupérés à partir de l'allèle modifié dans la lignée ATA67, et ce pour une raison inconnue.

Après une séparation génomique de 1 Mb du locus de départ, la fréquence de récupération de partenaires diminue considérablement. En fait, tous les partenaires plus distaux ne furent récupérés qu'une seule fois chacun, ce qui suggère que l'accessibilité à ces loci est plus ou moins aléatoire. Il est aussi notable que les premiers partenaires récupérés à l'extérieur de ce domaine génomique de 1 Mb de part et d'autre du locus de départ sont situés à des distances de 8 et 21 Mb. La distribution des partenaires de REM-3C dans les bandes chromosomiques du chromosome 2 est illustrée en Figure 20. Plus de la moitié des événements intra-chromosomiques de REM-3C récupérés (24 sur 46) impliquent des partenaires situés dans la même bande chromosomique que le locus de départ. De façon intéressante, bien que les partenaires de fixation situés dans cette bande chromosomique soient majoritairement situés en direction du télomère par rapport au locus de départ, les bandes chromosomiques situées dans cette direction ne furent que peu utilisées en comparaison des bandes situées vers le centromère dans la lignée IMR90, alors que les deux côtés semblent également utilisés pour la lignée ATA67. Ceci pourrait suggérer que les interactions plus distales entre le locus d'intérêt et son chromosome sont régulées différemment dans ces deux lignées cellulaires.

### **Distribution des partenaires de fixation intra-chromosomiques**

La distribution des partenaires situés à l'intérieur du domaine génomique proximal couvrant environ 1 Mb de chaque côté du site de départ fut investiguée avec plus de détails. Comme nous l'avons mentionné, les partenaires préférentiels de REM-3C sont tous situés à

l'intérieur de cette distance d'un côté ou de l'autre du locus de départ. Tous ces partenaires furent d'ailleurs récupérés à partir des deux lignées cellulaires étudiées, ce qui suggère que l'environnement spatial de ce locus possède des similarités dans ces deux lignées. Il est notable que ces partenaires préférentiels soient localisés à des distances de 500 kb à environ 1 Mb du locus de départ. Ceci pourrait être expliqué si aucun locus situé plus près du locus de départ ne présentait la configuration requise pour être amplifié par la méthode de REM-3C (homologie suffisante avec l'amorce L1, présence d'un site BglII, taille du produit de REM-PCR résultant comparable avec ceux des partenaires préférentiels). Nous avons donc vérifié si d'autres loci que ceux que nous avons récupérés rencontraient ces conditions à l'intérieur du domaine proximal. Le Tableau 8 présente donc les résultats d'une analyse informatique des régions couvrant environ 1 Mb de part et d'autre du locus d'intérêt.

Cette analyse montre que plusieurs partenaires potentiels sont présents dans cette région, et que certains d'entre eux sont plus rapprochés du locus de départ que certains partenaires préférentiels. Cependant, dans la plupart des cas la taille des produits de REM-PCR résultant de l'amplification de ces loci est plus grande que celle des produits obtenus pour la majorité des partenaires récupérés (voir aussi le Tableau 7) : l'amplification par PCR de ces partenaires est donc probablement défavorisée par rapport à ceux que nous avons récupérés. La faible fréquence de récupération de certains autres loci peut aussi s'expliquer par d'autres raisons : par exemple, le locus situé en position 39947030 ne fut pas récupéré malgré le fait que la taille du produit de REM-PCR correspondant soit comparable à celle des loci récupérés. Par contre, il possède une relativement faible homologie à l'amorce L1 (23/26), ce qui a potentiellement pu le rendre difficile à amplifier par notre méthodologie. Nous concluons donc que seuls certains loci peuvent être efficacement amplifiés par la stratégie de REM-3C, et qu'il existe probablement une taille limite d'environ 1 kb aux produits de REM-PCR, du moins dans les conditions de PCR que nous avons utilisées. Il nous est donc impossible de nous prononcer sur l'organisation spatiale interne du domaine génomique proximal couvrant 1 Mb de part et d'autre du locus de départ. Cependant, il semble clair que dans l'ensemble ces partenaires proximaux en terme de séquence linéaire sont favorisés par rapport à d'autres partenaires plus distants qui résultent en des produits de REM-PCR de taille semblable (Tableau 7). Notons aussi que nos résultats suggèrent qu'il n'y a pas de différences importantes dans la fréquence de fixation 3C entre un locus



situé à environ 500 kb du site de départ et un autre situé à 1 Mb (par exemple, 40403114 vs 39936915).

| Position<br>(Chr2)                | Amorce | PCR<br>(pb) | Distance<br>(Mb) |
|-----------------------------------|--------|-------------|------------------|
| 39936915                          | 26/26  | 813         | 0,94             |
| 39947030                          | 23/26  | 1039        | 0,93             |
| 40017798                          | 24/26  | 1778        | 0,86             |
| 40075863                          | 23/26  | 1063        | 0,80             |
| 40147931                          | 25/26  | 3390        | 0,72             |
| 40205376                          | 22/26  | 861         | 0,67             |
| 40403114                          | 25/26  | 1124        | 0,47             |
| 40552016                          | 25/26  | 2374        | 0,32             |
| 40564640                          | 25/26  | 2670        | 0,31             |
| 40820459                          | 23/26  | 1956        | 0,05             |
| 40846827                          | 24/26  | 2938        | 0,03             |
| <i>40872835 (Locus de départ)</i> |        |             |                  |
| 40921363                          | 26/26  | 1794        | 0,05             |
| 41227710                          | 23/26  | 1450        | 0,35             |
| 41233421                          | 25/26  | 2037        | 0,36             |
| 41251075                          | 23/26  | 3116        | 0,38             |
| 41364763                          | 25/26  | 2382        | 0,49             |
| 41413565                          | 26/26  | 1171        | 0,54             |
| 41480908                          | 25/26  | 2968        | 0,61             |
| 41572336                          | 26/26  | 2394        | 0,70             |
| 41588930                          | 23/26  | 2710        | 0,72             |
| 41656943                          | 25/26  | 2964        | 0,78             |
| 41742111                          | 26/26  | 665         | 0,87             |

**Tableau 8 : Partenaires potentiels de REM-3C situés dans une fenêtre génomique de 2 Mb autour du locus de départ**

La colonne « Amorce » présente l'homologie entre l'amorce L1 utilisée pour le REM-3C et la séquence génomique correspondante. Les rangées en gris représentent des partenaires qui furent récupérés au moins une fois dans nos expériences de REM-3C. Les partenaires potentiels résultants en des produits de REM-PCR de plus de 3 kb ne furent pas considérés pour cette analyse, étant donné que le temps d'élongation utilisé lors du REM-PCR (3 minutes) n'est pas compatible avec l'amplification de fragments d'ADN plus longs.

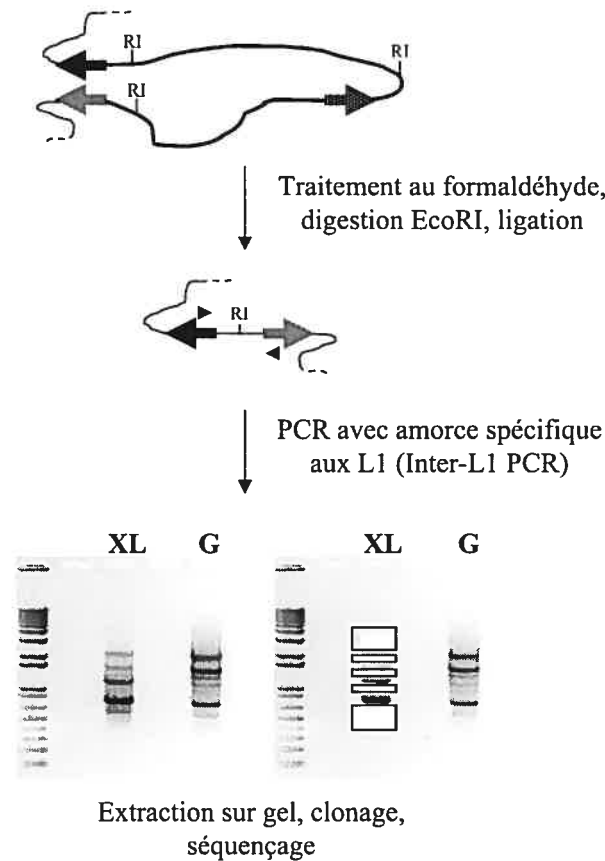
### **Distribution des partenaires de REM-3C inter-chromosomiques**

Des partenaires inter-chromosomiques furent aussi récupérés (Tableau 7). Ces partenaires furent cependant récupérés une seule fois chacun, proviennent de plusieurs chromosomes différents et ne présentent pas de spécificité évidente. Il est notable que trois loci situés sur le chromosome 7q furent récupérés, dont deux sur des bandes adjacentes (7q31.1 et 7q31.2). Par contre, il est statistiquement impossible de conclure que ces associations sont significativement plus fréquentes que des interactions aléatoires : une étude à plus grande échelle serait donc nécessaire pour révéler des interactions inter-chromosomiques spécifiques. De façon intrigante, la proportion d'événements inter-chromosomiques est significativement plus élevée dans la lignée ATA67 que dans la lignée IMR90, bien que le nombre d'événements clonés ne soit pas très élevé (9 inter et 18 intra pour ATA67 vs 3 inter et 28 intra pour IMR90; Test exact de Fisher,  $p=0,0265$ ). Ceci suggère que le locus de départ pourrait avoir un accès plus restreint au reste du génome dans la lignée normale IMR90 que dans lignée instable ATA67.

### **L'accessibilité spatiale privilégiée au Mb génomique adjacent est une caractéristique commune à plusieurs loci**

Les résultats présentés jusqu'à présent montrent que le locus de départ situé sur le chromosome 2p22.1 possède un environnement spatial couvrant préférentiellement les loci situés à l'intérieur d'une distance de 1 Mb. Pour vérifier si les caractéristiques de cet environnement spatial sont partagées par d'autres loci génomiques, nous avons réalisé l'expérience décrite en Figure 21. La procédure de 3C utilisée dans cette expérience est essentiellement la même que précédemment, à l'exception de l'étape de PCR. Ce PCR fut réalisé en utilisant uniquement une amorce complémentaire à la séquence consensus des L1. N'importe quels loci fixés qui contiennent des séquences L1 et un site de restriction en position adéquate (EcoRI dans ce cas) peuvent donc être amplifiés par cette méthode (que nous avons nommée inter-L1 3C). Les produits de PCR résultant de cette amplification sont clonés et séquencés. Les clones positifs sont finalement identifiés par la présence d'un site de restriction EcoRI séparant deux fragments d'ADN génomique non-contigus.

Cette expérience fut réalisée en utilisant les cellules ATA67 et les résultats sont présentés au Tableau 9. De façon concordante avec les résultats obtenus pour le locus du chromosome 2p22.1, les partenaires intra-chromosomiques (20/27) furent récupérés significativement ( $p=0,0004$ ; Test exact de Fisher) plus fréquemment que les partenaires inter-chromosomiques (7/27). De plus, la majorité des partenaires intra-chromosomiques sont situés à l'intérieur de 1 Mb les uns des autres. Ceci suggère donc que l'environnement immédiat d'une fraction importante des loci du génome est formé de leur Mb adjacent.



### Figure 21 : La méthodologie de l'inter-L1 3C

Les flèches représentent des éléments L1 et leur orientation. RI : site de restriction EcoRI. XL : Inter-L1 PCR sur de l'ADN traité par 3C. G : Inter-L1 PCR sur de l'ADN génomique. Les bandes majeures proviennent de loci génomiques contenant deux séquences L1 en orientation inverses et sont donc présentes dans les amplifications provenant des deux préparations d'ADN : ces bandes ne sont pas d'intérêt et sont évitées lors du clonage (zones encadrées du gel).

| Partenaire #1       | Partenaire #2      | Distance (kb) |
|---------------------|--------------------|---------------|
| Intra-chromosomique |                    |               |
| 9p24.3: 1769012     | 9p24.3: 1750825    | 18            |
| 10p14: 10171826     | 10p14: 10190252    | 18            |
| 8q24.22: 135430218  | 8q24.22: 135389485 | 41            |
| Xq21.2: 85622520    | Xq21.2: 85669537   | 47            |
| 5q31.3: 144245615   | 5q31.3: 144099343  | 146           |
| 2q11.2: 95905685    | 2q11.2: 96063873   | 158           |
| 1p31.1: 76636063    | 1p31.1: 76420702   | 215           |
| Xq25: 126227231     | Xq25: 125971293    | 256           |
| 12p11.1: 33623666   | 12p11.1: 33946290  | 322           |
| 5q34: 164400511     | 5q34: 164078371    | 322           |
| 8p22: 13551860      | 8p22: 13947021     | 395           |
| 11q14.3: 90732793   | 11q14.3: 90009554  | 723           |
| 1p21.1: 103631575   | 1p21.1: 102885532  | 746           |
| 4p13: 45109487      | 4p13: 44339883     | 770           |
| 19q13.43: 62515857  | 19q13.43: 61132425 | 1383          |
| 1q24.3: 168423674   | 1q25.1: 170211673  | 1788          |
| Xp11.22: 51117735   | Xp11.21: 54744301  | 3627          |
| 10p13: 15856833     | 10p15.1: 6439416   | 9417          |
| 4q31.1: 141223680   | 4q34.1: 173060421  | 31836         |
| 13q33.3: 107219511  | 13q21.33: 71453636 | 35766         |
| Inter-chromosomique |                    |               |
| Yp11.2: 6946826     | 2p25.1: 8132345    | -             |
| 10p13: 15856833     | Xp11.1: 56749282   | -             |
| 10q24.33: 105563907 | 20p12.1: 16708407  | -             |
| 7q31.33: 125835478  | 9p21.3: 23851151   | -             |
| 10p13: 15856833     | Xp11.1: 56749282   | -             |
| 8p23.2: 4854868     | 5p12: 44218612     | -             |
| 9p24.1: 5872745     | 14q21.1: 40080233  | -             |

**Tableau 9 : Inter-L1 3C sur les cellules ATA67**

La distance entre les partenaires de ligation est indiquée. Les clones positifs furent identifiés par la présence d'un site de restriction EcoRI séparant deux fragments d'ADN génomique non-contigus.

### 3. Modifications de l'environnement spatial génomique du gène HoxB1 lors de l'induction de son expression dans le contexte de la différenciation de cellules embryonnaires de souris induite par l'acide rétinoïque

#### Objectifs

Les résultats obtenus dans l'étude précédente suggèrent que l'environnement spatial d'un locus peut avoir une influence sur son choix de partenaires de recombinaison. Ils montrent d'autre part que les stratégies de 3C ouvertes peuvent être utilisées pour décrire l'environnement spatial d'un locus donné dans des cellules de mammifères. Cependant, la méthode décrite précédemment est laborieuse au niveau technique, surtout à cause du fait que les produits de REM-3C spécifiques doivent être identifiés par *colony lift* avant leur séquençage. De plus, une proportion relativement élevée des loci génomiques ne peut être récupérée par cette méthode, à cause de biais introduits par la séquence de l'amorce utilisée et à cause de la longueur des produits de PCR résultants. Nous avons donc décidé de développer une méthode de 3C ouverte plus efficace basée sur l'amplification par inverse-PCR des loci fixés à un locus d'intérêt. L'étude présentée plus tôt avait pour but l'investigation de l'environnement spatial d'un locus « statique », dont l'état ne change pas durant l'expérience. Or, comme nous l'avons décrit dans l'introduction, la transcription est particulièrement liée à des changements d'organisation spatiale de la chromatine. Nous nous sommes donc intéressés à la caractérisation des changements dynamiques subis par l'environnement spatial génomique d'un locus dont l'état transcriptionnel est modulable, soit le gène HoxB1 de souris. Les objectifs de ce projet sont donc les suivants :

-Adapter et améliorer la méthodologie REM-3C décrite précédemment en utilisant une étape d'inverse-PCR au lieu de l'étape limitante du REM-PCR

-Utiliser cette méthodologie pour caractériser l'environnement spatial dynamique du gène HoxB1 lors de l'induction de son expression dans le contexte de la différenciation de cellules souches embryonnaires de souris induite par l'acide rétinoïque

**Note :** Cette étude a fait l'objet d'une publication (Wurtele et Chartrand 2006)

## Design expérimental

Un des buts de cette étude était de modifier la technologie du REM-3C de façon à la rendre plus efficace et la moins biaisée possible. Nous avons donc remplacé l'étape du REM-PCR par une étape d'inverse-PCR. Les produits d'inverse-PCR ainsi clonés sont identifiés à partir des bases de données du génome de souris. Les loci qui sont fréquemment à proximité spatiale du locus d'intérêt sont amplifiés et clonés plus fréquemment à partir des inverse-PCR que les loci qui sont rarement à proximité de ce locus.

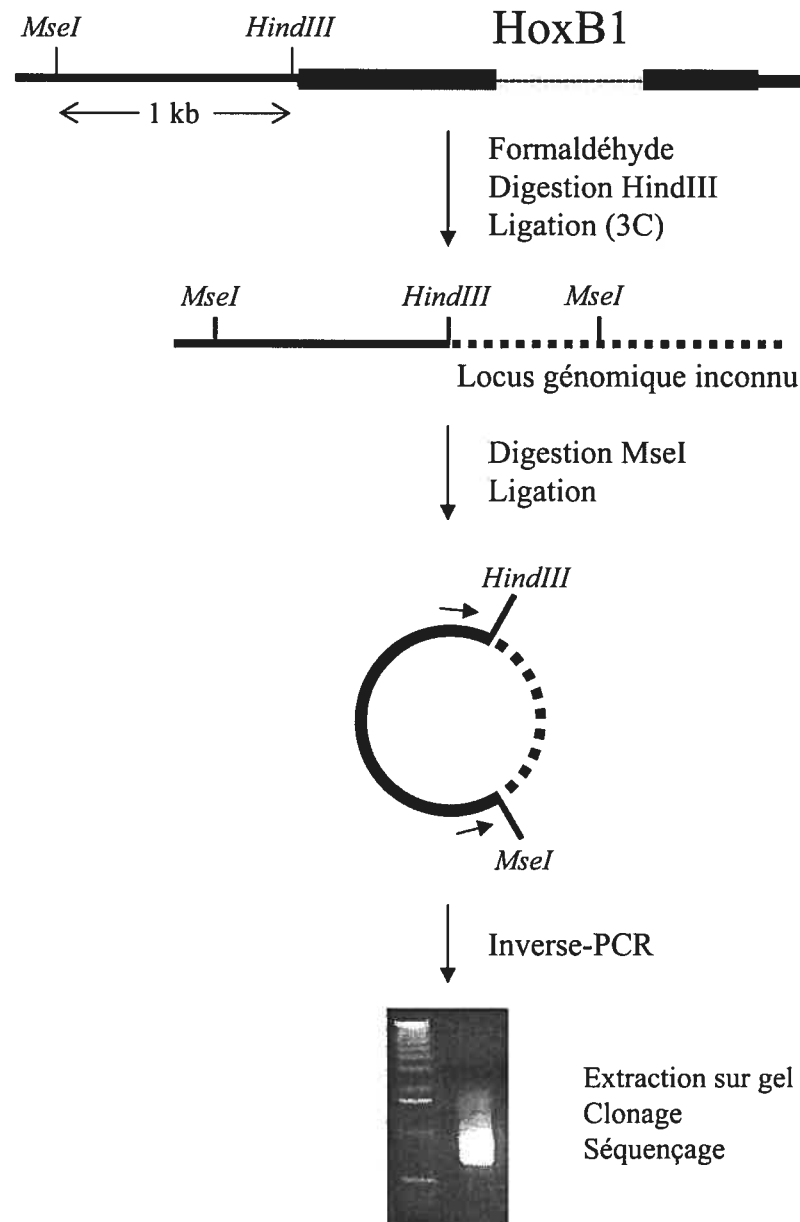
Nous étions intéressés à investiguer l'effet de l'induction de l'expression du gène HoxB1 sur son environnement spatial, dans le contexte de la différenciation de cellules souches embryonnaires (cellules ES) de souris OS25 induite par l'acide rétinoïque. Ce système cellulaire fut choisi parce qu'il fut utilisé par un autre groupe pour caractériser la décondensation à grande échelle et le repositionnement du locus du gène HoxB1 en relation avec son territoire chromosomique lors de son induction par l'acide rétinoïque (Chambeyron et Bickmore 2004). Nous nous attendions donc à ce que l'environnement spatial de ce gène change de façon marquée durant son induction. D'autre part, nous étions aussi intéressés à vérifier l'identité des loci associés au gène actif : cette caractérisation nous permettrait ainsi de vérifier si le gène HoxB1 est recruté vers des usines de transcription lors de son induction. Globalement, cette méthodologie nous donnait la possibilité de caractériser l'effet de la décondensation et de l'activation transcriptionnelle du locus HoxB1 sur ses interactions avec le reste du génome.

Le système cellulaire utilisé présente aussi d'autres avantages techniques. En effet, les cellules OS25 peuvent être maintenues dans l'état non-différencié en les cultivant dans un milieu contenant du G418, grâce à un gène beta-geo introduit par *knock-in* dans le gène Sox2. Ce gène est réprimé dans les cellules différenciées, ce qui rend les cellules sensibles au G418. D'autre part, le gène viral de la thymidine kinase (TK) fut lui aussi intégré par *knock-in* dans les séquences codantes du gène OCT4. Ce gène est fortement réprimé lors de la différenciation : les cellules différenciées deviennent donc résistantes au ganciclovir. Il est donc possible de sélectionner les cellules en différenciation ou les cellules non-différenciées, ce qui permet d'obtenir des populations cellulaires plus homogènes. Ceci est utile pour les expériences de 3C, qui décrivent l'état moyen des interactions entre régions

génomiques dans une population cellulaire donnée : il est donc préférable d'obtenir une population aussi homogène que possible lors de l'utilisation de cette méthode.

Les expériences d'inverse-PCR 3C furent réalisées tel qu'illustré à la Figure 22. Un site de restriction HindIII situé 16 pb en amont (5') du codon de départ du gène HoxB1 fut utilisé pour la partie 3C des expériences (formaldéhyde, digestion HindIII, ligation), alors qu'un site de restriction MseI situé à environ 1 kb en amont du site HindIII fut utilisé en tant que site de restriction d'inverse-PCR. MseI est une enzyme qui possède un site de reconnaissance de 4 pb très fréquent dans le génome : l'utilisation de cette enzyme devrait donc limiter la taille des produits d'inverse-PCR et faciliter leur amplification. La Figure 22 montre une électrophorèse sur gel d'agarose d'une expérience représentative d'inverse-PCR 3C. Tel qu'attendu, une trainée est observée sur le gel; étant donné que plusieurs loci sont fixés au locus HoxB1 et sont amplifiés par cette méthode, les produits de PCR résultants sont de tailles variées. Il est à noter que la taille de la plupart des produits se situe entre 650 pb et 1 kb (650 pb est la taille minimale possible pour un produit d'inverse-PCR à cause de la position du site de restriction MseI et des amorces de PCR utilisées). Les produits sont donc effectivement de tailles similaires et une grande proportion d'entre eux devraient être amplifiés avec une efficacité comparable. Les produits de PCR sont ensuite clonés et séquencés, puis leur localisation génomique est déterminée en utilisant les bases de données du génome de souris.

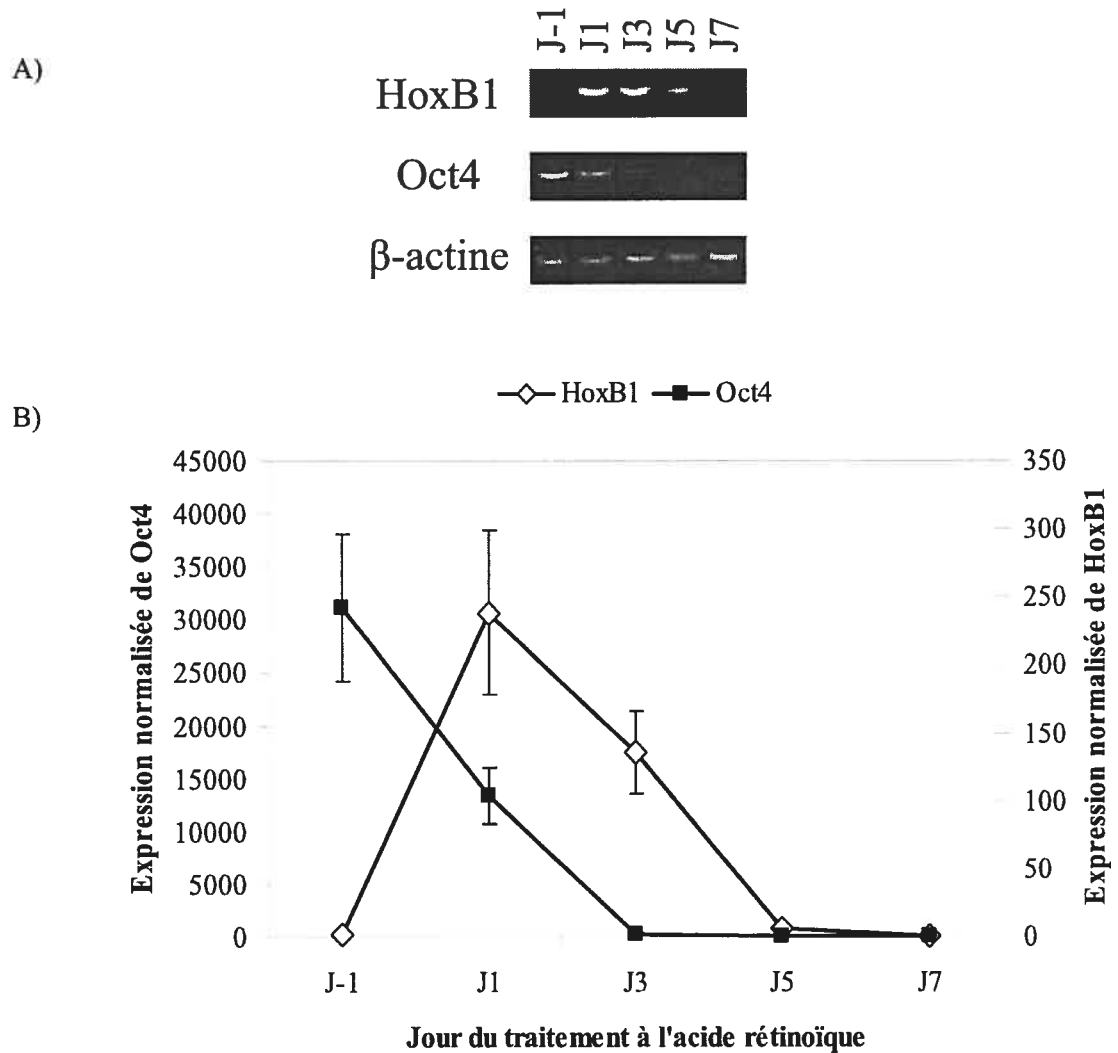
La Figure 23 montre que l'expression de HoxB1 est fortement induite après une journée de traitement à l'acide rétinoïque, reste élevée au jour 3 et est graduellement perdue au jour 5, de telle sorte que son expression est indétectable au jour 7. L'expression du gène Oct4, qui comme nous l'avons mentionné est caractéristique des cellules non-différenciées, fut utilisée comme contrôle : de façon contrastante avec HoxB1, ce gène est fortement exprimé dans les cellules non-différenciées et est réprimé après traitement à l'acide rétinoïque. Une analyse par PCR quantitatif a d'ailleurs confirmé une augmentation d'environ 250 et 150 fois dans le nombre de transcrits de HoxB1 au jour 1 et 3 respectivement du protocole de différenciation en comparaison avec les cellules non-traitées. Les inverse-PCR 3C furent donc réalisés au jour -1 (avant le traitement à l'acide rétinoïque et l'expression de HoxB1 concomitante) et aux jours 1 et 3 (après 1 ou 3 jours de traitement à l'acide rétinoïque et durant l'expression de HoxB1).



**Figure 22 : Design expérimental de l'inverse-PCR 3C réalisé sur le gène *HoxB1* de souris**

La méthodologie du Inverse-PCR 3C basée sur le Chromosome Conformation Capture développé par Dekker et collaborateurs (Dekker et al. 2002). Les loci qui sont fréquemment à proximité spatiale du locus de départ (dans notre cas le gène *HoxB1*) y sont fixés et ligaturés à haute fréquence. Ces loci sont donc fréquemment amplifiés par inverse-PCR et clonés. Le site de restriction *HindIII* utilisé dans la portion 3C de la méthode est situé 16 pb en amont de *HoxB1*.





**Figure 23 : Analyse par RT-PCR de l'expression des gènes HoxB1 et Oct4 durant le protocole de différenciation des cellules OS25 de souris par l'acide rétinoïque**

A) Analyse par PCR semi-quantitatif. Chaque puit correspond à un jour différent du protocole de différenciation induit par l'acide rétinoïque (voir Matériel et méthode). B) Analyse par PCR quantitatif en temps réel. Chaque point du graphique représente la moyenne des résultats obtenus à partir de quatre à huit traitements à l'acide rétinoïque indépendants. Les amplifications furent normalisées par rapport au signal  $\beta$ -actine. Les unités d'expression sont arbitraires : la valeur du jour-1 fut fixée à 1 pour l'expression de HoxB1 alors que la valeur du jour 7 fut fixée à 1 pour l'expression de Oct4. Barres d'erreur :  $\pm$  erreur moyenne.

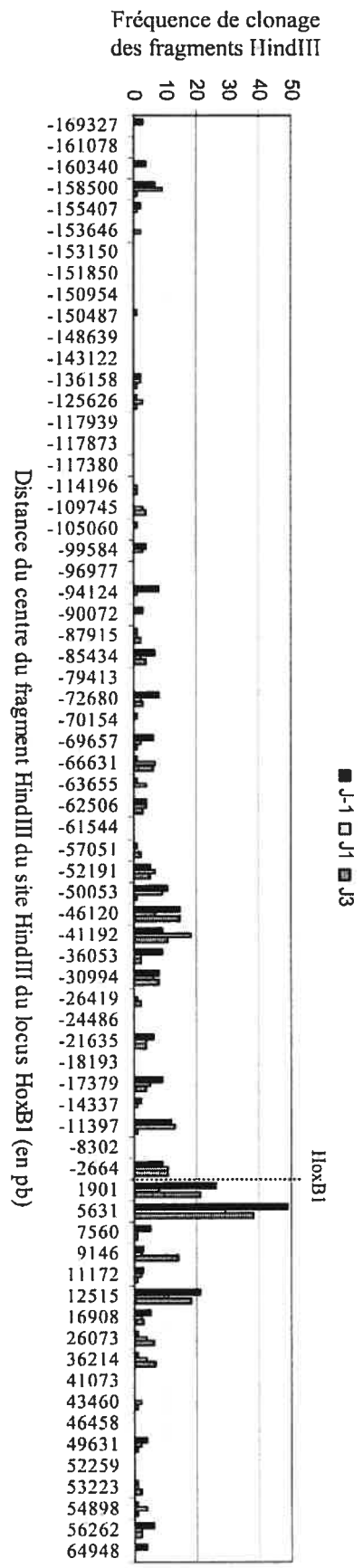
La Tableau 1 donne des détails sur les différents inverse-PCRs qui furent réalisés à chaque jour du traitement à l'acide rétinoïque (voir Matériel et méthode pour les détails). Pour chaque jour du protocole de différenciation, entre 653 et 904 clones furent séquencés et localisés dans les bases de données du génome de souris. Il est à noter que dans certains cas, plusieurs PCR furent regroupés avant le clonage. Cette procédure fut utilisée pour parer à l'éventualité de l'obtention d'une sur-représentation de certains événements dans un PCR donné, comme c'était le cas avec le REM-3C. Le Tableau 1 montre que nous avons cependant clonés plusieurs centaines de produits de PCR à partir de réactions individuelles indépendantes (par exemple, les PCR A à F furent réalisés sur de l'ADN du jour -1, préparation B1; nous avons cloné ces PCR individuellement, sans les mettre en commun). Ceci nous a permis de constater que la majorité des produits récupérés à partir de ces PCR individuels ne sont pas des doublets, c'est-à-dire qu'ils ne furent pas récupérés plusieurs fois lors du clonage de ce PCR particulier. Malgré tout, pour tenter d'éliminer au maximum ce genre de problème potentiel, nous avons généralement regroupé entre 7 et 10 inverse-PCR indépendants réalisés sur un ADN matrice donné avant l'étape de clonage, de façon à augmenter la diversité des événements clonés. Notons en outre que la plupart des conclusions de cette étude concernent des régions génomiques relativement étendues : la méthode est donc robuste par rapport à une suramplification occasionnelle d'un produit de ligation particulier, puisque nos analyses ne sont généralement pas basées sur la fréquence de clonage d'un produit en particulier, mais bien sur l'analyse de régions génomiques contenant un plus grand nombre d'événements de fixation indépendants.

### **Récupération à haute fréquence des loci proximaux par inverse-PCR 3C**

Des partenaires inter et intra-chromosomiques furent clonés à partir des expériences d'inverse-PCR 3C. Les loci proximaux intra-chromosomiques furent analysés en premier lieu. Les fréquences de clonage des fragments génomiques HindIII proximaux entourant le locus HoxB1 sont présentées en Figure 24. À cause de leur grande proximité linéaire avec HoxB1, nous nous attendions à ce que ces fragments génomiques proximaux soient récupérés à haute fréquence dans notre crible. Effectivement, presque tous les loci situés à l'intérieur de 40 kb en 3' et de 70 kb en 5' furent récupérés au moins une fois (seuls 4 de ces 31 fragments proximaux ne furent pas récupérés). Les résultats présentés en Figure 24 suggèrent que globalement, les fréquences de fixation avec HoxB1 sont plus hautes pour les

loci situés à l'intérieur de ces régions génomiques proximales et qu'en dehors de cet intervalle génomique, la distribution des partenaires de fixation devient plus éparse et ceux-ci sont en général récupérés moins fréquemment. Comme nous le documenterons avec plus de détails plus bas, ceci suggère que l'environnement spatial du gène HoxB1 peut être séparé en régions génomiques distinctes d'accessibilité. La Figure 24 suggère en outre que la fréquence de clonage des partenaires de fixation diminue plus lentement avec la distance dans la région proximale en 5' que dans celle située en 3' du gène. Finalement, la plupart des fragments semblent avoir été récupérés à des fréquences relatives similaires aux différents jours du traitement à l'acide rétinoïque : l'accessibilité du gène à son environnement spatial proximal ne semble donc pas modifiée de façon importante par ce traitement.

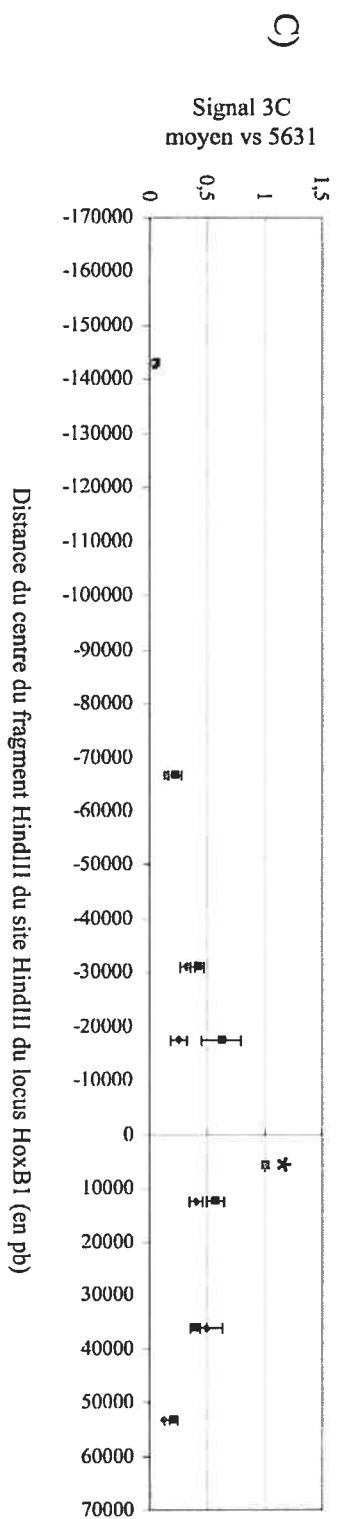
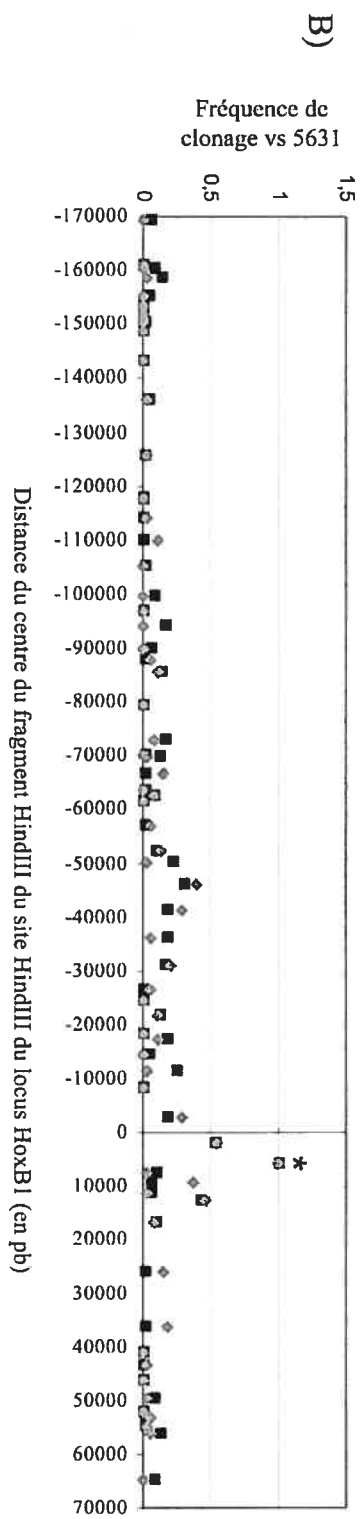
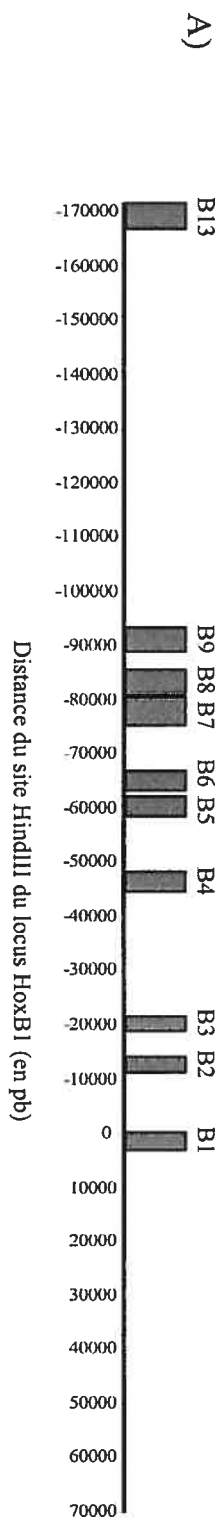
Les résultats présentés en Figure 24 montrent que certains fragments HindIII génomiques furent définitivement récupérés plus fréquemment que d'autres dans l'intervalle proximal. Par exemple, les fragments 5631, 12515 et -30994 furent clonés relativement fréquemment dans notre crible, alors que d'autres furent récupérés à faible fréquence ou pas du tout (par exemple -57051, -8302 et 11172). Bien que ces différences dans la fréquence de clonage de ces fragments puisse refléter de véritables différences dans les fréquences de fixation de ces loci au locus HoxB1, l'analyse de fragments HindIII génomiques individuels est compliquée par le fait qu'une faible fréquence de clonage pourrait aussi résulter d'une amplification inefficace par la procédure d'inverse-PCR que nous avons employée. Pour vérifier cette possibilité, nous avons déterminé la taille des produits d'inverse-PCR qui résultent de l'amplification des fragments proximaux (situés à l'intérieur de 70 kb en 5' et 40 kb en 3' ; résultats non-illustrés). Cette analyse a montré que la plupart des fragments génomiques qui ne furent pas récupérés fréquemment dans notre crible étaient probablement de piètres substrats d'inverse-PCR : soit ils ne possédaient pas de site MseI ou le produit résultant était de grande taille, et donc probablement défavorisé lors du PCR. Ceci suggère que certains fragments génomiques ne peuvent pas être récupérés efficacement par la méthodologie d'inverse-PCR 3C, sans égard à leur fréquence de fixation à HoxB1.



**Figure 24 : Fréquence de clonage des fragments HindIII génomiques proximaux qui entourent le locus HoxB1**

Étant donné qu'un fragment HindIII génomique peut être ligaturé au site HindIII du locus HoxB1 par une ou l'autre de ses extrémités, deux produits d'inverse-PCR distincts peuvent être générés à partir de chaque fragment particulier. La fréquence de clonage présentée ici est la somme des fréquences impliquant chacune des extrémités. La distance du centre de chaque fragment HindIII du site HindIII du locus HoxB1 est indiquée. Les distances négatives représentent des loci situés en 5' du gène HoxB1 alors que les distances positives représentent des loci situés en 3'. Ce graphique n'est pas à l'échelle. Le fragment -2664 provient de la recircularisation du locus de départ sur lui-même et ne représente donc pas un authentique événement de fixation et de ligation 3C.

Nous avons voulu vérifier si globalement, les résultats d'inverse-PCR 3C étaient compatibles avec une analyse par 3C classique. La Figure 25 montre les résultats d'une analyse par 3C classique des fréquences de fixation de certains fragments HindIII génomiques proximaux au locus HoxB1. En général, il existe une bonne corrélation entre les résultats obtenus par les deux méthodes, puisque la densité des partenaires récupérés et leur fréquence de clonage sont liées aux fréquences de fixation relatives observées par 3C classique (voir Figure 25B et C). Si l'on considère chaque côté du locus de départ de façon indépendante, les résultats montrent que les fréquences de fixation relatives par 3C des trois premiers fragments HindIII génomiques en 3' et des deux premiers en 5' sont similaires entre elles. Cependant, les loci qui sont plus distaux (par exemple les loci 53223 et -143122) sont clairement fixés moins fréquemment à HoxB1 que les autres loci qui sont situés du même côté du gène.



**Figure 25 : Analyse par 3C classique des fragments HindIII génomiques proximaux**

A) Localisation génomique des gènes du complexe HoxB. B) Fréquence relative de clonage des fragments HindIII génomiques individuels par rapport au fragment 5631 (mis en évidence par une étoile \*) dans l'intervalle génomique présenté en Figure 24. Les résultats du jour -1 et du jour 3 sont présentés. Les distances négatives représentent des loci situés en 5' de HoxB1 alors que les distances positives représentent des loci situés en 3'. C) Analyse par 3C classique de l'intervalle génomique présenté en Figure 24. Les signaux de 3C provenant de préparations d'ADN 3C indépendantes sont normalisés à l'interne au signal représentant la fixation de HoxB1 au fragment 5631, dont la valeur fut fixée à 1 (mis en évidence par une étoile \*). La moyenne de ces signaux relatifs est présentée pour chaque locus. Barres d'erreur :  $\pm$ erreur moyenne. Les distances négatives représentent des loci situés en 5' de HoxB1 alors que les distances positives représentent des loci situés en 3'. Les fragments HindIII génomiques qui furent utilisés sont nommés -143122, -66631, -30994, -17379, 5631, 12515, 36214 et 53223 dans la Figure 24.

En général, il existe aussi une bonne corrélation entre la fréquence de clonage des fragments HindIII génomiques individuels par inverse-PCR 3C et leur fréquence de fixation telle que révélée par 3C classique. Il est notable que le fragment qui fut récupéré le plus fréquemment à partir du crible par inverse-PCR 3C (fragment 5631 de la Figure 24) est aussi le fragment dont la fréquence de fixation est la plus élevée selon l'analyse de la région par 3C classique. Tel que suggéré par les résultats d'inverse-PCR 3C, les fréquences de fixation des fragments HindIII investigués sont aussi généralement similaires aux jours -1 et 3, à l'exception du fragment situé à 17379 pb en 5' de HoxB1, qui semble être fixé légèrement moins fréquemment au jour 3 qu'au jour 1 selon l'analyse de 3C classique. Ce fragment fut aussi cloné moins fréquemment au jour 3 qu'au jour -1 lors des expériences d'inverse-PCR 3C. Bien que cette différence dans la fréquence de clonage ne soit pas statistiquement significative, ceci suggère qu'un crible à plus grande échelle pourrait dans certains cas identifier des différences plus subtiles dans les fréquences de fixation de loci individuels.

Il existe aussi des différences entre les résultats des deux méthodes au niveau des fragments HindIII individuels, ce qui était attendu puisque certains fragments ne peuvent pas être récupérés efficacement par inverse-PCR 3C. Par exemple, le fragment 36214 situé en 3' du gène ne fut pas cloné aussi fréquemment que le fragment 12515, mais semble être fixé à HoxB1 à une fréquence similaire par analyse de 3C. Cependant, une analyse bioinformatique nous a révélé que le fragment 36214 pouvait surtout être récupéré par une de ses deux extrémités étant donné que le produit d'inverse-PCR 3C provenant de l'autre extrémité serait très long et donc probablement amplifié peu efficacement. Cette étude montre donc que les résultats d'inverse-PCR 3C concernant des fragments génomiques individuels doivent être validés par des analyses de 3C classique.

### **L'environnement nucléaire du gène HoxB1 peut être séparé en domaines génomiques distincts d'interactions spatiales**

La Figure 26 présente la proportion cumulative des clones de partenaires de fixation récupérés à diverses distances du site HindIII du locus HoxB1. Les proportions furent obtenues en divisant le nombre de clones représentant des partenaires situés à l'intérieur d'une distance donnée de HoxB1 par le nombre total de clones récupérés pour chaque jour du traitement à l'acide rétinoïque. Cette courbe permet donc la description de l'accumulation de clones de partenaires sur de vastes régions génomiques et tient compte de la densité des partenaires de fixation et de leur fréquence individuelle de récupération par inverse-PCR 3C.

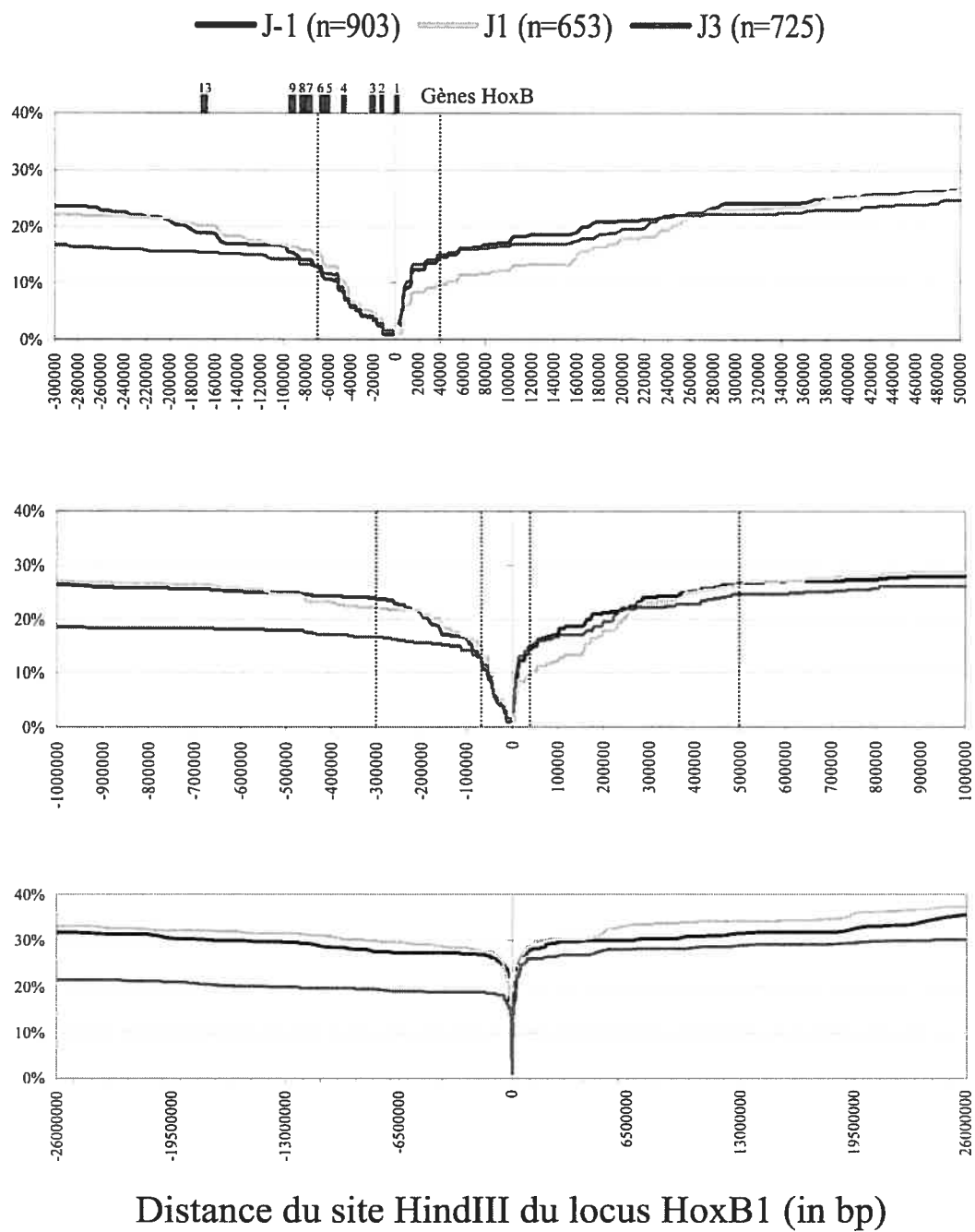
Étant donné que la courbe change abruptement à certaines distances génomiques de HoxB1 et en se basant sur les résultats présentés en Figure 24 et en Figure 25, nous avons séparé la courbe en régions distinctes présentant différents taux d'accumulation de clones de partenaires de fixation. Ainsi, la courbe est très abrupte dans les régions génomiques englobant 70 kb en 5' et 40 kb en 3', ce qui indique que ces régions possèdent une haute densité de partenaires de fixation qui sont récupérés fréquemment par inverse-PCR 3C. Dans la région génomique comptant les 70 kb proximaux en 5', la courbe semble être plutôt linéaire, ce qui indique une haute densité de partenaires de fixation distribués de façon relativement égale dans l'intervalle, et donc un taux d'accumulation de clones qui est généralement indépendant de la distance linéaire entre un site donné et HoxB1. En clair,



ceci suggère que plusieurs loci situés dans cette région sont fréquemment fixés à HoxB1 et que les fréquences de fixation diminuent plus lentement avec la distance de HoxB1 dans cette région que dans la région située en 3' du gène. En effet, dans la région proximale couvrant les 40 kb en 3' de HoxB1, la courbe est clairement logarithmique, ce qui suggère que la fréquence de fixation diminue relativement rapidement avec la distance dans cet intervalle génomique.

Après ces régions proximales, le taux d'accumulation de clones diminue, ce qui signifie que la densité de partenaires ainsi que leur fréquence de clonage sont plus basses que celles des domaines proximaux. Cependant, le fait que la courbe semble toujours abrupte indique qu'il y a encore une densité de partenaires de fixation relativement élevée dans ces régions jusqu'à un autre point de cassure de la courbe situé à environ 300 kb en 5' et 500 kb en 3' du gène. Cette portion de la courbe est logarithmique : la probabilité d'être fixé à HoxB1 diminue donc avec la distance génomique jusqu'aux frontières de ce domaine. Après ces distances, le taux d'accumulation de clones diminue de façon importante et la courbe semble être plus ou moins linéaire pour le reste de la distance mesurée. La probabilité d'être fixé à HoxB1 semble donc devenir plus stable sur de grandes distances génomiques après les frontières du deuxième domaine d'accessibilité situées à 300 kb en 5' et 500 kb en 3', ce qui résulte en une distribution plus éparse des partenaires de fixation et en une fréquence réduite de récupération par inverse-PCR 3C. Comme nous le verrons en Figure 30, l'accessibilité à HoxB1 semble néanmoins diminuer avec la séparation génomique après les limites de 300 kb en 5' et 500 kb en 3'. Cependant, le fait que cette diminution soit très progressive résulte en une courbe qui semble plus ou moins linéaire en Figure 26. Bien que les domaines décrits plus haut semblent exister peu importe le statut d'expression du gène HoxB1, leur importance relative varie en lien avec l'expression. Ces différences sont le sujet de la prochaine section.

Proportion des partenaires de fixation clonés totaux

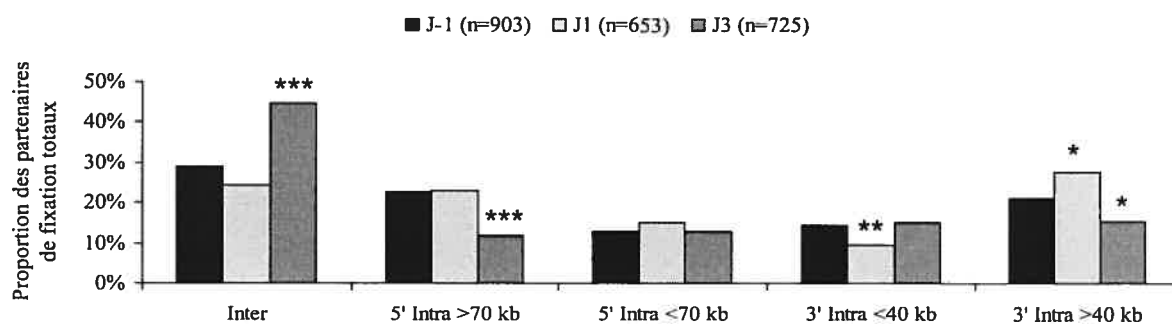


**Figure 26 : Proportion cumulative des partenaires de fixation récupérés par inverse-PCR 3C à différentes distances du locus HoxB1**

Les proportions sont obtenues en divisant le nombre de clones de partenaires situés entre une position génomique donnée et le site HindIII de HoxB1 par le nombre total de clones récupérés pour chaque jour du traitement à l'acide rétinoïque. Les distances négatives représentent des loci situés en 5' du locus HoxB1 alors que les loci situés en 3' sont représentés par des distances positives. Les lignes pointillées délimitent des régions particulières de la courbe (voir le texte pour les détails). La position des gènes du complexe HoxB est aussi présentée.

**L'environnement génomique spatial du gène HoxB1 change durant le traitement des cellules ES de souris à l'acide rétinoïque**

L'analyse présentée en Figure 26 suggère que la distribution des clones de partenaires de fixation dans les différents domaines génomiques d'accessibilité à HoxB1 n'est pas équivalente à chaque jour du traitement des cellules à l'acide rétinoïque. Par exemple, les partenaires de fixation situés dans le domaine distal situé en 5' du gène HoxB1 semblent être récupérés moins fréquemment au jour 3 qu'au jour -1 ou 1 du traitement. La Figure 27 présente la proportion des clones de partenaires de fixation qui tombent dans une ou l'autre de cinq régions génomiques définies selon les analyses décrites plus haut. Les régions génomiques proximales de 70 kb en 5' et de 40 kb en 3' contiennent au total entre 25% et 28% des clones récupérés pour chaque jour du traitement à l'acide rétinoïque. Ceci suggère qu'en général, l'accessibilité de HoxB1 à ces domaines proximaux n'est pas modifiée de façon importante par son statut d'expression. De façon intrigante, cependant, les partenaires de fixation situés dans le domaine proximal en 3' représentent une proportion significativement plus basse des clones totaux récupérés au jour 1 qu'au jour -1 ( $p=0,0097$ , Test Chi carré de Pearson), ce qui semble être lié à une plus grande proportion de partenaires dans le compartiment distal 3' intra >40kb ( $p=0,015$ , Test Chi carré de Pearson).



**Figure 27 : Distribution des partenaires de fixation clonés du gène HoxB1 dans différents intervalles génomiques**

Les proportions représentent le nombre de clones de partenaires de fixation situés dans un intervalle donné divisé par le nombre de clones récupérés pour un jour donné du traitement à l'acide rétinoïque. Inter : partenaires inter-chromosomiques. 5' intra >70 kb : partenaires intra-chromosomiques situés à plus de 70 kb en 5' du site HindIII de HoxB1. 5' intra <70 kb : partenaires intra-chromosomiques situés à moins de 70 kb en 5' du site HindIII de HoxB1. 3' intra >40 kb : partenaires intra-chromosomiques situés à plus de 40 kb en 3' du site HindIII de HoxB1. 3' intra <40 kb : partenaires intra-chromosomiques situés à moins de 40 kb en 3' du site HindIII de HoxB1. \*\*\*: significativement différent de la valeur du jour -1 ( $p < 0,0001$ ; Test Chi carré de Pearson). \*\*: significativement différent de la valeur du jour -1 ( $p = 0,0097$ ; Test Chi carré de Pearson). \*: significativement différent de la valeur du jour -1 ( $p = 0,015$ ; Test Chi carré de Pearson).

Comme les résultats de la Figure 26 le suggéraient, une proportion significativement plus faible de clones est située dans les régions distales 5' et 3' (5' intra >70 kb et 3' intra >40 kb) au jour 3 qu'au jour -1 ( $p < 0,0001$  and  $p = 0,015$ , respectivement, Test Chi carré de Pearson). En effet, ces régions contiennent au total 44 et 51% de tous les clones récupérés au jour -1 et au jour 1 respectivement, alors qu'elles ne contiennent que 27% des clones du jour 3. Notons au passage que cette différence semble beaucoup plus significative pour la région distale située en 5' du gène que pour celle située en 3'. De façon intéressante, des différences correspondantes dans la proportion des partenaires de fixation inter-chromosomiques sont observées entre les jours 3 et -1. Les interactions inter-chromosomiques représentent ainsi une minorité d'événements (environ 25 à 30%) au jour

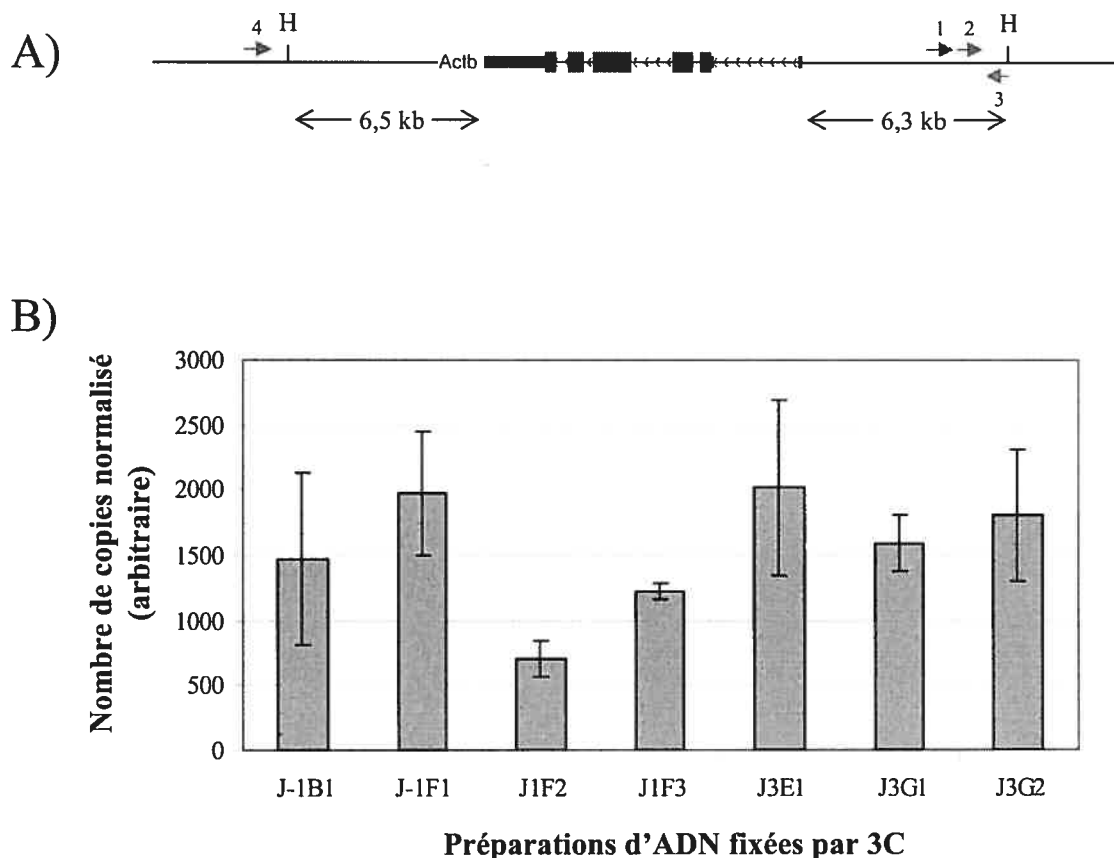
-1 et au jour 1. En contraste, ces interactions furent récupérées presque aussi fréquemment que les interactions intra-chromosomiques au jour 3 (45% des partenaires totaux) et donc de façon significativement plus fréquente qu'au jour -1 ( $p < 0,0001$ , Test Chi carré de Pearson). Ces résultats sont compatibles avec l'extrusion documentée du locus HoxB1 hors de son territoire chromosomique après un traitement de cellules ES à l'acide rétinoïque (Chambeyron et Bickmore 2004).

Une explication potentielle pour ces observations voudrait que l'étape de fixation de la procédure de 3C soit pour une raison quelconque moins efficace au jour 3 qu'au jour -1 ou au jour 1. Cette efficacité réduite de fixation pourrait résulter en une plus grande quantité d'événements de ligation aléatoires. Ces événements de ligation aléatoires devraient par définition impliquer majoritairement des fragments HindIII inter-chromosomiques, et leur récupération par inverse-PCR 3C pourrait mener à une sur-évaluation de la fréquence d'interactions inter-chromosomiques. Pour évaluer s'il existe effectivement d'importantes différences dans l'efficacité du protocole de 3C entre les différentes préparations d'ADN utilisées pour l'inverse-PCR 3C, nous avons réalisé l'expérience décrite en Figure 28. Le but de cette expérience était de comparer la fréquence de fixation de deux fragments HindIII génomiques adjacents entre les différentes préparations d'ADN. Le locus du gène  $\beta$ -actine fut choisi comme locus de référence pour ces expériences, puisque son profil d'expression n'est pas modifié par le traitement à l'acide rétinoïque, tel que confirmé par une comparaison avec le profil d'expression du gène GAPDH (résultats non-montrés). Le gène  $\beta$ -actine est d'ailleurs situé sur un chromosome différent de celui du complexe de gènes HoxB. Nous ne nous attendons donc pas à ce que la structure de la chromatine de ce locus soit modifiée de façon importante durant le traitement à l'acide rétinoïque : la fréquence de fixation de ces deux fragments HindIII génomiques ne devrait donc pas être modifiée durant le traitement sauf si celui-ci a un impact sur l'efficacité de la procédure de 3C. Le signal de PCR quantitatif reflétant la ligation des fragments HindIII adjacents fut normalisé par rapport au signal reflétant la quantité d'ADN dans chaque préparation de 3C.

Les résultats présentés en Figure 28 démontrent que pour une quantité donnée d'ADN traité par la procédure de 3C, le signal normalisé représentant la fixation des fragments HindIII adjacents ne varie pas de façon importante entre les différentes préparations d'ADN utilisées lors de l'inverse-PCR 3C. En effet, la plus grande différence existe entre les préparations

J3E1 et J1F2, et est seulement d'environ trois fois. Il n'y a donc pas de corrélation entre la progression dans le protocole de différenciation induit par l'acide rétinoïque et l'efficacité de la procédure de 3C. Ces résultats suggèrent que les variations dans l'efficacité de la procédure de 3C n'affectent pas spécifiquement la proportion des partenaires inter-chromosomiques de fixation récupérés par la méthodologie d'inverse-PCR 3C, et donc que les différences dans ces proportions observées entre les jours -1 et 3 sont biologiquement importantes.

Nos résultats suggèrent donc que le gène HoxB1 a un plus grand accès au reste du génome lorsqu'il est exprimé, et ce de façon concomitante avec la différenciation des cellules induite par l'acide rétinoïque. Les mécanismes impliqués dans le changement d'environnement spatial observé sont inconnus. Il fut en effet démontré que bien que l'état d'acétylation du complexe HoxB soit modifié lors du traitement à l'acide rétinoïque, une hyperacétylation artificielle du complexe induite par le TSA ne permet pas l'extrusion du locus HoxB1 (Chambeyron et Bickmore 2004). Par contre, certaines observations suggèrent que le processus de différenciation cellulaire peut rendre les territoires chromosomiques plus diffus (Stadler et al. 2004), et il est connu que les territoires chromosomiques se décondensent en G1, après la mitose (Shelby et al. 1996). Il est donc possible que des changements dus à la différenciation cellulaire, par exemple dans la durée des cycles cellulaires, puissent influencer le positionnement du locus HoxB1 par rapport à son territoire chromosomique en affectant le degré de décondensation des territoires chromosomiques. Le gène HoxB1 étant normalement positionné à la périphérie de son territoire dans les cellules non-différenciées, une plus grande décondensation de celui-ci en réponse à une plus grande durée de la phase G1 pourrait conduire à un positionnement plus fréquent hors du territoire chromosomique dans une population de cellules en différenciation.

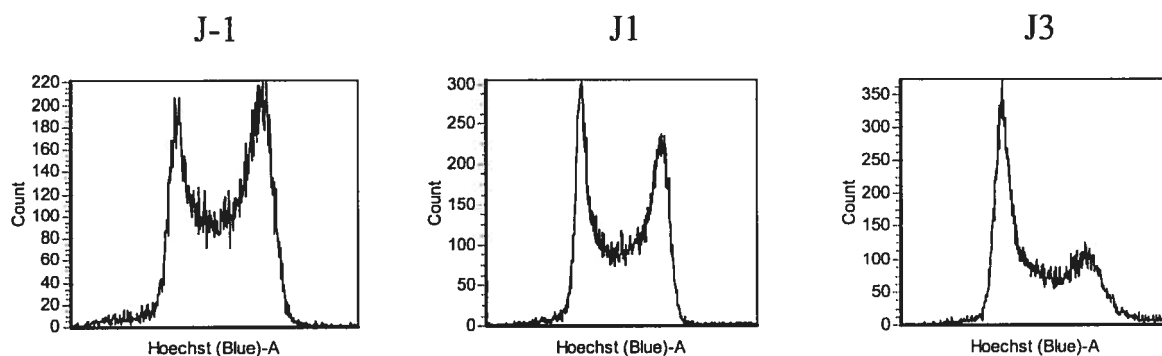


**Figure 28 : Contrôle d'efficacité du traitement 3C**

A) Configuration du locus du gène de la  $\beta$ -actine. Les flèches grises identifiées par 1, 2 et 4 représentent les amorces de PCR qui furent utilisées pour amplifier le produit de ligation résultant de la fixation des fragments génomiques HindIII adjacents. Les flèches grises identifiées par 2 et 3 représentent les amorces de PCR qui furent utilisées pour normaliser le signal selon la quantité d'ADN présente dans les diverses préparations d'ADN 3C. La figure n'est pas à l'échelle. B) Résultats de l'expérience décrite en A. Le signal 3C (provenant des amorces 1, 2 et 4) fut normalisé en utilisant le signal de quantité génomique (provenant des amorces 2 et 3). Le nom des préparations d'ADN 3C réfère aux noms du Tableau 1. Les nombres de copies normalisés furent arbitrairement définis pour permettre la comparaison des signaux résultant de la fixation des fragments HindIII adjacents. Barres d'erreur :  $\pm$ erreur moyenne.

Nous avons donc analysé l'état du cycle cellulaire des populations de cellules traitées à l'acide rétinoïque par cytométrie en flux avec marquage du contenu en ADN par le Hoechst

33342. La Figure 29 présente les résultats de cette analyse. Le nombre de cellules en phase S décroît au cours du protocole de différenciation, ce qui correspond bien avec le fait que la croissance des cellules est visiblement ralentie au cours du traitement à l'acide rétinoïque (observation non-illustrée). Les résultats montrent en effet qu'il existe une différence notable entre le jour -1 et le jour 3 au niveau du profil de cycle cellulaire des populations étudiées. Il est donc possible que la haute vitesse de dédoublement des cellules non-différenciées ait pour effet de raccourcir la phase G1 du cycle cellulaire, ce qui pourrait conduire à une moins grande décondensation des chromosomes et ainsi influencer le repositionnement observé du locus HoxB1 hors de son territoire chromosomique. Cette expérience n'exclut toutefois pas que des mécanismes plus directs puissent agir sur le positionnement nucléaire du locus HoxB1 durant la différenciation cellulaire.



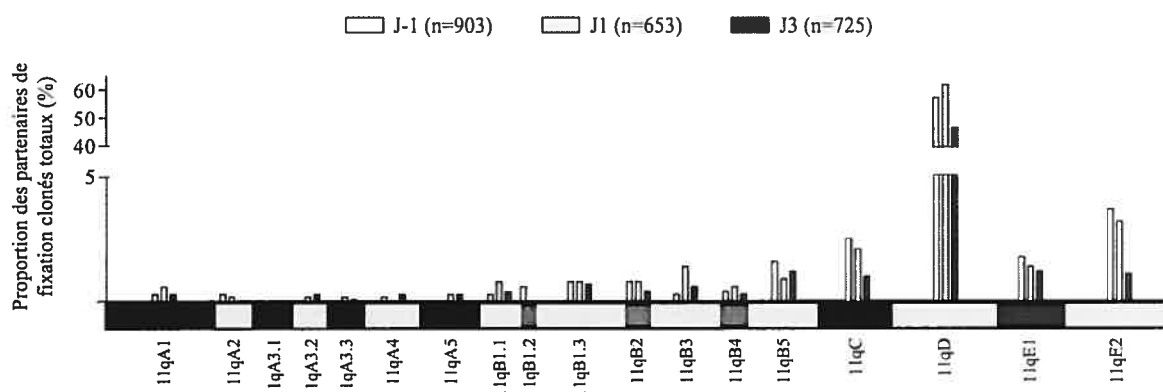
**Figure 29 : Analyse du cycle cellulaire des populations de cellules souches embryonnaires traitées à l'acide rétinoïque par cytométrie en flux avec le marqueur Hoechst 33342**

### **Distribution génomique des partenaires de fixation inter et intra-chromosomiques distaux**

La Figure 30 présente la distribution des partenaires de fixation intra-chromosomiques clonés tout au long du chromosome 11 de souris. Tel que nous pouvions nous y attendre à partir des analyses présentées plus haut, la bande chromosomique 11qD à l'intérieur de laquelle le complexe HoxB est situé est fortement représentée dans notre crible. Comme nous l'avons déjà brièvement mentionné, les proportions des clones de partenaires de fixation semblent diminuer lentement avec la séparation génomique de la bande 11qD.

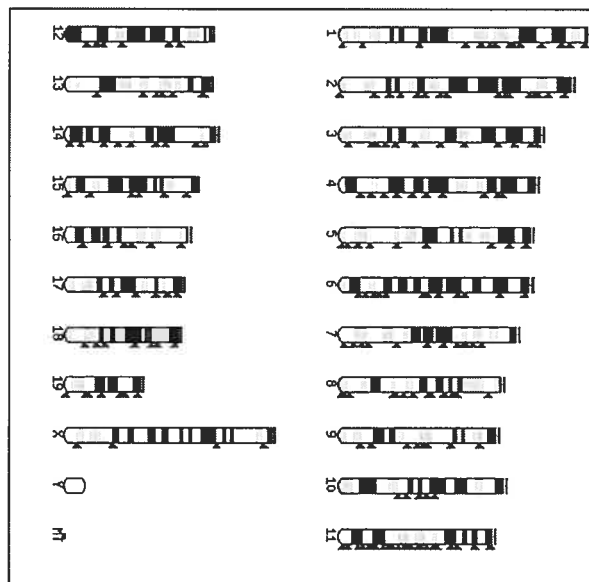


D'autre part, aucune région ou bande intra-chromosomique distale ne semble être significativement sur-représentée dans notre crible, peu importe son statut d'expression. La Figure 31 montre que les partenaires de fixation inter-chromosomiques semblent aussi être distribués sur les divers chromosomes en proportions qui sont approximativement compatibles avec la taille de ceux-ci. Les distributions des interactions du locus HoxB1 avec les loci distaux intra et inter-chromosomiques sont donc compatibles avec un ordre aléatoire.

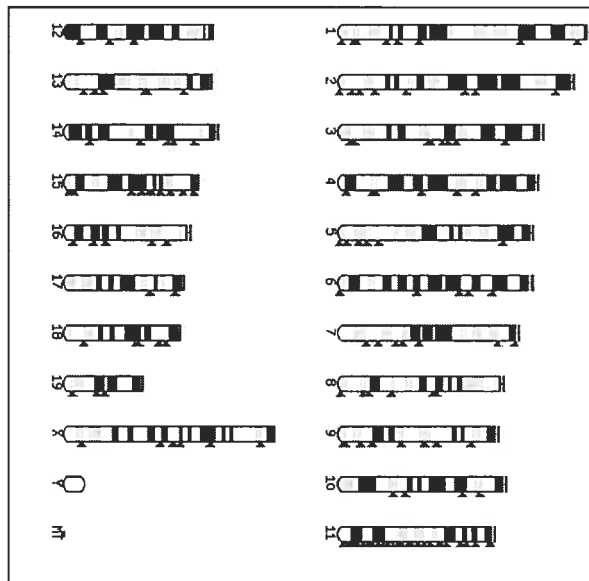


**Figure 30 : Distribution des proportions des clones de partenaires de fixation dans les bandes intra-chromosomiques**

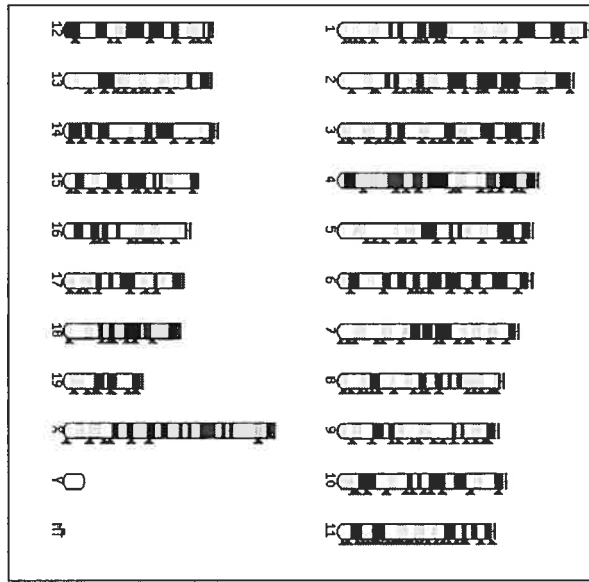
Les proportions sont calculées en divisant le nombre de clones de partenaires de fixation situés dans une bande donnée par le nombre total de clones récupérés pour chaque jour du traitement à l'acide rétinoïque. Le locus du gène HoxB1 est situé dans la bande 11qD.



J-1



J1



J3

**Figure 31 : Distribution des partenaires inter-chromosomiques de fixation récupérés par Inverse-PCR 3C sur le locus du gène HoxB1**

Chaque tête de flèche grise représente un partenaire récupéré par inverse-PCR 3C.

## DISCUSSION

### 1. Recombinaison entre séquences en répétition directe dans l'embryon de souris et son impact sur l'évolution des génomes de mammifères

#### **Description de l'essai**

Dans la première partie de cette thèse, nous avons décrit les activités de recombinaison homologues et illégitimes à l'œuvre dans l'embryon de souris grâce au développement d'un essai basé sur l'intégration au génome et la récupération subséquente d'un grand nombre de plasmides ayant ou non participé à des événements de recombinaison. Très peu d'études *in vivo* ont été réalisées à ce jour pour évaluer la fréquence et la nature des événements de recombinaison chez la souris. Il fut par exemple démontré que les événements de conversion génique sans crossing-over étaient fréquents chez les spermatozoïdes de souris (Murti et al. 1992), et que des activités efficaces de single-strand annealing existaient dans l'ovule de souris et dans l'embryon pré-implantation (Fiorenza et al. 2001; Pieper et al. 1992; Shimoda et al. 1991). D'autre part, l'existence d'arrangements de transgènes intégrés en répétition directe dans les lignées de souris transgéniques peut être expliqué par des événements de recombinaison homologue entre les transgènes avant ou durant leur intégration (Folger et al. 1982; Palmiter et Brinster 1986). Cependant, à notre connaissance, aucune quantification de la fréquence et de la nature des événements de recombinaison *in vivo* dans l'embryon de souris n'avait jusqu'à présent été rapportée dans la littérature. Il est aussi important de noter que notre essai examine les événements se produisant dans un embryon qui deviendra une souris viable et normale : notre description des événements est donc basée sur un essai qui ne perturbe pas indûment les processus cellulaires normaux.

L'essai présenté ici est basé sur l'intégration chromosomique d'un plasmide contenant deux fragments de séquences L1 arrangés en répétition directe dans l'embryon de souris. Nos résultats indiquent qu'avant ou pendant le processus d'intégration des plasmides au génome, ces derniers furent modifiés à haute fréquence par des activités de recombinaison homologue et/ou illégitime. L'essai nous a d'ailleurs permis de détecter tous les produits de recombinaison attendus. Il a été rapporté que la plupart des événements de recombinaison homologue dans les cellules en culture impliquant des plasmides micro-injectés se

produisent à l'intérieur d'une heure de l'injection (Folger et al. 1985). Il est donc probable que les événements décrits dans notre essai reflètent les activités présentes dans l'ovule fécondé ou durant les premières divisions cellulaires de l'embryon. De plus, la majorité des lignées transgéniques que nous avons générées ne sont pas des mosaïques, ce qui suggère aussi que les plasmides se sont intégrés tôt durant l'embryogénèse. Ces résultats sont compatibles avec le fait que globalement, moins de 30% des lignées transgéniques générées par micro-injection sont des mosaïques (Rijkers et al. 1994).

Notre essai ne se base pas sur la reconstitution d'un marqueur de sélection, contrairement à la plupart des essais de recombinaison rapportés dans la littérature, mais est basé sur la récupération par *plasmid rescue* de plasmides intégrés. Il permet donc théoriquement de récupérer un plus large spectre de recombinants, étant donné que des événements non-attendus peuvent aussi être récupérés. Cependant, certains plasmides fortement réarrangés, en particulier dans les séquences plasmidiques nécessaires au *plasmid rescue*, peuvent être indétectables par notre méthode. La procédure de *plasmid rescue* ne nous permet pas non plus de reconstruire les structures intégrées en tandem (par exemple, l'ordre dans lequel les plasmides sont organisés les uns par rapport aux autres) ou de déterminer avec certitude le nombre de sites d'intégration présents dans chaque souris. Le fait que nous ayons récupéré peu de jonctions (un fragment contenant à la fois des séquences génomiques et plasmidiques délimitant la fin de la structure intégrée) suggère par contre que la plupart des fondateurs ont un ou peu de sites d'intégration distincts. Quoi qu'il en soit, ces limitations n'affectent pas la détermination de la nature ou de la fréquence des différents produits de recombinaison générés lors de l'intégration des transgènes au génome des souris fondatrices.

Il convient de noter que notre essai n'introduit pas de CDB dans la région d'homologie entre les séquences L1 : le site de linéarisation des plasmides (avant la micro-injection dans les ovules fécondés) est en effet situé dans les séquences plasmidiques du pXH. Les événements que nous décrivons sont donc plus similaires à de la recombinaison spontanée qu'à de la recombinaison induite par une CDB. À cause de ce fait, notre essai ne biaise pas artificiellement l'étude en faveur de certains mécanismes de recombinaison, en particulier le SSA, comme nous en discuterons un peu plus bas.

### Fréquences de recombinaison homologue observées

Les résultats obtenus dans les souris fondatrices indiquent que les fréquences de recombinaison homologue sont très élevées avant ou pendant l'intégration des plasmides au génome des embryons de souris. Environ la moitié des molécules intégrées sont des recombinants homologues, avec des valeurs individuelles variant entre 7% et 70%, ce qui signifie que 100% des embryons de souris ont intégré des molécules recombinantes. Il est difficile de comparer adéquatement ces fréquences avec les études déjà publiées sur la recombinaison homologue extra-chromosomique dans des populations cellulaires de mammifère. En effet, les procédures expérimentales sont très variables d'une étude à une autre, et sont généralement différentes de celle que nous avons utilisée (Folger et al. 1985; Rubnitz et Subramani 1985; Rubnitz et Subramani 1986; Subramani et Rubnitz 1985; Wong et Capecchi 1986). Certaines comparaisons sont néanmoins possibles. Mentionnons d'entrée de jeu que la recombinaison extra-chromosomique se produisant avant ou pendant l'intégration d'un plasmide est beaucoup plus fréquente que la recombinaison chromosomique spontanée, qui se produit à des fréquences de l'ordre de  $10^{-6}$  à  $10^{-8}$  événements par cycle cellulaire chez la souris (Subramani et Rubnitz 1985). Pour ce qui est des fréquences de recombinaison extrachromosomiques rapportées dans la littérature, Subramani et Rubnitz ont montré qu'environ 15% des cellules de souris 3T6 transfectées avec un substrat de recombinaison étaient capables d'intégrer au moins une molécule recombinante, ce qui suggère une fréquence de recombinaison légèrement inférieure à celles que nous avons observées (Subramani et Rubnitz 1985). Le même groupe a aussi réalisé une autre étude dans laquelle un plasmide contenant deux séquences en répétition directe est transfecté dans des cellules de singe (COS) et est par la suite récupéré et transformé dans des bactéries (Rubnitz et Subramani 1985). Dans cet essai, environ 1 à 4% des molécules récupérées sont recombinantes après un passage de un à trois jours dans les cellules, ce qui semble aussi inférieur aux niveaux observés dans notre essai. Cependant, cet essai ne pouvait récupérer que les événements de délétions homologues, ce qui pourrait diminuer les fréquences observées. D'autre part, le plasmide transfecté était circulaire dans ces essais, ce qui pourrait aussi potentiellement rendre la recombinaison plus difficile que pour un plasmide linéarisé, et dont les extrémités sont reconnues par la machinerie de recombinaison cellulaire, comme c'est le cas pour notre essai.

Les résultats présentés plus haut semblent donc suggérer que les fréquences de recombinaison que nous avons observées dans l'ovule et l'embryon de souris sont particulièrement élevées. Par contre, le groupe de Folger et ses collaborateurs ont réalisé une étude dans laquelle l'événement de recombinaison étudié se produisait entre deux plasmides extrachromosomiques (Folger et al. 1985). Ces auteurs ont évalué que près de 100% des molécules micro-injectées participaient à au moins un événement de recombinaison extrachromosomique. Dans les faits, par contre, les événements de recombinaison recherchés se produisaient dans seulement 10% des cellules transfectées, et par extrapolation, les auteurs ont évalué que les autres séquences du plasmide utilisé devaient elles aussi participer à autant d'événements de recombinaison, ce qui leur a permis d'atteindre une fréquence de recombinaison théorique d'environ 100%. Si nous appliquons cette logique à nos résultats, il est fort probable que la grande majorité (sinon la totalité) des plasmides microinjectés dans les ovules fécondés dans notre essai ont participé à au moins un événement de recombinaison, bien qu'il soit difficile de démontrer ce fait hors de tout doute.

Une autre étude d'intérêt pour cette discussion fut publiée par le groupe de Fiorenza et collaborateurs (Fiorenza et al. 2001). Cette étude a étudié la recombinaison par single-strand annealing se produisant dans des ovules fécondés et des embryons de souris isolés. Leurs résultats montrent que les activités de recombinaison homologue par single-strand annealing sont environ dix fois plus actives dans l'ovule fécondé que dans l'embryon, alors que la formation de concatémères de plasmides, présumément par l'action de la recombinaison illégitime, ne se produit que dans l'embryon pré-implantation. Notons au passage que les concatémères pourraient aussi être générés par recombinaison homologue, comme ce fut déjà suggéré (Folger et al. 1982). Les fréquences de recombinaison par single-strand annealing furent évaluées à environ 12% des molécules dans l'ovule fécondé et à 1,5% dans l'embryon. Les auteurs expliquent ces différences par l'absence d'activité de recombinaison illégitime dans l'ovule, ce qui laisserait toute la place à la recombinaison homologue. Ces observations pourraient donc expliquer en partie les très hautes fréquences de recombinaison homologue observées dans notre essai en comparaison aux événements de recombinaison illégitime. En effet, la majorité des événements de recombinaison se produisent à l'intérieur d'une heure post-injection dans les cellules de souris, ce qui suggère que les événements que nous observons se sont produits majoritairement dans

l'ovule fécondé ou dans les premières divisions cellulaires de l'embryon, et donc potentiellement en absence de recombinaison illégitime (Folger et al. 1985). D'autre part, le single-strand annealing semble très actif dans l'ovule fécondé : ce mécanisme pourrait ainsi expliquer une partie des recombinants contenant des délétions homologues que nous avons récupérés (voir plus bas). Globalement, les fréquences de recombinaison observées dans notre essai semblent donc supérieures ou égales à ce qui a été rapporté dans la littérature.

Une raison qui pourrait expliquer la haute incidence de molécules recombinantes chez les plasmides intégrés pourrait être que les molécules qui s'intègrent au génome sont plus sujettes à participer à des événements de recombinaison homologue que celles qui ne s'intègrent pas. En effet, ce type d'effet a déjà été rapporté en cellules en culture (Dellaire et al. 2002; Folger et al. 1982). De plus, nous avons observé plusieurs molécules intégrées qui semblent avoir participé à plus d'un événement de recombinaison, ce qui suggère une fois de plus que les molécules qui s'intègrent ont un grand potentiel de recombinaison (réarrangements de la Figure 17A, B et C) ou inversement, que les molécules extrachromosomiques qui se recombinent sont plus susceptibles de s'intégrer. Il fut proposé que l'intégration de l'ADN étranger se produise généralement à des sites de dommage à l'ADN chromosomique et soit médiée par la machinerie cellulaire de réparation de l'ADN (Wurtele et al. 2003). Les facteurs de recombinaison attroupés autour de la lésion de l'ADN et affairés à sa réparation pourraient donc préférentiellement modifier les plasmides lors de leur intégration.

La variation de fréquence de recombinaison d'environ 10 fois qui existe entre les différents fondateurs semble indiquer que certains embryons possédaient un potentiel de recombinaison plus élevé que d'autres. La raison de ces différences est inconnue mais il est possible que certains embryons aient mis plus de temps avant de commencer à se diviser, ce qui aurait pour résultat de donner plus de temps à la recombinaison pour agir sur les plasmides sans subir de compétition de la part de la recombinaison illégitime qui commence à être active seulement dans l'embryon. D'autre part, dans les cas où la recombinaison des plasmides fut couplée à leur intégration, des différences dans la phase du cycle cellulaire au moment de l'intégration pourraient expliquer les variations observées. En effet, comme nous l'avons mentionné la phase du cycle cellulaire a un impact sur la



disponibilité des facteurs de recombinaison, la phase G2/S ayant le meilleur potentiel pour la recombinaison homologue (Johnson et Jasin 2000; Takata et al. 1998).

### **Fréquences des différents produits de recombinaison**

Les résultats présentés ici décrivent l'aboutissement d'une compétition entre les différents types d'activités de recombinaison présents dans l'ovule fécondé et l'embryon de souris. Plusieurs études conduites à ce jour dans la littérature et décrivant les activités de recombinaison présentes dans ces systèmes furent développées pour investiguer exclusivement certains mécanismes, comme le single-strand annealing ou la recombinaison illégitime (Fiorenza et al. 2001; Pieper et al. 1992; Shimoda et al. 1991). En effet, ces essais créent généralement une CDB dans la région d'homologie et/ou ne permettent que la récupération d'événements reconstituant un marqueur fonctionnel. Ceci biaise donc inévitablement l'étude et ne permet pas de recueillir tous les produits de recombinaison possibles : par exemple, un substrat de SSA conduira inévitablement à des délétions homologues. Il n'est donc pas possible de comparer directement notre étude à celles mentionnées précédemment à cause de ce fait.

L'approche que nous avons utilisée nous permet de comparer l'incidence relative des différents produits de recombinaison dans un seul et même essai. Globalement, pour les 15 fondateurs, les délétions homologues se sont produites plus fréquemment que les événements de conversion génique par un facteur de 5 pour 1. Cet état de fait est aussi apparent lorsque les fondateurs sont comparés sur une base individuelle : 11 fondateurs présentent une incidence plus élevée de délétions homologues que de conversion génique sans crossing-over alors que la situation inverse n'est retrouvée que dans 2 fondateurs. Ces résultats sont compatibles avec des études en cellules de souris en culture qui ont démontré que la recombinaison homologue extra-chromosomique se produisant entre des séquences en répétition directe donnait naissance à des délétions homologues 10 fois plus souvent qu'à des conversion géniques sans crossing-over (Rubnitz et Subramani 1985; Rubnitz et Subramani 1986). Dans un contexte chromosomique, il fut aussi rapporté que la recombinaison spontanée chromosomique entre séquences en répétition directe résultait généralement en des délétions homologues dans les cellules d'ovaire de hamster chinois (Bill et Nickoloff 2001). De façon intéressante, d'autres résultats publiés montrent aussi

que les événements de réparations induits par une brisure double-brin dans une séquence en répétition directe donnaient aussi majoritairement naissance à des délétions homologues (Liang et al. 1998). Cependant, une autre étude réalisée dans des cellules L de souris a suggéré que ce type de recombinaison spontanée résultait en des produits de conversion génique sans crossing-over dans plus de 80% des cas (Liskay et al. 1984). Les raisons de ces variations sont peu claires, mais nos résultats sont évidemment plus compatibles avec les premières études mentionnées plus haut. Il est fort possible que ces différences dépendent des systèmes cellulaires et des substrats de recombinaison utilisés. Par contre, notre essai possède certaines limitations qui sont en particulier liées au fait que la recombinaison spontanée est induite par un mécanisme inconnu de l'investigateur, et que l'endroit précis d'initiation de la recombinaison dans la séquence étudiée est par le fait même inconnu. Il est donc possible que certains événements de conversion génique impliquant les régions non-couvertes par les sites de restriction caractéristiques utilisés dans notre essai n'aient pu être détectés et que la fréquence de conversion sans crossing-over puisse avoir été sous-évaluée dans notre essai à cause de ce fait.

Les événements inter-moléculaires d'amplification ne représentent pour leur part qu'environ 1% des événements totaux. Ces observations suggèrent que la recombinaison intramoléculaire fut favorisée largement dans notre essai; ceci à des implications au niveau des mécanismes possibles ayant généré les recombinants. En effet, les événements de single-strand annealing et d'invasion inter-moléculaires devraient normalement conduire à un nombre plus ou moins équivalent de délétions homologues et d'amplifications. La faible fréquence d'événements inter-moléculaires est donc une des raisons qui explique le fait que nous ne croyons pas que le SSA soit le mécanisme le plus important dans notre essai, comme nous en discuterons plus bas.

La haute fréquence de délétions homologues observée *in vivo* dans l'embryon peut sembler suprenante à première vue, étant donné que ces événements peuvent résulter en des réarrangements délétères surtout s'ils impliquent des séquences homologues non-alléliques comme des éléments répétitifs. Par contre, la recombinaison chromosomique est beaucoup moins fréquente que la recombinaison extra-chromosomique. À cause de ce fait, même si les événements de délétion homologue étaient plus fréquents que les événements de conversion génique sans crossing-over dans l'ovule fécondé et l'embryon précoce, les

effets problématiques de ce biais seraient probablement mis en échec par leurs fréquences absolues faibles. Mentionnons aussi que la présence d'un très grand nombre de séquences homologues en répétition directe dans le noyau après la micro-injection, et par la suite dans la structure intégrée aux chromosomes, est certainement un facteur qui favorise une haute fréquence de recombinaison dans notre système. Ceci étant dit, des études récentes indiquent qu'il existe des variations à grande échelle dans le nombre de copies de plusieurs régions génomique chez l'humain et la souris, résultant apparemment de délétions et d'amplifications (Iafrate et al. 2004; Li et al. 2004). Ceci démontre bien que le génome peut être surprenamment dynamique et s'adapter à des pertes ou gains importants de matériel.

De façon contrastante avec l'incidence élevée de recombinants homologues, nous avons observé peu de réarrangements non-homologues parmi les plasmides intégrés, et plusieurs de ces réarrangements sont en fait des jonctions transgène-génome. Cette faible fréquence de réarrangements de plasmides observée dans notre système est cependant compatible avec le fait que les activités de recombinaison non-homologue semblent peu actives dans l'ovule fécondé et devenir plus actives dans l'embryon précoce (Fiorenza et al. 2001). Cependant, cette évaluation pourrait être une sous-estimation étant donné que la méthodologie de *plasmid rescue* que nous avons utilisée ne nous permet pas de récupérer des plasmides dans lesquels les séquences plasmidiques furent fortement réarrangées ou délétées. Il est aussi opportun de noter que bien qu'il ait été proposé que la recombinaison homologue soit impliquée dans la formation d'arrangements en répétitions directes de transgènes intégrés, nos résultats ne nous permettent pas d'exclure que la recombinaison illégitime ait participé dans la formation des structures d'intégration de transgènes.

### **Mécanismes de recombinaison**

La nature exacte des mécanismes et des intermédiaires de recombinaison impliqués dans la recombinaison spontanée est inconnue. Cependant, la plupart des molécules micro-injectées dans notre essai contiennent des brisures simple-brin qui résultent de leur préparation. Il est donc plausible qu'un mécanisme initié par les brisures simple-brin ou *nicks* tel que le modèle classique de Meselson-Radding (Meselson et Radding 1975) puisse avoir été impliqué (voir Figure 5). Ce modèle peut d'ailleurs expliquer une grande proportion des produits obtenus par recombinaison intra-moléculaire. D'autres mécanismes peuvent être

invoqués, mais demandent tous l'ajout d'étapes additionnelles, comme la recircularisation des plasmides micro-injectés et la création de nouvelles brisures double-brin dans la région d'homologie pour initier la recombinaison. De plus, ces mécanismes doivent nécessairement être inter-moléculaires. Ainsi, il est possible qu'après une dégradation plus ou moins extensive d'un des côtés du plasmide, une séquence L1 du plasmide puisse servir à l'invasion d'une autre séquence homologue, qu'elle soit sur un plasmide linéaire ou sur un chromosome. Ce type de mécanisme, soit le *one-sided invasion*, peut effectivement avoir joué un rôle dans notre essai puisque nous avons retrouvé un événement de cette nature impliquant une séquence L1 endogène et un plasmide dans la souris 4475 (Figure 17). Par contre, ce type d'événement, lorsqu'il implique des interactions entre plasmides, devrait normalement conduire à des fréquences similaires d'amplifications et de délétions étant donné que l'invasion peut se produire dans une ou l'autre des séquences L1 du plasmide « donneur ». Les amplifications inter-moléculaires ne sont donc pas plus complexes à générer que les délétions homologues inter-moléculaires par ces mécanismes d'invasion inter-moléculaire. Il est donc difficile d'envisager que ces événements puissent expliquer la majorité des événements observés puisque nos résultats suggèrent que la majorité des événements que nous avons observés sont intra-moléculaires.

Une autre possibilité voudrait que le single-strand annealing soit responsable d'une majorité des événements de délétion homologue que nous avons observés (Lin et al. 1984; Lin et al. 1990a; Lin et al. 1990b) (voir Figure 3). Cependant, le SSA ne peut pas générer simplement les événements de conversion génique sans crossing-over que nous avons observés, en particulier les événements PHH (voir Figure 14). De plus, comme nous l'avons mentionné précédemment et par le même raisonnement que pour le *one-sided invasion*, si le SSA avait été le mécanisme le plus important dans notre système, nous nous serions attendus à obtenir environ le même nombre d'amplifications que de délétions. Or, nous n'avons pas observé cette équivalence. Finalement, le SSA présuppose une dégradation extensive des plasmides avant leur appariement, mais nous avons retrouvé peu de plasmides extensivement dégradés dans les structures intégrées (les plasmides utilisés peuvent être dégradés de plusieurs kb sur un des côtés sans déléter de séquences nécessaires au *plasmid rescue*). Ces considérations suggèrent donc que bien que le SSA ait pu jouer un rôle, il n'est probablement pas le seul mécanisme ayant généré les événements de délétions homologues que nous avons observés. Il est donc probable que plusieurs mécanismes

distincts de recombinaison homologue aient pu participer à la génération des patrons de recombinants observés.

### **Recombinaison avec L1 endogènes**

L'événement d'intégration qui s'est produit dans la souris fondatrice 4475 est d'un intérêt particulier, parce qu'il implique une interaction homologue entre un des fragments L1 du plasmide pXH et une séquence L1 chromosomique (Figure 17). Le mécanisme le plus plausible pour expliquer ce recombinant est le *one-sided invasion* (Belmaaza et Chartrand 1994; Richard et al. 1994) (voir Figure 6). Ce mécanisme permet à une extrémité simple-brin d'un plasmide extra-chromosomique d'envahir une séquence double-brin homologue et d'amorcer une synthèse d'ADN. Après la synthèse, qui peut continuer au-delà de la région d'homologie entre les séquences impliquées, l'ADN envahisseur peut soit s'intégrer au site d'invasion ou être relâché, auquel cas l'intégration peut se produire à un autre endroit dans le génome, probablement avec l'aide de la machinerie de recombinaison non-homologue. La réparation de brisures double-brin chromosomiques par ce mécanisme fut d'ailleurs observée dans les cellules souches embryonnaires de souris (Richardson et al. 1998). Ces résultats montrent que les séquences répétitives L1 peuvent participer à des événements de recombinaison homologue dans l'embryon de souris, et suggèrent que le *one-sided invasion* entre séquences répétitives est possible *in vivo*. Étant donné que ce mécanisme permet le transfert de séquences à l'intérieur et à l'extérieur de la région d'homologie, la recombinaison homologue *in vivo* par ce mécanisme dans l'embryon pourrait avoir joué un rôle dans l'évolution concertée des éléments L1 et dans le transfert de séquences d'un endroit à l'autre dans le génome.

### **Modification de la structure intégrée au fil des générations et impact sur l'évolution des génomes de mammifères**

Nos résultats montrent que les structures d'intégration en tandem peuvent être modifiées par recombinaison homologue au fil des générations de souris. Bien qu'il soit difficile d'évaluer quantitativement la fréquence de ces événements en utilisant notre essai, le fait qu'il existe des différences statistiquement significatives entre les distributions de fréquences des différentes souris dans trois lignées sur cinq suggère que ces modifications

chromosomiques sont relativement fréquentes. Le fait que certaines structures intégrées soient instables est documenté dans les cellules en culture et dans les lignées de souris transgéniques (Wurtele et al. 2003). La recombinaison homologue entre les différentes copies d'un arrangement en tandem a aussi été suggérée comme étant capable de jouer un rôle dans la fluctuation dans le nombre de copies de transgènes dans les différentes cellules d'une population (McBurney et al. 2001). De plus, il fut suggéré que le phénomène de « *repeat-induced silencing* » ou répression induite par les séquences répétées soit possiblement dépendant des interactions homologues entre les copies de transgènes intégrées (Marahrens 1999). Dans une de nos lignées transgéniques (fondateur 4400), les différences entre les arrangements intégrés des souris de la lignée semblent être dues à des événements de délétion homologue. Ces résultats indiquent donc que des L1 intégrés aux chromosomes peuvent se recombiner soit dans les cellules germinales des fondateurs ou dans l'embryon des descendants pour modifier les arrangements de plasmides intégrés de façon héritable.

## Conclusions

Nous avons mis au point un nouvel essai qui nous a permis de caractériser la recombinaison se produisant *in vivo* dans un embryon de souris viable et nous avons montré que tous les produits attendus de recombinaison pouvaient se produire et être détectés par cet essai. Les fréquences de recombinaison observées sont très élevées et les événements de délétions homologues sont 5 fois plus fréquents que les événements conservatifs de conversion génique sans crossing-over. Finalement, les L1 endogènes et transgéniques peuvent participer à des événements de recombinaison qui donnent lieu à des modifications héréditaires du génome. Il pourrait être très intéressant de reprendre ces essais avec des souris possédant des mutations dans divers gènes impliqués dans le cancer ou la réponse aux dommages à l'ADN tels que WRN, BLM ou p53. Nous pourrions alors comparer la recombinaison se produisant *in vivo* dans l'embryon entre les souris normales (cette étude) et les mutants. Des souris mutantes pourraient ainsi présenter des patrons de recombinants quantitativement ou qualitativement différents de ceux des souris normales. Cet essai est donc un outil puissant qui permet d'investiguer des cellules embryonnaires rares dans lesquelles il est traditionnellement difficile de réaliser des expériences.

Tous les événements de recombinaison décrits plus haut sont probablement favorisés par la proximité entre les séquences impliquées dans les événements de recombinaison, proximité artificiellement induite dans le cas de ce système expérimental. Cette proximité entre séquences homologues influence donc le choix d'un partenaire de recombinaison. La prochaine section discute donc des résultats de REM-3C qui nous permettent de penser que dans le noyau de cellules humaines, le choix d'un partenaire de recombinaison est influencé non seulement par son environnement proximal linéaire mais aussi par son environnement spatial génomique.

## **2. Lien entre l'environnement spatial d'un locus et le choix de ses partenaires de recombinaison**

### **La méthodologie du REM-3C**

Les mouvements de chromatine dans l'espace nucléaire sont restreints, ce qui implique que certains loci génomiques devraient être plus fréquemment à proximité spatiale d'un locus donné que d'autres. La plupart des interactions fonctionnelles impliquant ce locus, comme les événements de recombinaison, devraient donc être relativement restreintes à cet environnement spatial. Nous avons utilisé une méthodologie ouverte permettant de déterminer quels loci sont retrouvés à proximité d'un locus situé sur le chromosome 2 humain pour lequel les partenaires fréquents de recombinaison sont connus. Cette partie de thèse visait donc d'une part à vérifier si l'environnement spatial d'un locus pouvait effectivement avoir un impact sur le choix des partenaires de recombinaison. D'autre part, un objectif important de cette étude visait à développer des technologies permettant l'analyse ouverte de l'environnement spatial d'un locus donné par 3C. Cette étape de développement nous a donné l'opportunité de définir comment ce type d'analyse pouvait être utilisé et quelles en étaient les limitations. Nous discuterons donc des implications fonctionnelles des résultats obtenus dans cette section, mais étant donné le nombre restreint d'événements clonés, nous réserverons une discussion plus en profondeur des modèles d'organisation nucléaire compatibles avec ces observations pour la section concernant l'analyse de l'environnement spatial dynamique du gène HoxB1. En effet, les résultats que nous avons obtenus sont conceptuellement comparables pour ces deux sections, mais

l'analyse que nous avons faite du gène HoxB1 est beaucoup plus étendue et permet de ce fait une discussion fonctionnelle plus approfondie des résultats.

### **Choix du système d'étude**

Nous avons utilisé deux lignées cellulaires distinctes dans les expériences de REM-3C. Étant donné qu'un des buts de cette étude était d'investiguer les liens existant entre l'environnement spatial d'un locus donné et son choix de partenaires de recombinaison, nous avons utilisé la lignée cellulaire ATA67 qui fut récemment le sujet d'une étude investigant le choix des partenaires de recombinaison d'un locus génomique donné publiée par notre laboratoire (D'Anjou et al. 2004). Ce clone provient de la lignée de fibroblastes humains GM005849, qui est déficiente pour le gène ATM. Cette protéine, comme nous l'avons vu en introduction, est impliquée dans plusieurs étapes de reconnaissance et de signalisation de la réponse aux dommages à l'ADN. Conséquemment, sa déficience résulte en une dérégulation de la réparation de l'ADN et en une instabilité génomique. Entre autre, la fréquence de recombinaison homologue est presque 40 fois plus élevée dans ces cellules (Meyn 1993). Le clone utilisé comporte aussi l'intégration d'un plasmide contenant une séquence L1 modifiée pour pouvoir y introduire spécifiquement une brisure double-brin par l'introduction de l'enzyme I-SceI dans les cellules. Cette enzyme possède un site de reconnaissance de 18 pb, qui est normalement absent du génome des mammifères, ce qui résulte donc en une CDB site-spécifique. L'introduction de ce vecteur nous a permis de différencier les deux allèles du locus investigué dans cette lignée cellulaire. D'autre part, nous avons voulu vérifier si le même locus présent dans une lignée « normale » présenterait des différences d'accessibilité par rapport à la lignée ATM-déficiente ATA67. En effet, une hypothèse plausible voudrait que l'instabilité génomique observée dans la lignée ATA67 puisse être liée à une plus grande accessibilité au génome, résultant en une plus grande diversité de partenaires de recombinaison. Nous avons donc, pour les besoins de la comparaison, utilisé une autre lignée de fibroblastes normale et non-immortalisée, la lignée humaine IMR90.



### Localisation des partenaires de REM-3C

Les résultats que nous présentons en Figure 19 montrent que certains produits de ligation 3C sont amplifiés dans plus d'un PCR indépendant. Étant donné que seuls deux produits de ligation peuvent être générés par cellule, ceci signifie que certains produits spécifiques proviennent de plusieurs cellules distinctes de la population étudiée. Ces résultats suggèrent donc que l'environnement immédiat d'un locus est relativement défini et est partagé par plusieurs cellules individuelles d'une population. Donc, bien que la chromatine possède une certaine mobilité, l'identité des loci entrant en contact le plus fréquemment avec un locus donné est conservée dans une population cellulaire. Au niveau strictement technique, ce résultat suggère que la méthodologie ne donne pas majoritairement lieu à des amplifications aléatoires impliquant le locus étudié, étant donné que plusieurs produits d'amplification spécifiques sont obtenus de façon reproductible. Une application potentielle du REM-3C pourrait être de créer des « codes-barre » de l'environnement spatial de divers loci génomiques : il pourrait par exemple être intéressant de comparer ces loci dans plusieurs situations cellulaires (phase du cycle, expression de gènes ou autres) à l'aide de cette méthodologie. Ceci pourrait faciliter le repérage de modifications importantes dans l'environnement spatial d'un locus donné.

Nos résultats montrent que les interactions intra-chromosomiques sont favorisées par rapport aux interactions inter-chromosomiques dans les deux lignées cellulaires. De plus, les loci qui sont le plus fréquemment à proximité spatiale du locus investigué sont situés dans le même bras chromosomique que ce dernier. Finalement, la bande chromosomique du locus investigué, et plus particulièrement un domaine couvrant environ 1 Mb de chaque côté du locus d'intérêt, lui semble particulièrement accessible. Nous avons vérifié si cette préférence pour les interactions avec des loci situés à l'intérieur d'une distance génomique d'environ 1 Mb était une caractéristique répandue à travers le génome par inter-L1 3C. De façon concordante avec les résultats obtenus par REM-3C sur le locus du chromosome 2p22.1, 16 interactions intra-chromosomiques sur 20 impliquaient des partenaires situés à moins de 1 Mb l'un de l'autre. De plus, une majorité des interactions récupérées par inter-L1 3C étaient intra-chromosomiques. Encore une fois, ces résultats suggèrent fortement que la méthodologie développée est spécifique et non aléatoire, étant donné que des ligations

aléatoires ne devraient que très rarement impliquer des loci proximaux intra-chromosomiques.

### **Modèles d'organisation de la chromatine en domaines spatiaux**

La distribution des partenaires intra-chromosomiques que nous observons dans cette étude est compatible avec plusieurs observations et modèles proposés dans la littérature. En effet, les chromosomes et certains domaines, tels que les bras et les bandes chromosomiques, occupent des territoires distincts dans l'espace nucléaire (Cremer et Cremer 2001; Dietzel et al. 1998). À plus haute résolution, des modèles postulent l'existence de domaines ou boucles de chromatine d'environ 1 ou 2 Mb (Munkel et al. 1999; Yokota et al. 1995). Nos résultats sont donc en accord avec ces observations. De plus, ils concordent en général avec les observations réalisées par inverse-PCR 3C au locus *HoxB1* de souris. Nous réserverons donc une discussion plus en détail de ces modèles pour cette prochaine section.

L'organisation interne des boucles de chromatine d'une taille de l'ordre de 1 Mb est considérée comme étant aléatoire dans le modèle de Yokota et collaborateurs (Yokota et al. 1995). À cause du fait que le REM-3C dépend de la présence de séquences L1 suffisamment homologues à l'amorce consensus utilisée, et de la présence d'un site *BglII* convenablement positionné, il nous est impossible de déterminer si tel est effectivement le cas dans notre système expérimental. En effet, seul un nombre restreint de loci situés à l'intérieur du domaine proximal génomique observé pouvaient être récupérés par REM-3C (Tableau 8). Pour les mêmes raisons, il ne nous est pas possible non plus de déterminer si ce domaine est organisé en boucle et de déterminer avec précision sa taille génomique. Cependant, l'existence d'un domaine d'interactions et d'accessibilité préférentielle d'une taille de l'ordre 1 ou 2 Mb est supportée par le fait que les contacts intra-chromosomiques plus distaux sont beaucoup moins fréquents que ceux observés pour ces loci. En fait, aucun des partenaires situés en dehors du domaine couvrant les mégabases adjacents au site investigué ne fut récupéré dans plus d'un PCR indépendant. D'autre part, le partenaire le plus rapproché que nous ayons récupéré après le domaine de 2 Mb est situé à 8 Mb de distance, alors que plusieurs partenaires potentiels situés entre ces deux régions ne furent pas récupérés. Nous notons aussi que les partenaires situés à une distance de 500 kb environ du site de départ ne furent pas récupérés plus fréquemment que d'autres situés à près de 1

Mb : ceci suggère que plusieurs loci situés dans cette région proximale sont relativement également accessibles au site de départ. La méthodologie du REM-3C est donc apte à décrire les grandes lignes de l'organisation spatiale d'une région génomique, mais ne possède pas une résolution qui permette de définir précisément celle-ci pour des régions ayant des tailles en deçà de 500 kb à 1 Mb. Certaines régions très riches en séquences L1 pourraient cependant être plus facilement investiguées par cette méthode.

L'environnement génomique des partenaires de REM-3C préférentiels situés à l'intérieur de 1 Mb du locus investigué fut analysé. Cependant, ces séquences ne possèdent aucune caractéristique particulière ou commune (contenu GC, gènes, etc...). Il est donc fort probable que l'élément limitant la récupération de régions génomiques est la présence de séquences L1 en bonne configuration. Il est cependant notable que les partenaires préférentiels que nous avons récupérés soient les mêmes dans les deux lignées cellulaires. Ceci suggère que l'environnement immédiat du locus investigué est similaire dans ces deux lignées. Ces deux lignées sont d'ailleurs des lignées de fibroblastes, ce qui pourrait expliquer les similitudes. L'étude de Yokota et collaborateurs (Yokota et al. 1995) a investigué plusieurs loci et a démontré que l'organisation de la chromatine en domaines de 1 ou 2 Mb était remarquablement similaire pour ces différentes régions. Notre étude va dans le même sens, et nos résultats d'inter-L1 3C suggèrent que l'organisation en boucle d'environ 1 à 2 Mb est répandue à travers le génome et est partagée par plusieurs lignées cellulaires.

### **Partenaires distaux et inter-chromosomiques**

Aucun partenaire récurrent ou préférentiel du locus 2p22.1 ne fut observé en dehors des loci du domaine de 2 Mb. Par contre, le nombre restreint d'observations ne nous permet pas vraiment de conclure quant à la spécificité (ou son absence) des interactions distales intra ou inter-chromosomiques. Nos résultats à cet égard ne sont cependant pas différents de ce qui serait attendu d'un comportement aléatoire. Ceci serait compatible avec plusieurs observations rapportées dans la littérature suggérant une organisation généralement aléatoire des chromosomes les uns par rapport aux autres (Cornforth et al. 2002; Walter et al. 2003). De façon intrigante, nos résultats suggèrent aussi que le locus investigué pourrait avoir accès à une plus grande fraction du génome dans les cellules ATA67 que

dans les cellules IMR90. En effet, la lignée ATA67 présente une plus grande proportion de partenaires inter-chromosomiques que la lignée IMR90. Les cellules ATA67 sont dérivées de la lignée cellulaire déficiente en ATM GM005849. Ces cellules sont caractérisées par des fréquences de recombinaison élevées, de l'aneuploïdie et de l'instabilité génomique. Le fait que le locus analysé ait un accès moins restrictif au reste du génome dans cette lignée que dans une lignée normale et non-immortalisée (IMR90) suggère que cette instabilité génomique pourrait potentiellement être liée à une accessibilité accrue à une grande variété de régions génomiques. Évidemment, ces observations gagneraient à être substantiées par des nombres d'événements plus élevés : les différences entre les deux lignées au chapitre des proportions d'interactions inter-chromosomiques atteignent de justesse un degré suffisant de significativité statistique de  $p=0,0265$  (test exact de Fisher). Il pourrait donc être intéressant de reprendre ces expériences avec la méthode de inverse-PCR 3C, qui est moins lourde techniquement, et qui permettrait d'obtenir un nombre d'observations plus important.

### **Impact de l'environnement spatial d'un locus sur ses choix des partenaires de recombinaison**

Les résultats que nous avons obtenus par la méthodologie du REM-3C soutiennent en général l'idée voulant que l'environnement spatial d'un locus génomique puisse influencer le choix des partenaires de recombinaison. En effet, l'étude réalisée précédemment sur ce locus a révélé que les partenaires intra-chromosomiques étaient plus fréquents que les partenaires inter-chromosomiques, que la majorité des partenaires de recombinaison étaient situés dans le même bras chromosomique que le site de brisure et que la distribution des partenaires inter-chromosomiques était compatible avec un ordre aléatoire (D'Anjou et al. 2004). De plus, la plupart des partenaires préférentiels de recombinaison étaient généralement situés à l'intérieur d'une distance de environ 500 kb à 1 Mb du locus endommagé. Nos résultats concordent avec ces observations : en effet, les loci les plus fréquemment utilisés pour la réparation de cassures double-brin sont situés dans la même région génomique que les partenaires préférentiels récupérés dans notre crible (Tableau 7). Ces partenaires sont en majorité situés du côté télomérique du site de départ, ce qui est aussi compatible avec les résultats obtenus par D'Anjou et collaborateurs (D'Anjou et al. 2004). Le fait que les partenaires préférentiels que nous avons identifiés et que les

partenaires préférentiels de recombinaison identifiés soient en majorité situés entre 470 et 940 kb de distance du site de départ suggère aussi que cet intervalle chromosomique est effectivement favorisé en ce qui a trait aux interactions avec le site de départ.

Un des partenaires préférentiels de l'étude de recombinaison ne fut pas récupéré par notre crible (partenaire 29272637). L'étude de recombinaison de d'Anjou et collaborateurs fut réalisée dans la lignée ATA67. Or, cette lignée possède un karyotype fortement réarrangé : il est donc possible que la région génomique contenant le partenaire 29272637 ait été en proximité linéaire avec le site de départ à cause d'un tel réarrangement. Étant donné que des résultats que nous avons obtenus dans cette lignée concernent en majorité l'allèle non-modifié de ATA67 et les allèles présents dans la lignée IMR90, il est possible que ce locus distal nous soit apparu comme étant moins accessible à cause de ce fait. La raison expliquant la faible fréquence de récupération d'événements impliquant l'allèle modifié d'ATA67 est par contre inconnue.

Ces résultats suggèrent donc que l'organisation spatiale « pré-brisure » d'une région génomique peut effectivement avoir une influence sur le choix des partenaires de recombinaison, et sont ainsi compatibles avec plusieurs études qui montrent que les loci impliqués dans des translocations sont régulièrement à proximité les uns des autres dans l'espace nucléaire (Kozubek et al. 1999; Kozubek et al. 1997; Nikiforova et al. 2000; Parada et al. 2004; Roix et al. 2003). Il pourrait être intéressant d'induire une brisure et de répéter l'expérience pour vérifier si l'environnement spatial du locus est modifié par ce traitement. En effet, en contraste avec le modèle qui suggère que la proximité entre loci est un prérequis à leurs interactions fréquentes par recombinaison, certaines études (Aten et al. 2004; Lisby et al. 2003) ont montré que cette proximité entre loci génomiques pouvait être induite par les dommages à l'ADN (voir Figure 11). Il est cependant probable que les loci qui sont préférentiellement regroupés dans une « usine de réparation » due à l'induction de dommages à l'ADN soient généralement déjà en relative proximité les uns des autres au préalable : les deux mécanismes ont donc possiblement un rôle à jouer dans la détermination de l'identité des partenaires de recombinaison d'un locus donné.

Ces observations et celles réalisées dans l'embryon de souris (section précédente) ont des implications importantes pour l'organisation linéaire du génome. En effet, comme nous

l'avons mentionné dans l'introduction, plusieurs gènes présentant des fonctions et patrons similaires sont souvent regroupés dans le génome. Ceci suggère qu'ils pourraient posséder des séquences régulatrices ou des exons communs. Or, il est probable que des événements de recombinaison entre ces séquences qui sont à proximité linéaire ou spatiale soient favorisés, et que ces interactions favorisent le partage de séquences en vase clos. Il s'agirait donc d'une forme de mécanisme d'homogénéisation régionale du génome. Une telle hypothèse suggérerait ainsi que certains complexes de gènes similaires possèdent des environnements spatiaux communs; il serait intéressant de vérifier si tel est effectivement le cas, en particulier dans les types cellulaires évolutionnairement pertinents comme les cellules embryonnaires et germinales. Dans cette ligne de pensée, il est déjà connu que les séquences L1 sont homogénéisées dans une espèce donnée. Il pourrait être intéressant de vérifier si en plus de cette homogénéisation globale des séquences répétitives, des homogénéisations de nature plus régionales existent. Cette façon de faire pourrait révéler l'existence de « voisinages spatiaux » à même la séquence du génome.

### **Usages et limitations de la méthodologie du REM-3C**

La méthode que nous avons utilisée est basée sur l'amplification de séquences inconnues qui sont fréquemment à proximité spatiale du locus investigué. Ce type de méthodologie n'avait, à notre connaissance, encore jamais été décrit précédemment. Au niveau technique, nos résultats indiquent que les produits de REM-3C ne sont pas aléatoires, et que certains domaines chromosomiques sont différentiellement accessibles au locus investigué. Cette méthode peut donc être utilisée pour obtenir une image globale des interactions se produisant entre un locus donné et le reste du génome, et ce sans avoir de connaissances préalables de l'environnement spatial de ce locus. Elle est donc utile pour générer des hypothèses quant à l'organisation spatiale d'un locus, hypothèses qui peuvent être testées par la suite par des méthodologies plus « dirigées » comme le FISH ou le 3C « classique ».

Cette méthode n'est cependant pas sans faille. Au niveau conceptuel, les problèmes les plus importants rencontrés lors de l'utilisation de cette technique viennent du fait que le nombre de loci récupérables par REM-3C est limité. En effet, un nombre restreint de séquences L1 ont une configuration adéquate pour être amplifiée par REM-PCR. D'une part, les séquences L1 génomiques n'étant pas parfaitement homologues entre elles, l'homologie à

l'amorce consensus L1 utilisée est souvent imparfaite. De plus, la position du site utilisé pour l'étape de 3C (BglII dans notre cas) est souvent limitante : ceci peut résulter en des produits de REM-PCR qui sont trop longs pour être amplifiés efficacement. En somme, toutes ces variables font que même si un grand nombre de loci peuvent être amplifiés, ceux-ci ne le seront pas avec une efficacité équivalente. Il peut donc être problématique de comparer la fréquence de récupération des différents loci entre eux pour cette raison. Il convient donc d'utiliser cette méthode pour décrire l'accessibilité de plus vastes régions génomiques à un locus d'intérêt. Finalement, au niveau technique, cette méthodologie est relativement lourde, étant donné que les clones positifs doivent être identifiés par hybridation *in situ*.

## Conclusions

La méthodologie du REM-3C est fonctionnelle et nous a permis d'évaluer les possibilités et les limitations potentielles des méthodologies de 3C ouvertes. Le REM-3C est particulièrement utile pour décrire les différences d'accessibilité entre de grandes régions du génome. Dans le cas qui nous occupe, nous avons pu remarquer qu'une région située entre 470 et 940 kb de distance en direction du télomère de notre site de départ était fréquemment à proximité spatiale de celui-ci. Les loci situés dans le Mb adjacent au locus de départ dans l'autre direction semblent aussi être plus fréquemment à proximité de celui-ci que les loci plus distaux. Les interactions distales intra et inter-chromosomiques sont elles aussi possibles, mais à une fréquence réduite. Nos résultats suggèrent donc que le locus investigué dans les cellules humaines possède un environnement spatial défini, dans l'ordre, par les Mb adjacents et par son bras chromosomique. Les résultats que nous avons obtenus par inter-L1 3C suggèrent que l'existence de domaines d'accessibilité d'environ 1 Mb est une caractéristique commune à plusieurs loci génomiques. Finalement, nos résultats suggèrent que l'existence de différents domaines d'accessibilité spatiale se traduit par des fréquences de recombinaison plus élevées entre le locus de départ et les loci génomiques situés dans un domaine proximal d'environ 2 Mb distribué autour du locus de départ. Ce locus recombine ensuite préférentiellement avec d'autres loci situés dans le même bras chromosomique que lui, alors que les loci situés dans l'autre bras chromosomique et le reste du génome lui sont accessibles de façon apparemment aléatoire.

Au niveau méthodologique, le nombre restreint des séquences L1 génomiques compatibles avec la méthodologie du REM-3C diminue la résolution de la technique et pourrait biaiser les résultats. De plus, la lourdeur technique associée au bruit de fond provenant du REM-PCR ne permet pas d'envisager une utilisation à grande échelle de cette méthode. Dans le but de pallier à ces inconvénients, nous avons décidé de mettre au point une nouvelle méthodologie moins dépendante d'un contexte génomique favorable et moins lourde au niveau technique. La discussion de la mise au point d'une telle méthode et de son application est le focus de la prochaine section.

### **3. Modifications de l'environnement spatial génomique du gène HoxB1 lors de l'induction de son expression dans le contexte de la différenciation de cellules embryonnaires de souris induite par l'acide rétinoïque**

#### **Développement de la méthodologie de l'inverse-PCR 3C**

Nous avons développé et utilisé une méthodologie de 3C ouverte basée sur l'inverse-PCR pour investiguer l'environnement spatial du gène HoxB1 durant l'induction de son expression dans le contexte de la différenciation de cellules souches embryonnaires de souris induite par l'acide rétinoïque. Cette méthodologie présente plusieurs améliorations significatives par rapport au REM-3C. L'utilisation de l'inverse-PCR réalisé avec une enzyme à site de reconnaissance fréquent (comme MseI) permet en effet la récupération d'un éventail beaucoup plus large de loci que ce qui est possible avec le REM-3C. Ceci est dû au fait que les produits d'inverse-PCR générés par cette méthode sont tous relativement courts à cause de la fréquence du site de reconnaissance de l'enzyme de restriction : il existe donc une très grande quantité de loci qui permettent ce type d'amplification. De plus, contrairement au REM-3C, les amorces utilisées sont complémentaires à la partie « fixe » du produit d'inverse-PCR : l'homologie entre les amorces et leurs cibles est donc toujours absolue. À cause de ceci, l'inverse-PCR ne donne lieu qu'à une très faible quantité d'amplification non-spécifique, ce qui fait que la grande majorité des produits clonés sont d'intérêt, contrairement au cas du REM-3C qui nécessite une étape d'hybridation *in situ* des colonies de bactéries. La méthode est donc très efficace au niveau technique.



## Choix du système d'étude

Nous avons décidé d'utiliser le système de cellules souches embryonnaires de souris OS25 pour plusieurs raisons. Étant donné que nous avons décrit une situation statique dans l'étude par REM-3C, un de nos objectifs était de caractériser les modifications à l'environnement spatial d'un locus qui étaient potentiellement liées à des changements dans son statut d'expression. L'induction de la différenciation des cellules OS25 par l'acide rétinoïque est bien caractérisée dans la littérature et résulte en l'expression du gène HoxB1 dans une fenêtre temporelle bien définie, ce qui représente évidemment un avantage indéniable : il nous est possible de réaliser une progression dans le temps (*time-course*), de façon à investiguer le gène et son environnement dans différents états transcriptionnels. D'autre part, il fut démontré très clairement que l'induction de l'expression de HoxB1 dans ce système était accompagnée par une décondensation de la chromatine du locus et par son repositionnement hors du territoire chromosomique (Chambeyron et Bickmore 2004). Ce système se comporte donc d'une façon relativement prévisible, ce qui facilite l'interprétation des résultats, et qui nous donne une forme de « contrôle positif » interne. En effet, étant donné que les méthodologies de 3C ouvertes n'avaient jusqu'à présent jamais été décrites dans la littérature, il nous semblait important de choisir un système qui pourrait nous confirmer la validité de l'approche. La composition dynamique de l'environnement du locus ne fut cependant pas caractérisée en détail dans l'étude de Chambeyron et Bickmore (2004) ayant décrit la décondensation du locus HoxB1 : il nous était donc possible d'apporter des précisions importantes à ce niveau tout en ayant certaines hypothèses vérifiables quant à la composition globale de cet environnement au long du traitement à l'acide rétinoïque.

Ce système cellulaire présente aussi l'avantage que les cellules souches embryonnaires sont généralement considérées « normales », ce qui suggère que les observations que nous avons réalisées sont probablement valables au niveau biologique et se rapprochent de la situation ayant cours *in vivo*. D'ailleurs, la décondensation et la relocalisation du complexe HoxB a été observée au cours de l'embryogénèse de souris (Chambeyron et al. 2005), ce qui suggère effectivement que plusieurs observations réalisées sur ce locus en cellules ES sont valables *in vivo*, comparativement à des systèmes de cellules en culture immortalisées par activation d'oncogènes viraux par exemple. Finalement, le système OS25 permet de

sélectionner les cellules indifférenciées ou différenciées à l'aide des agents G418 et ganciclovir, respectivement. Étant donné que les expériences de 3C sont réalisées sur des populations cellulaires, il est important que ces populations soient aussi homogènes que possible; la possibilité de sélectionner pour l'état de différenciation des cellules est donc intéressante dans cette optique.

### **Description de l'environnement spatial de base du gène HoxB1**

Nos résultats montrent qu'indépendamment de son statut d'expression, les interactions entre le locus HoxB1 et les loci intra-chromosomiques sont distribuées dans des domaines génomiques d'accessibilité différentielles. Le premier domaine est composé d'un intervalle d'environ 110 kb distribué autour du locus HoxB1 dans lequel la majorité des loci ont une accessibilité élevée au gène. Les interactions plus distales suivent une distribution biphasique dans laquelle le taux d'accumulation des partenaires diminue avec la distance génomique à l'intérieur d'un domaine de 800 kb et devient plus ou moins constant dans les environs des frontières de ce domaine situées à 300 kb en 5' et 500 kb en 3' du gène. Ceci suggère qu'après ces distances, un grand nombre de loci distribués sur de grands intervalles génomiques sont plus ou moins également et aléatoirement accessibles au locus HoxB1.

Les résultats que nous avons obtenus pour le domaine proximal d'interaction de HoxB1 sont globalement compatibles avec les données de 3C réalisées chez la levure (Dekker et al. 2002). En effet, la fréquence de fixation des loci semble généralement diminuer progressivement jusqu'à une distance de l'ordre de 75 à 100 kb du locus de départ. Cependant, il est intéressant de noter qu'aucune étude n'a, à notre connaissance, été en mesure de documenter globalement l'accessibilité de régions génomiques distales à un locus donné chez les mammifères. En effet, la majorité des études existantes visaient à tester des hypothèses bien précises sur des loci proximaux ou des loci distaux particuliers (par exemple (Osborne et al. 2004; Tolhuis et al. 2002)). D'autre part, l'étude originale réalisée chez la levure observait de très faibles fréquences de ligation entre un locus donné et des loci situés à plus de 100 kb. Après une certaine séparation du locus de départ, il devient en effet difficile de quantifier la fréquence de fixation des loci, étant donné la rareté des produits de ligation. Dans cette optique, notre méthode est intéressante puisqu'elle tient compte non seulement de la fréquence de ligation de loci individuels au locus d'intérêt,

mais aussi de leur densité dans une région génomique donné. Cette agglomération de deux variables nous permet d'obtenir une distribution des interactions qui existent entre le locus d'intérêt et le reste du génome.

Nos observations sont d'autre part généralement compatibles avec des données de FISH qui montrent que les distances spatiales entre loci ont une relation discontinue avec leur séparation génomique linéaire. Elles sont donc aussi compatibles avec les modèles de boucles de chromatine de 1 Mb et les modèles multi-boucles qui furent proposés pour expliquer ces résultats (Munkel et al. 1999; Sachs et al. 1995; Yokota et al. 1995). En effet, nous avons observé un domaine préférentiel de 110 kb situé à l'intérieur d'un domaine plus grand de 800 kb, ce qui est en accord avec la présence de sous-boucles à l'intérieur d'un domaine d'environ 1 Mb. De façon intéressante, nous avons récupéré une proportion similaire de clones de partenaires pour les côtés 5' et 3' du domaine proximal, mais sur une distance de 70 kb en 5' et de seulement 40 kb en 3'. Ceci suggère que la position du gène HoxB1 à l'intérieur du domaine proximal n'est pas tout à fait centrée. La même chose peut être dite du domaine distal de 800 kb, puisqu'il s'étend sur une distance de seulement 300 kb en 5' en comparaison de 500 kb en 3'. Nos résultats ne nous ont pas permis de confirmer si la chromatine de ces domaines est effectivement organisée en boucles. Par contre, il faut noter que des variations dans la compaction locale de la chromatine pourraient aussi avoir pour résultat de rendre certains loci plus rapprochés et plus accessibles au locus HoxB1 sans nécessairement invoquer un modèle de boucles (Yokota et al. 1997). Il pourrait être intéressant de reprendre les expériences d'inverse-PCR 3C ou de 3C classique en utilisant des loci situés en bordure des domaines observés : ceci pourrait nous éclairer sur la nature de l'organisation spatiale de ces domaines.

L'environnement génomique des frontières des domaines d'accessibilité à HoxB1 fut examiné. De façon intéressante, deux éléments de réponse à l'acide rétinoïque impliqués dans la régulation de HoxB1 sont situés dans le domaine proximal en 3' du gène, dans les deux fragments HindIII qui furent clonés le plus fréquemment dans notre essai (fragments 1901 et 5631 dans la Figure 24) (Huang et al. 2002; Marshall et al. 1996). Il existe aussi un élément de réponse à l'acide rétinoïque à l'intérieur du fragment HindIII qui fut utilisé comme « appât » dans nos expériences (Figure 22). Il est donc possible que la présence de ces éléments régulateurs puisse être impliquée dans l'accessibilité accrue du domaine

proximal 3' au gène HoxB1. En effet, les récepteurs à l'acide rétinoïque peuvent être liés à l'ADN en l'absence de ligand : il est possible de penser que des interactions protéines-protéines médiées par ces récepteurs pourraient favoriser les contacts entre ces loci. Il pourrait être d'intérêt de reprendre ces expériences dans des cellules n'exprimant pas les récepteurs à l'acide rétinoïque, ou dans des lignées possédant des mutations dans les éléments de reconnaissance mentionnés plus haut pour voir quel est le rôle de ces protéines dans la formation de l'environnement proximal du gène HoxB1. De façon intrigante, nous notons d'autre part que le gène HoxB6 est situé à seulement 3 kb de la frontière 5' du domaine génomique proximal. Il a été démontré chez le poulet que le complexe HoxB était séparé en deux groupes distincts : les gènes HoxB1-5 sont sensibles à l'acide rétinoïque alors que les gènes HoxB6-13 sont surtout régulés par le FGF (Fibroblast Growth Factor-Facteur de croissance des fibroblastes) (Bel-Vialar et al. 2002). Une possibilité excitante serait donc que le complexe de gènes HoxB soit séparé en domaines spatiaux fonctionnellement distincts, et que le gène HoxB6 soit situé à la frontière de ces domaines. Contrairement au domaine proximal, les limites des domaines de 300 kb en 5' et 500 kb en 3' ne semblent pas posséder de particularités évidentes. Il faut par contre mentionner que le nombre relativement restreint de clones de partenaires de fixation récupérés près de ces limites ne nous permet pas de les localiser avec une grande précision. Un crible à plus grande échelle pourrait donc être informatif à cet égard.

Nous avons vérifié la présence d'autres séquences fonctionnelles autour des frontières des domaines d'accessibilité que nous avons définies. En particulier, l'utilisation du logiciel MARscan (Rice et al. 2000) suggère la présence de MAR (Matrix attachment region) potentiels autour des limites des domaines génomiques proximaux et distaux. Cependant, étant donné leurs séquences dégénérées, des sites MAR potentiels furent aussi découverts à d'autres endroits génomiques situés à l'intérieur des domaines proximaux et distaux. La pertinence fonctionnelle de ces sites potentiels et le mécanisme impliqué dans la formation et la maintenance des domaines d'accessibilité du gène HoxB1 sont donc peu clairs pour le moment.

## **Modifications de l'environnement spatial du gène HoxB1 suite au traitement des cellules ES à l'acide rétinoïque**

L'analyse présentée en Figure 27 démontre que la distribution des clones de partenaires de fixation de HoxB1 à l'intérieur des domaines décrits précédemment est influencée par le protocole de différenciation induite par l'acide rétinoïque. L'accessibilité de HoxB1 au domaine proximal n'est pas dramatiquement affectée par son statut d'expression. Néanmoins, il semble y avoir une proportion légèrement plus basse de clones situés dans le domaine proximal 3' au jour 1 par rapport au jour -1 ou 3. Ceci semble être corrélé avec une augmentation dans l'accessibilité au domaine 3' distal. Il est donc possible que le domaine distal devienne temporairement plus accessible au gène HoxB1 peu de temps après son induction. Nous notons aussi que des mesures du coefficient de corrélation ( $R_2$ ) de la relation linéaire existant entre la distance de HoxB1 et la proportion de clones récupérés suggèrent que l'extrémité 5' du domaine proximal est moins bien définie au jour -1 qu'au jour 1 ou 3. En fait, la relation linéaire demeure robuste ( $R_2 > 0,98$ ) jusqu'à environ 100 kb en 5' pour le jour -1 alors qu'elle perd de sa pertinence après la limite de 70 kb en 5' pour les jours 1 et 3. Ceci suggère que la région située entre 70 et 100 kb en 5' de HoxB1 pourrait être plus accessible au gène au jour -1 qu'aux jours 1 ou 3. En accord avec ces résultats, cette région génomique contient le gène HoxB9, qui devient nettement séparé de HoxB1 lors de l'induction de ce dernier (Chambeyron et Bickmore 2004).

L'accessibilité de HoxB1 aux domaines distaux est clairement affectée par le traitement à l'acide rétinoïque. En fait, la proportion des clones représentant des partenaires de fixation situés à l'intérieur du domaine intra-chromosomique distal est significativement plus basse au jour 3 qu'au jour -1 ou 1, et cette diminution se traduit par une proportion significativement plus élevée d'interactions inter-chromosomiques. En contraste avec la situation observée au jour -1 ou 1, le locus HoxB1 est donc aussi souvent à proximité de loci inter-chromosomiques que de loci intra-chromosomiques au jour 3. Ces résultats sont compatibles avec la relocalisation documentée du gène HoxB1 à l'extérieur de son territoire chromosomique durant la différenciation de cellules ES induites par l'acide rétinoïque et durant le développement embryonnaire de souris (Chambeyron et Bickmore 2004; Chambeyron et al. 2005). Il est intéressant de noter que même dans les cellules non-différenciées, près du tiers des clones récupérés représentent des partenaires de fixation

inter-chromosomiques : ceci pourrait être dû au fait que le gène HoxB1 est généralement retrouvé en périphérie de son territoire chromosomique même lorsqu'il est inactif en transcription (Chambeyron et Bickmore 2004).

L'induction du gène HoxB1 précède l'augmentation de l'accessibilité aux loci inter-chromosomiques dans notre essai, puisque nous ne l'observons pas au jour 1 bien que HoxB1 soit clairement induit à ce moment (Figure 23). Cependant, la relocalisation du gène HoxB1 par rapport à son territoire chromosomique observée par FISH fut démontrée comme étant maximale seulement après 4 jours de traitement à l'acide rétinoïque (Chambeyron et Bickmore 2004). De plus, HoxB1 semble demeurer hors de son territoire chromosomique même après son expression. Étant donné que les résultats de 3C représentent une moyenne de l'organisation spatiale chromosomique d'une population cellulaire, il est possible que la proportion de cellules dans lesquelles le locus HoxB1 est situé hors de son territoire chromosomique ne devienne assez élevée pour influencer significativement nos résultats qu'après 3 jours de traitement, à un moment où une majorité de gènes HoxB1 sont exprimés ou l'ont été récemment. Pour tenter de répondre à ces questions, il pourrait être intéressant de reprendre les expériences aux jours 2, 4 et 7 du protocole de différenciation, et de voir si des valeurs intermédiaires sont atteintes au jour 2, et si la proportion d'interactions inter-chromosomiques tend à plafonner après la période d'expression maximale de HoxB1.

Il ne nous est pas possible de déterminer si l'effet que nous observons (l'augmentation de la proportion d'interactions inter-chromosomiques au jour 3) est directement lié à l'expression de HoxB1. En effet, les mécanismes impliqués dans la décondensation du complexe HoxB et le repositionnement nucléaire du gène HoxB1 ne sont pas entièrement clairs. Par exemple, bien que l'acétylation de la chromatine du locus (H3 lysine 9) soit augmentée de façon concomitante avec la décondensation, un traitement au TSA, qui résulte aussi en une hyperacétylation de la chromatine de cette région ne résulte pas en une visible décondensation du locus HoxB1. Il est intéressant de noter qu'une région beaucoup plus large que le gène HoxB1 est acétylée en réponse au traitement des cellules à l'acide rétinoïque. En effet, il fut démontré que l'acétylation était présente jusqu'au gène HoxB9 après 4 jours de traitement à l'acide rétinoïque, bien que ce gène ne soit pas exprimé à ce moment. Il est donc possible que cet état soit préparatoire à l'expression et puisse faciliter

celle-ci. Une autre possibilité voudrait que des activités spécifiques de décondensation soient impliquées: ces activités pourraient être similaires à celles qui agissent sur des gènes répondant aux facteurs BRCA1 ou au récepteur aux œstrogènes (Carpenter et al. 2004; Carpenter et Belmont 2004; Nye et al. 2002; Ye et al. 2001).

D'autres possibilités mécanistiques existent pour expliquer la décondensation et le repositionnement du locus HoxB1 hors de son territoire chromosomique lors du traitement à l'acide rétinoïque. Par exemple, étant donné que tous les gènes HoxB du complexe sont sensibles à des degrés divers à l'acide rétinoïque, il est probable que les gènes rapprochés de HoxB1 (B2-B5 par exemple) deviennent actifs peu de temps après HoxB1 lors du traitement à l'acide rétinoïque. Le positionnement de ces gènes par rapport à leur territoire chromosomique en réponse à l'induction de leur expression n'est pas connu. Il est cependant plausible de penser que ces gènes puissent être à leur tour repositionnés hors de leur territoire chromosomique. Le repositionnement de plusieurs gènes du complexe pourrait ainsi créer un effet d'entraînement et mener à des différences observables dans le positionnement nucléaire de HoxB1. Le fait que HoxB1 reste hors de son territoire après son expression pourrait d'ailleurs soutenir ce modèle.

Il fut par ailleurs démontré que les territoires chromosomiques de poulet deviennent plus diffus durant la différenciation cellulaire, ce qui résulte en un repositionnement de plusieurs loci à des positions plus externes de leurs territoires chromosomiques, et ce indépendamment de leur expression (Stadler et al. 2004). Il est donc possible que certains phénomènes liés à la différenciation cellulaire et non à l'expression de HoxB1 en tant que tel soient impliqués dans l'augmentation de la proportion des interactions inter-chromosomiques que nous observons au jour 3. Parmi ces phénomènes, nous notons que le rythme de croissance des cellules diminue au fur et à mesure de la différenciation cellulaire (Figure 29). Ce ralentissement semble se traduire par une augmentation de la proportion des cellules en G1 dans la population cellulaire du jour 3 par rapport aux jours -1 et 1. Ceci suggère que la phase G1 est potentiellement plus longue dans ces cellules. Or, la décondensation des chromosomes après la mitose se produit généralement au début de la phase G1 (Shelby et al. 1996). Il est donc possible que le chromosome contenant le complexe HoxB soit globalement plus décondensé au jour 3 qu'aux jours -1 et 1. Étant donné que le gène HoxB1 est initialement positionné à la périphérie de son territoire dans

les cellules non-différenciées, une décondensation globale de la chromatine de son chromosome pourrait faciliter l'extrusion du gène hors de son territoire chromosomique. Ceci n'exclut cependant pas que d'autres mécanismes plus directs puissent agir sur le positionnement du gène HoxB1. En effet, certaines observations suggèrent que cet effet ne peut expliquer tous nos résultats, notamment le fait que la diminution dans les interactions intra-chromosomiques distales au jour 3 ne soit pas également distribuée des deux côtés du gène (la diminution est clairement plus prononcée en 5'). Si la décompaction non-spécifique de la chromatine due à la phase G1 était seule responsable de la proportion plus élevée d'interactions inter-chromosomiques dans les cellules du jour 3, nous nous serions attendus à ce que la diminution dans les interactions intra-chromosomique soit distribuée de façon relativement égale dans tous les domaines. Ce n'est pas ce qui fut observé : il semble donc y avoir une certaine directionnalité dans cet effet. Pour obtenir plus d'informations sur ces différentes possibilités, il pourrait être intéressant de trier les cellules en utilisant un ou des marqueurs des différentes phases du cycle cellulaire qui seraient couplés à la GFP, pour obtenir des fractions correspondant à ces diverses phases du cycle, et de reprendre les expériences sur ces fractions.

Étant donné que HoxB1 est repositionné pendant ou après son expression, il faut se demander quel est le but poursuivi par le repositionnement du locus dans un nouvel environnement. Comme nous l'avons mentionné, il est possible que ce repositionnement ne soit qu'une conséquence passive de la décondensation du complexe HoxB et de son chromosome en général. En effet, la chromatine décondensée devrait être plus mobile, ce qui pourrait favoriser des contacts avec une plus grande variété de loci. Cette augmentation de mobilité résulterait en outre en une plus grande proportion des cellules dans lesquelles le gène serait dans l'espace inter-chromatine, où il pourrait interagir plus facilement avec des facteurs nécessaires à son expression. Il est donc possible que le repositionnement du gène HoxB1 hors de son territoire chromosomique soit une fin en soi, et que les interactions plus fréquentes avec les loci inter-chromosomiques ne soient effectivement qu'une conséquence « passive » au niveau mécanistique, en ce sens qu'elle ne serait pas dirigée vers une destination nucléaire particulière ou pour favoriser des contacts inter-loci spécifiques. Ce repositionnement serait néanmoins évolutionnairement justifié à cause de la facilitation de l'expression du gène résultant de son déplacement vers l'espace inter-chromatine.



Nos observations contrastent avec certaines études qui montrent que le repositionnement des loci par rapport à leur territoire chromosomique représente un état « préparatoire » (*poised state*) à l'expression, et se produit avant celle-ci (Ragoczy et al. 2003). Ce modèle plait à l'esprit : une décondensation de la chromatine avant l'expression pourrait permettre la relocalisation du locus en question vers l'espace inter-chromatine où il pourrait interagir avec des facteurs de transcription qui permettraient l'induction de l'expression. Par contre, il est aussi possible d'envisager une situation où l'induction de l'expression d'un gène pourrait permettre la nucléation d'une machinerie transcriptionnelle, et que celle-ci favoriserait graduellement le repositionnement du gène vers des positions de plus en plus externes par rapport au territoire chromosomique de façon à faciliter l'exportation des ARNm ou les contacts avec des facteurs de transcription et d'épissage.

Une autre hypothèse voudrait que le gène HoxB1 actif soit recruté vers des environnements plus permissifs à l'expression des gènes tels que des « usines de transcription ». Ainsi, même un recrutement tardif vers ces environnements pourrait permettre non seulement de soutenir l'expression de HoxB1, mais aussi de favoriser l'expression subséquente des autres gènes du complexe. Nous avons donc vérifié si le locus HoxB1 inactif était associé à des loci situés dans des bandes plus compactes et marquées plus intensément au Giemsa (et donc généralement pauvres en gènes) que le gène actif : nos résultats ne supportent pas ces hypothèses. Nous avons aussi vérifié si les partenaires du gène HoxB1 étaient plus rapprochés de gènes, ou plus fréquemment à proximité de ceux-ci lorsque que HoxB1 est actif (jour-1 vs jour 1 et 3). Nous n'avons noté aucune différence significative entre les jours du traitement à l'acide rétinoïque. Il est cependant possible que les gènes qui sont associés avec un HoxB1 activé soient eux même actifs, ce qui pourrait supporter l'hypothèse d'un recrutement vers une « usine de transcription ». Nous avons donc tenté de déterminer le degré d'activité des gènes les plus rapprochés des partenaires de fixation de HoxB1. Pour ce faire, nous avons vérifié l'expression de ces gènes en utilisant des données d'analyse par micro-puce d'ADN disponible (Logiciel SymAtlas, Genomics institute of the Novartis research foundation). Nous n'avons pas constaté de différences entre le degré d'expression des gènes proches des partenaires de fixation du gène au jour -1 et aux jours 1 ou 3. Par contre, une limitation de cette analyse provient du fait que les données d'expression disponibles concernent des embryons en développement, et pourraient donc ne pas refléter fidèlement la situation existant lors du traitement de cellules souches

embryonnaires à l'acide rétinoïque. Finalement, nous avons aussi vérifié si les gènes rapprochés des partenaires de fixation étaient impliqués dans des fonctions communes ou partagées aux différents jours du traitement à l'acide rétinoïque, par comparaison des termes de « *Gene Ontology* » qui leur sont associés. Nous n'avons décelé aucune différence entre les différents jours du traitement à l'acide rétinoïque dans cette analyse. Nos résultats ne soutiennent donc pas l'hypothèse qui voudrait que le gène HoxB1 activé soit recruté vers une « usine de transcription » spécifique.

Nos résultats sont donc compatibles avec l'idée voulant que les interactions distales entre HoxB1 et le reste du génome soient aléatoires. Bien que certaines études aient rapporté qu'un ordre chromosomique stable puisse être partagé dans une population cellulaire (Gerlich et al. 2003; Nagele et al. 1999), d'autres résultats suggèrent que l'arrangement des chromosomes dans l'espace nucléaire n'est pas conservé dans une population cellulaire, et qu'il est donc plus ou moins aléatoire (Cornforth et al. 2002; Walter et al. 2003). Les régions intra-chromosomiques distales semblent aussi être associées au locus HoxB1 de façon aléatoire, sans lien avec son statut d'expression. Nos résultats suggèrent donc que le locus HoxB1 peut interagir avec une grande diversité de loci inter et intra-chromosomiques dans la population cellulaire étudiée. Cependant, malgré le fait que nous ayons cloné et séquencé plus de 2300 clones, notre crible représente un relativement faible échantillonnage des possibilités d'interactions offertes par le génome complet. Il est donc possible que les interactions distales ou inter-chromosomiques n'aient pas été suffisamment représentées dans notre étude pour que nous puissions déceler une spécificité d'interaction statistiquement significative. Il ne nous est donc pas possible d'exclure que certaines interactions spécifiques existent ou que le gène HoxB1 soit recruté à une « usine de transcription ». Un crible à plus grande échelle serait donc nécessaire pour obtenir une puissance statistique suffisante pour nous permettre de vérifier ces possibilités.

### **Validations, interprétations et limitations des méthodologies de 3C ouvertes développées**

Nos résultats en général (inverse-PCR 3C, REM-3C et inter-L1 3C) sont compatibles avec l'idée que la chromatine soit mobile dans l'espace nucléaire, mais que cette mobilité soit restreinte (Gasser 2002; Marshall et al. 1997). En effet, nous avons montré que dans une

population cellulaire donnée, un locus particulier peut être retrouvé en proximité spatiale avec une grande diversité d'autres loci génomiques. Les modèles courants d'organisation de la chromatine dans le noyau postulent que l'organisation d'un locus est dynamique, et qu'elle est définie par des règles statistiques dictant des probabilités de proximité avec d'autres loci (Parada et al. 2003). Notre étude sur le gène HoxB1 est compatible avec cette vision des choses : par exemple, dans environ 25% des cellules, le locus HoxB1 semble en proximité spatiale avec des loci qui sont situés dans un domaine de 110 kb l'entourant alors qu'il est en contact avec d'autres loci dans le reste de la population cellulaire investiguée. Notre méthodologie permet donc d'évaluer la fréquence à laquelle un locus d'intérêt sera en proximité spatiale avec une région génomique donnée. D'ailleurs, il est probable que dans une cellule donnée, plusieurs loci sont situés à proximité d'un locus donné, et que ceux-ci soient tous fixés au locus d'intérêt. En effet, nous nous faisons généralement une idée très « propre » du processus de fixation au formaldéhyde. Cependant, la réalité est moins idéalisée : il existe ainsi probablement plusieurs loci situés dans une région donnée qui sont fixés au locus d'intérêt pour former un amas de chromatine. La ligation de plusieurs fragments individuels fixés au locus d'intérêt est donc probablement souvent réalisée avec une efficacité similaire. De plus, étant donné que chaque locus individuel sera lui aussi ligaturé à ses propres loci proximaux au niveau spatial, il est possible que certaines interactions indirectes soient observées par les méthodologies de 3C ouvertes. Par exemple, le gène HoxB1 pourrait interagir spécifiquement avec un locus donné, qui lui-même interagit avec un autre locus. Nous pourrions dans ce cas nous attendre à cloner cet autre locus comme partenaire d'interaction d'HoxB1, bien qu'il n'y ait à proprement parler qu'une interaction indirecte entre ces deux loci.

Il convient d'autre part de distinguer nos observations par 3C de la notion de distance spatiale moyenne entre deux loci qui est observable par FISH. En effet, les technologies de 3C déterminent la fréquence à laquelle un locus donné est en grande proximité spatiale avec un autre locus. Elle ne donne donc qu'une évaluation indirecte de la distance spatiale moyenne entre deux loci. Par exemple, un locus très mobile et décondensé pourrait, en moyenne, être situé à une plus grande distance spatiale d'un locus donné qu'un autre locus moins mobile attaché à la périphérie nucléaire. Paradoxalement, dans une telle situation, le locus peu mobile pourrait ne jamais entrer en proximité suffisante avec le locus d'intérêt pour être récupéré par 3C : il y serait ainsi très rarement fixé et ligaturé. Au contraire, le

locus mobile pourrait, du fait d'une plus grande incertitude statistique sur son positionnement et d'une distribution de distance spatiale plus étendue, être retrouvé en grande proximité plus fréquemment avec le locus d'intérêt, et ce malgré le fait qu'il soit en moyenne plus distant de celui-ci par FISH. Les comparaisons entre les résultats obtenus par les deux méthodes doivent donc être réalisées avec précautions, puisqu'elles ne mesurent pas les mêmes paramètres de l'organisation du noyau.

Les résultats présentés ici indiquent que l'approche 3C que nous avons développée rend possible l'analyse de l'organisation spatiale d'un locus donné dans plusieurs conditions. Au niveau technique, cette méthodologie est efficace et permet une analyse à haute résolution de l'environnement spatial d'un locus. Nous avons aussi démontré que cette méthodologie était suffisamment sensible pour réaliser l'analyse d'un environnement spatial dynamique. Nous avons cependant pu constater certaines limitations inhérentes à cette nouvelle méthodologie. En effet, l'analyse de la fréquence de fixation de fragments génomiques précis est rendue problématique par le fait que bien que les problèmes liés aux différences d'efficacité d'amplification par inverse-PCR soient beaucoup moins importants que ceux rencontrés lors de l'utilisation d'amplification par REM-PCR, ils existent et peuvent influencer les résultats.

Étant donné que certains fragments ne sont pas récupérables par inverse-PCR 3C, nous avons voulu vérifier les résultats que nous avons obtenus par une analyse de 3C classique. Les résultats du crible que nous avons réalisé par inverse-PCR 3C sont généralement compatibles avec l'analyse du locus par 3C classique (Figure 25). Une comparaison de la Figure 25B et C montre que globalement, les régions génomiques qui ne sont pas fréquemment fixées au locus HoxB1 ne furent pas récupérées fréquemment par inverse-PCR 3C et vice-versa. Cependant, cette analyse démontre aussi que deux variables doivent être considérées lorsque les résultats d'inverse-PCR 3C sont comparés avec ceux du 3C classique : la densité des partenaires de fixation au long d'un intervalle génomique donné et leur fréquence de récupération. Donc, bien qu'il existe des différences dans les résultats obtenus avec les deux méthodes au niveau des fragments génomiques individuels, les résultats sont remarquablement compatibles lorsque de plus grands intervalles génomiques sont considérés. Nous notons d'autre part que la plupart des études ayant utilisé le 3C ont observé une diminution assez rapide de la fréquence de fixation des loci en lien avec leur

séparation du locus d'intérêt (en particulier à l'intérieur d'environ 50 à 100 kb) (Dekker et al. 2002; Palstra et al. 2003; Tolhuis et al. 2002). Nos résultats d'inverse-PCR 3C et de 3C classique sont globalement en accord avec ces observations. Par contre, nos résultats suggèrent qu'il existe des différences dans le taux de diminution entre les deux côtés du gène HoxB1, le côté 5' présentant une diminution moins importante que le côté 3'. Nous avons aussi pu constater que certains loci sont difficilement récupérables par inverse-PCR 3C, et ce de façon indépendante de leur fréquence de fixation au locus de départ. Il est donc important de valider la fréquence de fixation d'un fragment génomique particulier au locus de départ observée par inverse-PCR 3C avec des analyses de 3C classique.

Nos résultats suggèrent qu'il est préférable d'analyser des régions plus vastes par inverse-PCR 3C, et de comparer les fréquences relatives de récupération de loci situés dans des zones génomiques plus étendues. Ceci permet ainsi d'obtenir une mesure du regroupement de partenaires dans ces régions. Comme nous l'avons mentionné, plusieurs loci sont probablement fixés par le formaldéhyde au locus d'intérêt : il est donc probable que des fragments génomiques HindIII adjacents au locus en interaction directe avec le locus d'intérêt soient souvent fixés à fréquence similaire à celui-ci. À cause de ce fait, la récupération de plusieurs partenaires provenant d'une région génomique relativement restreinte suggère qu'un ou plusieurs fragments situés dans cette région interagissent spécifiquement en interaction fréquente avec le locus d'intérêt. Il est donc ensuite souhaitable de confirmer ces interactions par 3C classique pour déterminer lesquels de ces fragments groupés interagissent plus spécifiquement au locus d'intérêt.

Ceci nous mène à une deuxième problématique, qui n'est pas liée à la stratégie d'inverse-PCR 3C en soi, mais plutôt à la façon dont nous avons recueilli les résultats. Bien que nous ayons cloné, séquencé et identifié un nombre considérable de clones pour chacune des conditions, il est certain qu'une fois dépassé une certaine distance du site HoxB1, le nombre d'événements clonés devient trop faible pour pouvoir déterminer avec certitude qu'une région génomique particulière est sur-représentée de façon statistiquement significative. Il faudrait donc augmenter le nombre de clonage de façon assez importante pour obtenir une puissance statistique suffisante dans ces régions distales ou inter-chromosomique, et pour ainsi pouvoir observer des regroupements significatifs de clones dans des régions génomiques de taille restreinte. Cependant, il peut rapidement devenir peu

pratique techniquement et très coûteux de réaliser tous ces séquençages. Pour ces raisons, nous avons choisi de représenter nos résultats par une courbe d'accumulation, qui permet d'obtenir une image globale de l'accessibilité du locus HoxB1 au reste de son chromosome. En effet, une représentation des fréquences de clonage absolues nous semblait moins informative, étant donné le faible nombre de partenaires récupérés dans les loci distaux.

Le couplage de notre méthode d'inverse-PCR 3C avec les technologies de micro-puce d'ADN pourrait potentiellement améliorer son efficacité et faciliter une analyse à grand débit de ces loci distaux. Il s'agirait donc de marquer les produits d'inverse-PCR 3C et de les hybrider à des micro-puces contenant des BACs génomiques par exemple. Ceci diminuerait d'autant la lourdeur technique associée au clonage et au séquençage des événements individuels. Malgré tout, il reste que pour identifier le fragment précis qui est impliqué dans l'interaction avec le locus d'intérêt et pour confirmer cette interaction, les analyses par 3C classiques demeurent essentielles.

## CONCLUSIONS GÉNÉRALES

Dans cette thèse, nous avons exploré les interactions entre loci dans l'espace nucléaire et leurs liens avec le fonctionnement et l'organisation du génome des mammifères. Nos résultats et ceux retrouvés dans une littérature de plus en plus abondante suggèrent fortement que les activités des loci génomiques sont directement liées à leur environnement nucléaire. Ceci rejoint d'ailleurs un concept de plus en plus accepté en biologie : le fait que la création de micro-environnements spécifiques puisse faciliter la régulation et l'efficacité des processus biologiques. Au-delà de la séparation de la cellule eucaryote en organelles, plusieurs exemples démontrent ainsi que certaines protéines jouent le rôle « d'échafaudages » qui facilitent les interactions entre deux autres protéines particulières, ou que le recrutement de certaines molécules dans des structures particulières (par exemple les corps nucléaires) peut augmenter l'efficacité de plusieurs processus biologiques. Il s'agit en fait d'une idée très logique : augmenter la concentration locale de facteurs impliqués dans une réaction biologique quelconque devrait généralement conduire à une plus grande efficacité et à une moins grande dépendance vis-à-vis de la diffusion aléatoire de ces molécules.

Nous avons premièrement vu que la proximité linéaire entre régions d'ADN pouvait rendre très fréquentes leurs interactions. En particulier, nous avons démontré que les interactions entre séquences en répétition directe et situées en proximité linéaire les unes des autres peuvent conduire à des modifications héréditaires du génome, en participant fréquemment à des événements de recombinaison homologue et non-homologue dans l'embryon précoce ou les cellules germinales. D'autre part, nous avons aussi observé que l'existence d'un environnement spatial relativement bien défini résulte en certaines contraintes dans le choix des partenaires de recombinaison entre séquences L1 chez l'humain. Cette partie de l'étude suggère donc que l'organisation linéaire et spatiale du génome peut influencer les interactions entre régions génomiques par recombinaison, et ainsi conduire à des remodelages fréquents à l'intérieur de ces environnements au cours de l'évolution. Étant donné le grand nombre de séquences répétitives des génomes de mammifères, il est possible de penser que l'organisation linéaire du génome a pu être dramatiquement influencée par ces processus. Il pourrait être intéressant de combiner ces observations et de vérifier si en général, les séquences répétitives d'une région génomique donnée ont au

cours de l'évolution effectivement participé à des événements de recombinaison de façon plus fréquente avec des partenaires situés dans leur environnement spatial et linéaire immédiat. Comme nous l'avons brièvement mentionné plus haut, il est possible que l'homogénéisation des séquences L1 par conversion génique ne soit pas distribuée aléatoirement dans le génome, mais que les L1 situés dans une région génomique quelconque soient généralement plus similaires entre eux que s'ils sont comparés avec un échantillonnage aléatoire de L1. Si ceci était vrai, il serait potentiellement possible de déterminer la localisation des voisinages spatiaux génomiques distincts simplement par l'analyse de la séquence du génome. Il faut d'ailleurs noter que ces processus sont probablement circulaires : l'augmentation de l'homogénéité entre séquences dans une région génomique donnée rend probablement leurs interactions plus faciles grâce à leur homologie plus élevée, renforçant ainsi l'existence de cet environnement.

Nous avons d'autre part observé que les événements de *one-sided invasion* étaient possibles entre L1 dans l'embryon de souris. Les événements médiés par ce mécanisme pourraient potentiellement avoir été impliqués dans les phénomènes de « brassage d'exons » (*exon shuffling*). Comme nous l'avons vu dans l'introduction, les gènes co-régulés ou impliqués dans des voies de signalisation ou des fonctions particulières sont souvent regroupés dans l'organisation du génome. Il est possible que des événements de *one-sided invasion* entre séquences répétitives situées dans ces environnements génomiques linéaires et spatiaux restreints aient favorisé le partage de domaines protéiques ou de séquences régulatrices entre ces différents gènes. Ces mécanismes pourraient donc avoir eu un rôle à jouer dans le regroupement génomique de gènes ayant des fonctions ou régulations similaires.

L'environnement spatial et linéaire statique d'un locus ont donc un impact sur les interactions qu'il aura avec d'autres loci génomiques. Cependant, certains environnements spatiaux génomiques sont plus dynamiques. Nous avons entre autre observé que l'activation d'un gène, en l'occurrence HoxB1, pouvait conduire à des modifications de son environnement spatial génomique, et que ses possibilités d'interactions avec le reste du génome pouvaient être modifiées par les processus d'activation de la transcription. Ceci suggère donc que non seulement l'environnement de base d'un locus peut influencer son fonctionnement, mais qu'il existe aussi des mécanismes permettant à la cellule de modifier cet environnement spatial. Cependant, un débat demeure concernant ces observations et



plusieurs autres : est-ce que les modifications dans l'environnement spatial d'un gène sont importantes pour ses fonctions, ou sont-elles simplement des conséquences non-spécifiques de son activation? Nos résultats ne nous permettent pas de statuer sur cette question. Par contre, nous pouvons noter que les modifications d'environnement spatial du gène HoxB1 se produisent clairement après qu'il soit induit. Il est cependant possible qu'un recrutement vers un environnement nucléaire particulier de façon subséquente à l'activation initiale puisse favoriser le maintien de l'expression à de hauts niveaux. D'autre part, une décondensation progressive du locus HoxB1 dans l'espace nucléaire pourrait résulter en un repositionnement hors de son territoire sans que ce changement d'environnement n'ait nécessairement d'importance fonctionnelle. Cependant, comme nous l'avons mentionné dans l'introduction, il existe des exemples clairs dans la littérature où le recrutement de certains loci vers certains environnements spatiaux spécifiques se produit de façon concomitante avec l'expression (Osborne et al. 2004). De plus, il fut démontré que le recrutement artificiel d'un locus vers certaines régions d'hétérochromatine comme à la périphérie du noyau par exemple pouvait résulter en son inactivation transcriptionnelle (Andrulis et al. 1998). Il est donc probable que ces événements doivent être considérés au cas par cas, et que dans certains cas le recrutement d'un locus vers un environnement spatial particulier soit important pour moduler ses fonctions, alors que dans d'autres situations les modifications à l'environnement spatial soient des conséquences relativement indirectes d'un changement d'état de ce locus.

Les méthodologies que nous avons décrites dans cette thèse sont toutes originales et n'avaient jamais été décrites auparavant dans la littérature. Peu de méthodologies *in vivo* existent pour investiguer la recombinaison homologue se produisant dans l'ovule fécondé et l'embryon. La méthodologie que nous avons développée pourrait être appliquée dans certaines lignées de souris mutantes pour certaines protéines de réparation par exemple, pour étudier leur effet *in vivo*. En fait, le point majeur d'intérêt de cette méthode réside dans le principe que nous avons utilisé : l'intégration de plusieurs plasmides au génome des souris à cause de la micro-injection permet de prendre un « instantané » de la population de plasmides qui s'y trouve au moment de l'intégration. Comme nous l'avons mentionné, l'utilisation de cette méthode dans diverses lignées de souris mutantes pourrait clarifier le rôle de plusieurs protéines de réparation et de signalisation dans l'ovule fécondé et l'embryon de souris.

Les méthodologies de 3C ouvertes sont elles aussi novatrices. À notre connaissance, nous avons été les premiers à publier les résultats d'une étude présentant le développement et l'utilisation de ce type d'essai. Les possibilités offertes par ces méthodes, en particulier l'inverse-PCR 3C, sont extrêmement intéressantes. En effet, un des problèmes majeurs de l'étude de l'organisation spatiale des génomes de mammifères provient de leur grande taille, ce qui rend l'investigation systématique de tous les loci impossibles, contrairement à ce qui a été réalisé pour un des chromosomes de levure (Dekker et al. 2002). Le fait de pouvoir poser une question ouverte permet d'investiguer d'un seul coup toutes les interactions impliquant un locus donné, ce qui rend possible la formulation d'hypothèses d'intérêt sur l'organisation spatiale de ce locus. Le couplage de ces méthodes avec les technologies de micro-puces d'ADN pourrait d'ailleurs dramatiquement faciliter l'étude de l'organisation spatiale des génomes de mammifères dans une population cellulaire en permettant l'étude à grand débit de l'organisation spatiale des génomes. Ces méthodes devraient donc permettre de découvrir de nombreuses nouvelles interactions entre régions génomiques, qui pourront par la suite être validées par d'autres méthodologies plus ciblées. Par contre, dans un premier temps le séquençage d'un grand nombre de produits, comme nous l'avons fait, a permis de comprendre la méthode et pourra servir de référence au développement d'améliorations subséquentes à ces méthodologies.

Nos études soutiennent donc une vision globale du noyau dans laquelle non-seulement la séquence des chromosomes et la concentration des différents facteurs sont importantes pour les fonctions du génome, mais aussi son organisation spatiale dans le noyau. Cette vision de cette organelle cruciale est sans aucun doute appelée à être reconnue comme étant un des déterminants majeurs du fonctionnement nucléaire.

## RÉFÉRENCES

- Abney, J. R., Cutler, B., Fillbach, M. L., Axelrod, D., Scalettar, B. A. (1997) Chromatin dynamics in interphase nuclei and its implications for nuclear structure. *J Cell Biol* 137: 1459-1468.
- Alani, E., Reenan, R. A., Kolodner, R. D. (1994) Interaction between mismatch repair and genetic recombination in *Saccharomyces cerevisiae*. *Genetics* 137: 19-39.
- Allan, J., Harborne, N., Rau, D. C., Gould, H. (1982) Participation of core histone "tails" in the stabilization of the chromatin solenoid. *J Cell Biol* 93: 285-297.
- Andrulis, E. D., Neiman, A. M., Zappulla, D. C., Sternglanz, R. (1998) Perinuclear localization of chromatin facilitates transcriptional silencing. *Nature* 394: 592-595.
- Aten, J. A., Stap, J., Krawczyk, P. M., van Oven, C. H., Hoebe, R. A., Essers, J., Kanaar, R. (2004) Dynamics of DNA double-strand breaks revealed by clustering of damaged chromosome domains. *Science* 303: 92-95.
- Bakkenist, C. J., Kastan, M. B. (2003) DNA damage activates ATM through intermolecular autophosphorylation and dimer dissociation. *Nature* 421: 499-506.
- Baumann, P., Benson, F. E., West, S. C. (1996) Human Rad51 protein promotes ATP-dependent homologous pairing and strand transfer reactions in vitro. *Cell* 87: 757-766.
- Bednar, J., Horowitz, R. A., Grigoryev, S. A., Carruthers, L. M., Hansen, J. C., Koster, A. J., Woodcock, C. L. (1998) Nucleosomes, linker DNA, and linker histone form a unique structural motif that directs the higher-order folding and compaction of chromatin. *Proc Natl Acad Sci U S A* 95: 14173-14178.
- Bel-Vialar, S., Itasaki, N., Krumlauf, R. (2002) Initiating Hox gene expression: in the early chick neural tube differential sensitivity to FGF and RA signaling subdivides the HoxB genes in two distinct groups. *Development* 129: 5103-5115.

- Belmaaza, A., Chartrand, P. (1994) One-sided invasion events in homologous recombination at double-strand breaks. *Mutat Res* 314: 199-208.
- Belmont, A. S., Bruce, K. (1994) Visualization of G1 chromosomes: a folded, twisted, supercoiled chromonema model of interphase chromatid structure. *J Cell Biol* 127: 287-302.
- Berezney, R., Coffey, D. S. (1974) Identification of a nuclear protein matrix. *Biochem Biophys Res Commun* 60: 1410-1417.
- Berrios, M., Osheroff, N., Fisher, P. A. (1985) In situ localization of DNA topoisomerase II, a major polypeptide component of the *Drosophila* nuclear matrix fraction. *Proc Natl Acad Sci U S A* 82: 4142-4146.
- Bill, C. A., Nickoloff, J. A. (2001) Spontaneous and ultraviolet light-induced direct repeat recombination in mammalian cells frequently results in repeat deletion. *Mutat Res* 487: 41-50.
- Boulikas, T. (1995) Chromatin domains and prediction of MAR sequences. *Int Rev Cytol* 162A: 279-388.
- Boyle, S., Gilchrist, S., Bridger, J. M., Mahy, N. L., Ellis, J. A., Bickmore, W. A. (2001) The spatial organization of human chromosomes within the nuclei of normal and emerimutant cells. *Hum Mol Genet* 10: 211-219.
- Bridger, J. M., Herrmann, H., Munkel, C., Lichter, P. (1998) Identification of an interchromosomal compartment by polymerization of nuclear-targeted vimentin. *J Cell Sci* 111 ( Pt 9): 1241-1253.
- Brown, K. E., Guest, S. S., Smale, S. T., Hahm, K., Merckenschlager, M., Fisher, A. G. (1997) Association of transcriptionally silent genes with Ikaros complexes at centromeric heterochromatin. *Cell* 91: 845-854.

Buluwela, L., Forster, A., Boehm, T., Rabbitts, T. H. (1989) A rapid procedure for colony screening using nylon filters. *Nucleic Acids Res* 17: 452.

Burma, S., Chen, B. P., Murphy, M., Kurimasa, A., Chen, D. J. (2001) ATM phosphorylates histone H2AX in response to DNA double-strand breaks. *J Biol Chem* 276: 42462-42467.

Capelson, M., Corces, V. G. (2004) Boundary elements and nuclear organization. *Biol Cell* 96: 617-629.

Carpenter, A. E., Ashouri, A., Belmont, A. S. (2004) Automated microscopy identifies estrogen receptor subdomains with large-scale chromatin structure unfolding activity. *Cytometry A* 58: 157-166.

Carpenter, A. E., Belmont, A. S. (2004) Direct visualization of transcription factor-induced chromatin remodeling and cofactor recruitment in vivo. *Methods Enzymol* 375: 366-381.

Carvalho, C., Pereira, H. M., Ferreira, J., Pina, C., Mendonca, D., Rosa, A. C., Carmo-Fonseca, M. (2001) Chromosomal G-dark bands determine the spatial organization of centromeric heterochromatin in the nucleus. *Mol Biol Cell* 12: 3563-3572.

Cary, R. B., Peterson, S. R., Wang, J., Bear, D. G., Bradbury, E. M., Chen, D. J. (1997) DNA looping by Ku and the DNA-dependent protein kinase. *Proc Natl Acad Sci U S A* 94: 4267-4272.

Casolari, J. M., Brown, C. R., Komili, S., West, J., Hieronymus, H., Silver, P. A. (2004) Genome-wide localization of the nuclear transport machinery couples transcriptional status and nuclear organization. *Cell* 117: 427-439.

Celeste, A., Fernandez-Capetillo, O., Kruhlak, M. J., Pilch, D. R., Staudt, D. W., Lee, A., Bonner, R. F., Bonner, W. M., Nussenzweig, A. (2003) Histone H2AX phosphorylation is dispensable for the initial recognition of DNA breaks. *Nat Cell Biol* 5: 675-679.

Chambeyron, S., Bickmore, W. A. (2004) Chromatin decondensation and nuclear reorganization of the HoxB locus upon induction of transcription. *Genes Dev* 18: 1119-1130.

Chambeyron, S., Da Silva, N. R., Lawson, K. A., Bickmore, W. A. (2005) Nuclear reorganisation of the Hoxb complex during mouse embryonic development. *Development* 132: 2215-2223.

Chambon, P. (1996) A decade of molecular biology of retinoic acid receptors. *Faseb J* 10: 940-954.

Chen, J. D., Evans, R. M. (1995) A transcriptional co-repressor that interacts with nuclear hormone receptors. *Nature* 377: 454-457.

Chen, J. J., Silver, D., Cantor, S., Livingston, D. M., Scully, R. (1999) BRCA1, BRCA2, and Rad51 operate in a common DNA damage response pathway. *Cancer Res* 59: 1752s-1756s.

Christensen, M. O., Larsen, M. K., Barthelmes, H. U., Hock, R., Andersen, C. L., Kjeldsen, E., Knudsen, B. R., Westergaard, O., Boege, F., Mielke, C. (2002) Dynamics of human DNA topoisomerases IIalpha and IIbeta in living cells. *J Cell Biol* 157: 31-44.

Chubb, J. R., Boyle, S., Perry, P., Bickmore, W. A. (2002) Chromatin motion is constrained by association with nuclear compartments in human cells. *Curr Biol* 12: 439-445.

Cohen, B. A., Mitra, R. D., Hughes, J. D., Church, G. M. (2000) A computational analysis of whole-genome expression data reveals chromosomal domains of gene expression. *Nat Genet* 26: 183-186.

Cook, P. R. (1999) The organization of replication and transcription. *Science* 284: 1790-1795.

Cooper, M. P., Machwe, A., Orren, D. K., Brosh, R. M., Ramsden, D., Bohr, V. A. (2000) Ku complex interacts with and stimulates the Werner protein. *Genes Dev* 14: 907-912.

Cornforth, M. N., Greulich-Bode, K. M., Loucas, B. D., Arsuaga, J., Vazquez, M., Sachs, R. K., Bruckner, M., Molls, M., Hahnfeldt, P., Hlatky, L., Brenner, D. J. (2002) Chromosomes are predominantly located randomly with respect to each other in interphase human cells. *J Cell Biol* 159: 237-244.

Craig, J. M. (2005) Heterochromatin--many flavours, common themes. *Bioessays* 27: 17-28.

Cremer, T., Cremer, C. (2001) Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat Rev Genet* 2: 292-301.

Croft, J. A., Bridger, J. M., Boyle, S., Perry, P., Teague, P., Bickmore, W. A. (1999) Differences in the localization and morphology of chromosomes in the human nucleus. *J Cell Biol* 145: 1119-1131.

Csink, A. K., Henikoff, S. (1996) Genetic modification of heterochromatic association and nuclear organization in *Drosophila*. *Nature* 381: 529-531.

Csink, A. K., Henikoff, S. (1998) Large-scale chromosomal movements during interphase progression in *Drosophila*. *J Cell Biol* 143: 13-22.

D'Anjou, H., Chabot, C., Chartrand, P. (2004) Preferential accessibility to specific genomic loci for the repair of double-strand breaks in human cells. *Nucleic Acids Res* 32: 6136-6143.

de Laat, W., Grosveld, F. (2003) Spatial organization of gene expression: the active chromatin hub. *Chromosome Res* 11: 447-459.

DeFazio, L. G., Stansel, R. M., Griffith, J. D., Chu, G. (2002) Synapsis of DNA ends by DNA-dependent protein kinase. *Embo J* 21: 3192-3200.

Dehghani, H., Dellaire, G., Bazett-Jones, D. P. (2005) Organization of chromatin in the interphase mammalian cell. *Micron* 36: 95-108.

Dekker, J., Rippe, K., Dekker, M., Kleckner, N. (2002) Capturing chromosome conformation. *Science* 295: 1306-1311.

Dellaire, G., Bazett-Jones, D. P. (2004) PML nuclear bodies: dynamic sensors of DNA damage and cellular stress. *Bioessays* 26: 963-977.

Dellaire, G., Yan, J., Little, K. C., Drouin, R., Chartrand, P. (2002) Evidence that extrachromosomal double-strand break repair can be coupled to the repair of chromosomal double-strand breaks in mammalian cells. *Chromosoma* 111: 304-312.

Dernburg, A. F., Broman, K. W., Fung, J. C., Marshall, W. F., Philips, J., Agard, D. A., Sedat, J. W. (1996) Perturbation of nuclear architecture by long-distance chromosome interactions. *Cell* 85: 745-759.

Dewannieux, M., Esnault, C., Heidmann, T. (2003) LINE-mediated retrotransposition of marked Alu sequences. *Nat Genet* 35: 41-48.

Dietzel, S., Jauch, A., Kienle, D., Qu, G., Holtgreve-Grez, H., Eils, R., Munkel, C., Bittner, M., Meltzer, P. S., Trent, J. M., Cremer, T. (1998) Separate and variably shaped chromosome arm domains are disclosed by chromosome arm painting in human cell nuclei. *Chromosome Res* 6: 25-33.

Dou, Y., Mizzen, C. A., Abrams, M., Allis, C. D., Gorovsky, M. A. (1999) Phosphorylation of linker histone H1 regulates gene expression in vivo by mimicking H1 removal. *Mol Cell* 4: 641-647.

Downs, J. A., Allard, S., Jobin-Robitaille, O., Javaheri, A., Auger, A., Bouchard, N., Kron, S. J., Jackson, S. P., Cote, J. (2004) Binding of chromatin-modifying activities to phosphorylated histone H2A at DNA damage sites. *Mol Cell* 16: 979-990.

Drissen, R., Palstra, R. J., Gillemans, N., Splinter, E., Grosveld, F., Philipsen, S., de Laat, W. (2004) The active spatial organization of the beta-globin locus requires the transcription factor EKLF. *Genes Dev* 18: 2485-2490.



Dronkert, M. L., Beverloo, H. B., Johnson, R. D., Hoeijmakers, J. H., Jasin, M., Kanaar, R. (2000) Mouse RAD54 affects DNA double-strand break repair and sister chromatid exchange. *Mol Cell Biol* 20: 3147-3156.

Dudley, D. D., Chaudhuri, J., Bassing, C. H., Alt, F. W. (2005) Mechanism and control of V(D)J recombination versus class switch recombination: similarities and differences. *Adv Immunol* 86: 43-112.

Durocher, D., Jackson, S. P. (2001) DNA-PK, ATM and ATR as sensors of DNA damage: variations on a theme? *Curr Opin Cell Biol* 13: 225-231.

Earnshaw, W. C., Heck, M. M. (1985) Localization of topoisomerase II in mitotic chromosomes. *J Cell Biol* 100: 1716-1725.

Eberharter, A., Becker, P. B. (2002) Histone acetylation: a switch between repressive and permissive chromatin. Second in review series on chromatin dynamics. *EMBO Rep* 3: 224-229.

Elliott, B., Jasin, M. (2002) Double-strand breaks and translocations in cancer. *Cell Mol Life Sci* 59: 373-385.

Esnault, C., Maestre, J., Heidmann, T. (2000) Human LINE retrotransposons generate processed pseudogenes. *Nat Genet* 24: 363-367.

Essers, J., van Cappellen, W. A., Theil, A. F., van Drunen, E., Jaspers, N. G., Hoeijmakers, J. H., Wyman, C., Vermeulen, W., Kanaar, R. (2005) Dynamics of relative chromosome position during the cell cycle. *Mol Biol Cell* 16: 769-775.

Fan, J. Y., Gordon, F., Luger, K., Hansen, J. C., Tremethick, D. J. (2002) The essential histone variant H2A.Z regulates the equilibrium between different chromatin conformational states. *Nat Struct Biol* 9: 172-176.

Ferreira, J., Paoletta, G., Ramos, C., Lamond, A. I. (1997) Spatial organization of large-scale chromatin domains in the nucleus: a magnified view of single chromosome territories. *J Cell Biol* 139: 1597-1610.

Fiorenza, M. T., Bevilacqua, A., Bevilacqua, S., Mangia, F. (2001) Growing dictyate oocytes, but not early preimplantation embryos, of the mouse display high levels of DNA homologous recombination by single-strand annealing and lack DNA nonhomologous end joining. *Dev Biol* 233: 214-224.

Fletcher, T. M., Hansen, J. C. (1995) Core histone tail domains mediate oligonucleosome folding and nucleosomal DNA organization through distinct molecular mechanisms. *J Biol Chem* 270: 25359-25362.

Folger, K. R., Thomas, K., Capecchi, M. R. (1985) Nonreciprocal exchanges of information between DNA duplexes coinjected into mammalian cell nuclei. *Mol Cell Biol* 5: 59-69.

Folger, K. R., Wong, E. A., Wahl, G., Capecchi, M. R. (1982) Patterns of integration of DNA microinjected into cultured mammalian cells: evidence for homologous recombination between injected plasmid DNA molecules. *Mol Cell Biol* 2: 1372-1387.

Furano, A. V. (2000) The biological properties and evolutionary dynamics of mammalian LINE-1 retrotransposons. *Prog Nucleic Acid Res Mol Biol* 64: 255-294.

Gall, J. G. (2000) Cajal bodies: the first 100 years. *Annu Rev Cell Dev Biol* 16: 273-300.

Gall, J. G. (2001) A role for Cajal bodies in assembly of the nuclear transcription machinery. *FEBS Lett* 498: 164-167.

Garcia-Ramirez, M., Rocchini, C., Ausio, J. (1995) Modulation of chromatin folding by histone acetylation. *J Biol Chem* 270: 17923-17928.

Gasser, S. M. (2002) Visualizing chromatin dynamics in interphase nuclei. *Science* 296: 1412-1416.

Gerasimova, T. I., Byrd, K., Corces, V. G. (2000) A chromatin insulator determines the nuclear localization of DNA. *Mol Cell* 6: 1025-1035.

Gerlich, D., Beaudouin, J., Kalbfuss, B., Daigle, N., Eils, R., Ellenberg, J. (2003) Global chromosome positions are transmitted through mitosis in mammalian cells. *Cell* 112: 751-764.

Giannattasio, M., Lazzaro, F., Plevani, P., Muzi-Falconi, M. (2005) The DNA damage checkpoint response requires histone H2B ubiquitination by Rad6-Bre1 and H3 methylation by Dot1. *J Biol Chem* 280: 9879-9886.

Gilbert, N., Boyle, S., Fiegler, H., Woodfine, K., Carter, N. P., Bickmore, W. A. (2004) Chromatin architecture of the human genome: gene-rich domains are enriched in open chromatin fibers. *Cell* 118: 555-566.

Gordon, J. W., Ruddle, F. H. (1983) Gene transfer into mouse embryos: production of transgenic mice by pronuclear injection. *Methods Enzymol* 101: 411-433.

Gorisch, S. M., Wachsmuth, M., Toth, K. F., Lichter, P., Rippe, K. (2005) Histone acetylation increases chromatin accessibility. *J Cell Sci* 118: 5825-5834.

Grigoryev, S. A. (2001) Higher-order folding of heterochromatin: protein bridges span the nucleosome arrays. *Biochem Cell Biol* 79: 227-241.

Guarente, L. (1999) Diverse and dynamic functions of the Sir silencing complex. *Nat Genet* 23: 281-285.

Hansen, J. C. (2002) Conformational dynamics of the chromatin fiber in solution: determinants, mechanisms, and functions. *Annual Review of Biophysics & Biomolecular Structure* 31: 361-392.

Hardies, S. C., Martin, S. L., Voliva, C. F., Hutchison, C. A., 3rd, Edgell, M. H. (1986) An analysis of replacement and synonymous changes in the rodent L1 repeat family. *Mol Biol Evol* 3: 109-125.

Hebbes, T. R., Clayton, A. L., Thorne, A. W., Crane-Robinson, C. (1994) Core histone hyperacetylation co-maps with generalized DNase I sensitivity in the chicken beta-globin chromosomal domain. *Embo J* 13: 1823-1830.

Hecht, A., Laroche, T., Strahl-Bolsinger, S., Gasser, S. M., Grunstein, M. (1995) Histone H3 and H4 N-termini interact with SIR3 and SIR4 proteins: a molecular model for the formation of heterochromatin in yeast. *Cell* 80: 583-592.

Hizume, K., Yoshimura, S. H., Takeyasu, K. (2005) Linker histone H1 per se can induce three-dimensional folding of chromatin fiber. *Biochemistry* 44: 12978-12989.

Horlein, A. J., Naar, A. M., Heinzl, T., Torchia, J., Gloss, B., Kurokawa, R., Ryan, A., Kamei, Y., Soderstrom, M., Glass, C. K., et al. (1995) Ligand-independent repression by the thyroid hormone receptor mediated by a nuclear receptor co-repressor. *Nature* 377: 397-404.

Horn, P. J., Crowley, K. A., Carruthers, L. M., Hansen, J. C., Peterson, C. L. (2002) The SIN domain of the histone octamer is essential for intramolecular folding of nucleosomal arrays. *Nat Struct Biol* 9: 167-171.

Horn, P. J., Peterson, C. L. (2002) Molecular biology. Chromatin higher order folding--wrapping up transcription. *Science* 297: 1824-1827.

Huang, D., Chen, S. W., Gudas, L. J. (2002) Analysis of two distinct retinoic acid response elements in the homeobox gene *Hoxb1* in transgenic mice. *Dev Dyn* 223: 353-370.

Huang, J., Dynan, W. S. (2002) Reconstitution of the mammalian DNA double-strand break end-joining reaction reveals a requirement for an Mre11/Rad50/NBS1-containing fraction. *Nucleic Acids Res* 30: 667-674.

Huang, L. C., Clarkin, K. C., Wahl, G. M. (1996) Sensitivity and selectivity of the DNA damage sensor responsible for activating p53-dependent G1 arrest. *Proc Natl Acad Sci U S A* 93: 4827-4832.

Huang, S., Beresten, S., Li, B., Oshima, J., Ellis, N. A., Campisi, J. (2000) Characterization of the human and mouse WRN 3'-->5' exonuclease. *Nucleic Acids Res* 28: 2396-2405.

Huyen, Y., Zgheib, O., Ditullio, R. A., Jr., Gorgoulis, V. G., Zacharatos, P., Petty, T. J., Sheston, E. A., Mellert, H. S., Stavridi, E. S., Halazonetis, T. D. (2004) Methylated lysine 79 of histone H3 targets 53BP1 to DNA double-strand breaks. *Nature* 432: 406-411.

Iafraite, A. J., Feuk, L., Rivera, M. N., Listewnik, M. L., Donahoe, P. K., Qi, Y., Scherer, S. W., Lee, C. (2004) Detection of large-scale variation in the human genome. *Nat Genet* 36: 949-951.

Iborra, F. J., Jackson, D. A., Cook, P. R. (2001) Coupled transcription and translation within nuclei of mammalian cells. *Science* 293: 1139-1142.

Iborra, F. J., Pombo, A., Jackson, D. A., Cook, P. R. (1996) Active RNA polymerases are localized within discrete transcription 'factories' in human nuclei. *J Cell Sci* 109 ( Pt 6): 1427-1436.

Ikeda, H. (1994) DNA topoisomerase-mediated illegitimate recombination. *Adv Pharmacol* 29A: 147-165.

Imai, S., Nishibayashi, S., Takao, K., Tomifuji, M., Fujino, T., Hasegawa, M., Takano, T. (1997) Dissociation of Oct-1 from the nuclear peripheral structure induces the cellular aging-associated collagenase gene expression. *Mol Biol Cell* 8: 2407-2419.

Jackson, D. A., Iborra, F. J., Manders, E. M., Cook, P. R. (1998) Numbers and organization of RNA polymerases, nascent transcripts, and transcription units in HeLa nuclei. *Mol Biol Cell* 9: 1523-1536.

Jacobs, E. Y., Frey, M. R., Wu, W., Ingledue, T. C., Gebuhr, T. C., Gao, L., Marzluff, W. F., Matera, A. G. (1999) Coiled bodies preferentially associate with U4, U11, and U12 small nuclear RNA genes in interphase HeLa cells but not with U6 and U7 genes. *Mol Biol Cell* 10: 1653-1663.

- Jenuwein, T., Allis, C. D. (2001) Translating the histone code. *Science* 293: 1074-1080.
- Johnson, C. N., Adkins, N. L., Georgel, P. (2005) Chromatin remodeling complexes: ATP-dependent machines in action. *Biochem Cell Biol* 83: 405-417.
- Johnson, R. D., Jasin, M. (2000) Sister chromatid gene conversion is a prominent double-strand break repair pathway in mammalian cells. *Embo J* 19: 3398-3407.
- Kalendar, R., Tanskanen, J., Immonen, S., Nevo, E., Schulman, A. H. (2000) Genome evolution of wild barley (*Hordeum spontaneum*) by BARE-1 retrotransposon dynamics in response to sharp microclimatic divergence. *Proc Natl Acad Sci U S A* 97: 6603-6607.
- Kazazian, H. H., Jr. (1998) Mobile elements and disease. *Curr Opin Genet Dev* 8: 343-350.
- Kazazian, H. H., Jr., Goodier, J. L. (2002) LINE drive, retrotransposition and genome instability. *Cell* 110: 277-280.
- Kazazian, H. H., Jr., Moran, J. V. (1998) The impact of L1 retrotransposons on the human genome. *Nat Genet* 19: 19-24.
- Keeney, S. (2001) Mechanism and control of meiotic recombination initiation. *Curr Top Dev Biol* 52: 1-53.
- Kobayashi, J., Tauchi, H., Sakamoto, S., Nakamura, A., Morishima, K., Matsuura, S., Kobayashi, T., Tamai, K., Tanimoto, K., Komatsu, K. (2002) NBS1 localizes to gamma-H2AX foci through interaction with the FHA/BRCT domain. *Curr Biol* 12: 1846-1851.
- Kosak, S. T., Skok, J. A., Medina, K. L., Riblet, R., Le Beau, M. M., Fisher, A. G., Singh, H. (2002) Subnuclear compartmentalization of immunoglobulin loci during lymphocyte development. *Science* 296: 158-162.
- Kozubek, S., Lukasova, E., Mareckova, A., Skalnikova, M., Kozubek, M., Bartova, E., Kroha, V., Krahulcova, E., Slotova, J. (1999) The topological organization of chromosomes

9 and 22 in cell nuclei has a determinative role in the induction of t(9,22) translocations and in the pathogenesis of t(9,22) leukemias. *Chromosoma* 108: 426-435.

Kozubek, S., Lukasova, E., Ryznar, L., Kozubek, M., Liskova, A., Govorun, R. D., Krasavin, E. A., Horneck, G. (1997) Distribution of ABL and BCR genes in cell nuclei of normal and irradiated lymphocytes. *Blood* 89: 4537-4545.

Krebs, J. E., Fry, C. J., Samuels, M. L., Peterson, C. L. (2000) Global role for chromatin remodeling enzymes in mitotic gene expression. *Cell* 102: 587-598.

Labrador, M., Corces, V. G. (2002) Setting the boundaries of chromatin domains and nuclear organization. *Cell* 111: 151-154.

Lachner, M., O'Carroll, D., Rea, S., Mechtler, K., Jenuwein, T. (2001) Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature* 410: 116-120.

Lamond, A. I., Earnshaw, W. C. (1998) Structure and function in the nucleus. *Science* 280: 547-553.

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., *et al.* (2001) Initial sequencing and analysis of the human genome. *Nature* 409: 860-921.

Le Douarin, B., Zechel, C., Garnier, J. M., Lutz, Y., Tora, L., Pierrat, P., Heery, D., Gronemeyer, H., Chambon, P., Losson, R. (1995) The N-terminal part of TIF1, a putative mediator of the ligand-dependent activation function (AF-2) of nuclear receptors, is fused to B-raf in the oncogenic protein T18. *Embo J* 14: 2020-2033.

Lee, D. Y., Hayes, J. J., Pruss, D., Wolffe, A. P. (1993) A positive role for histone acetylation in transcription factor access to nucleosomal DNA. *Cell* 72: 73-84.

Lee, G. S., Neiditch, M. B., Salus, S. S., Roth, D. B. (2004) RAG proteins shepherd double-strand breaks to a specific pathway, suppressing error-prone repair, but RAG nicking initiates homologous recombination. *Cell* 117: 171-184.

Lee, J. M., Sonnhammer, E. L. (2003) Genomic gene clustering analysis of pathways in eukaryotes. *Genome Res* 13: 875-882.

Lehnertz, B., Ueda, Y., Derijck, A. A., Braunschweig, U., Perez-Burgos, L., Kubicek, S., Chen, T., Li, E., Jenuwein, T., Peters, A. H. (2003) Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin. *Curr Biol* 13: 1192-1200.

Lercher, M. J., Urrutia, A. O., Hurst, L. D. (2002) Clustering of housekeeping genes provides a unified model of gene order in the human genome. *Nat Genet* 31: 180-183.

Lesko, S. A., Callahan, D. E., LaVilla, M. E., Wang, Z. P., Ts'o, P. O. (1995) The experimental homologous and heterologous separation distance histograms for the centromeres of chromosomes 7, 11, and 17 in interphase human T-lymphocytes. *Exp Cell Res* 219: 499-506.

Leuba, S. H., Bustamante, C., van Holde, K., Zlatanova, J. (1998a) Linker histone tails and N-tails of histone H3 are redundant: scanning force microscopy studies of reconstituted fibers. *Biophys J* 74: 2830-2839.

Leuba, S. H., Bustamante, C., Zlatanova, J., van Holde, K. (1998b) Contributions of linker histones and histone H3 to chromatin structure: scanning force microscopy studies on trypsinized fibers. *Biophys J* 74: 2823-2829.

Li, J., Jiang, T., Mao, J. H., Balmain, A., Peterson, L., Harris, C., Rao, P. H., Havlak, P., Gibbs, R., Cai, W. W. (2004) Genomic segmental polymorphisms in inbred mouse strains. *Nat Genet* 36: 952-954.

Liang, F., Han, M., Romanienko, P. J., Jasin, M. (1998) Homology-directed repair is a major double-strand break repair pathway in mammalian cells. *Proc Natl Acad Sci U S A* 95: 5172-5177.



Lin, F. L., Sperle, K., Sternberg, N. (1984) 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.

Lin, F. L., Sperle, K., Sternberg, N. (1990a) Intermolecular recombination between DNAs introduced into mouse L cells is mediated by a nonconservative pathway that leads to crossover products. *Mol Cell Biol* 10: 103-112.

Lin, F. L., Sperle, K., Sternberg, N. (1990b) Repair of double-stranded DNA breaks by homologous DNA fragments during transfer of DNA into mouse L cells. *Mol Cell Biol* 10: 113-119.

Lisby, M., Barlow, J. H., Burgess, R. C., Rothstein, R. (2004) Choreography of the DNA damage response: spatiotemporal relationships among checkpoint and repair proteins. *Cell* 118: 699-713.

Lisby, M., Mortensen, U. H., Rothstein, R. (2003) Colocalization of multiple DNA double-strand breaks at a single Rad52 repair centre. *Nat Cell Biol* 5: 572-577.

Liskay, R. M., Stachelek, J. L., Letsou, A. (1984) Homologous recombination between repeated chromosomal sequences in mouse cells. *Cold Spring Harb Symp Quant Biol* 49: 183-189.

Loeb, D. D., Padgett, R. W., Hardies, S. C., Shehee, W. R., Comer, M. B., Edgell, M. H., Hutchison, C. A., 3rd (1986) The sequence of a large L1Md element reveals a tandemly repeated 5' end and several features found in retrotransposons. *Mol Cell Biol* 6: 168-182.

Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F., Richmond, T. J. (1997) Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* 389: 251-260.

Lukasova, E., Kozubek, S., Kozubek, M., Kjeronska, J., Ryznar, L., Horakova, J., Krahulcova, E., Horneck, G. (1997) Localisation and distance between ABL and BCR genes in interphase nuclei of bone marrow cells of control donors and patients with chronic myeloid leukaemia. *Hum Genet* 100: 525-535.

Ma, Y., Pannicke, U., Schwarz, K., Lieber, M. R. (2002) Hairpin opening and overhang processing by an Artemis/DNA-dependent protein kinase complex in nonhomologous end joining and V(D)J recombination. *Cell* 108: 781-794.

Maconochie, M., Nonchev, S., Morrison, A., Krumlauf, R. (1996) Paralogous Hox genes: function and regulation. *Annu Rev Genet* 30: 529-556.

Mahy, N. L., Perry, P. E., Bickmore, W. A. (2002a) Gene density and transcription influence the localization of chromatin outside of chromosome territories detectable by FISH. *J Cell Biol* 159: 753-763.

Mahy, N. L., Perry, P. E., Gilchrist, S., Baldock, R. A., Bickmore, W. A. (2002b) Spatial organization of active and inactive genes and noncoding DNA within chromosome territories. *J Cell Biol* 157: 579-589.

Malkova, A., Ivanov, E. L., Haber, J. E. (1996) Double-strand break repair in the absence of RAD51 in yeast: a possible role for break-induced DNA replication. *Proc Natl Acad Sci U S A* 93: 7131-7136.

Malone, R. E., Ward, T., Lin, S., Waring, J. (1990) The RAD50 gene, a member of the double strand break repair epistasis group, is not required for spontaneous mitotic recombination in yeast. *Curr Genet* 18: 111-116.

Marahrens, Y. (1999) X-inactivation by chromosomal pairing events. *Genes Dev* 13: 2624-2632.

Margueron, R., Trojer, P., Reinberg, D. (2005) The key to development: interpreting the histone code? *Curr Opin Genet Dev* 15: 163-176.

Marshall, H., Morrison, A., Studer, M., Popperl, H., Krumlauf, R. (1996) Retinoids and Hox genes. *Faseb J* 10: 969-978.

Marshall, W. F., Straight, A., Marko, J. F., Swedlow, J., Dernburg, A., Belmont, A., Murray, A. W., Agard, D. A., Sedat, J. W. (1997) Interphase chromosomes undergo constrained diffusional motion in living cells. *Curr Biol* 7: 930-939.

Masumoto, H., Hawke, D., Kobayashi, R., Verreault, A. (2005) A role for cell-cycle-regulated histone H3 lysine 56 acetylation in the DNA damage response. *Nature* 436: 294-298.

McBurney, M. W., Lau, S., Jardine, K., Yang, X., Davies, B. (2001) Reexpression of a cluster of silenced transgenes is associated with their rearrangement. *Genes Chromosomes Cancer* 32: 311-323.

McIlwraith, M. J., Vaisman, A., Liu, Y., Fanning, E., Woodgate, R., West, S. C. (2005) Human DNA polymerase eta promotes DNA synthesis from strand invasion intermediates of homologous recombination. *Mol Cell* 20: 783-792.

Meselson, M. S., Radding, C. M. (1975) A general model for genetic recombination. *Proc Natl Acad Sci U S A* 72: 358-361.

Meyn, M. S. (1993) High spontaneous intrachromosomal recombination rates in ataxia-telangiectasia. *Science* 260: 1327-1330.

Minami, M., Poussin, K., Brechot, C., Paterlini, P. (1995) A novel PCR technique using Alu-specific primers to identify unknown flanking sequences from the human genome. *Genomics* 29: 403-408.

Misteli, T., Caceres, J. F., Spector, D. L. (1997) The dynamics of a pre-mRNA splicing factor in living cells. *Nature* 387: 523-527.

Modrich, P. (1991) Mechanisms and biological effects of mismatch repair. *Annu Rev Genet* 25: 229-253.

Moen, P. T., Jr., Johnson, C. V., Byron, M., Shopland, L. S., de la Serna, I. L., Imbalzano, A. N., Lawrence, J. B. (2004) Repositioning of muscle-specific genes relative to the periphery of SC-35 domains during skeletal myogenesis. *Mol Biol Cell* 15: 197-206.

Mohaghegh, P., Karow, J. K., Brosh Jr, R. M., Jr., Bohr, V. A., Hickson, I. D. (2001) The Bloom's and Werner's syndrome proteins are DNA structure-specific helicases. *Nucleic Acids Res* 29: 2843-2849.

Moore, S. C., Ausio, J. (1997) Major role of the histones H3-H4 in the folding of the chromatin fiber. *Biochem Biophys Res Commun* 230: 136-139.

Moran, J. V., DeBerardinis, R. J., Kazazian, H. H., Jr. (1999) Exon shuffling by L1 retrotransposition. *Science* 283: 1530-1534.

Morrish, T. A., Gilbert, N., Myers, J. S., Vincent, B. J., Stamato, T. D., Taccioli, G. E., Batzer, M. A., Moran, J. V. (2002) DNA repair mediated by endonuclease-independent LINE-1 retrotransposition. *Nat Genet* 31: 159-165.

Morrison, A. J., Highland, J., Krogan, N. J., Arbel-Eden, A., Greenblatt, J. F., Haber, J. E., Shen, X. (2004) INO80 and gamma-H2AX interaction links ATP-dependent chromatin remodeling to DNA damage repair. *Cell* 119: 767-775.

Munkel, C., Eils, R., Dietzel, S., Zink, D., Mehring, C., Wedemann, G., Cremer, T., Langowski, J. (1999) Compartmentalization of interphase chromosomes observed in simulation and experiment. *J Mol Biol* 285: 1053-1065.

Murti, J. R., Bumbulis, M., Schimenti, J. C. (1992) High-frequency germ line gene conversion in transgenic mice. *Mol Cell Biol* 12: 2545-2552.

Nagele, R. G., Freeman, T., McMorrow, L., Thomson, Z., Kitson-Wind, K., Lee, H. (1999) Chromosomes exhibit preferential positioning in nuclei of quiescent human cells. *J Cell Sci* 112 ( Pt 4): 525-535.

Narlikar, G. J., Fan, H. Y., Kingston, R. E. (2002) Cooperation between complexes that regulate chromatin structure and transcription. *Cell* 108: 475-487.

Nelson, D. L., Ballabio, A., Victoria, M. F., Pieretti, M., Bies, R. D., Gibbs, R. A., Maley, J. A., Chinault, A. C., Webster, T. D., Caskey, C. T. (1991) Alu-primed polymerase chain reaction for regional assignment of 110 yeast artificial chromosome clones from the human X chromosome: identification of clones associated with a disease locus. *Proc Natl Acad Sci U S A* 88: 6157-6161.

New, J. H., Sugiyama, T., Zaitseva, E., Kowalczykowski, S. C. (1998) Rad52 protein stimulates DNA strand exchange by Rad51 and replication protein A. *Nature* 391: 407-410.

Ng, H. H., Robert, F., Young, R. A., Struhl, K. (2003) Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity. *Mol Cell* 11: 709-719.

Nicolas, A. L., Young, C. S. (1994) Characterization of DNA end joining in a mammalian cell nuclear extract: junction formation is accompanied by nucleotide loss, which is limited and uniform but not site specific. *Mol Cell Biol* 14: 170-180.

Nikiforova, M. N., Stringer, J. R., Blough, R., Medvedovic, M., Fagin, J. A., Nikiforov, Y. E. (2000) Proximity of chromosomal loci that participate in radiation-induced rearrangements in human cells. *Science* 290: 138-141.

Nye, A. C., Rajendran, R. R., Stenoien, D. L., Mancini, M. A., Katzenellenbogen, B. S., Belmont, A. S. (2002) Alteration of large-scale chromatin structure by estrogen receptor. *Mol Cell Biol* 22: 3437-3449.

Orlando, V. (2003) Polycomb, epigenomes, and control of cell identity. *Cell* 112: 599-606.

Osborne, C. S., Chakalova, L., Brown, K. E., Carter, D., Horton, A., Debrand, E., Goyenechea, B., Mitchell, J. A., Lopes, S., Reik, W., Fraser, P. (2004) Active genes dynamically colocalize to shared sites of ongoing transcription. *Nat Genet* 36: 1065-1071.

- Ostertag, E. M., Kazazian, H. H., Jr. (2001) Biology of mammalian L1 retrotransposons. *Annu Rev Genet* 35: 501-538.
- Palmiter, R. D., Brinster, R. L. (1986) Germ-line transformation of mice. *Annu Rev Genet* 20: 465-499.
- Palstra, R. J., Tolhuis, B., Splinter, E., Nijmeijer, R., Grosveld, F., de Laat, W. (2003) The beta-globin nuclear compartment in development and erythroid differentiation. *Nat Genet* 35: 190-194.
- Paques, F., Haber, J. E. (1999) Multiple pathways of recombination induced by double-strand breaks in *Saccharomyces cerevisiae*. *Microbiol Mol Biol Rev* 63: 349-404.
- Parada, L. A., McQueen, P. G., Misteli, T. (2004) Tissue-specific spatial organization of genomes. *Genome Biol* 5: R44.
- Parada, L. A., Roix, J. J., Misteli, T. (2003) An uncertainty principle in chromosome positioning. *Trends Cell Biol* 13: 393-396.
- Park, P. C., De Boni, U. (1996) Transposition of DNase hypersensitive chromatin to the nuclear periphery coincides temporally with nerve growth factor-induced up-regulation of gene expression in PC12 cells. *Proc Natl Acad Sci U S A* 93: 11646-11651.
- Park, P. C., De Boni, U. (1998) A specific conformation of the territory of chromosome 17 locates ERBB-2 sequences to a DNase-hypersensitive domain at the nuclear periphery. *Chromosoma* 107: 87-95.
- Pastwa, E., Blasiak, J. (2003) Non-homologous DNA end joining. *Acta Biochim Pol* 50: 891-908.
- Patrinos, G. P., de Krom, M., de Boer, E., Langeveld, A., Imam, A. M., Strouboulis, J., de Laat, W., Grosveld, F. G. (2004) Multiple interactions between regulatory regions are required to stabilize an active chromatin hub. *Genes Dev* 18: 1495-1509.

Paull, T. T., Gellert, M. (1999) Nbs1 potentiates ATP-driven DNA unwinding and endonuclease cleavage by the Mre11/Rad50 complex. *Genes Dev* 13: 1276-1288.

Paull, T. T., Gellert, M. (2000) A mechanistic basis for Mre11-directed DNA joining at microhomologies. *Proc Natl Acad Sci U S A* 97: 6409-6414.

Pavlicek, A., Jabbari, K., Paces, J., Paces, V., Hejnar, J. V., Bernardi, G. (2001) Similar integration but different stability of Alus and LINEs in the human genome. *Gene* 276: 39-45.

Pederson, T. (1998) Thinking about a nuclear matrix. *J Mol Biol* 277: 147-159.

Phair, R. D., Misteli, T. (2000) High mobility of proteins in the mammalian cell nucleus. *Nature* 404: 604-609.

Pieper, F. R., de Wit, I. C., Pronk, A. C., Kooiman, P. M., Strijker, R., Krimpenfort, P. J., Nuyens, J. H., de Boer, H. A. (1992) Efficient generation of functional transgenes by homologous recombination in murine zygotes. *Nucleic Acids Res* 20: 1259-1264.

Platani, M., Goldberg, I., Lamond, A. I., Swedlow, J. R. (2002) Cajal body dynamics and association with chromatin are ATP-dependent. *Nat Cell Biol* 4: 502-508.

Pokholok, D. K., Harbison, C. T., Levine, S., Cole, M., Hannett, N. M., Lee, T. I., Bell, G. W., Walker, K., Rolfe, P. A., Herbolsheimer, E., *et al.* (2005) Genome-wide map of nucleosome acetylation and methylation in yeast. *Cell* 122: 517-527.

Politz, J. C., Tuft, R. A., Pederson, T., Singer, R. H. (1999) Movement of nuclear poly(A) RNA throughout the interchromatin space in living cells. *Curr Biol* 9: 285-291.

Powers, P. A., Smithies, O. (1986) Short gene conversions in the human fetal globin gene region: a by-product of chromosome pairing during meiosis? *Genetics* 112: 343-358.

Puget, N., Gad, S., Perrin-Vidoz, L., Sinilnikova, O. M., Stoppa-Lyonnet, D., Lenoir, G. M., Mazoyer, S. (2002) Distinct BRCA1 rearrangements involving the BRCA1 pseudogene suggest the existence of a recombination hot spot. *Am J Hum Genet* 70: 858-865.

Pusarla, R. H., Bhargava, P. (2005) Histones in functional diversification. Core histone variants. *Febs J* 272: 5149-5168.

Ragoczy, T., Telling, A., Sawado, T., Groudine, M., Kosak, S. T. (2003) A genetic analysis of chromosome territory looping: diverse roles for distal regulatory elements. *Chromosome Res* 11: 513-525.

Reymond, A., Marigo, V., Yaylaoglu, M. B., Leoni, A., Ucla, C., Scamuffa, N., Caccioppoli, C., Dermitzakis, E. T., Lyle, R., Banfi, S., *et al.* (2002) Human chromosome 21 gene expression atlas in the mouse. *Nature* 420: 582-586.

Rice, P., Longden, I., Bleasby, A. (2000) EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet* 16: 276-277.

Rich, T., Allen, R. L., Wyllie, A. H. (2000) Defying death after DNA damage. *Nature* 407: 777-783.

Richard, M., Belmaaza, A., Gusew, N., Wallenburg, J. C., Chartrand, P. (1994) Integration of a vector containing a repetitive LINE-1 element in the human genome. *Mol Cell Biol* 14: 6689-6695.

Richardson, C., Moynahan, M. E., Jasin, M. (1998) Double-strand break repair by interchromosomal recombination: suppression of chromosomal translocations. *Genes Dev* 12: 3831-3842.

Rijkers, T., Peetz, A., Ruther, U. (1994) Insertional mutagenesis in transgenic mice. *Transgenic Res* 3: 203-215.



Robert, F., Pokholok, D. K., Hannett, N. M., Rinaldi, N. J., Chandy, M., Rolfe, A., Workman, J. L., Gifford, D. K., Young, R. A. (2004) Global position and recruitment of HATs and HDACs in the yeast genome. *Mol Cell* 16: 199-209.

Rogakou, E. P., Pilch, D. R., Orr, A. H., Ivanova, V. S., Bonner, W. M. (1998) DNA double-stranded breaks induce histone H2AX phosphorylation on serine 139. *J Biol Chem* 273: 5858-5868.

Roix, J. J., McQueen, P. G., Munson, P. J., Parada, L. A., Misteli, T. (2003) Spatial proximity of translocation-prone gene loci in human lymphomas. *Nat Genet* 34: 287-291.

Roth, D. B., Wilson, J. H. (1986) Nonhomologous recombination in mammalian cells: role for short sequence homologies in the joining reaction. *Mol Cell Biol* 6: 4295-4304.

Rubnitz, J., Subramani, S. (1985) Rapid assay for extrachromosomal homologous recombination in monkey cells. *Mol Cell Biol* 5: 529-537.

Rubnitz, J., Subramani, S. (1986) Extrachromosomal and chromosomal gene conversion in mammalian cells. *Mol Cell Biol* 6: 1608-1614.

Sachs, R. K., van den Engh, G., Trask, B., Yokota, H., Hearst, J. E. (1995) A random-walk/giant-loop model for interphase chromosomes. *Proc Natl Acad Sci U S A* 92: 2710-2714.

Sadoni, N., Langer, S., Fauth, C., Bernardi, G., Cremer, T., Turner, B. M., Zink, D. (1999) Nuclear organization of mammalian genomes. Polar chromosome territories build up functionally distinct higher order compartments. *J Cell Biol* 146: 1211-1226.

Sadoni, N., Zink, D. (2004) Nascent RNA synthesis in the context of chromatin architecture. *Chromosome Res* 12: 439-451.

Saintigny, Y., Makienko, K., Swanson, C., Emond, M. J., Monnat, R. J., Jr. (2002) Homologous recombination resolution defect in werner syndrome. *Mol Cell Biol* 22: 6971-6978.

Sargent, R. G., Brenneman, M. A., Wilson, J. H. (1997) Repair of site-specific double-strand breaks in a mammalian chromosome by homologous and illegitimate recombination. *Mol Cell Biol* 17: 267-277.

Sargent, R. G., Meservy, J. L., Perkins, B. D., Kilburn, A. E., Intody, Z., Adair, G. M., Nairn, R. S., Wilson, J. H. (2000) Role of the nucleotide excision repair gene ERCC1 in formation of recombination-dependent rearrangements in mammalian cells. *Nucleic Acids Res* 28: 3771-3778.

Saxton, J. A., Martin, S. L. (1998) Recombination between subtypes creates a mosaic lineage of LINE-1 that is expressed and actively retrotransposing in the mouse genome. *J Mol Biol* 280: 611-622.

Scheer, U., Hock, R. (1999) Structure and function of the nucleolus. *Curr Opin Cell Biol* 11: 385-390.

Shelby, R. D., Hahn, K. M., Sullivan, K. F. (1996) Dynamic elastic behavior of alpha-satellite DNA domains visualized in situ in living human cells. *J Cell Biol* 135: 545-557.

Shiels, C., Islam, S. A., Vatcheva, R., Sasieni, P., Sternberg, M. J., Freemont, P. S., Sheer, D. (2001) PML bodies associate specifically with the MHC gene cluster in interphase nuclei. *J Cell Sci* 114: 3705-3716.

Shimoda, K., Cai, X., Kuhara, T., Maejima, K. (1991) Reconstruction of a large DNA fragment from coinjected small fragments by homologous recombination in fertilized mouse eggs. *Nucleic Acids Res* 19: 6654.

Shinohara, A., Ogawa, T. (1998) Stimulation by Rad52 of yeast Rad51-mediated recombination. *Nature* 391: 404-407.

Shinohara, A., Shinohara, M., Ohta, T., Matsuda, S., Ogawa, T. (1998) Rad52 forms ring structures and co-operates with RPA in single-strand DNA annealing. *Genes Cells* 3: 145-156.

Shopland, L. S., Byron, M., Stein, J. L., Lian, J. B., Stein, G. S., Lawrence, J. B. (2001) Replication-dependent histone gene expression is related to Cajal body (CB) association but does not require sustained CB contact. *Mol Biol Cell* 12: 565-576.

Shopland, L. S., Johnson, C. V., Byron, M., McNeil, J., Lawrence, J. B. (2003) Clustering of multiple specific genes and gene-rich R-bands around SC-35 domains: evidence for local euchromatic neighborhoods. *J Cell Biol* 162: 981-990.

Sims, R. J., 3rd, Nishioka, K., Reinberg, D. (2003) Histone lysine methylation: a signature for chromatin function. *Trends Genet* 19: 629-639.

Smith, K. P., Moen, P. T., Wydner, K. L., Coleman, J. R., Lawrence, J. B. (1999) Processing of endogenous pre-mRNAs in association with SC-35 domains is gene specific. *J Cell Biol* 144: 617-629.

Sonza, S., Maerz, A., Deacon, N., Meanger, J., Mills, J., Crowe, S. (1996) Human immunodeficiency virus type 1 replication is blocked prior to reverse transcription and integration in freshly isolated peripheral blood monocytes. *J Virol* 70: 3863-3869.

Stadler, S., Schnapp, V., Mayer, R., Stein, S., Cremer, C., Bonifer, C., Cremer, T., Dietzel, S. (2004) The architecture of chicken chromosome territories changes during differentiation. *BMC Cell Biol* 5: 44.

Strahl, B. D., Allis, C. D. (2000) The language of covalent histone modifications. *Nature* 403: 41-45.

Subramani, S., Rubnitz, J. (1985) Recombination events after transient infection and stable integration of DNA into mouse cells. *Mol Cell Biol* 5: 659-666.

Sugiyama, T., Zaitseva, E. M., Kowalczykowski, S. C. (1997) A single-stranded DNA-binding protein is needed for efficient presynaptic complex formation by the *Saccharomyces cerevisiae* Rad51 protein. *J Biol Chem* 272: 7940-7945.

- Sun, H. B., Yokota, H. (1999) Correlated positioning of homologous chromosomes in daughter fibroblast cells. *Chromosome Res* 7: 603-610.
- Sun, Y., Durrin, L. K., Krontiris, T. G. (2003) Specific interaction of PML bodies with the TP53 locus in Jurkat interphase nuclei. *Genomics* 82: 250-252.
- Sung, P. (1994) Catalysis of ATP-dependent homologous DNA pairing and strand exchange by yeast RAD51 protein. *Science* 265: 1241-1243.
- Sung, P. (1997) Function of yeast Rad52 protein as a mediator between replication protein A and the Rad51 recombinase. *J Biol Chem* 272: 28194-28197.
- Szostak, J. W., Orr-Weaver, T. L., Rothstein, R. J., Stahl, F. W. (1983) The double-strand-break repair model for recombination. *Cell* 33: 25-35.
- Takata, M., Sasaki, M. S., Sonoda, E., Morrison, C., Hashimoto, M., Utsumi, H., Yamaguchi-Iwai, Y., Shinohara, A., Takeda, S. (1998) Homologous recombination and non-homologous end-joining pathways of DNA double-strand break repair have overlapping roles in the maintenance of chromosomal integrity in vertebrate cells. *Embo J* 17: 5497-5508.
- Tanabe, H., Habermann, F. A., Solovei, I., Cremer, M., Cremer, T. (2002a) Non-random radial arrangements of interphase chromosome territories: evolutionary considerations and functional implications. *Mutat Res* 504: 37-45.
- Tanabe, H., Muller, S., Neusser, M., von Hase, J., Calcagno, E., Cremer, M., Solovei, I., Cremer, C., Cremer, T. (2002b) Evolutionary conservation of chromosome territory arrangements in cell nuclei from higher primates. *Proc Natl Acad Sci U S A* 99: 4424-4429.
- Thoma, F., Koller, T. (1977) Influence of histone H1 on chromatin structure. *Cell* 12: 101-107.

Thoma, F., Koller, T., Klug, A. (1979) Involvement of histone H1 in the organization of the nucleosome and of the salt-dependent superstructures of chromatin. *J Cell Biol* 83: 403-427.

Thomas, J. O., Rees, C., Finch, J. T. (1992) Cooperative binding of the globular domains of histones H1 and H5 to DNA. *Nucleic Acids Res* 20: 187-194.

Thomson, I., Gilchrist, S., Bickmore, W. A., Chubb, J. R. (2004) The radial positioning of chromatin is not inherited through mitosis but is established de novo in early G1. *Curr Biol* 14: 166-172.

Tolhuis, B., Palstra, R. J., Splinter, E., Grosveld, F., de Laat, W. (2002) Looping and interaction between hypersensitive sites in the active beta-globin locus. *Mol Cell* 10: 1453-1465.

Toth, K. F., Knoch, T. A., Wachsmuth, M., Frank-Stohr, M., Stohr, M., Bacher, C. P., Muller, G., Rippe, K. (2004) Trichostatin A-induced histone acetylation causes decondensation of interphase chromatin. *J Cell Sci* 117: 4277-4287.

Tremblay, A., Jasin, M., Chartrand, P. (2000) 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.

Trujillo, K. M., Yuan, S. S., Lee, E. Y., Sung, P. (1998) Nuclease activities in a complex of human recombination and DNA repair factors Rad50, Mre11, and p95. *J Biol Chem* 273: 21447-21450.

Tse, C., Hansen, J. C. (1997) Hybrid trypsinized nucleosomal arrays: identification of multiple functional roles of the H2A/H2B and H3/H4 N-termini in chromatin fiber compaction. *Biochemistry* 36: 11381-11388.

Tse, C., Sera, T., Wolffe, A. P., Hansen, J. C. (1998) Disruption of higher-order folding by core histone acetylation dramatically enhances transcription of nucleosomal arrays by RNA polymerase III. *Mol Cell Biol* 18: 4629-4638.

Tumbar, T., Sudlow, G., Belmont, A. S. (1999) Large-scale chromatin unfolding and remodeling induced by VP16 acidic activation domain. *J Cell Biol* 145: 1341-1354.

Valerie, K., Povirk, L. F. (2003) Regulation and mechanisms of mammalian double-strand break repair. *Oncogene* 22: 5792-5812.

van Attikum, H., Fritsch, O., Hohn, B., Gasser, S. M. (2004) Recruitment of the INO80 complex by H2A phosphorylation links ATP-dependent chromatin remodeling with DNA double-strand break repair. *Cell* 119: 777-788.

Van Dyck, E., Stasiak, A. Z., Stasiak, A., West, S. C. (2001) Visualization of recombination intermediates produced by RAD52-mediated single-strand annealing. *EMBO Rep* 2: 905-909.

Verneau, O., Catzeflis, F., Furano, A. V. (1998) Determining and dating recent rodent speciation events by using L1 (LINE-1) retrotransposons. *Proc Natl Acad Sci U S A* 95: 11284-11289.

Verschure, P. J., van Der Kraan, I., Manders, E. M., van Driel, R. (1999) Spatial relationship between transcription sites and chromosome territories. *J Cell Biol* 147: 13-24.

Vettese-Dadey, M., Grant, P. A., Hebbes, T. R., Crane-Robinson, C., Allis, C. D., Workman, J. L. (1996) Acetylation of histone H4 plays a primary role in enhancing transcription factor binding to nucleosomal DNA in vitro. *Embo J* 15: 2508-2518.

Vilenchik, M. M., Knudson, A. G. (2003) Endogenous DNA double-strand breaks: production, fidelity of repair, and induction of cancer. *Proc Natl Acad Sci U S A* 100: 12871-12876.

Visser, A. E., Jaunin, F., Fakan, S., Aten, J. A. (2000) High resolution analysis of interphase chromosome domains. *J Cell Sci* 113 ( Pt 14): 2585-2593.

Volpi, E. V., Chevret, E., Jones, T., Vatcheva, R., Williamson, J., Beck, S., Campbell, R. D., Goldsworthy, M., Powis, S. H., Ragoussis, J., *et al.* (2000) Large-scale chromatin

organization of the major histocompatibility complex and other regions of human chromosome 6 and its response to interferon in interphase nuclei. *J Cell Sci* 113 ( Pt 9): 1565-1576.

Wade, P. A. (2001) Methyl CpG-binding proteins and transcriptional repression. *Bioessays* 23: 1131-1137.

Walker, J. R., Corpina, R. A., Goldberg, J. (2001) Structure of the Ku heterodimer bound to DNA and its implications for double-strand break repair. *Nature* 412: 607-614.

Walter, J., Schermelleh, L., Cremer, M., Tashiro, S., Cremer, T. (2003) Chromosome order in HeLa cells changes during mitosis and early G1, but is stably maintained during subsequent interphase stages. *J Cell Biol* 160: 685-697.

Wang, J., Shiels, C., Sasieni, P., Wu, P. J., Islam, S. A., Freemont, P. S., Sheer, D. (2004) Promyelocytic leukemia nuclear bodies associate with transcriptionally active genomic regions. *J Cell Biol* 164: 515-526.

Wang, X., He, C., Moore, S. C., Ausio, J. (2001) Effects of histone acetylation on the solubility and folding of the chromatin fiber. *J Biol Chem* 276: 12764-12768.

Wang, Y., Cortez, D., Yazdi, P., Neff, N., Elledge, S. J., Qin, J. (2000) BASC, a super complex of BRCA1-associated proteins involved in the recognition and repair of aberrant DNA structures. *Genes Dev* 14: 927-939.

Waterston, R. H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J. F., Agarwal, P., Agarwala, R., Ainscough, R., Alexandersson, M., An, P., *et al.* (2002) Initial sequencing and comparative analysis of the mouse genome. *Nature* 420: 520-562.

Wechsler, M. A., Kladde, M. P., Alfieri, J. A., Peterson, C. L. (1997) Effects of Sin-versions of histone H4 on yeast chromatin structure and function. *Embo J* 16: 2086-2095.

Wei, W., Gilbert, N., Ooi, S. L., Lawler, J. F., Ostertag, E. M., Kazazian, H. H., Boeke, J. D., Moran, J. V. (2001) Human L1 retrotransposition: cis preference versus trans complementation. *Mol Cell Biol* 21: 1429-1439.

Weierich, C., Brero, A., Stein, S., von Hase, J., Cremer, C., Cremer, T., Solovei, I. (2003) Three-dimensional arrangements of centromeres and telomeres in nuclei of human and murine lymphocytes. *Chromosome Res* 11: 485-502.

Weimer, R., Haaf, T., Kruger, J., Poot, M., Schmid, M. (1992) Characterization of centromere arrangements and test for random distribution in G0, G1, S, G2, G1, and early S' phase in human lymphocytes. *Hum Genet* 88: 673-682.

Weston, A. D., Blumberg, B., Underhill, T. M. (2003) Active repression by unliganded retinoid receptors in development: less is sometimes more. *J Cell Biol* 161: 223-228.

Williams, R. R., Broad, S., Sheer, D., Ragoussis, J. (2002) Subchromosomal positioning of the epidermal differentiation complex (EDC) in keratinocyte and lymphoblast interphase nuclei. *Exp Cell Res* 272: 163-175.

Wold, M. S. (1997) Replication protein A: a heterotrimeric, single-stranded DNA-binding protein required for eukaryotic DNA metabolism. *Annu Rev Biochem* 66: 61-92.

Wong, E. A., Capecchi, M. R. (1986) Analysis of homologous recombination in cultured mammalian cells in transient expression and stable transformation assays. *Somat Cell Mol Genet* 12: 63-72.

Woodcock, C. L., Grigoryev, S. A., Horowitz, R. A., Whitaker, N. (1993) A chromatin folding model that incorporates linker variability generates fibers resembling the native structures. *Proc Natl Acad Sci U S A* 90: 9021-9025.

Wurtele, H., Chartrand, P. (2006) Genome-wide scanning of HoxB1-associated loci in mouse ES cells using an open-ended Chromosome Conformation Capture methodology. *Chromosome Res* 14: 477-495.



Wurtele, H., Gusew, N., Lussier, R., Chartrand, P. (2005) Characterization of in vivo recombination activities in the mouse embryo. *Mol Genet Genomics* 273: 252-263.

Wurtele, H., Little, K. C., Chartrand, P. (2003) Illegitimate DNA integration in mammalian cells. *Gene Ther* 10: 1791-1799.

Wurtele, H., Verreault, A. (2006) Histone post-translational modifications and the response to DNA double-strand breaks. *Curr Opin Cell Biol* 18: 137-144.

Yang, H., Jeffrey, P. D., Miller, J., Kinnucan, E., Sun, Y., Thoma, N. H., Zheng, N., Chen, P. L., Lee, W. H., Pavletich, N. P. (2002) BRCA2 function in DNA binding and recombination from a BRCA2-DSS1-ssDNA structure. *Science* 297: 1837-1848.

Ye, Q., Hu, Y. F., Zhong, H., Nye, A. C., Belmont, A. S., Li, R. (2001) BRCA1-induced large-scale chromatin unfolding and allele-specific effects of cancer-predisposing mutations. *J Cell Biol* 155: 911-921.

Yokota, H., Singer, M. J., van den Engh, G. J., Trask, B. J. (1997) Regional differences in the compaction of chromatin in human G0/G1 interphase nuclei. *Chromosome Res* 5: 157-166.

Yokota, H., van den Engh, G., Hearst, J. E., Sachs, R. K., Trask, B. J. (1995) Evidence for the organization of chromatin in megabase pair-sized loops arranged along a random walk path in the human G0/G1 interphase nucleus. *J Cell Biol* 130: 1239-1249.

Zachar, Z., Kramer, J., Mims, I. P., Bingham, P. M. (1993) Evidence for channeled diffusion of pre-mRNAs during nuclear RNA transport in metazoans. *J Cell Biol* 121: 729-742.

Zhou, C., Yang, Y., Jong, A. Y. (1990) Mini-prep in ten minutes. *Biotechniques* 8: 172-173.

Zhou, G. L., Xin, L., Song, W., Di, L. J., Liu, G., Wu, X. S., Liu, D. P., Liang, C. C. (2006) Active chromatin hub of the mouse alpha-globin locus forms in a transcription factory of clustered housekeeping genes. *Mol Cell Biol* 26: 5096-5105.

Zink, D., Cremer, T., Saffrich, R., Fischer, R., Trendelenburg, M. F., Ansorge, W., Stelzer, E. H. (1998) Structure and dynamics of human interphase chromosome territories in vivo. *Hum Genet* 102: 241-251.

Zirbel, R. M., Mathieu, U. R., Kurz, A., Cremer, T., Lichter, P. (1993) Evidence for a nuclear compartment of transcription and splicing located at chromosome domain boundaries. *Chromosome Res* 1: 93-106.

