

# Thèse

en vue de l'obtention du titre de

**Docteur de l'université Paris VI - Pierre et Marie Curie**

Spécialité : **Sciences de la Vie**

École Doctorale iViv - Interdisciplinaire pour le Vivant

Présentée et soutenue publiquement le 7 février 2014 par  
Jean OLLION

## **Dynamique de l'organisation nucléaire des séquences d'ADN répétées centromériques humaines au cours du cycle cellulaire**

Laboratoire "Régulation et dynamique des génomes", Muséum National  
d'Histoire Naturelle RDDM, CNRS UMR 7196, INSERM U565, Paris

Jury :

Dr. Nathalie BEAUJEAN	Rapporteur
Dr. Christophe ZIMMER	Rapporteur
Dr. Frédéric BANTIGNIES	Examinateur
Dr. Nathalie DOSTATNI	Examinateur
Dr. Alexandre MÉJAT	Examinateur
Dr. Christophe ESCUDÉ	Co-directeur de Thèse
Dr. Thomas BOUDIER	Co-directeur de Thèse



*À tous les miens*



# Remerciements

J'aimerais tout d'abord remercier le Dr. Nathalie Beaujean, le Dr. Christophe Zimmer, le Dr. Frédéric Bantignies, le Dr. Nathalie Dostatni et le Dr. Alexandre Méjat pour avoir jugé mon travail de thèse. Merci pour la discussion enrichissante lors de la phase de questions, je suis certain qu'elle permettra notamment à ceux qui prendront la suite de ce travail de l'améliorer grandement.

Je souhaite également remercier les Drs. Carine Giovannangeli et Jean-François Riou pour m'avoir accueilli au laboratoire.

Je remercie le Dr. Christophe Escudé pour m'avoir encadré durant toutes ces années. Nous nous sommes engagés dans une direction qui était nouvelle pour nous, et qui a demandé beaucoup de travail d'exploration. Ce travail a été très enrichissant et stimulant parce qu'il m'a permis de dialoguer avec des spécialistes de domaines très différents; il a été possible parce que tu m'a fait confiance et que tu m'a donné des moyens. Je te remercie aussi pour m'avoir toujours soutenu malgré mes doutes, et pour avoir passé beaucoup de temps à m'aider dans tous les travaux de présentation, depuis les premières conférences à la thèse même. J'espère que ce travail exploratoire sera continué dans l'équipe, et qu'il vous donnera des clés pour approfondir vos questions. Je remercie le Dr. Thomas Boudier<sup>1</sup>, pour m'avoir encadré et supporté<sup>2</sup>, je crois que nous sommes arrivés à construire un bel outil, et j'espère qu'il continuera à s'améliorer<sup>3</sup>.

Je voudrais particulièrement remercier Julien (aka le Juju) et François (aka le Franz), sans qui ce travail ne serait pas arrivé à ce point. Cela a été un vrai plaisir de travailler avec vous, merci pour votre soutien sans failles, pour votre bonne humeur, votre enthousiasme, vos blagues pouraves (dans le bus)... Je voudrais saluer votre conscience professionnelle qui vous a poussé à poursuivre le travail jusqu'à des heures indécentes lors de nos "debriefs" au Brewberry. D'ailleurs, c'est quand le prochain debrief? Merci Laurianne pour ton aide précieuse, notamment pour la correction du manuscrit. J'ai trouvé tes remarques vraiment pertinentes et intelligentes, même lorsque le sujet était très éloigné de ton domaine d'expertise. Au passage, j'ai beaucoup appris sur la syntaxe de la langue française! Merci à Jean-Baptiste Boulé pour tes conseils et les discussions (toujours soit ultra perchées, soit très pratiques), qui m'ont souvent apporté beaucoup d'enthousiasme, et aidé à prendre du recul. Merci à Philippe Andrey, pour tous les conseils et les discussions. J'ai appris beaucoup de choses à tes côtés, non seulement dans des domaines techniques mais aussi sur la méthode scientifique. Cela a été un plaisir de faire mon monitorat avec toi.

---

1. Ceci est une note de bas de page...

2. Aux deux sens du terme!

3. A quand la 4D?

Merci à tous les membres du laboratoire pour tous les gestes du quotidien qui en font un lieu de travail très agréable. Merci pour votre accueil, votre bienveillance, votre bonne humeur, votre aide et vos conseils. Dans le désordre : merci à Erika pour tes conseils (c'est toi qui m'a appris à cultiver des cellules entre autres!), Loïc P<sup>32</sup> pour ton aide sur les manip (surtout du temps où je faisais des gels) et sur les techniques conserve/culture de légumes (un jour ça me servira!), Loïc pour les discussions (stats et autres), Anne pour les conseils de sexiness pour ma soutenance (désolé ça m'a marqué!), Gilles pour les discussions de guitare et de culture de plantes, Kristov pour la musique ancienne (et pas ton métal ça ne m'intéresse pas!), Lorena pour toutes les discussions et le soutien moral, Fara pour avoir toujours su rattraper mes maladresses administratives avec le sourire (je crois que je ne t'ai jamais demandé de m'inscrire à un congrès plus d'une semaine avant la date limite, et jamais en dehors des périodes de vacances du comptable de l'INSERM!), Patritzzzia notamment pour le deal sur les tomates séchées (contre cacao cru), Charlottes pour tes blagues nulles (et aussi pour comprendre les miennes), François Strauss, Judith, Tula, Anne-Laure, Sacha pour toutes les discussions et notamment au moment des répétitions de soutenance, les djeun's (cf juju), Dona, Vivien, Marion, Hind, et tous les autres.

Je vous souhaite à tous beaucoup d'épanouissement dans votre travail de recherche et.. surtout dans tous les autres aspects de la vie.

Je voudrai enfin remercier mes proches. Ma famille : Pop, Mum, Charlot, Math, Vava, Sug, Vincent, Dominique pour tous les bon moments qu'on passe ensemble, pour votre aide, vos conseils, votre soutien, vos encouragements et toutes vos attentions. Et je n'oublie pas mes petites nièces que j'adore, Zizou et Ondjjjjineu, étincelles de vie, je suis continuellement ébloui de vous voir grandir. Mes amis : Christo (hermano!), Rémy-la-fritouze, Cléclé, Marion, Caro, Manelle, le Modjo, le Zoutou & la Juju, Bouboule, Tata, PJ (bièrejoint), Cam, Val & Yo, Madame Poe, Mano, Clem & Gabs, Fab & May et tous les autres. Quel bonheur de vous connaître. Et puis l'abuelo et tous ceux qui dansent autour : Sergio, Célinette, Solal, Yann-marc, Raül, Karla, Xavier, Gaby, Lara ... et bien-sur DT, pour tous ces bon moments passés et à venir, aho!

# Résumé

Le noyau des cellules est une structure très organisée, dont l'organisation joue un rôle important dans la régulation de l'expression des gènes. La compréhension des mécanismes à l'origine de cette organisation est donc essentielle à la compréhension du fonctionnement des génomes.

De nombreuses expériences conduites chez la souris ont montré que les régions centromériques (RC) des chromosomes jouent un rôle dans l'organisation du noyau. L'organisation spatiale des RCs humaines est beaucoup moins étudiée, principalement à cause de la complexité des séquences qui les composent, qui rend plus difficile leur détection.

Nous avons développé des outils de traitement et d'analyse quantitative d'image, qui, combinés à des nouveaux marqueurs des RCs humaines, nous ont permis de mieux décrire deux aspects de leur organisation spatiale. D'une part nous avons montré qu'elles se positionnent préférentiellement en périphérie du noyau ou aux bords des nucléoles, avec des fréquences qui dépendent des chromosomes. D'autre part nous avons montré qu'elles s'agrègent dans le noyau pour former un compartiment d'hétérochromatine, qui présente des caractéristiques similaires à celui observé dans d'autres espèces telles que la souris. Ces deux aspects sont tous deux inter-dépendants et varient au cours du cycle cellulaire.

Cette description nouvelle met sur la piste de mécanismes responsables de l'organisation particulière des RCs, qui pourront être étudiés grâce à la méthode d'analyse et aux observables que nous avons développées. L'étude de ces mécanismes permettra de mieux comprendre la fonction des RCs humaines dans l'organisation du noyau.

**Mots clés :** organisation nucléaire, centromères, analyse d'image, cycle cellulaire, hétérochromatine, nucléole



# Abstract

The cell nucleus is a highly organized structure, playing an important role in gene regulation. Understanding the underlying mechanisms is therefore essential for understanding genome function.

Numerous studies conducted in mouse cells have shown that centromeric regions (RC) of chromosomes play a role in nuclear organization. The spatial organization of human RCs is less studied, mainly because of the complexity of the underlying DNA sequences that make them hard to detect.

We have developed image processing and analysis tools, that, combined with new markers for human RCs, have allowed us to draw a better description of two features of their spatial organization. On the one hand, we have shown that they are preferentially located close to the nuclear periphery or nucleoli borders, with chromosome-dependent frequencies. On the other hand, we have shown that they cluster to form a heterochromatic compartment that displays similar properties as the one observed in other species such as mouse. Both features are inter-dependent, and vary throughout the cell-cycle.

This new description puts on the track of mechanisms responsible for the peculiar organization of RCs. Those mechanisms could be studied using the methodology and the observables we have developed. The study of those mechanisms will provide a better understanding of human RC function in nuclear organization.

**Keywords :** nuclear organization, centromeres, image analysis, cell-cycle, heterochromatin, nucleolus



# Table des matières

<b>A</b>	<b>Introduction</b>	<b>17</b>
A.1	L'architecture tridimensionnelle du noyau . . . . .	18
A.1.1	Différents niveaux hiérarchiques d'organisation de la chromatine	19
A.1.2	La compartimentation nucléaire . . . . .	31
A.1.3	Positionnement nucléaire du génome . . . . .	37
A.1.4	Dynamique de l'architecture nucléaire . . . . .	45
A.1.5	Modélisation de l'architecture nucléaire . . . . .	50
A.2	Structure et fonction des régions centromériques . . . . .	54
A.2.1	Généralités sur les régions centromériques . . . . .	54
A.2.2	Structure chromatinienne des régions centromériques . . . . .	56
A.2.3	Séquences d'ADN centromérique humaines . . . . .	56
A.2.4	Lien entre structure et fonction . . . . .	61
A.3	Organisation nucléaire des régions centromériques . . . . .	65
A.3.1	Positionnement nucléaire des régions centromériques . . . . .	65
A.3.2	Associations de régions centromériques . . . . .	68
A.3.3	Mécanismes moléculaires impliqués dans l'organisation nu- cléaire des régions centromériques . . . . .	75
A.4	Méthodes d'études de l'architecture nucléaire . . . . .	81
A.4.1	Approches expérimentales . . . . .	81
A.4.2	Déterminisme dans l'organisation spatiale . . . . .	85
A.4.3	Analyse spatiale des signaux de microscopie . . . . .	87
A.5	Projet de thèse . . . . .	98
<b>B</b>	<b>Matériels et Méthodes</b>	<b>99</b>
B.1	Culture Cellulaire . . . . .	99
B.2	Marquages <i>in situ</i> . . . . .	99
B.2.1	Préparation des échantillons . . . . .	99
B.2.2	Immuno-FISH . . . . .	100
B.3	Acquisition des images . . . . .	103
B.4	Analyse statistique . . . . .	104
B.4.1	Test statistique . . . . .	104
B.4.2	Distributions cumulatives . . . . .	104
<b>C</b>	<b>Résultats</b>	<b>105</b>
C.1	Développement d'outils pour l'analyse d'image quantitative . . . . .	105
C.1.1	TANGO : un outil générique pour l'analyse d'image 3D haut- débit . . . . .	105
C.1.2	Procédures de segmentations . . . . .	113
C.1.3	Mesures . . . . .	118
C.1.4	Analyse quantitative du cycle cellulaire . . . . .	121

C.1.5	Analyse du signal Ki67 . . . . .	122
C.1.6	Discussion et conclusion . . . . .	126
C.2	Positionnement des régions centromériques humaines . . . . .	128
C.2.1	Introduction . . . . .	129
C.2.2	Résultats . . . . .	131
C.2.3	Discussion . . . . .	140
C.2.4	Matériel supplémentaire . . . . .	143
C.3	Agrégations des régions centromériques humaines . . . . .	148
C.3.1	Résultats . . . . .	148
C.3.2	Discussion . . . . .	163
<b>D</b>	<b>Discussion et Conclusion</b>	<b>169</b>
D.1	Méthodologie d'étude de l'organisation nucléaire . . . . .	169
D.1.1	Sources de variabilité inter-cellulaire . . . . .	169
D.1.2	Portée des analyses . . . . .	170
D.2	Organisation nucléaire des régions centromériques humaines . . . . .	171
D.2.1	Liens entre positionnement et agrégation . . . . .	171
D.2.2	Dynamique de l'organisation spatiale des régions centromériques . . . . .	172
D.3	Conclusion . . . . .	176
<b>E</b>	<b>Annexes</b>	<b>177</b>
	<b>Bibliographie</b>	<b>181</b>

# Table des figures

A.1	Différents niveaux hiérarchiques d'organisation de la chromatine dans un noyau de mammifère . . . . .	20
A.2	Structure cristallographique du nucléosome . . . . .	21
A.3	Principales modifications post-traductionnelles de la queue des histones . . . . .	22
A.4	Segmentation linéaire des chromosomes en domaines . . . . .	26
A.5	Domaines topologiquement associés . . . . .	27
A.6	Preuves expérimentales de l'existence des territoires chromosomiques . . . . .	29
A.7	Principes de la conformation spatiale du génome dans le noyau . . . . .	30
A.8	Fonctions des compartiments nucléaires . . . . .	32
A.9	Vue d'ensemble des rôles connus des protéines se liant aux lamines . . . . .	34
A.10	Mécanisme d'affinité permettant d'expliquer l'établissement des profils d'organisation spatiale des chromosomes . . . . .	38
A.11	Comparaison des trois types de dynamiques de la chromatine . . . . .	46
A.12	Modèle d'auto-organisation du génome . . . . .	51
A.13	Modèle décrivant les interactions entre structure et fonction . . . . .	53
A.14	Structure génomique des régions centromériques entre les espèces . . . . .	55
A.15	Modèle de structure secondaire de la chromatine centromérique . . . . .	57
A.16	Modèle d'organisation des séquences $\alpha$ -satellites au niveau des régions centromériques humaines . . . . .	58
A.17	Structure du LNA . . . . .	59
A.18	Modèle d'évolution des régions centromériques par expansion . . . . .	60
A.19	Modèle de mécanisme de formation de néocentromères au cours de l'évolution . . . . .	62
A.20	Organisation spatiale des centromères et péricentromères murins . . . . .	70
A.21	Observation des chromocentres murins dans différents types cellulaires . . . . .	71
A.22	Associations de régions centromériques chez l'homme . . . . .	72
A.23	Architecture inversée de la chromatine chez la souris . . . . .	77
A.24	Différentes mesures du positionnement radial . . . . .	90
A.25	Méthodologie de statistiques spatiales . . . . .	93
C.1	Chaîne de traitement du signal de TANGO . . . . .	109
C.2	Paramétrage des chaînes de traitement du signal . . . . .	110
C.3	Interface de TANGO . . . . .	111
C.4	Différents modes de représentation des images . . . . .	111
C.5	Interface entre TANGO et R . . . . .	112
C.6	Environnements utilisés par TANGO . . . . .	113
C.7	Étapes de la procédure de segmentation des noyaux . . . . .	115
C.8	Étapes de la procédure de segmentation des foyers . . . . .	116

C.9 Répartition des volumes des foyers CenpA par noyaux . . . . .	117
C.10 Mesure de l'EVF depuis différentes structures de références . . . . .	119
C.11 Mesure d'orientation . . . . .	121
C.12 Patterns caractéristiques des différents marqueurs du cycle cellulaire	122
C.13 Analyse quantitative du cycle cellulaire . . . . .	123
C.14 Reproductibilité de la mesure du positionnement nucléaire du centromère du chromosome 18 entre des cultures cellulaires indépendantes	132
C.15 Reproductibilité des profils de positionnement entre les types cellulaires GM06990 et GM12878 . . . . .	135
C.16 Différences de positionnement nucléaire des centromères individuels en fonction du cycle cellulaire . . . . .	137
C.17 Comparaison du positionnement nucléaire de centromère avec des régions génomiques non centromériques . . . . .	138
C.18 Positionnement des centromères au sein des territoires chromosomiques	139
C.19 Reproductibilité des volumes des noyaux et des nucléoles entre des cultures indépendantes . . . . .	145
C.20 Profil de positionnement des centromères des chromosomes 8 et 22 .	146
C.21 Influence du cycle cellulaire sur le positionnement nucléaire des centromères . . . . .	147
C.22 Attribution du nombre de centromères par foyer cenpA . . . . .	150
C.23 Détection de foyers d'hétérochromatine par immunocytochimie dans les cellules GM12878 . . . . .	151
C.24 Détection des séquences péricentromériques humaines par FISH . . .	152
C.25 Deux descriptions de l'agrégation globale des RCs . . . . .	154
C.26 Composition des foyers PCY . . . . .	156
C.27 Variation de l'agrégation globale des RCs au cours du cycle cellulaire	157
C.28 Variation de la composition des foyers PCY au cours du cycle cellulaire.	159
C.29 Différences de composition des foyers PCY en fonction de la localisation sub-nucléaire. . . . .	160
C.30 Orientation des foyers CenpA par rapport aux foyers PCY et à la périphérie nucléaire . . . . .	161
C.31 Détection des séquences péricentromériques humaines par immunofluorescence 3D sur cellules mésenchymateuses souches . . . . .	162
C.32 Comparaison de l'agrégation des types cellulaires hMES et GM12878 au cours du cycle cellulaire . . . . .	162
C.33 Modèle décrivant les variations de l'agrégation au cours du cycle cellulaire . . . . .	165
D.1 Les régions centromériques peuvent se trouver en dehors de leur territoire chromosomique . . . . .	175

# Liste des tableaux

A.1 Répartition des principales séquences satellites sur les régions péri-centromériques des différents chromosomes humains . . . . .	58
A.2 Localisation nucléaire des centromères portés par des chromosomes spécifiques . . . . .	67
C.1 Séquences des sondes utilisées pour la détections de régions centromériques de chromosomes spécifiques . . . . .	144
E.1 Modules de traitement d'images inclus dans la première version de TANGO . . . . .	177
E.2 Modules d'analyse quantitative inclus dans la première version de TANGO . . . . .	179
E.3 Logiciels tiers utilisés par TANGO . . . . .	180



# Liste des abréviations

<b>Sigle</b>	<b>Signification</b>
2D	Bidimensionnel
3D	Tridimensionnel
3C	Capture de Conformation Chromosomique
5mC	5-méthyl-cytosine
ADN	Acide désoxyribonucléique
ADNr	gènes d'ARN ribosomaux
ATP	Adénosine Triphosphate
ARN	Acide ribonucléique
ARNr	ARN ribosomaux
BAC	Bacterial Artificial Chromosome (chromosome artificiel bactérien)
ChIP	Chromatin Immunoprecipitation (immunoprécipitation de chromatine)
CTCF	CCCTC-binding factor (facteur se liant au motif CCCTC)
DAPI	4',6-diamino-2-phenylindole
DNMT	DNA methyl-transferase (méthyl-transférase d'ADN)
EN	Enveloppe Nucléaire
ESC	Embryonic Stem Cell (cellule souche embryonnaires)
EVF	Eroded Volume Fraction (fraction de volume érodée)
FISH	Fluorescence <i>In Situ</i> Hybridization (hybridation de fluorescence <i>in situ</i> )
HDAC	Histone Deacetylase (déacétylase d'histone)
HGPS	Hutchinson-Gilford Progeria Syndrom (syndrome de progeria d'Hutchinson-Gilford)
HiC	High-throughput Chromosome conformation capture (capture de conformation chromosomique sur l'ensemble du génome)
hMES	Human Mesenchymal Stem Cells (cellules souches mésenchymateuses humaines)
HOR	Higher-Order Repeat (répétitions d'ordre supérieur)
HP1	Heterochromatin Protein 1
K-S <sub>1</sub>	test de Kolmogorov-Smirnov à un échantillon
K-S <sub>2</sub>	test de Kolmogorov-Smirnov à deux échantillons
Kb	Kilobase (10 <sup>3</sup> pb)
LAD	Lamina Associated Domain (domaine associé à la lamina)
LBR	lamin B receptor
LINE	Long Interspersed Element (élément intersticiel long)
LNA	Locked Nucleid Acid (acide nucléique bloqué)

Continué sur la page suivante

---

<b>Sigle</b>	<b>Signification</b>
Mb	Mégabase = ( $10^6$ pb)
MNE	Membrane Nucléaire Externe
MNI	Membrane Nucléaire Interne
pb	Paire de Base
PcG	Polycomb Group (groupe Polycomb)
PHC	Pericentromeric Heterochromatin Compartment (compartiment d'hétérochromatine péricentromérique)
PML	Promyelocytic leukaemia
RC	Région Centromérique
RIDGEs	Region of Increased Gene Expression
SINE	Short Interspersed Element (élément intersticiel court)
TAD	Topologically Associated Domain (domaine topologiquement associé)
TC	Territoire Chromosomique
TF	Transcription Factory (usine à transcription)
$W_2$	test de Welch à deux échantillons

---

# Introduction

## Sommaire

<b>A.1</b>	<b>L'architecture tridimensionnelle du noyau</b>	<b>18</b>
A.1.1	Différents niveaux hiérarchiques d'organisation de la chromatine	19
A.1.2	La compartimentation nucléaire	31
A.1.3	Positionnement nucléaire du génome	37
A.1.4	Dynamique de l'architecture nucléaire	45
A.1.5	Modélisation de l'architecture nucléaire	50
<b>A.2</b>	<b>Structure et fonction des régions centromériques</b>	<b>54</b>
A.2.1	Généralités sur les régions centromériques	54
A.2.2	Structure chromatinienne des régions centromériques	56
A.2.3	Séquences d'ADN centromérique humaines	56
A.2.4	Lien entre structure et fonction	61
<b>A.3</b>	<b>Organisation nucléaire des régions centromériques</b>	<b>65</b>
A.3.1	Positionnement nucléaire des régions centromériques	65
A.3.2	Associations de régions centromériques	68
A.3.3	Mécanismes moléculaires impliqués dans l'organisation nucléaire des régions centromériques	75
<b>A.4</b>	<b>Méthodes d'études de l'architecture nucléaire</b>	<b>81</b>
A.4.1	Approches expérimentales	81
A.4.2	Déterminisme dans l'organisation spatiale	85
A.4.3	Analyse spatiale des signaux de microscopie	87
<b>A.5</b>	<b>Projet de thèse</b>	<b>98</b>

L'acide désoxyribonucléique (ADN), présent dans chaque cellule, constitue le support de l'information génétique. L'expression de cette information est souvent modulée lorsque la cellule se spécialise, et afin d'adapter sa réponse à l'environnement. Le noyau constitue le contexte dans lequel s'expriment les gènes, et de plus en plus de résultats montrent qu'il intervient dans cette modulation. Cet organe est très organisé : l'ADN y est compacté à un très haut degré, de manière très ordonnée, et il contient de nombreux sous-compartiments qui sont le siège de fonctions spécifiques. Les mécanismes qui sous-tendent cette organisation sont encore peu connus et font intervenir de nombreux acteurs.

Les régions centromériques des chromosomes (RCs<sup>1</sup>), jouant un rôle dans la division cellulaire, adoptent une organisation spatiale particulière dans le noyau en interphase ; cela suggère qu’elles pourraient aussi avoir un rôle lié à l’organisation spatiale du génome dans le noyau. On observe notamment qu’elles ne se localisent pas aléatoirement dans le noyau, et qu’elles ont tendance à s’agréger dans de nombreux organismes tels que la souris. Elles forment ainsi un compartiment qui joue un rôle dans la régulation de l’expression des gènes. L’organisation des RCs a été peu étudiée chez l’homme, principalement à cause des difficultés de détection dues à la complexité de leur séquence, mais aussi à cause de difficultés liées à l’étude de l’organisation spatiale de manière plus générale. Par exemple, l’absence de points de repère et le caractère probabiliste de l’organisation nucléaire sont des difficultés majeures lorsque l’on adopte une approche faisant intervenir l’imagerie.

Dans la partie résultats, différents aspects de l’organisation spatiale des RCs seront décrits par une approche d’imagerie de fluorescence, ainsi que la méthodologie développée. Néanmoins, l’introduction sera plus générale et définira le cadre de cette étude. Dans un premier temps, un tableau général de l’organisation nucléaire sera dressé ; principalement chez l’homme, mais en s’appuyant aussi sur d’autres organismes modèles tels que la souris (§A.1). Les RCs seront volontairement laissées de côté dans cette partie. La structure génomique et chromatinienne des RCs humaines sera ensuite décrite, ainsi que le lien entre leur structure et leur fonction dans la division cellulaire (§A.2) ; ce avant d’aborder leur organisation spatiale chez l’homme et la souris (§A.3). Enfin, les méthodes d’analyse de l’organisation nucléaire seront abordées, en insistant particulièrement sur les méthodes faisant intervenir l’imagerie (§A.4).

## A.1 L’architecture tridimensionnelle du noyau

Les chromosomes ont été découverts il y a plus d’un siècle lorsque Flemming observa, juste avant la division cellulaire, la formation de corps marqués par un colorant de l’ADN (Flemming (1965)). Au début du 20<sup>ème</sup> siècle, l’observation de chromosomes mitotiques de plantes a conduit Heitz à définir deux types de chromatine (Heitz (1928)) : l’*hétérochromatine* et l’*euchromatine*. Le premier se distingue du second par une coloration au carmin acétique plus forte, et par le fait qu’il reste condensé après la mitose. Dans les décennies suivantes, le principe dualiste d’euchromatine et d’hétérochromatine n’a plus été suffisant pour rendre compte des propriétés dynamiques de la chromatine. Une définition structurale est donc proposée, définissant la chromatine comme “un état plutôt qu’une substance”<sup>2</sup>. Le terme *hétérochromatine facultative* est introduit pour désigner les régions d’hétérochromatine potentiellement capables de devenir des régions d’euchromatine (Brown (1966)), tandis que le terme *hétérochromatine constitutive* désigne l’hétérochromatine qui ne change pas d’état. Les termes *euchromatine*, *hétérochromatine facul-*

1. Cette abréviation que nous choisissons n’est pas standard dans la littérature. Voir §A.2 pour une définition plus précise du terme.

2. Traduit de l’anglais : “a state rather than a substance” (Brown (1966))

*tative* et *hétérochromatine constitutive* sont encore utilisés aujourd'hui, mais leur caractérisation a été précisée par la découverte de la structure moléculaire de la chromatine.

Au cours du 20<sup>ème</sup> siècle, l'attention s'est tournée vers l'ADN qui fut identifié comme support de l'information génétique. Cette découverte a conduit à l'idée que les organismes sont déterminés par leur génome, et donc qu'un décodage de l'information génétique permettrait à lui seul de comprendre le fonctionnement d'un organisme. Cependant, l'étude de l'expression de l'ADN dans les cellules et organismes a montré que l'expression n'est pas uniquement déterminée par le génome : l'information contenue dans la séquence primaire est souvent modulée. L'identification de mécanismes dit *épigénétiques* a permis de comprendre comment des modifications chimiques réversibles de la chromatine pouvaient moduler l'expression des gènes, sans modification de la séquence d'ADN. Depuis la fin du 20<sup>ème</sup> siècle, le développement des techniques de microscopie et de techniques moléculaires pour l'étude de l'organisation spatiale du génome dans le noyau, a permis de mettre en évidence l'implication de la topologie du génome dans la modulation de l'expression des gènes. La compréhension du fonctionnement des génomes nécessite donc de connaître leur séquence, mais également leur organisation structurelle dans le noyau des cellules.

Dans cette partie, une description générale de l'organisation nucléaire du génome sera donnée, en décrivant tout d'abord la structure de la chromatine qui présente plusieurs niveaux de compaction (§A.1.1). Les différents compartiments nucléaires seront ensuite présentés (§A.1.2), afin de pouvoir décrire le positionnement du génome au sein de l'espace du noyau et par rapport à ces compartiments (§A.1.3). Ce tableau statique sera complété par une description des aspects dynamiques de l'architecture nucléaire (§A.1.4), puis une vue d'ensemble sera dressée, abordant la question de la relation entre structure et fonction en s'appuyant sur la modélisation (§A.1.5).

### A.1.1 Différents niveaux hiérarchiques d'organisation de la chromatine

La chromatine correspond à l'association de l'ADN, de protéines et de molécules d'ARN. Cette organisation permet la compaction d'un polymère d'ADN d'environ deux mètres, dans un espace nucléaire dont les dimensions ne dépassent pas quelques micromètres dans le cas de l'homme. La chromatine est très organisée dans le noyau, et cette organisation résulte de différents niveaux hiérarchiques de compaction qui sont présentés dans la figure A.1. Le premier niveau correspond à l'enroulement de l'ADN autour des histones, formant le nucléofilament (§A.1.1.1) constitué de nucléosomes. Les nucléosomes interagissent entre eux pour former une fibre de 30 nm (§A.1.1.2), qui s'organise en domaines de l'ordre de la Mb (§A.1.1.3). Enfin, ces domaines interagissent entre eux et forment les territoires chromosomiques (§A.1.1.4).

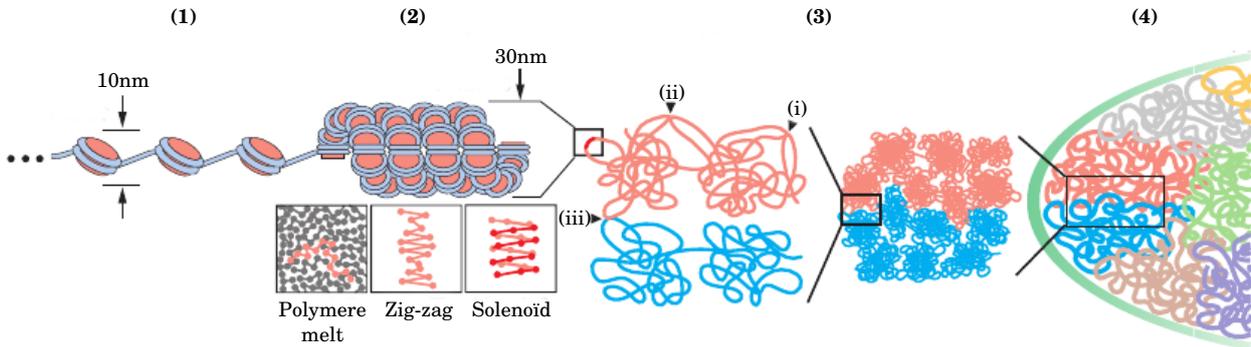


FIGURE A.1: Les différents niveaux hiérarchiques d'organisation de la chromatine dans un noyau de mammifère (d'après Hübner *et al.* (2012)). (1) Le nucléofilament constitué de nucléosomes. (2) Organisation d'ordre supérieure des nucléosomes. La chromatine pourrait former une fibre de 30 nm avec une organisation en solénoïde, en zig-zag ou en *polymer melt*. (3) La chromatine est organisée en domaines de l'ordre de la Mb. La fibre de chromatine interagit (i) fréquemment au sein du même domaine, (ii) plus rarement entre domaines du même chromosome, et (iii) très rarement avec des domaines adjacents portés par d'autres chromosomes. (4) Les chromosomes sont organisés en territoires chromosomiques.

### A.1.1.1 Le nucléofilament

A l'échelle de la dizaine de paires de bases (pb), la chromatine peut être décrite comme un filament d'ADN enroulé autour de protéines, les histones, structure parfois appelée *collier de perles*, fibre de 10 nm ou nucléofilament. A cette échelle sont encodées des informations qui vont influencer entre-autres l'état de compaction de la chromatine aux échelles supérieures. Le support moléculaire de cette information est multiple; il se trouve à la fois dans la composition des nucléosomes, les histones canoniques pouvant être substituées par des variants, dans les modifications post-traductionnelles des histones mais aussi dans la méthylation de l'ADN. Ces marques sont souvent qualifiées d'*épigénétiques*<sup>1</sup>. Les différentes marques sont associées à des voies de régulation de l'expression des gènes qui coopèrent entre elles.

**Le Nucléosome.** La molécule d'ADN s'enroule autour d'un octamère d'histones comprenant deux dimères d'histones H<sub>3</sub>-H<sub>4</sub> et H<sub>2</sub>A-H<sub>2</sub>B. Cette structure d'un diamètre de 10 nm et d'une hauteur de 5.5 nm est appelée nucléosome (Olins et Olins (1974)) et constitue l'unité fondamentale de la chromatine. Le fragment d'ADN contenu dans un nucléosome a une longueur d'environ 146 pb et s'enroule en faisant 1,65 tours orientés à gauche (pour revue : Olins et Olins (2003)). La structure du nucléosome a pu être caractérisée de manière très précise par cristallographie aux rayons X (Luger *et al.* (1997)) (voir Fig. A.2). Le nucléosome constitue le premier niveau de compaction de l'ADN : les 146 pb enroulées autour de l'octamère d'histones sont six fois plus compactes qu'un fragment d'ADN nu de même longueur (Olins et Olins (2003)). Les nucléosomes sont séparés les uns des autres par un seg-

1. Nous utiliserons ce terme pour les désigner car c'est le terme usuel consacré, mais sans nécessairement se référer à la définition formelle du terme qui implique en principe la notion de transmission de l'information.

ment d'ADN de liaison de longueur variant entre 10 et 80 pb. L'histone de liaison  $H_1$  s'associe à un fragment de l'ADN de liaison de taille variable, en général de 20 pb, fixant ainsi deux tours d'ADN en interagissant avec la surface de l'octamère d'histones, et formant une particule nommée chromatosome. Contrairement à ce qui est observé dans le cas des histones  $H_2$ ,  $H_3$ , et  $H_4$ , l'inactivation de l'histone  $H_1$  n'est pas létale, mais induit une altération globale de la structure de la chromatine, suggérant un important rôle de  $H_1$  en tant que composant structural (voir pour revue : [Khochbin \(2001\)](#)).



FIGURE A.2: Structure cristallographique du nucléosome ([Luger \*et al.\* \(1997\)](#)). Les histones sont représentées en orange ( $H_2A$ ), rose ( $H_2B$ ), bleu ( $H_3$ ) et vert ( $H_4$ ). L'ADN est représenté en marron et vert.

**Les variants d'histones.** Les classes d'histones  $H_2A$ ,  $H_2B$  et  $H_3$  comprennent plusieurs sous-types appelés variants d'histones, et retrouvés dans des contextes chromatiniens particuliers. A titre d'exemple, le variant d'histone CenpA remplace l'histone  $H_3$  au niveau des centromères des chromosomes et joue un rôle essentiel dans l'assemblage du kinétochore au cours de la mitose ; ce variant sera décrit plus en détails au §A.2.2. L'un des variants d'histones les plus étudiés est sans doute  $H_2A.X$  dont la forme phosphorylée sur la sérine 139 ( $\gamma$ - $H_2A.X$ ) est trouvée spécifiquement aux sites de cassures double-brin de l'ADN et permet la reconnaissance et la réparation de ces dommages (voir pour revue : [Sarma et Reinberg \(2005\)](#)).

**Les modifications post-traductionnelles d'histones.** Les extrémités (ou queues) N-terminales des histones sortent nettement de la structure du nucléosome et sont sujettes à de nombreuses modifications post-traductionnelles (Fig. A.3). Les conséquences fonctionnelles de ces modifications dépendent du type de groupement chimique apposé (méthyle, acétyle, phosphate, ubiquitine, glycosyle, sumo, etc.) ainsi que du type de résidu ciblé (lysine K, arginine R, sérine S ...) et de sa position le long de la queue N-terminale. La diversité des groupements et des résidus qui peuvent être ciblés donne lieu à une combinatoire potentiellement très riche. Ces modifications influent, d'une part, sur l'interaction entre l'ADN et les histones et par conséquent sur la compaction de la fibre de chromatine ; et d'autre part, sur l'interaction de l'ADN avec les autres protéines associées à la chromatine. Ainsi, elles agissent sur les fonctions liées au génome, telles que la transcription ou la ré-

paration. Les deux modifications les plus étudiées, l'acétylation et la méthylation sont détaillées ici.

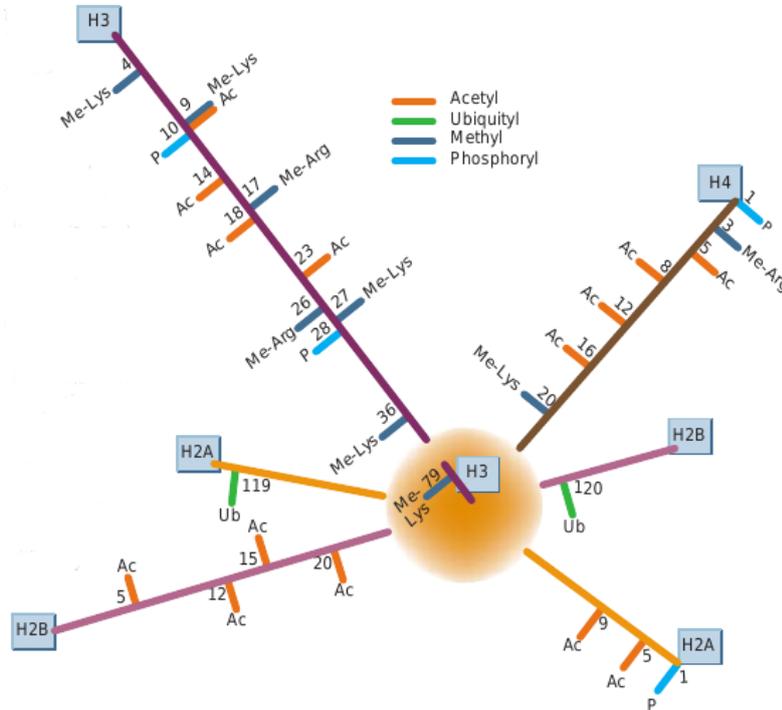


FIGURE A.3: Principales modifications post-traductionnelles de la queue des histones (Felsenfeld et Groudine (2003)).

**L'acétylation.** L'acétylation des lysines est généralement associé à un état chromatinien décompacté et permissif pour la transcription (Hebbes *et al.* (1988)). La neutralisation de la charge positive du groupement  $\text{NH}_3^+$  permet de relâcher l'interaction ADN-histone et induit la décompaction de la chromatine. L'acétylation des lysine résulte d'un équilibre entre deux activités antagonistes : l'activité histone-acétyltransférase (HAT) et l'activité histone-désacétylase (HDAC). L'acétylation est reconnue par des protéines à bromo-domaine.

**La méthylation.** La méthylation des histones est plus complexe ; elle dépend du type de résidu (K ou R), de la position de celui-ci ( $\text{K}_4$ ,  $\text{K}_9$ ,  $\text{K}_{20}$ ...) et du nombre de méthyles (de 1 à 3) apposés par les HMTs (Histone Méthyl-transférases) (pour revue : Campos et Reinberg (2009)). Les marques  $\text{H}_3\text{K}_{27}\text{me}$ ,  $\text{H}_3\text{K}_9\text{me}$ ,  $\text{H}_3\text{K}_4\text{me}_{1/2/3}$ ,  $\text{H}_3\text{K}_{36}\text{me}_3$  et  $\text{H}_4\text{K}_{20}\text{me}$  sont associées à la transcription (Barski *et al.* (2007)).

Par ailleurs, il existe deux marques principales associées à des états chromatiniens compacts et transcriptionnellement silencieux. Elles sont régulées par deux voies de répression (pour revue : Beisel et Paro (2011)).

- La di/tri-méthylation de la lysine 9 de l'histone H<sub>3</sub> ( $\text{H}_3\text{K}_9\text{me}_{2/3}$ ) est reconnue par la protéine à chromo-domaine HP1 (Heterochromatin Protein 1)

qui recrute les enzymes apposant cette modification, incluant par exemple G9a/GLP, SETDB1 et Suv39h1, et permet la propagation de la structure.

- La marque  $H_3K_{27}me_3$  est apposée par les protéines du groupe Polycomb (PcG) qui forment deux complexes principaux, les complexes répressifs polycomb 1 (PRC1) et 2 (PRC2). Cette marque est associée à la mise en silence de nombreux gènes du développement et de la différenciation.

**Le code histone.** La caractérisation systématique des modifications d'histone a conduit à proposer l'existence d'un *code histone* qui associerait chaque modification à un état transcriptionnel donné (Jenuwein et Allis (2001)). Cependant, ce code reste très controversé; bien que certaines modifications soient souvent associées à un état transcriptionnel précis, il existe de nombreuses exceptions et beaucoup ont un rôle non élucidé. Il semble par ailleurs que l'effet dû à la combinaison de plusieurs marques sur des domaines chromatiniens soit plus important que celui de marques isolées, ce qui complexifie l'étude des conséquences de ces modifications individuelles.

**La méthylation de l'ADN.** Chez les vertébrés, la méthylation de l'ADN a lieu quasi exclusivement sur le carbone 5 des cytosines (C) qui précède une guanine (G), on parle alors de dinucléotides CpG, et la cytosine méthylée est désignée par 5-méthyl-cytosine (5mC). Durant la réplication, la méthyl-transférase d'ADN 1 (DNMT1) propage le profil de méthylation sur le brin néo-synthétisé.

La méthylation de l'ADN est généralement associée à la répression transcriptionnelle, agissant par encombrement stérique du grand sillon de l'hélice d'ADN et inhibant le recrutement des facteurs de transcription (Watt et Molloy (1988)). Il a en effet été montré que certains facteurs de transcription se lient sélectivement aux CpG non méthylés (Voo *et al.* (2000)). L'activité répressive de la méthylation peut aussi être indirecte, en coopération avec d'autres voies de répression transcriptionnelle (voir plus bas).

Les sites CpG ne sont pas distribués aléatoirement le long du génome, mais sont concentrés au niveau de régions génomiques de plus de 500 pb que l'on nomme îlots CpG et qui se localisent notamment au niveau de la région 5' de 60% des gènes (Antequera et Bird (1993)). Au sein des îlots CpG, les dinucléotides CpG sont généralement déméthylés, et l'état de méthylation corrèle avec l'activité du gène en aval<sup>1</sup>. La méthylation de l'ADN est notamment impliquée dans la régulation de gènes associés à la pluripotence, la différenciation, de gènes soumis à l'empreinte parentale ainsi que dans la mise en silence des gènes sur le chromosome X inactif (Mohn *et al.* (2008)).

**Coopération entre les voies de répression.** La méthylation de l'ADN semble agir en synergie avec d'autres voies de répression transcriptionnelle, via le recru-

1. Lorsque la région 5' est méthylée, le gène en aval a tendance à être réprimé, inversement lorsqu'elle est déméthylée il a tendance à être actif.

tement de protéines qui se lient à la 5mC (MeCP2, MBD1-4) et qui interagissent avec des HDACs et des HMTs (Sasai et Defossez (2009)). Par exemple, la protéine MeCP2 reconnaît la 5mC et interagit avec les modificateurs chromatiniens HDAC1, HDAC2 et Suv39h1 ainsi que la protéine HP1 (Agarwal *et al.* (2007), Sasai et Defossez (2009)). Ceci suggère que l'information codée par la méthylation de l'ADN est lue par des protéines se liant à la 5mC, et traduite par des modifications chromatiniennes qui permettent de consolider l'état chromatinien répressif.

Par ailleurs, des états chromatiniens répressifs peuvent être stabilisés par la méthylation pour établir une répression transcriptionnelle plus stable de gènes spécifiques. Lorsque les cellules s'engagent dans une voie de différenciation, le système PcG semble coopérer avec la méthylation de l'ADN : en effet, il a été montré que les promoteurs qui présentent la marque  $H_3K_{27}me_3$  (associée aux protéines CpG) sont plus fréquemment méthylés *de novo* que les autres promoteurs. De plus les DNMT interagissent avec PRC2 pour être recrutées au niveau des sites de répression polycomb (Viré *et al.* (2006)). Cela suggère que les gènes réprimés par le système polycomb au cours du développement sont prédisposés à être méthylés *de novo*.

La méthylation de l'ADN étant une marque épigénétique stable (bien que réversible), elle pourrait donc jouer un rôle de verrou transcriptionnel.

**Un nombre limité d'états chromatiniens.** L'analyse de la co-occurrence de différentes protéines de la chromatine, par des approches intégratives<sup>1</sup> sur des données provenant d'immuno-précipitation de chromatine (ChIP, voir §A.4.1.1), a mené plusieurs équipes à classer la chromatine en états caractérisés par la présence ou l'absence de protéines de la chromatine. Plusieurs études convergent sur l'idée qu'il n'existe que quelques états de chromatine répressive, alors que la très grande diversité des modifications post-traductionnelles d'histones pourrait mener à une combinatoire très riche. Ceci suggère que les systèmes associés à la régulation des marques coopèrent entre eux.

Les états répressifs identifiés dans plusieurs modèles cellulaires<sup>2</sup> correspondent à la chromatine caractérisée par la marque épigénétique  $H_3K_9me_3$ , la chromatine polycomb, ainsi qu'à un troisième type de chromatine qui n'a pas d'enrichissement particulier pour les marques utilisées dans les analyses (Filion *et al.* (2010), Ernst *et al.* (2011)). En ce qui concerne la chromatine active (ouverte), plusieurs états ont été décrits, mais il n'existe pas de consensus pour le moment. Il est possible de distinguer au moins quatre états qui sont associés aux régions *activatrices* (*enhancers* en anglais), *promotrices*, *transcrites* et aux *insulateurs* (Dunham *et al.* (2012)). Ces résultats permettent d'affiner la caractérisation historique de la chromatine en *euchromatine*, *hétérochromatine facultative* et *hétérochromatine constitutive*.

1. Ces études utilisent des méthodes telles que l'analyse en composante principale.

2. Notamment chez la drosophile, la souris et l'homme.

### A.1.1.2 La fibre de chromatine

*In vitro*, la fibre de 10 nm forme une structure secondaire via les interactions entre histones ; elle se condense en une fibre de 30 nm qui se forme grâce aux interactions entre les nucléosomes via l'histone H<sub>1</sub>. Cette structure a aussi été observée dans les cellules de systèmes non-mammifères très spécialisés (Woodcock (1994)), qui ont tous la caractéristique d'avoir une histone H<sub>1</sub> plus chargée, peu de protéines décorant la chromatine qui ne soient pas des histones, et en moyenne une plus grande longueur d'ADN entre deux nucléosomes. La réorganisation structurelle de la fibre de 10 nm en fibre de 30 nm correspond à une compaction de la chromatine, elle est donc supposée correspondre à la transition entre chromatine active et chromatine silencieuse. Plusieurs modèles de structure ont été proposés pour la fibre de 30 nm. Dans un des modèles, la fibre de 10 nm est enroulé en solénoïde où deux nucléosomes consécutifs se retrouvent côte à côte dans une hélice simple (Kruithof *et al.* (2009)). L'autre modèle principal propose une structure d'hélice en zig-zag à deux départs, dans laquelle un nucléosome est en contact avec ses seconds voisins et non avec les nucléosomes consécutifs (Horowitz *et al.* (1994)).

Les travaux qui mettent en évidence l'existence d'une fibre de 30 nm reposent sur des expériences conduites *in vitro*, mais cette fibre n'a jamais été observée dans les cellules. Des travaux récents faisant appel à la combinaison de spectroscopie électronique et de tomographie remettent en cause l'existence de la fibre de 30 nm dans le noyau en interphase de cellules somatiques de souris (Fussner *et al.* (2012)).

### A.1.1.3 Organisation d'ordre supérieure de la fibre de chromatine

**Domaines chromatiniens.** Différentes observations ont permis de mettre en évidence que les chromosomes pouvaient être divisés en domaines de tailles variables présentant des caractéristiques homogènes.

D'une part, l'analyse des profils d'expression des gènes a montré que les régions très exprimées forment des domaines dénommés RIDGEs (Region of Increased Gene Expression) ; inversement les régions très peu exprimées forment des domaines nommés anti-RIDGEs (Fig. A.4A). Les RIDGEs sont en général riches en gènes et les anti-RIDGEs pauvres en gènes (Caron *et al.* (2001)).

D'autre part, les marques épigénétiques précédemment décrites ne sont pas réparties uniformément sur le génome. Le développement de techniques telles que la ChIP a permis de dresser des cartes de répartition de protéines sur l'ensemble du génome avec une haute résolution. L'analyse de ces cartes a montré que les chromosomes linéaires sont segmentés en des centaines de domaines (Fig. A.4B) de composition protéique homogène, dont certains peuvent être mis en relation avec les états chromatiniens précédemment décrits (pour revue : Bickmore *et van Steensel* (2013)).

**Conformation spatiale des domaines chromatiniens.** Des analyses récentes de la conformation 3D du génome conduites chez la mouche, la souris et l'homme

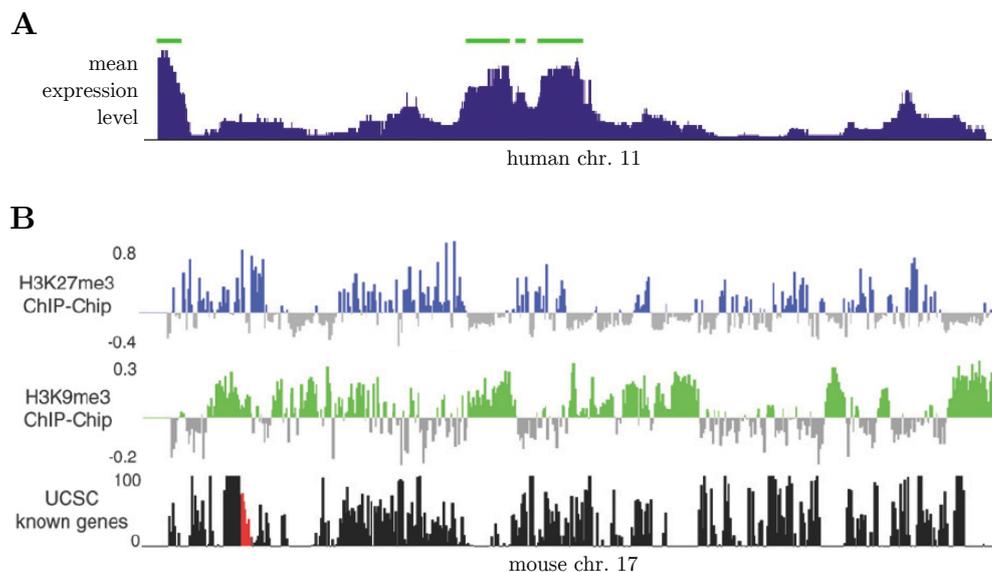


FIGURE A.4: Segmentation linéaire des chromosomes en domaines. A : Niveau d'expression moyen dans plusieurs tissus (d'après [Caron et al. \(2001\)](#)). Le profil montré correspond au chromosome 11 humain. Les régions riches en gènes exprimés (RIDGES) forment des domaines indiqués par des traits verts. B : Deux types de domaines identifiés par les cartes de répartition de modifications d'histones sur le chromosome 17 murin (d'après [Pauler et al. \(2009\)](#)). Les profils d'enrichissement sont obtenus par CHIP sur puce à ADN, et correspondent au logarithme du signal normalisé. De haut en bas : profil d'enrichissement de la marque H<sub>3</sub>K<sub>27</sub>me<sub>3</sub> (bleu), de la marque H<sub>3</sub>K<sub>9</sub>me<sub>3</sub> (vert), densité en gènes. Chaque profil d'enrichissement forme des domaines; il y a une anti-corrélation entre les domaines verts et bleus, et une corrélation entre les domaines bleus et la densité en gènes.

ont montré que les domaines précédemment décrits (définis par l'analyse de l'épigénome) correspondent à des domaines physiques qui occupent un espace restreint dans le noyau. Ces domaines d'une taille de l'ordre de la mégabase (Mb, 10<sup>6</sup> pb), dénommés *domaines topologiquement associés* (TAD : Topologically Associated Domains), ont été identifiés à partir de données de HiC (capture de conformation chromosomique (3C) sur l'ensemble du génome, voir §A.4.1.1) ([Dixon et al. \(2012\)](#), [Sexton et al. \(2012\)](#)). Ils sont caractérisés par une fréquence d'interaction très élevée au sein du domaine, et plus faible avec le reste du chromosome (Fig. A.5). Les frontières de ces domaines semblent être définies génétiquement : la délétion d'une frontière au sein du centre d'inactivation du chromosome X a pour conséquence une fusion des deux TADs adjacents ([Nora et al. \(2012\)](#)). Ce qui caractérise ces frontières n'est pas clair ; elles contiennent notamment des sites de démarrage de la transcription et des sites de liaison pour la protéine insulatrice CTCF, elles sont enrichies en cohésines et en d'autres protéines liant les insulateurs ([Dixon et al. \(2012\)](#), [Gibcus et Dekker \(2013\)](#)). Ces frontières correspondent aux régions qui contraignent l'expansion de l'hétérochromatine. Ces domaines sont conservés entre espèces pour un même type cellulaire. La comparaison de différents types cellulaires a montré que le type de chromatine de certains domaines bascule au cours de la différenciation, suggérant que ces domaines peuvent constituer des unités fonctionnelles qui sont contrôlées par un même mécanisme de régulation.

Avant la découverte des TADs, des expériences de microscopie avaient montré

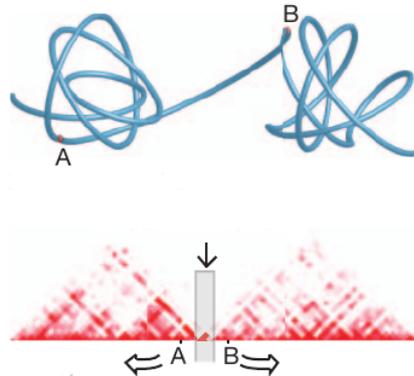


FIGURE A.5: Domaines topologiquement associés (Dixon *et al.* (2012)). En haut : représentation schématique de la conformation spatiale de la fibre de chromatine qui forme deux domaines spatialement distincts. En bas : matrice de HiC montrant que les interactions au sein du domaine sont plus fréquentes que les interactions entre les deux domaines, y compris pour des loci proches sur le génome linéaire tels que A et B. Pour ces points, la proximité spatiale ne reflète pas la proximité sur le génome. La flèche indique l'emplacement supposé de la frontière entre les deux domaines.

que la chromatine était segmentée en domaines chromosomiques structurés dans le noyau, d'une taille variant de 100 Kb jusqu'à quelques Mb (Cremer et Cremer (2010)). Ces domaines peuvent être observés par un marquage de l'ADN répliqué grâce au BrdU : il forment des foyers discrets dans le noyau correspondant aux sites de réplication (Ma *et al.* (1998)). Goetze *et al.* (2007) ont observé des domaines en utilisant des sondes FISH dirigées contre des RIDGEs et des anti-RIDGEs. Ils observent très peu de recouvrement entre les différents domaines marqués, ce qui suggère que la fibre de chromatine est organisée en domaines distincts dans le noyau. Les auteurs de ce travail ont aussi montré que les RIDGEs sont en général moins condensées dans le noyau et ont des formes plus irrégulières que les anti-RIDGEs. Ces résultats sont cohérents avec l'idée d'une fibre de chromatine structurée en domaines dans le noyau. La taille de ces domaines est proche de celle des TADs, ce qui suggère qu'ils correspondent aux mêmes structures (Gibcus et Dekker (2013)).

**Les boucles intra-chromosomiques.** A la même échelle spatiale que les domaines, la fibre de chromatine forme différents types de boucles qui ont chacune un rôle distinct dans la régulation de l'expression des gènes (pour revue : Hou et Corces (2012)). A ce jour, quatre types de boucles ont été identifiés :

- Les boucles intra-géniques qui joignent les extrémités 5' et 3' des gènes, et qui pourraient permettre le recyclage de l'ARN Pol. II et faciliter la maintenance de la directionnalité de la transcription.
- Les boucles entre activateur et promoteur, qui sont médiées par les facteurs de transcription spécifiques de séquence, et probablement par des ARNnc ou par des protéines liant l'ADN telles que CTCF ou des cohésines. Ces boucles induisent l'activation de la transcription.
- Les boucles entre les régions cibles des protéines du PcG (PRE) et les promoteurs des gènes, qui empêchent le recrutement de l'ARN Pol. II et qui pourraient gêner l'élongation de la transcription par les ARN polymérases

liées aux promoteurs.

- Les boucles médiées par des insulateurs, qui pourraient isoler des loci individuels, contenant la partie codante d'un gène, et ses régions régulatrices de leur environnement génomique direct.

#### A.1.1.4 Associations de domaines et territoires chromosomiques

Au niveau d'organisation suivant, des domaines portés par plusieurs chromosomes s'assemblent dans le noyau, influençant la conformation spatiale du génome. Les domaines d'un chromosome interagissent principalement avec d'autres domaines du même chromosome, ce qui entraîne la formation du territoire chromosomique (TC). Par ailleurs, il existe aussi des interactions entre domaines portés par des chromosomes différents, qui ont lieu principalement aux frontières des TCs.

#### Preuves expérimentales de l'existence des territoires chromosomiques.

Les chromosomes occupent un domaine restreint dans le noyau dénommé territoire chromosomique (TC). La démonstration de l'existence des TCs remonte aux années 80, et repose sur une expérience très ingénieuse, dans laquelle une lésion est infligée sur une zone restreinte d'un noyau en interphase par un microlaser. Les dommages à l'ADN sont ensuite analysés sur les chromosomes métaphasiques. Si les chromosomes étaient distribués aléatoirement dans le noyau, beaucoup de chromosomes auraient été endommagés. Au contraire, l'irradiation n'a causé de dommages qu'à un nombre restreint de chromosomes, montrant que chacun d'eux occupe un espace délimité dans le noyau (Zorn *et al.* (1979), Fig. A.6A-B). Les TCs ont par la suite été visualisés dans des expériences de FISH utilisant des peintures chromosomiques (Fig. A.6C). Les expériences de HiC permettent aussi de mettre en évidence les TCs : en effet, le trait dominant des expériences de HiC est que la grande majorité des interactions entre régions du génome a lieu en *cis* plutôt qu'en *trans* (Fig. A.6D, Lieberman-Aiden *et al.* (2009)).

**Associations de domaines chromatiniens.** Les analyses d'expériences de type 3C révèlent que différents TADs interagissent entre eux au sein d'un chromosome. D'une part, les TADs inactifs ont tendance à n'interagir qu'avec d'autres domaines inactifs portés par le même bras chromosomique, ce qui suggère que le centromère constitue en quelque sorte une barrière qui atténue les associations entre les séquences se trouvant sur les deux bras opposés du chromosome (Kalhor *et al.* (2012)). D'autre part, les TADs actifs interagissent préférentiellement avec d'autres TADs actifs, à la fois portés par le même bras, le bras opposé ou bien portés par d'autres chromosomes (Sexton *et al.* (2012), Kalhor *et al.* (2012)). Ces observations suggèrent que les domaines réprimés pourraient former le cœur du TC, alors que les domaines actifs pourraient s'étendre en dehors du TC et contribuer aux interactions inter-chromosomes. Cette idée est supportée par une analyse morphologique (Boyle *et al.* (2011)) qui utilise des sondes oligonucléotidiques de haute complexité qui ciblent l'ensemble des régions exoniques d'un chromosome, et qui permettent de

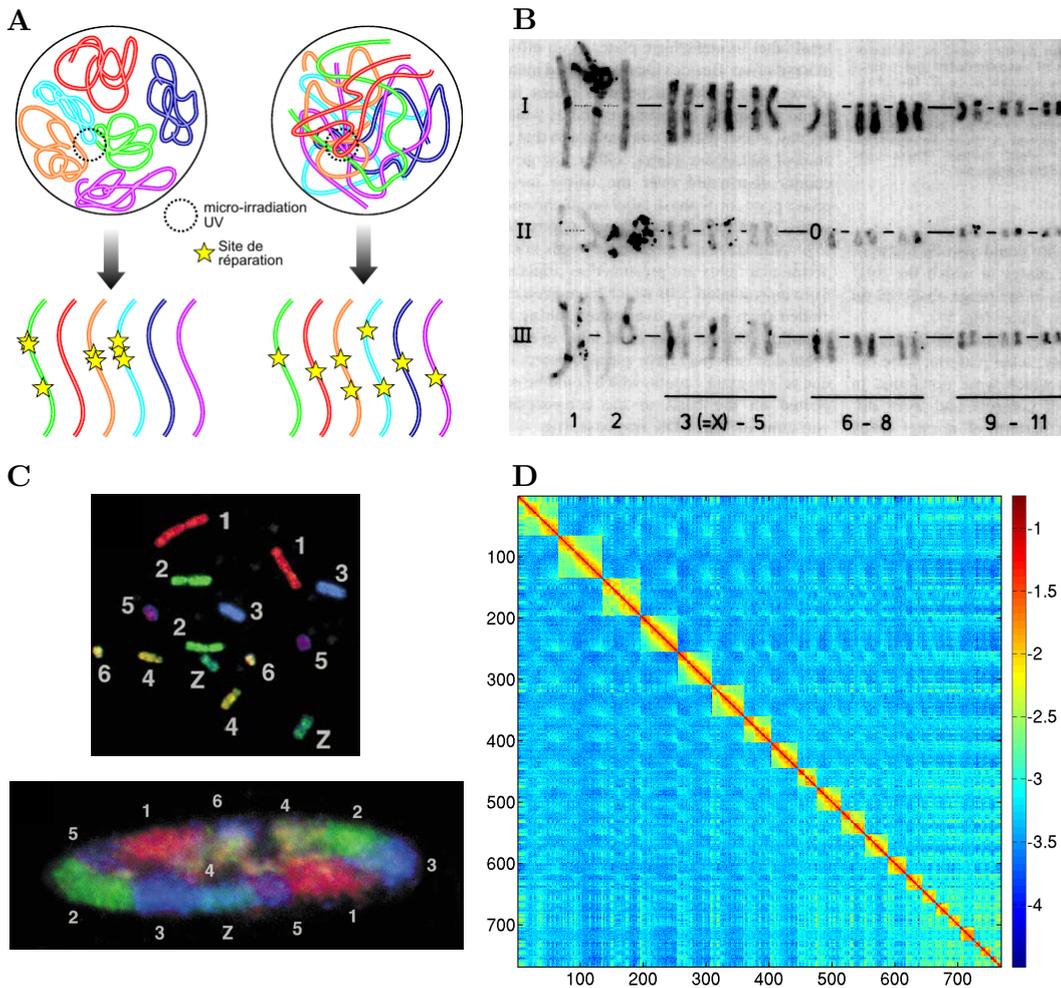


FIGURE A.6: Preuves expérimentales de l'existence des territoires chromosomiques. A : modèle prédictif des conséquences d'une micro-irradiation laser en fonction de l'organisation des chromosomes dans le noyau (d'après Meaburn et Misteli (2007), voir texte principal). B : Caryotype obtenu lorsque l'expérience (A) est effectuée sur des fibroblastes de hamster chinois (d'après Zorn *et al.* (1979)). C : Marquage fluorescent simultané des macrochromosomes de poulet par des peintures chromosomiques (d'après Habermann *et al.* (2001)). En haut, étalement de métaphase avec le marquage de 7 chromosomes ; en bas : noyau en interphase avec les 7 chromosomes marqués. D : Matrice de HiC tracée par l'équipe du Dr. Mozziconacci à partir de données expérimentales provenant de cellules lymphoblastoïdes (Kalhor *et al.* (2012)). Les couleurs sont associées aux fréquences d'interaction normalisées, via une échelle logarithmique. L'échelle des axes de la matrice correspond à 1000 fragments de restrictions d'environ 3Kb chacun. La caractéristique la plus évidente de cette matrice est la présence de carrés sur la diagonale, qui correspondent aux différents chromosomes, et montrent que chaque chromosome a tendance à interagir préférentiellement avec lui-même plutôt qu'avec le reste du génome ; cela reflète l'organisation en TCs (matrice reproduite avec l'aimable autorisation du Dr. Mozziconacci).

mettre en évidence un compartiment dénommé *corrona*. Ce compartiment se trouve en périphérie du cœur du TC, révélé par les peintures chromosomiques traditionnelles. Cela reflète certainement le fait que les domaines inactifs ont une moins grande liberté de mouvement que les domaines actifs, et sont donc contraints à rester dans le cœur du TC (Kalhor *et al.* (2012)). Cela a été directement observé pour le cluster de gènes HOX qui, lorsqu'il est inactif, se localise uniquement à

l'intérieur du territoire chromosomique; son activation lui confère une plus grande liberté de mouvement, il se localise alors à l'intérieur ou à l'extérieur du territoire chromosomique (Morey *et al.* (2009)).

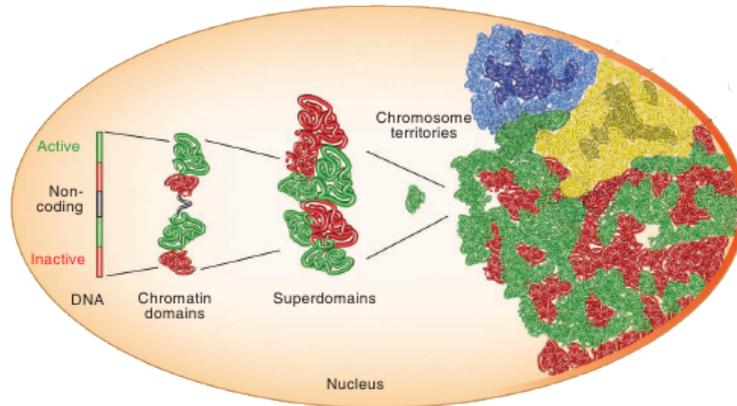


FIGURE A.7: Principes de la conformation spatiale du génome dans le noyau (d'après Cavalli et Misteli (2013)). La conformation spatiale des domaines chromosomiques est déterminée par l'activité transcriptionnelle (vert et rouge) de la région génomique. Les domaines de même activité interagissent entre eux et forment les territoires chromosomiques (jaune et bleu). Les domaines répressifs ont tendance à interagir avec les domaines répressifs du même bras chromosomique et se localiser à l'intérieur du TC (jaune sombre et bleu sombre), alors que les domaines actifs peuvent interagir avec des domaines actifs portés par le bras opposé, ainsi qu'avec des domaines actifs portés par d'autres chromosomes (jaune clair et bleu clair).

## Conclusion

Plusieurs concepts importants ressortent de cette description :

- La chromatine est repliée selon différents niveaux hiérarchiques dans le noyau, qui sont inter-dépendants.
- Les chromosomes sont organisés en domaines qui sont à la fois caractérisés par des combinaisons de marques épigénétiques, et par leur structuration 3D.
- Les domaines actifs s'associent principalement avec d'autres domaines actifs et les domaines répressifs avec d'autres domaines répressifs, formant deux compartiments distincts dans le noyau (Lieberman-Aiden *et al.* (2009), Gibcus et Dekker (2013), Fig. A.7).
- S'il existe des associations entre domaines portés par des chromosomes différents, la très grande majorité des associations a tout de même lieu entre domaines portés par le même chromosome, formant le TC. Cela montre que la liaison génomique est le facteur principal déterminant les associations dans l'espace du noyau.

A cette échelle spatiale, on observe aussi dans le noyau la formation de compartiments qui participent à l'organisation spatiale des domaines chromosomiques. Dans les parties suivantes, ces compartiments seront présentés, avant de décrire l'organisation du génome dans l'espace du noyau et notamment par rapport à ces compartiments.

### A.1.2 La compartimentation nucléaire

L'espace du noyau héberge de nombreux compartiments, qui constituent des environnements distincts de leurs alentours et sont définis cytologiquement et caractérisés par de fortes concentrations locales de protéines, d'ADN ou d'ARN. Après avoir discuté du phénomène de compartimentation en général, nous donnerons une description de la structure et des fonctions des principaux compartiments nucléaires, en distinguant les compartiments périphériques et les compartiments internes du noyau.

#### A.1.2.1 Généralités sur le phénomène de compartimentation

Certains compartiments nucléaires permettent de faciliter des réactions (et sont parfois nécessaires à ces réactions) ; c'est notamment le cas des nucléoles, des corps polycomb, des usines à transcription, des speckles nucléaires (Fig. A.8). Ce phénomène de compartimentation ressemble à celui observé dans le cytoplasme, où des organites sont le siège de processus métaboliques particuliers, à ceci près que les compartiments nucléaires ne sont pas séparés de leurs alentours par une membrane. Comme nous allons le voir au travers des différents exemples, les compartiments peuvent avoir d'autres fonctions que la facilitation de processus métaboliques : ils peuvent par exemple séquestrer des facteurs nucléaires pour réguler leur concentration nucléaire, ou encore servir de plate-forme permettant la co-régulation de plusieurs gènes (Fig. A.8). La plupart des compartiments sont des structures dynamiques, dont l'assemblage résulte de processus d'auto-organisation plutôt que d'un assemblage séquentiel ordonné (Mao *et al.* (2011)). Ces structures sont formées en sortie de mitose, et maintenues jusqu'au début de la prochaine mitose.

#### A.1.2.2 L'enveloppe nucléaire

**L'enveloppe nucléaire (EN).** L'EN est formée de deux membranes concentriques : les membranes *interne* (MNI) et *externe* (MNE), séparées par le lumen, et qui fusionnent pour former des orifices contenant les pores nucléaires qui permettent les mouvements passifs et actifs de molécules entre cytoplasme et nucléoplasme. La MNE est topologiquement liée au réticulum endoplasmique. Les protéines de la MNI à domaine SUN interagissent avec des protéines de la MNE à domaine KASH, formant des complexes LINK (Linkers of the Nucleoskeleton to the cytoskeleton). Les protéines à domaine KASH interagissent notamment avec des structures cytoplasmiques telles que des filaments d'actine, de tubuline, certains filaments intermédiaires et le centrosome. Les protéines à domaine SUN interagissent avec la lamina et avec des protéines de la MNI (Fig. A.9). Les complexes LINK permettent donc d'établir un couplage mécanique entre le nucléosquelette et le cytosquelette (pour revue : Méjat et Misteli (2010)). Cela confère un rôle à l'EN dans la motilité du noyau, la division cellulaire, et la transduction de signaux mécaniques (Broers *et al.* (2004)). L'EN possède donc paradoxalement à la fois le rôle de séparer le noyau du cytoplasme, et celui de les relier, via les pores nucléaires et le couplage du

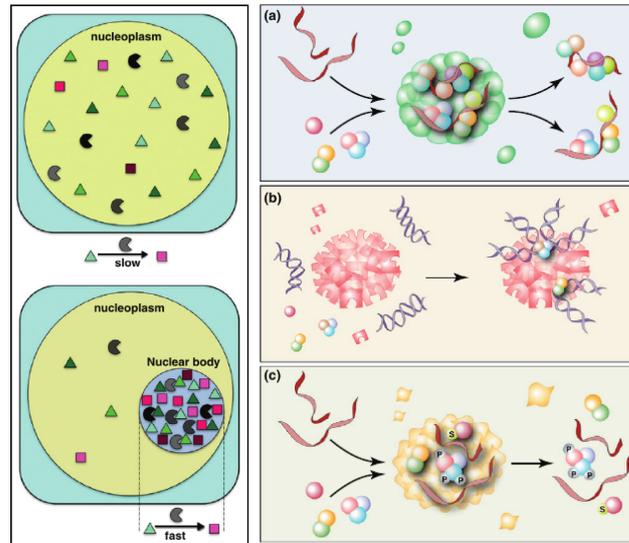


FIGURE A.8: Différents rôles identifiés pour les compartiments nucléaires. À gauche : La compartimentation facilite les réactions entre substrats et enzymes (d'après [Dundr \(2012\)](#)). Certains compartiments génèrent une forte concentration locale en enzymes ou molécules réactives (représentées par des *Pac-Mans* et en substrats (triangles) comparée à la concentration nucléoplasmique. Cela permet d'augmenter la vitesse des réactions, et éventuellement leur spécificité. À droite : Représentation schématique de différentes fonctions des compartiments nucléaires (d'après [Mao et al. \(2011\)](#)). Certains compartiments sont le siège de réactions (a), d'autres peuvent être vus comme des plate-formes de régulation de l'expression des gènes. Par exemple les corps polycomb recrutent de multiples loci pour stabiliser leurs interactions et ainsi réguler leur expression. D'autres encore sont considérés comme des usines à transcription spécialisées qui sont impliquées dans la co-régulation positive de gènes. Enfin, certains facteurs sont séquestrés dans des compartiments (c) ce qui permet la régulation de leur concentration nucléoplasmique; certains compartiments comme les speckles sont aussi le siège de modifications des molécules stockées.

nucléosquelette au cytosquelette. L'EN et ses composants sont impliqués dans de nombreux processus cellulaires, comme le témoignent les nombreuses pathologies associées à des mutations de ses composants; et notamment celles de gène codant pour la lamine A/C qui donne lieu à des maladies collectivement regroupées sous le terme de laminopathies (pour revue : [Méjat et Misteli \(2010\)](#)).

**La lamina.** La lamina nucléaire est une structure protéique associée au côté nucléoplasmique de la MNI. Elle se désassemble en début de mitose et se reforme en fin de mitose par un mécanisme impliquant des réactions de phosphorylation et déphosphorylation. Ses composants majeurs sont les lamines de type A et B, qui sont des filaments intermédiaires de type V et qui forment des maillages fibrillaires denses, de 10 à 20 nm d'épaisseur ([Dechat et al. \(2008\)](#)). La lamine C est une isoforme de la lamine A. Les lamines de type A/C et B forment deux réseaux distincts mais interagissant, et qui semblent avoir des fonctions différentes ([Shimi et al. \(2008\)](#)). Toutes les cellules des métazoaires expriment des lamines de type B, alors que les lamines de types A/C sont exprimées seulement dans les cellules différenciées des métazoaires et leur expression est régulée au cours du développement ([Dechat et al. \(2008\)](#)). Seule la lamine B est essentielle à la viabilité de la cellule. La lamina a un rôle structural, donnant sa forme au noyau et rigidifiant l'enveloppe nucléaire

(pour revue : [Dahl et al. \(2008\)](#)). Il a notamment été montré que les cellules qui n'expriment pas les lamines A/C sont moins rigides et plus déformables ([Broers et al. \(2004\)](#)). La lamina joue aussi un rôle crucial dans le positionnement correct de nombreuses protéines de la MNI. Réciproquement, des protéines de la MNI (possédant un domaine transmembranaire) telles que LBR (Lamin B Receptor) et des protéines à domaines LEM (l'emerine, LAP2 $\alpha/\beta$  LEM2 et MAN1), permettent de connecter la lamina à la MNI et stabiliser le réseau de lamines. De plus, les lamines de type B pourraient contribuer à l'assemblage de l'enveloppe nucléaire à la fin de la mitose ([Ho et Lammerding \(2012\)](#)). Par ailleurs, une fraction minoritaire des lamines se localise dans le nucléoplasme ([Shimi et al. \(2008\)](#)), et pourrait avoir un rôle dans la réparation, la réplication et la transcription du génome.

La lamina abrite de nombreuses protéines se liant aux lamines, qui possèdent souvent un domaine trans-membranaire, et qui sont les médiateurs des diverses fonctions de l'enveloppe nucléaire, comme l'illustre la figure [A.9](#) (pour revue : [Schirmer et Foisner \(2007\)](#), [Wilson et Foisner \(2010\)](#)). La lamina semble notamment avoir un rôle dans le positionnement du génome et dans la répression transcriptionnelle (voir §A.1.3.2).

### A.1.2.3 Les compartiments internes

**Le nucléole.** Le nucléole<sup>1</sup> est formé au niveau de plusieurs régions constituées de répétitions en tandem de gènes d'ARN ribosomiaux (ADNr) appelées régions organisatrices des nucléoles (NORs : Nucleolar Organizing Regions). Chez l'homme, des expériences d'hybridation *in situ* ont montré que les NORs sont localisées sur le bras court des cinq chromosomes acrocentriques, les chromosomes 13, 14, 15, 21, et 22 ([Henderson et al. \(1972\)](#)). La transcription des rDNA reprend dès la télophase, lorsque des corps pré-nucléolaires se forment au niveau des NORs, fusionnant en début de phase G<sub>1</sub> pour former un nombre réduit de nucléoles ([Hernandez-Verdun \(2006\)](#)). La manière dont les NORs actives sont organisées topologiquement au sein du nucléole n'est pas encore précisément connue ([Németh et Längst \(2011\)](#)). En plus de ces régions génomiques, le nucléole est constitué de nombreux ARN et protéines. Sa structure est complexe, composée du centre fibrillaire, lui-même entouré du composant fibrillaire dense, tous deux inclus dans le composant granulaire. Le nucléole est entouré d'une couche d'hétérochromatine très compacte nommée région péri-nucléolaire, qui inclut des régions centromériques (voir §A.3.1.2), notamment celles se trouvant à proximité des NORs, et les répétitions d'ADNr inactives (pour revue : [Pollock et Huang \(2010\)](#)). Ces sous-régions du nucléole sont chacune associées à des fonctions particulières.

Le nucléole est le site de transcription des ADNr par l'ARN Polymérase I, et leur niveau d'expression est régulé en fonction du besoin en protéines de la cellule. Cela lui confère un rôle dans la régulation de la croissance cellulaire. La transcription des ADNr nécessite donc une régulation fine qui implique des acteurs multiples

---

1. Certains noyaux peuvent contenir plusieurs nucléoles distincts. Nous utilisons le terme au singulier pour désigner le compartiment dans son ensemble.

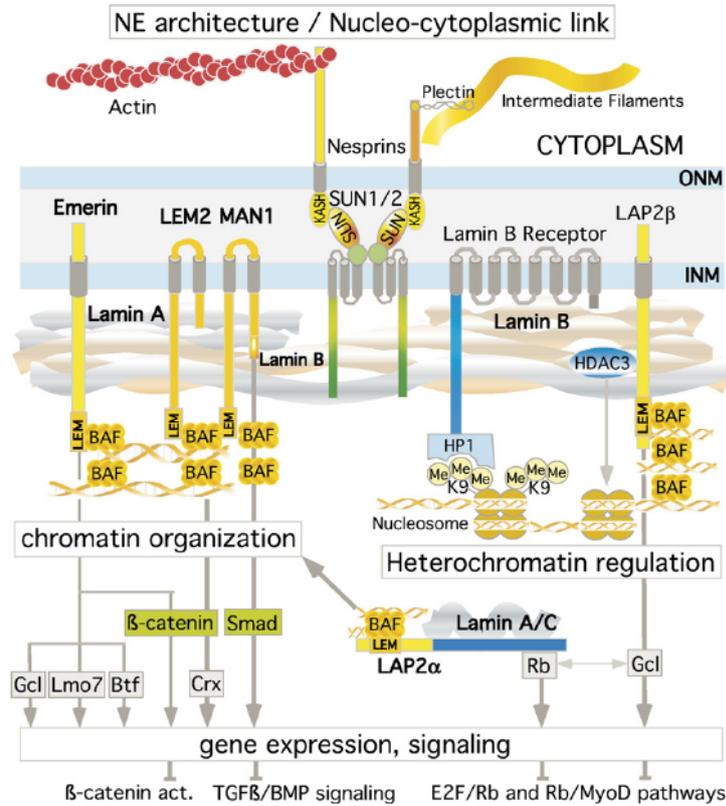


FIGURE A.9: Vue d'ensemble des rôles connus des protéines se liant aux lamines (d'après Schirmer et Foisner (2007)). L'emerine, MAN1 et LEM2, qui sont associées à la lamine A/C (et à la lamines B pour MAN1), interagissent avec BAF et pourraient être impliquées dans le positionnement du génome. LAP2β et LBR, qui sont associées à la lamine B, s'associent à BAF et HP1 respectivement et fournissent des sites d'ancrage de l'hétérochromatine portant la marque H<sub>3</sub>K<sub>9</sub>me<sub>3</sub>. LAP2β s'associe aussi à l'histone déacétylase HDAC3 qui contribue à la formation d'hétérochromatine. LAP2β et LBR sont donc aussi impliquées dans le positionnement du génome. Les protéines à domaines SUN de la MNI interagissent avec les protéines à domaines KASH de la MNE, permettant le couplage mécanique entre la lamina et le cytosquelette. Certaines protéines de la MNI interagissent avec des facteurs de transcription et des répresseurs transcriptionnels, et pourraient de cette manière être impliquées dans la régulation transcriptionnelle. Enfin, la fraction nucléoplasmique des lamines A/C interagit avec BAF, et pourrait ainsi affecter le positionnement du génome.

(voir pour revue : McStay et Grummt (2008)). Les ARNr sont par la suite modifiés au niveau des régions fibrillaires denses, et les ribosomes assemblés au niveau des composants granulaires Boisvert *et al.* (2007).

Il a été montré que le nucléole est aussi le site de modifications d'ARN et d'assemblage de multiples complexes ribonucléoprotéiques tels que la télomérase, la RNase P et de nombreux microARN. Le nucléole est aussi impliqué dans la régulation de modifications post-traductionnelles telles que la sumoylation et la phosphorylation de protéines nucléaires (Pederson (2010)).

Le nucléole est capable de séquestrer des protéines spécifiques pour inhiber leur activité au cours du cycle cellulaire, et de coordonner différents aspects de la progression du cycle cellulaire (Pederson (2010)). Le nucléole joue aussi un rôle essentiel dans la réponse à différentes formes de stress (pour revue : Boulon *et al.* (2010)).

Enfin, la région péri-nucléolaire semble avoir un rôle dans le positionnement du

génomique et la répression transcriptionnelle (voir §A.1.3.3).

**Les corps polycomb.** Les corps polycomb sont des structures nucléaires de diamètre variant entre 0.3 et 1  $\mu\text{m}$ , enrichies en protéines du PcG localisées dans l'ensemble de l'espace du noyau. Ils correspondent à une très forte accumulation locale de fibres d'hétérochromatine (enrichie en la marque  $\text{H}_3\text{K}_{27}\text{me}_3$ ) comme le montre l'analyse récente de leur ultra-structure (Smigová *et al.* (2011)). Au sein des corps polycomb, la chromatine forme de multiples boucles entre des régions régulatrices, des promoteurs et l'extrémité 3' de gènes, (Bantignies et Cavalli (2011)), et, à un ordre supérieur, des interactions à longue distance (voir §A.1.3.3). Il a été suggéré que ces corps puissent constituer des compartiments nucléaires appauvris en ARN polymérase II, qui servent de centres de répression pour établir et maintenir une répression stable de la chromatine grâce à une forte concentration locale de protéines du PcG.

**Les usines à transcription.** Chez les mammifères, les gènes transcrits sont presque toujours associés à des domaines discrets d'une taille comprise entre 0.2 et 0.5  $\mu\text{m}$  de diamètre, fortement enrichis en ARN polymérase de type II, appelées usines à transcription (TF : Transcription Factory), et qui sont distribuées dans l'ensemble de l'espace du noyau (Jackson et Hassan (1993)). Il a été suggéré que ces compartiments résultent de l'agrégation de complexes contenant l'ARN polymérase II et de facteurs de transcriptions (Razin *et al.* (2011)). Les TF sont des structures relativement immobiles dont le nombre varie entre les types cellulaires (Chakalova et Fraser (2010)), et sont visitées par les gènes actifs lorsqu'ils doivent être transcrits ; des expériences de RNA-FISH suggèrent que la transcription a lieu quasiment exclusivement dans les TFs (Osborne *et al.* (2004)). Cette mise en commun de la transcription permet certainement d'augmenter le rendement de la transcription. Les usines à transcription présentent une forte concentration en complexes de transcription pré-assemblés, estimée à plus de 1000 fois supérieure que dans le reste du noyau. Ces complexes peuvent facilement être recyclés à la fin de la transcription ; soit au niveau du même gène, soit au niveau d'autres gènes présents dans la même usine (Cook (2002)), ce qui a pour effet d'augmenter le rendement de la transcription.

Les usines à transcription pourraient être spécialisées dans la transcription de groupes de gènes particuliers, par exemple en concentrant des facteurs de transcriptions impliqués dans la régulation des gènes du même groupe. Cela concerne notamment les gènes de la  $\beta$ -globine (pour revue : Bartlett *et al.* (2006)).

La notion d'usine à transcription est généralement utilisée pour désigner les corps qui concentrent l'ARN polymérase de type II, mais cette notion peut être étendue aux autres polymérases : la polymérase de type I est enrichie dans les nucléoles, et la polymérase de type III qui se localise dans des foyers nucléoplasmiques dédiés à celle-ci et distincts des foyers contenant la polymérase de type II (Pombo *et al.* (1999)).

**Les speckles nucléaires.** Les speckles nucléaires sont généralement considérés comme le site de stockage et de modification des facteurs d'épissage des ARN pré-messagers. Il contiennent aussi des ARN non codants tels que Neat1/2 et Mala1, mais pas d'ADN. Ils forment des foyers de taille variable, composés de granules de 20-25 nm (Austin et Bellini (2010)). L'association de gènes au niveau des speckles pourrait augmenter l'efficacité de leur épissage ; les speckles pourraient donc aussi influencer l'expression des gènes (pour revue : Mao *et al.* (2011)).

**Les paraspeckles.** Les paraspeckles sont des structures de diamètre compris entre 0.5 et 1 µm qui sont localisées à proximité des speckles. Ils sont caractérisés par la présence de la protéine PSP1 (Paraspeckle Protein 1) et n'ont été observés que chez les mammifères. Ils jouent un rôle dans la régulation de l'expression des gènes via la séquestration de certains ARNm. Ce sont des structures dynamiques qui ne sont pas présentes dans tous les types cellulaires et qui se forment à partir de l'ARN long non codant Men $_{\epsilon/\beta}$  (aussi appelé Neat1), à proximité du locus du gène (pour revue : Bond et Fox (2009)). Ils constituent un paradigme pour le mécanisme de biogenèse des corps nucléaires à partir d'une *graine*<sup>1</sup> (en l'occurrence l'ARN Neat1) (Caudron-Herger et Rippe (2012)).

**Les corps PML.** Des observations en microscopie électronique ont montré que les corps PML ont un diamètre compris entre 0.2 et 1 µm, et une forme de tore. L'anneau est composé de protéines, notamment la protéine PML (Promyelocytic Leukaemia), et l'on observe de l'ARN et de l'ADN à sa périphérie. Les corps PML ont longtemps été considérés comme des sites de séquestration de protéines, mais des résultats récents suggèrent qu'ils ont aussi un rôle dans la réparation des dommages à l'ADN, la régulation de la transcription, la sénescence ainsi que l'apoptose (pour revue : Dundr (2012)).

**Les corps de Cajal.** Les corps de Cajal sont des structures nucléaires présentes chez tous les eucaryotes, et caractérisées par la présence de la protéine coilin. Elles sont impliquées dans la biogenèse (assemblage et modifications post-traductionnelles) et le recyclage de complexes permettant la transformation des ARN, tels que les pRNPn du spliceosome (Nizami *et al.* (2010)). Les corps de Cajal contiennent de nombreux facteurs impliqués dans l'épissage des pré-ARNm, dans les modifications post-transcriptionnelles des pré-ARNr et de l'extrémité 3' des pré-ARNm des histones, ainsi que dans la maintenance des télomères. Ils semblent résulter d'un processus d'auto-organisation qui ne suit pas un processus d'association séquentiel de leurs composants (Austin et Bellini (2010)).

## Conclusion

Le noyau abrite de nombreux compartiments qui sont le siège de fonctions particulières. Le phénomène de compartimentation permet en général d'optimiser des

---

1. *Seed* en anglais.

réactions métaboliques en concentrant des facteurs dans une zone restreinte. Nous allons voir par la suite que certains compartiments constituent des environnements ayant un rôle activateur ou répressif pour la transcription lorsque des régions génomiques se positionnent à proximité.

### A.1.3 Positionnement nucléaire du génome

Le génome ne se positionne pas aléatoirement<sup>1</sup> dans le noyau ; des arguments appuyant cette idée seront présentés tout au long de cette section. Le positionnement du génome a souvent été quantifié en mesurant la position radiale de TCs ou de gènes (A.1.3.1). Cette approche est utile pour donner une description du positionnement global du génome dans l'espace du noyau et pour montrer des réorganisations.

Cependant, le noyau étant un espace très hétérogène (voir A.1.2), il peut être important d'adopter une description plus fine de la localisation d'une région génomique en caractérisant son environnement local. Celui-ci peut en effet avoir une importance fonctionnelle directe. Le positionnement du génome sera donc aussi décrit par rapport à différents compartiments nucléaires, situés soit en périphérie du noyau (A.1.3.2) soit à l'intérieur du noyau (A.1.3.3), en détaillant à chaque fois les hypothèses de mécanismes ainsi que les fonctions associées au positionnement.

#### A.1.3.1 Positionnement radial du génome

Les chromosomes et les gènes n'occupent pas des positions aléatoires dans l'espace du noyau. Ce fait a d'abord été démontré en mesurant le positionnement radial de gènes et de TCs afin de les localiser dans de l'espace du noyau ; les différentes approches seront discutées au §A.4.3.2.

**Positionnement radial des territoires chromosomiques.** Les TCs occupent des positions radiales préférentielles dans le noyau. Des règles générales ont été établies à partir de corrélations : les chromosomes riches en gènes et transcriptionnellement plus actifs ont tendance à se localiser à l'intérieur du noyau, par opposition aux chromosomes pauvres en gènes et moins actifs qui se localisent à la périphérie (Croft *et al.* (1999), Boyle *et al.* (2001)). Dans certains types cellulaires, c'est la taille des chromosomes qui est le facteur expliquant le mieux leur position (Bolzer *et al.* (2005)). Cette organisation peut aussi être modulée par l'état physio-pathologique de la cellule. Par exemple, sous l'effet de déprivation de sérum, certains TCs se relocalisent rapidement vers l'intérieur ou vers la périphérie selon les chromosomes (voir §A.1.4.3). Enfin cette organisation peut varier entre les types cellulaires, par exemple, chez la souris, le chromosome 5 se localise préférentiellement à l'intérieur du noyau de cellules de foie et en périphérie de noyaux de cellules de poumons (Parada *et al.* (2004)). Ces observations suggèrent que la localisation des TCs dans le noyau est liée aux fonctions du génome.

---

1. La notion de *non aléatoire* est utilisé ici pour désigner la présence de régularités statistiques. Il est utilisé dans cette section car c'est le terme consacré dans le domaine, une discussion est proposée au §A.4.2 pour préciser cette notion.

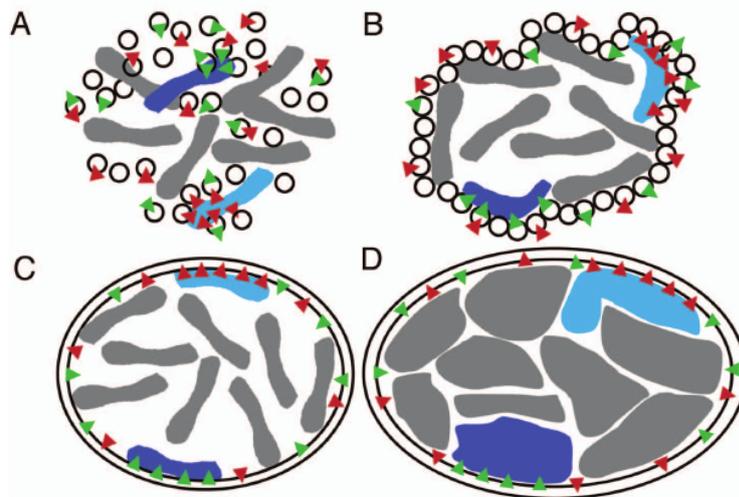


FIGURE A.10: Mécanisme spéculatif d'affinité permettant d'expliquer l'établissement des profils d'organisation spatiale des chromosomes (d'après Zuleger *et al.* (2011)). (A) Au cours de la mitose, la membrane nucléaire pourrait soit se fissionner en vésicules EN/RE, soit diffuser dans le RE. Le modèle de vésicules est montré ici. Différentes vésicules contiennent des composants spécifiques (les triangles rouges et verts représentent différentes protéines de la MNI). Les composants de certaines vésicules interagissent préférentiellement avec certains chromosomes condensés. (B) À la fin de la mitose, l'EN commence à se reformer à partir des vésicules, et certains chromosomes restent accrochés spécifiquement à certaines vésicules. (C) L'EN s'est reformée, et certains chromosomes sont piégés à la membrane à cause de l'affinité particulière de certains composants spécifiques de l'EN pour ces chromosomes. À ce stade, les chromosomes sont encore très condensés. (D) Au cours de l'interphase, les chromosomes se décondensent et occupent des espaces distincts au sein du noyau. À ce stade, un grand nombre d'interactions moins spécifiques et de moins forte affinité avec la lamina pourraient maintenir le positionnement établi au cours de la mitose.

La densité en gènes et la taille des chromosomes ne sont pas les seuls facteurs permettant d'expliquer les positions préférentielles des TCs, puisque l'on observe des différences entre types cellulaires pour lesquels ces caractéristiques sont identiques. Une étude récente a apporté de nouveaux éléments de réponse à cette question, en analysant l'impact, sur la localisation des TCs, de protéines nouvellement identifiées de la MNI se liant à la lamina. Les auteurs ont notamment montré que certaines de ces protéines pouvaient promouvoir spécifiquement la relocalisation de TCs vers l'intérieur ou vers la périphérie du noyau (Zuleger *et al.* (2013)). Ils suggèrent donc que le profil d'expression des protéines de la MNI se liant à la lamina influence l'organisation radiale des TCs. Ainsi, les variations observées de profils d'expression des protéines de la MNI entre différents types cellulaires pourraient expliquer, au moins en partie, les différences de profils de localisation des TCs. Les auteurs de cette étude proposent un modèle d'établissement de l'organisation spatiale des TCs expliqué par les affinités spécifiques de protéines de membrane pour les différents chromosomes condensés (Zuleger *et al.* (2011), Fig. A.10). Ce travail suggère que les vésicules formées par la membrane pendant la mitose contiennent une information qui est restituée au moment de la formation du noyau et qui influence l'organisation nucléaire.

La membrane nucléaire apparaît donc comme un nouveau support d'une information cellulaire persistante, transmissible d'une cellule mère aux cellules

filles; on pourrait alors parler d'*épigénétique de membrane*. Ce travail attire l'attention sur le fait que l'étude précise de la formation du noyau en fin de mitose et début de phase G<sub>1</sub> apportera beaucoup d'éléments sur la mise en place de l'organisation de la chromatine.

L'étude de pathologies liées aux protéines de la lamina<sup>1</sup> a aussi apporté des informations sur l'implication de la lamina dans le positionnement radial des TCs. Une des pathologies les plus étudiées est le syndrome de progéria d'Hutchinson-Gilford (HGPS : Hutchinson-Gilford Progeria Syndrom)<sup>2</sup>, qui résulte de mutations des protéines de la lamina<sup>3</sup>. Dans les cellules HGPS, le positionnement radial des TCs est altéré (Meaburn *et al.* (2007)). De plus, le traitement de ces cellules par des inhibiteurs de farnésyl-transférase (FTI)<sup>4</sup> restaure le profil de positionnement radial des TCs (Mehta *et al.* (2011)).

**Positionnement radial des gènes.** Comme celle des chromosomes, la position radiale des gènes n'est pas aléatoire et a été corrélée à leur activité. Par exemple le locus IgH est associé préférentiellement à la périphérie du noyau dans des progéniteurs de lymphocytes B, lorsqu'il est silencieux; et celui-ci se relocalise vers l'intérieur du noyau lorsqu'il est activé dans des précurseurs de lymphocytes B (Kosak *et al.* (2002)). Même si d'autres exemples existent, la position radiale d'un gène n'est cependant pas toujours corrélée à son activité; il existe des contre-exemples dans lesquels on observe le positionnement d'un gène actif à la périphérie du noyau (Zink *et al.* (2004)). De plus, on observe que les deux allèles d'un gène n'occupent pas toujours la même position radiale, même lorsque leur activité est la même (Roix *et al.* (2003)).

Le positionnement périphérique d'un gène peut être le reflet de l'association de celui-ci à des composants de l'EN, mais cela ne peut pas être directement corrélé à l'activité du gène car l'EN est un compartiment hétérogène (§A.1.2.2). D'autre part, le positionnement central peut correspondre à une association du gène à des compartiments centraux; mais au vue de la diversité des compartiments nucléaires et de leur impact sur la transcription (voir §A.1.2), on peut s'attendre à ce que le positionnement central d'un gène ne soit pas strictement corrélé à son activité.

### A.1.3.2 Association de régions génomiques à la périphérie

**Les domaines associés à la lamina.** La lamina constitue un site d'ancrage pour la chromatine. Chez les mammifère, la technique du DamID (voir §A.4.1.1) a permis de mettre en évidence environ 1300 domaines associés à la lamina (LAD : Lamina Associated Domains) qui sont en contact avec la lamine B1 (Guelen *et al.*

---

1. Pour une revue sur les pathologies liées à la lamina, voir : Dechat *et al.* (2008).

2. Cette pathologie rare est très étudiée, notamment parce qu'elle reproduit le phénotype du vieillissement physiologique (Scaffidi *et Misteli* (2006)).

3. Dans les cellules HGPS, la structure de la lamina est anormale à cause de l'accumulation de la progérine, qui est une forme tronquée de la lamine A qui ne pouvant être dé-farnésylée.

4. Les FTI réduisent l'accumulation de progérine dans la lamina.

(2008)). Ces domaines couvrent des régions génomiques de 100 Kb à 10 Mb, avec une taille médiane de 0.5 Mb, et représentent environ 40% du génome. Ils sont caractérisés par des marques de l'hétérochromatine et une faible densité en gènes. Certains LADs sont dit constitutifs car partagés entre tous les types cellulaires, et d'autres sont variables.

**Sites d'ancrages de la chromatine à la périphérie.** La lamina constitue un site d'ancrage de la chromatine à la périphérie du noyau, soit par une interaction directe des lamines avec l'ADN (Stierlé *et al.* (2003)) ou les histones (Taniura *et al.* (1995)), soit via différentes protéines se liant aux lamines. Les protéines à domaines LEM médient l'association de la chromatine à la périphérie du noyau via l'association à la protéine BAF (Barrier to Autointegration Factor). BAF est une protéine liant la chromatine très conservée chez les métazoaires, capable de ponter l'ADN et d'interagir avec les histones (pour revue : Margalit *et al.* (2007)). Il existe beaucoup de protéines à domaine LEM et il semble qu'elles soient, au moins en partie, fonctionnellement redondantes. Certaines de ces protéines peuvent aussi lier directement l'ADN ou d'autres protéines liant la chromatine, telles que la protéine HA95. En plus des protéines à domaine LEM, la protéine LBR s'associe à la lamine de type B et peut ancrer la chromatine à la périphérie ; ce, via son interaction avec la protéine HP1 (Ye et Worman (1996)), qui interagit avec l'hétérochromatine portant la marque H<sub>3</sub>K<sub>9</sub>me<sub>3</sub>. Il a aussi été montré que LBR peut interagir avec MeCP2, une protéine liant l'ADN méthylé (Guarda *et al.* (2009)).

Par ailleurs, la lamina est interrompue au niveau des pores nucléaires, et ceux-ci peuvent aussi servir de sites d'ancrage à la chromatine chez la levure. Chez les mammifères ce type de mécanismes pourrait aussi exister, comme le montre le fait que la chromatine enrichie dans la marque H<sub>4</sub>K<sub>5</sub>ac (acétylation de la lysine 5 sur l'histone H<sub>4</sub>) se localise à la périphérie du noyau, et interagit avec la protéine de pore nucléaire Nup83 (Brown *et al.* (2008)).

**Mécanisme de positionnement des LADs.** L'adressage des LADs à la lamina est encore peu compris. Il pourrait faire intervenir des protéines de la lamina se liant à l'ADN, et reconnaissant des caractéristiques de séquences particulières. Une étude récente montre que chez la souris, certains LADs sont enrichis en un motif GAGA qui interagit avec un complexe de protéines associé à la lamina (Zullo *et al.* (2012)). Cet enrichissement n'étant pas observé pour tous les LADs, il est peu probable que ce mécanisme soit général. En revanche, les LADs constitutifs sont caractérisés par la présence de longs domaines riches en paires A/T. Cela suggère un mécanisme d'adressage à la lamina *par défaut* impliquant des isochores A/T, et des mécanismes de régulation spécifiques pour les gènes dont l'expression est modulée au cours de la différenciation (Meuleman *et al.* (2013)). L'adressage des LADs à la lamina pourrait aussi impliquer leur état chromatinien, puisqu'il existe des protéines de la lamina qui interagissent avec des protéines de la chromatine. Un tel mécanisme a déjà été mis en évidence chez le nématode *Caenorhabditis elegans*, où deux HMTs capables de méthyliser H<sub>3</sub>K<sub>9</sub> induisent la localisation périphérique

d'un transgène hétérochromatinien ainsi que de bras chromosomiques endogènes riches en H<sub>3</sub>K<sub>9</sub>me<sub>3</sub> (Towbin *et al.* (2012)).

**Implications fonctionnelles des associations à la périphérie.** Les expériences dans lesquelles la relocalisation artificielle de gènes de l'intérieur du noyau vers la lamina mène à leur répression transcriptionnelle, ou les rend réfractaires à l'activation (Reddy *et al.* (2008)), montrent que la lamina participe directement à la répression transcriptionnelle des gènes. De plus, l'adressage à la lamina d'un transgène peut aussi induire la répression de certains gènes alentours (Finlan *et al.* (2008)), ce qui montre l'implication de l'environnement local dans la régulation de la transcription. Ces expériences constituent la preuve la plus forte du lien entre le positionnement nucléaire et la régulation de l'expression transcriptionnelle (pour revue : Geyer *et al.* (2011)).

La lamina favorise probablement l'induction d'un état chromatinien répressif; d'une part grâce à l'interaction de LBR avec HP1, qui permet la propagation de la marque H<sub>3</sub>K<sub>9</sub>me<sub>3</sub>, et d'autre part, la protéine LAP2 $\beta$  qui interagit avec HDAC3 (Somech *et al.* (2005)) et induirait la désacétylation de l'histone H<sub>4</sub>. Ce dernier mécanisme a récemment été mis en évidence chez la drosophile (Milon *et al.* (2012)). La lamina peut aussi avoir un rôle dans la régulation de l'expression des gènes en séquestrant des facteurs de transcription ou en liant des répresseurs transcriptionnels (pour revue : Schirmer et Foisner (2007)). De plus, il a été montré que la lamine A (et non la lamine B) peut réprimer l'expression des gènes dans des cellules de mammifères par une interaction directe avec la chromatine (Lee *et al.* (2009)). Il est possible que la lamina soit impliquée dans la stabilisation de la répression des gènes, par le biais d'une boucle de rétroaction : d'une part la lamina contribue à l'établissement d'un état chromatinien répressif, et d'autre part l'état chromatinien répressif augmente l'affinité de la chromatine pour la lamina. A l'appui de cette hypothèse, il a été montré que les gènes de pluripotence interagissent plus fréquemment avec la lamina dans les cellules différenciées que dans les cellules souches embryonnaires (ESC : Embryonic Stem Cell), ce qui suggère que la lamina participe à verrouiller l'état chromatinien de ces gènes au cours de la différenciation (Peric-Hupkes *et al.* (2010)).

Les LADs constitutifs pourraient participer à l'organisation basale du génome. La présence de LAD non constitutifs suggère quant à elle que la lamina est impliquée dans la régulation des gènes au cours de la différenciation. Les LADs sont globalement stables au cours de la différenciation de neurones murins, avec des altérations fines et séquentielles au fur et à mesure des stades de différenciation qui corrélerent avec l'activité des gènes : les gènes ou unités de gènes qui se détachent de la lamina sont activés (parfois seulement à un stade de différenciation ultérieur), et ceux qui s'y associent sont réprimés (Peric-Hupkes *et al.* (2010)).

Il a été observé que l'adressage artificiel d'un locus à la périphérie ne mène pas systématiquement à sa répression transcriptionnelle (Kumaran et Spector (2008)). Cela suggère que l'EN pourrait constituer un environnement bivalent pour la transcription, comme chez la levure; mais il n'existe pas de preuves expérimentales de

mécanismes de surexpression induits par une relocalisation à la périphérie à ce jour.

### A.1.3.3 Association de régions génomiques à des compartiments internes du noyau

**Association aux nucléoles.** Plusieurs résultats expérimentaux suggèrent que les nucléoles pourraient jouer un rôle dans le positionnement du génome. De larges domaines de taille comprise entre 100 Kb et 10 Mb, pauvres en gènes et s'associant avec les nucléoles ont été identifiés, et dénommés NADs (Nucleolus-Associated Domains). Ces régions sont particulièrement enrichies en gènes réprimés et en répétitions satellite spécifiques<sup>1</sup>, en séquences riches en A/T, et montrent un enrichissement pour des marques d'histones répressives, en particulier H<sub>3</sub>K<sub>9</sub>me<sub>3</sub> (Németh *et al.* (2010), Koningsbruggen *et al.* (2010)). La protéine CTCF, qui se lie au motif CCCTC, interagit aussi avec la protéine nucléolaire B23 (nucléophosmine); elle est donc capable de recruter ses sites de liaison à la périphérie des nucléoles (Yusufzai *et al.* (2004)). D'autres médiateurs d'interactions entre domaines chromatiniens et nucléoles ont été proposés (pour revue Németh et Längst (2011)).

Durant l'inactivation du chromosome X, le centre d'inactivation de l'X cible le chromosome X inactif (Xi) au niveau de la région péri-nucléolaire. Cette association dépend de l'ARN long non codant Xist. Il a été suggéré que le Xi doive continuellement visiter la région péri-nucléolaire au moment de sa réplication, entre le milieu et la fin de la phase S, afin de répliquer son état épigénétique répressif (Zhang *et al.* (2007)). La région péri-nucléolaire aurait donc un rôle dans la propagation et la maintenance d'états chromatiniens, notamment durant la réplication (pour revue : Guetg et Santoro (2012)). Cette idée est aussi supportée par le fait que les NADs soient transcriptionnellement inactifs.

**Association aux corps polycomb.** Les corps polycomb correspondent à des regroupements de loci réprimés, ils sont caractérisés cytologiquement par des foyers contenant une forte concentration en marques d'hétérochromatine H<sub>3</sub>K<sub>27</sub>me<sub>3</sub>. D'une part, ils sont le reflet de l'association globale de la chromatine polycomb avec elle-même. D'autre part, ils peuvent être considérés comme des centres de répression transcriptionnelle. En effet, il a été montré que les protéines du PcG sont capables de favoriser des interactions longue-distance inter-chromosomiques chez la drosophile (Bantignies *et al.* (2003)); de plus, les gènes Hox Antp et Adb-B, qui sont co-réprimés et colocalisent au niveau de corps polycomb, sont relocalisés dans des compartiments séparés lors de l'activation d'un des deux gènes (Bantignies *et al.* (2011)). Chez l'homme, un crible sur l'ensemble du génome a montré que EZH2 est un médiateur direct d'interactions longue-distance inter- et intra-chromosomiques qui régule la répression transcriptionnelle de multiples gènes en facilitant le regroupement spatial entre des régions génomiques éloignées, et en ciblant la machinerie polycomb (Tiwari *et al.* (2008)).

1. Les séquences satellites sont des séquences répétées trouvées majoritairement au niveau des régions centromériques, voir §A.2.1 pour plus de détails.

Trois types de mécanismes permettant le recrutement de loci par le système polycomb ont été proposés (pour revue : [Beisel et Paro \(2011\)](#)) :

- Des protéines liant l'ADN au niveau de régions génomiques particulières, pouvant être des facteurs de transcription globaux ou spécifiques de gènes ; ces facteurs interagissant avec des protéines du PcG. Certains éléments spécifiques de ces régions génomiques ont été identifiés, tels que PRE-kr ou D11.12, ainsi que les îlots CpG.
- La modification d'histones  $H_3K_{27}me_3$ , apposée par les protéines EZH1 et EZH2 et reconnue par les protéines CBX2/7 et EED.
- Des mécanismes impliquant des ARN non codants tels que Xist ou HOTAIR.

Il a été montré que certains de ces mécanismes impliquent des boucles d'ADN dont la formation dépend d'insulateurs tels que CTCF, et qui favorisent des interactions longue-distance entre loci régulés par le système polycomb (pour revue : [Bantignies et Cavalli \(2011\)](#))

**Association aux usines à transcription.** Les TFs étant moins nombreuses que le nombre de gènes actifs d'une cellule, il a été proposé que chacune d'entre-elles contiennent plusieurs gènes exprimés. Ceci a été montré dans plusieurs systèmes. Par exemple, chez la souris, un crible sur l'ensemble du génome des séquences partageant les mêmes TFs que les gènes actifs d' $\alpha$ - et  $\beta$ -globine a mis en évidence un réseau d'interaction contenant des centaines d'autres loci transcrits. La co-localisation de ces gènes dans les TFs se fait non seulement en *cis* mais aussi en *trans*. Ce réseau est très enrichi en gènes régulés par le facteur de transcription Klf1. Ces derniers s'associent préférentiellement dans un nombre limité de TF caractérisés par une forte concentration en Klf1 ([Schoenfelder \*et al.\* \(2010b\)](#)). Ce résultat suggère que certaines TF peuvent correspondre à des centres de transcription spécialisés pour un réseau de gènes en particulier.

Un mécanisme d'auto-organisation a été proposé pour expliquer l'enrichissement en gènes co-régulés au sein d'une même TF : l'association d'un facteur de transcription étant transitoire, il peut être libéré et s'associer à un autre gène du même réseau de régulation présent dans la même TF, initiant la transcription de celui-ci et stabilisant sa présence au sein d'une TF. A l'inverse, les gènes transcrits en l'absence d'autres gènes régulés par les mêmes facteurs sont plus enclins à se dissocier de la TF, puisque celle-ci présente alors une concentration moins élevée de ces facteurs ([Schoenfelder \*et al.\* \(2010a\)](#)).

De la même manière que pour les corps polycomb, la co-localisation de gènes co-régulés dans des TFs peut avoir une influence sur la conformation spatiale du génome (pour revue : [Schoenfelder \*et al.\* \(2010a\)](#)).

**Associations à d'autres compartiments.** D'autres compartiments nucléaires fournissent des sites d'ancrage de régions génomiques, ou se localisent au niveau de régions particulières du génome. Parmi ceux-ci nous pouvons citer :

- Les gènes réprimés qui s'associent aux foyers d'hétérochromatine péricentromériques, qui seront traités au §A.3.2.3.

- Les gènes U1/U2/U7 et les gènes codant pour les histones qui s’associent respectivement aux corps de Cajal et au HLB (Histone Locus Body) (pour revue : [Nizami \*et al.\* \(2010\)](#))
- Les régions génomiques actives qui s’associent aux corps PML ([Wang \*et al.\* \(2004\)](#))
- Les granules de stress se formant au niveau des régions péricentromériques des chromosomes 9, 12 et 15 ([Denegri \*et al.\* \(2002\)](#))
- Les paraspeckles sont localisés au niveau du locus Neat1 sur le chromosome 11 ([Bond et Fox \(2009\)](#)).

#### A.1.3.4 Conclusion sur la relation entre positionnement et fonction du génome

Les exemples qui viennent d’être cités démontrent l’existence de régularités dans le positionnement du génome, à la fois dans le positionnement radial des TCs et dans le positionnement de régions génomiques par rapport à des compartiments nucléaires. Il y a une corrélation entre le positionnement radial d’un locus et son activité transcriptionnelle : les gènes localisés à l’intérieur du noyau sont en général actifs, et les gènes situés en périphérie réprimés ; mais il existe des contre-exemples. La position radiale d’un locus ne détermine pas son activité ; en revanche, le repositionnement vers un compartiment répressif, qu’il se trouve à la périphérie ou à l’intérieur du noyau, est toujours corrélé à la répression de ce locus. Il est donc important d’adopter une description plus précise que le positionnement radial pour relier la position d’un gène à son activité.

La compartimentation nucléaire induit la formation d’environnements nucléaires activateurs tels que les TFs ou les pores nucléaires, et de quatre environnements répressifs où se localise l’hétérochromatine : la lamina, la périphérie du nucléole, les corps polycomb, et les foyers d’hétérochromatine constitutive qui seront décrits au §A.3.2.

**Lien entre les différents environnements répressifs.** Il est possible que certains compartiments fournissent un environnement fonctionnellement similaire, et que le génome s’associe indifféremment à l’un ou à l’autre. Ceci pourrait expliquer le caractère probabiliste du positionnement du génome.

Plusieurs résultats expérimentaux soutiennent cette idée dans le cas de la lamina et de la périphérie des nucléoles. D’une part, on constate un fort taux de recouvrement entre les LADs et les NADs, et les tailles de ces deux types de domaines sont du même ordre de grandeur ([Németh \*et al.\* \(2010\)](#), [Koningsbruggen \*et al.\* \(2010\)](#)). D’autre part, deux études utilisant des stratégies permettant de marquer spécifiquement les régions associées à la lamina dans une cellule mère, et d’étudier leur position dans les cellules filles ([Koningsbruggen \*et al.\* \(2010\)](#), [Kind \*et al.\* \(2013\)](#)). Les auteurs de ces deux études observent que les régions associées à la lamina sont redistribuées soit à la lamina, soit au niveau des nucléoles après la division cellu-

laire<sup>1</sup>.

Il est tentant de postuler que les nucléoles et la lamina sont deux plate-formes organisatrice du génome, fournissant un environnement répressif similaire ; et qu'au moins une partie des NADs et des LADs sont susceptibles de s'associer soit à l'une soit à l'autre de manière aléatoire.

**Relation de causalité entre positionnement et activité.** Si la corrélation entre le repositionnement d'un gène et la modification de son activité a été établie dans de nombreux cas, il est difficile de conclure sur la relation de causalité entre ces deux phénomènes. Les expériences de ciblage de loci à la lamina présentées au §A.1.3.2 montrent que dans certains cas, le positionnement peut influencer directement l'activité. En revanche, une étude a montré que la localisation nucléaire du gène CFTR est modifiée après altération de son niveau de transcription, alors que son activité n'est pas affectée lorsque sa relocalisation est induite vers un compartiment nucléaire différent ; cela montre que dans ce cas, c'est l'activité transcriptionnelle qui affecte le positionnement nucléaire et non l'inverse (Zink *et al.* (2004)). De plus, il est possible que l'environnement chromatinien local en *cis* ait une influence plus forte sur la transcription que des effets de positionnement dans l'espace du noyau, comme le montre le fait qu'une construction contenant un opéron lac intégrée dans une région riche en gènes sur le chromosome 1 humain se localise toujours à l'intérieur du noyau et ne se repositionne pas suite à l'activation transcriptionnelle (Janicki *et al.* (2004)).

Il est possible que, dans certains cas, le repositionnement d'un gène vers un compartiment nucléaire particulier ne soit pas nécessaire au changement de l'activité d'un gène, mais intervienne comme un niveau de régulation s'ajoutant aux niveaux de régulation génétique (liée aux facteurs de transcription) et épigénétique (liée à la compaction de la chromatine), afin d'optimiser l'activité du gène dans son ensemble, celle-ci incluant les modifications des ARN et leur export vers le cytoplasme.

#### A.1.4 Dynamique de l'architecture nucléaire

Nous avons donné jusqu'ici une description statique de l'organisation spatiale du noyau. Les développements techniques dans la microscopie des dernières années ont permis d'étudier les mouvements de la chromatine à courte échelle temporelle, avec une résolution pouvant aller jusqu'à quelques millisecondes. Celle-ci présente en fait un fort degré de dynamique ; à la fois à des échelles temporelles courtes auxquelles il a été observé que des régions chromatiniennes peuvent suivre des mouvements aléatoires à courte distance, et à des échelles temporelles plus élevées auxquelles des mouvements dirigés à longue portée ont été observés. Enfin, la structure du noyau subit des réorganisations en réponse à des signaux comme des stress ou

---

1. Kind *et al.* (2013) observent qu'une faible fraction des domaines associés à la lamina peut se relocaliser à l'intérieur du noyau sans être en contact avec les nucléoles ; il est possible que ces domaines soient alors associés aux foyers d'hétérochromatine constitutive, mais cette étude ne les prend pas en compte.

la différenciation. Les caractéristiques de ces différents types de dynamiques sont résumées dans la figure A.11.

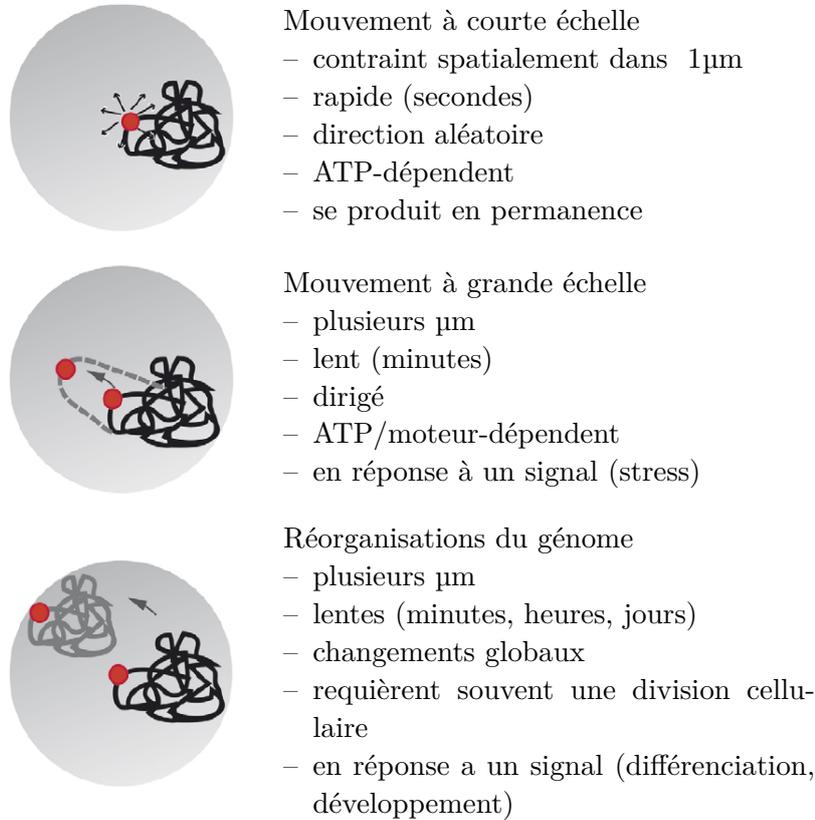


FIGURE A.11: Comparaison des trois types de dynamiques de la chromatine (d'après Soutoglou et Misteli (2007)). Mouvement à courte échelle : un locus particulier (rouge) suit un mouvement contraint localement sur des distances de l'ordre du  $\mu\text{m}$ . Mouvement à grande échelle : un locus peut suivre un mouvement sur une grande distance, certainement dirigé et dépendant de moteurs moléculaires. Réorganisations globales : des mouvements de réorganisation globale du génome via le repositionnement de chromosomes entiers ou de régions chromosomiques peuvent avoir lieu au cours de la différenciation ou du développement.

#### A.1.4.1 Mouvements à courte échelle.

**Mouvements de loci chromatiniens.** Des expériences de FRAP (Fluorescence Recovery After Photobleaching) ont montré que les mouvements de la chromatine restent contraints dans des zones d'environ  $0.4\ \mu\text{m}$ , sur des périodes de plus d'une heure (Abney *et al.* (1997)). Mais la faible résolution temporelle de cette approche ne permet pas d'exclure des mouvements de la chromatine à des échelles plus courtes encore. Des expériences de microscopie bi-photonique à haute résolution ont permis d'étudier la dynamique d'un locus artificiel avec une résolution spatiale de  $20\ \text{nm}$  et une résolution temporelle de  $30\ \text{ms}$  (Levi *et al.* (2005)). Cette étude montre que le locus marqué suit globalement un mouvement de diffusion contrainte dans un espace de rayon d'environ  $40\ \text{nm}$ . Comparée à la diffusion de l'ADN en solution, la diffusion de la chromatine dans le noyau est plus lente de plusieurs ordres de grandeur.

La diffusion de la chromatine est probablement influencée par la compaction de la chromatine, l'encombrement moléculaire, et l'interaction dynamique de régions génomiques avec des compartiments nucléaires plus stables<sup>1</sup>.

Le mouvement de diffusion est parfois interrompu par des sauts abrupts sur des distances d'environ 150 nm, qui durent entre 0.3 et 2 s (Levi *et al.* (2005)). Ce second régime est sensible à la déplétion en ATP (Adénosine Triphosphate), ce qui suggère qu'il ne résulte pas d'un processus de diffusion passive ; bien que cela ne soit pas une preuve formelle, la déplétion en ATP pouvant avoir des effets indirects<sup>2</sup>. Ce type de mouvement est généralement interprété comme un mouvement actif aléatoire (non dirigé).

**Mouvements des territoires chromosomiques.** Les territoires chromosomiques ont été visualisés dans des cellules vivantes en incorporant des analogues de nucléotides au cours de la phase S. Après plusieurs mitoses les chromosomes marqués sont dilués, les chromosomes néo-répliqués ne l'étant pas, et peuvent alors être distingués. Ce type d'approche a conduit à la conclusion que les TCs suivent un mouvement de diffusion Brownienne dans un espace assez restreint, avec une constante de diffusion estimée à  $0.1 \mu\text{m}^2.\text{s}^{-1}$  dans des cellules HeLa ; valeur pouvant cependant varier selon les types cellulaires (Bornfleth *et al.* (1999), Edelmann *et al.* (2001)).

#### A.1.4.2 Mouvements à grande échelle.

Par ailleurs, des mouvements à plus longue distance ont été observés en suivant la position d'un transgène artificiel détecté par la stratégie de l'opéron lac/répresseur lac fusionné à la GFP (Chuang *et al.* (2006)). En sélectionnant une lignée dans laquelle le transgène est préférentiellement localisé à la périphérie du noyau, les auteurs observent la relocalisation du locus vers l'intérieur du noyau suite au ciblage du domaine d'activation du facteur de transcription VP16. Le mouvement observé est dirigé vers l'intérieur du noyau, à une vitesse moyenne de  $0.3 \mu\text{m}.\text{min}^{-1}$ . De plus il est ralenti suite à la surexpression d'un mutant de myosine nucléaire, ce qui suggère que ce mouvement dépend du système actine/myosine. Dans une étude non publiée, présentée au congrès *19th International Chromosome Conference*, l'équipe du professeur Belmont a observé la relocalisation d'un transgène placé en aval d'un promoteur dépendant au choc thermique vers les speckles nucléaires. Le mouvement est rapide, environ  $1.2 \mu\text{m}.\text{min}^{-1}$  ; la mesure des angles entre les différentes positions montre qu'il n'y a pas de retour en arrière. Ils ont par ailleurs observé une accumulation des transcrits au moment du contact avec le speckle. Les auteurs interprètent ce mouvement comme un mouvement actif dirigé ; et soulèvent la question de ce qui détermine la directionnalité et le mouvement,

---

1. A l'appui de cette hypothèse, il a été démontré que la mobilité de loci est significativement plus faible à proximité des nucléoles ou de la périphérie du noyau par comparaison au reste du nucléoplasme (Chubb *et al.* (2002)).

2. En effet, il est possible que ces mouvements apparemment aléatoires soient par exemple causés par des remodelages de la structure chromatinienne dépendant de l'ATP.

celui-ci implique-t-il un rail, un moteur ? En plus du système actine/myosine, il a été proposé que la transcription elle-même soit responsable des mouvements de la chromatine<sup>1</sup>.

Il est tentant de postuler que cette mobilité puisse permettre à un locus donné de sonder son environnement et interagir avec d'autres loci chromatiniens ou des compartiments nucléaires.

#### A.1.4.3 Réorganisations du génome.

Le troisième type de dynamique que nous décrivons dans cette section concerne les réorganisations globales du génome. Ces réorganisations ont souvent lieu en réponse à un signal, sont relativement lentes, peuvent se dérouler sur des temps allant de plusieurs minutes à plusieurs jours, et nécessitent parfois un passage par la mitose.

**Relocalisations de chromosomes en conditions de stress.** La relocalisation de TCs a été mise en évidence dans des conditions de déprivation de sérum (Mehta *et al.* (2010)). Contrairement aux autres exemples discutés dans cette section, l'approche expérimentale n'implique pas d'imagerie sur cellules vivantes, mais une analyse indirecte sur une population de cellules fixées avant et après l'application du stress qui permet d'inférer la dynamique des chromosomes. Les auteurs observent une relocalisation rapide des chromosomes après seulement 15 min de déprivation de sérum. Certains chromosomes sont relocalisés vers l'intérieur du noyau, alors que d'autres sont relocalisés vers la périphérie. Ce repositionnement n'est ni observé dans un milieu sans ATP, ni suite à la déplétion de la myosine nucléaire  $1\beta$  ; ceci suggère qu'il s'agit d'un mouvement actif (Mehta *et al.* (2010)). Les mêmes auteurs ont complété ce travail par une étude dans des cellules HGPS, dans lesquelles ils observent que cette relocalisation est altérée, et restaurée après traitement par des FTIs (Mehta *et al.* (2011)) ; ce qui suggère que la lamina est impliquée dans ce repositionnement.

**Réorganisations au cours de la différenciation.** De nombreux travaux portant sur les cellules ES suggèrent que la chromatine des cellules ES est dans un état de plasticité plus important que dans les cellules différenciées. Par exemple, dans les cellules ES, l'association des protéines structurales (notamment HP1 $\alpha$ , H<sub>1</sub> et les histones de cœur) est moins stable (Meshorer *et al.* (2006)) ; la transcription intergénomique est plus répandue (Efroni *et al.* (2008)) ; le niveau global de certaines modifications d'histones varie ; et l'état transcriptionnel bivalent de certains domaines chromatiniens est résolu en état répressif ou permissif au cours de la différenciation (voir §A.1.1.3). Cet état de plasticité est certainement à mettre en relation avec

---

1. En effet, les ARN polymérase font partie des moteurs cellulaires les plus puissants, capables de développer des forces bien supérieures à celles que développent les myosines/kynesines (Marko et Poirier (2003)).

la pluripotence des cellules ES, cette plasticité étant perdue à mesure que les cellules s'engagent dans une voie de différenciation et perdent leur pluripotence (pour revue : [Mattout et Meshorer \(2010\)](#)). Ces changements sont accompagnés de modifications morphologiques du noyau, notamment dues à l'expression de la lamine A dans certains types différenciés (voir §A.1.2.2) ainsi que de modifications dans l'organisation nucléaire. Certaines observations indiquent que plusieurs compartiments nucléaires, tels que la lamina, le nucléole, les foyers d'hétérochromatine (voir §A.3.2.1), ou les speckles, subissent des modifications morphologiques au cours de la différenciation ([Meshorer et Misteli \(2006\)](#)). De plus, le positionnement radial des TCs varie aussi dans certaines voies de différenciation. Un des facteurs qui influence grandement le positionnement radial des TCs est la forme du noyau, facteur dépendant notamment de l'expression de la lamine A. Certains gènes (ou chromosomes) peuvent être repositionnés lorsqu'ils sont impliqués dans la voie de différenciation (ou contiennent des gènes qui le sont) (pour revue : [Meshorer et Misteli \(2006\)](#)).

Les régions centromériques sont aussi repositionnées au cours de la différenciation (voir §A.3.1). De nombreuses études ont montré que des variations de la topologie du génome sont étroitement liées au processus de différenciation et à la régulation de gènes maîtres du processus de différenciation. La différenciation et l'activation des lymphocytes a notamment été très étudiée (pour revue : [Hübner et Spector \(2010\)](#)).

Sachant que le positionnement de la chromatine est globalement assez stable ([Walter et al. \(2003\)](#)), de telles réorganisations globales sont probablement souvent dirigées par des réarrangements de régions chromatiniennes lorsque la cellule passe par la mitose. A l'appui de cette idée, plusieurs études ont montré que le positionnement des chromosomes n'est pas hérité de la cellule mère mais plutôt établi en début de phase G<sub>1</sub> ([Thomson et al. \(2004\)](#)). D'un autre côté, l'exemple du paragraphe précédent montre que des réorganisations importantes peuvent avoir lieu dans des temps courts, donc sans passage par la mitose.

## Conclusion

Le noyau est une structure qui présente plusieurs formes de dynamique. On observe des mouvements aléatoires de la chromatine sur de courtes distances, mais aussi des mouvements dirigés sur de plus longues distances. Il est tentant de supposer que cette mobilité confère à la chromatine la possibilité d'explorer le noyau afin de trouver un micro-environnement particulier. D'autre part on observe des réorganisations plus globales du noyau, en réponse à des stimuli (par exemple stress ou différenciation) qui nécessitent parfois un passage par la mitose. Ces réorganisations permettent certainement au génome d'adapter sa réponse à une situation particulière. Une des questions fondamentales qui se pose est celle de la relation de causalité entre la forme (la conformation spatiale du génome) et la fonction (expression des gènes). Nous allons aborder cette question dans la partie suivante en présentant des aspects de modélisation du noyau.

### A.1.5 Modélisation de l'architecture nucléaire

Plusieurs caractéristiques fondamentales du noyau ressortent de cette description. Le noyau est tout d'abord une structure ordonnée ; il est composé de compartiments et l'organisation spatiale du génome présente des régularités statistiques à l'échelle de la population de noyaux. Son organisation est aussi dynamique ; elle varie à plusieurs échelles temporelles, et possède un certain degré de stochasticité.

La structure du noyau résulte de nombreuses interactions dynamiques entre des molécules ou assemblages de molécules qui ont lieu à plusieurs échelles spatiales. Une des difficultés majeures réside dans la définition d'un cadre conceptuel permettant à la fois d'intégrer ces différentes échelles et de rendre compte des caractéristiques structurales, stochastiques et dynamiques du noyau. Une des hypothèses les plus abouties permettant de rendre compte des caractéristiques de structure du noyau est l'hypothèse selon laquelle l'organisation nucléaire serait dirigée par des processus d'auto-organisation (Dundr et Misteli (2001)).

#### A.1.5.1 L'auto-organisation dans le noyau

L'auto-organisation est un phénomène de mise en ordre croissant qui permet de rendre compte de l'émergence d'une structure ordonnée à une échelle donnée, uniquement à partir d'un très grand nombre d'interactions locales entre composants d'une échelle inférieure. La structure est une propriété émergente du système plutôt qu'une propriété imposée sur le système par une influence extérieure (Ashby (1947)). Si la structure globale ne contrôle pas les interactions locales, elle peut cependant influencer rétroactivement les composants locaux. Les changements locaux vont à leur tour influencer la structure globale. Ainsi, une augmentation de l'ordre global est atteinte progressivement par des rétroactions allant du local vers le global et inversement (Rajapakse et Groudine (2011)). Ce cadre conceptuel permet d'expliquer comment des configurations stables peuvent être atteintes à partir d'interactions locales dynamiques et transitoires.

Dans le noyau, les protéines sont extrêmement mobiles ; pourtant elles participent à la formation de structures stables<sup>1</sup> (Dundr et Misteli (2001)) :

- La composition des complexes protéiques semblent globalement stable, mais chaque composant peut être périodiquement échangé.
- La chromatine est formée d'interactions transitoires avec des protéines.
- Les compartiments nucléaires semblent avoir une composition stable, mais celle-ci n'est pas maintenue par une membrane : elle résulte d'interactions transitoires entre des protéines.

Il existe de nombreux exemples de structures nucléaires dont l'assemblage résulte d'un processus d'auto-organisation ; il a par exemple été montré que les corps de Cajal sont des structures auto-organisées, pouvant se former *de novo* plutôt que résultant d'un assemblage hiérarchique (Kaiser *et al.* (2008)).

---

1. Ce principe peut aussi s'appliquer pour d'autres molécules, tels que des ARN lorsqu'ils qui sont des composants structuraux de complexes, de la chromatine ou de compartiments.

Le positionnement spatial du génome peut lui aussi s'expliquer par des propriétés d'auto-organisation. Chaque région du génome est soumise à différents types d'interactions, soit avec des compartiments nucléaires (tels que la lamina ou le nucléole) soit avec d'autres régions génomiques (voir Fig. A.12). Ces interactions dépendent notamment de l'état chromatinien de la région : des gènes actifs peuvent co-localiser dans des usines à transcription, les régions d'hétérochromatine ont tendance à s'agréger dans le noyau, etc. La somme de ces interactions influence la conformation spatiale du génome et contraint ses mouvements. Ceci permet de rendre compte à la fois du caractère stochastique et de la présence de structures stables et de profils d'organisation préférentiels. Le caractère dynamique et transitoire des interactions locales sous-tend la structure globale émergente, ainsi que la plasticité et la dynamique du génome à une échelle globale.

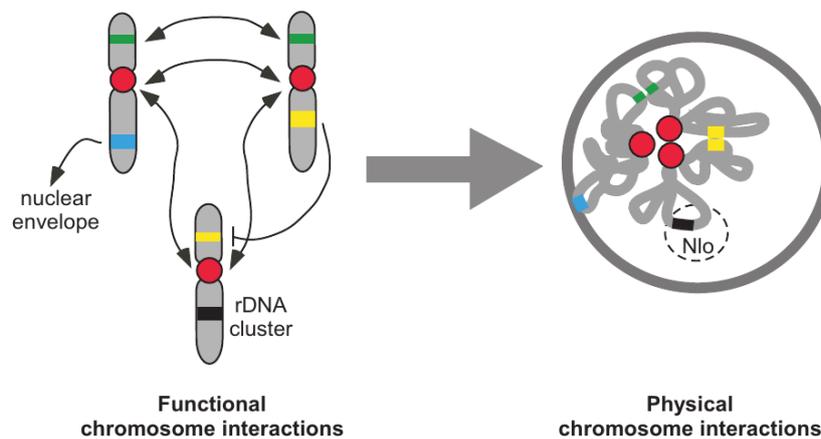


FIGURE A.12: Modèle d'auto-organisation du génome (d'après [Misteli \(2009\)](#)). Les interactions fonctionnelles au niveau de différentes régions des chromosomes peuvent influencer le pattern de localisation des chromosomes. Parmi ces interactions, citons les associations de centromères (rouge) et de rDNA (noir), les associations de gènes co-régulés dans les usines à transcription (vert), les associations d'éléments régulateurs avec leur gènes cibles (jaune), l'interaction d'une région génomique avec des compartiments tels que l'enveloppe nucléaire (bleu). Chaque interaction contraint la motilité du chromosome, et la somme de l'ensemble des interactions influence la position d'un chromosome par rapport aux autres chromosomes et par rapport à des structures nucléaires telles que la périphérie du noyau ou les nucléoles (Nlo).

Le modèle d'auto-organisation est souvent opposé à un modèle dit *déterministe* ([Misteli \(2007\)](#)) dans lequel le génome est organisé par des interactions spécifiques avec des structures stables, telles que la lamina ou un hypothétique nucléosquelette formé notamment d'actine nucléaire. La présence de structures stables ne nous semble pas en opposition avec le modèle d'auto-organisation ; il est très probable que ces structures dites stables résultent aussi de processus d'auto-organisation, la nature des interactions locales leur conférant une stabilité relative plus importante que d'autres composants nucléaires. Ces structures plus stables pourraient servir de plates-formes d'assemblage ou de points d'ancrage à des structures plus dynamiques.

### A.1.5.2 Liens entre structure et fonction.

De plus en plus de résultats expérimentaux mettent en évidence des situations dans lesquelles des modifications de l'organisation du noyau sont corrélées à des modifications de l'expression des gènes ; cependant, il est difficile d'établir une causalité entre ces deux types de phénomènes.

Plusieurs exemples ont montrés que l'organisation spatiale de certains gènes module leur expression, la preuve la plus forte provenant d'expériences dans lesquelles l'induction du repositionnement de gènes de l'intérieur du noyau vers la lamina a pour conséquence leur répression (voir §A.1.3.4). D'autre part les profils d'expression des gènes semblent moduler l'organisation spatiale du noyau, comme le suggère le fait que le génome est compartimenté physiquement en domaines actifs et inactifs (voir §A.1.1.4). Ces principes concernent aussi les compartiments nucléaires, comme le montre le fait que l'intégrité morphologique du nucléole dépend strictement de son activité transcriptionnelle : l'inhibition de la transcription désassemble le nucléole, et l'assemblage du nucléole dépend de l'activation de la transcription des séquences NORs à la fin de la mitose (Dundr et Misteli (2001)). Il y a donc un lien réciproque entre la structure et la fonction du noyau dans certaines situations. Une tentative a été faite pour montrer une relation entre les réseaux dynamiques de régulation de gènes et leurs interactions physiques dans le noyau à une échelle globale (Rajapakse *et al.* (2009)). Cette étude montre que les variations dans les profils d'expression sont corrélées aux variations d'arrangement spatial des chromosomes dans un modèle de différenciation de cellules murines. Cela pose la question de savoir si les variations de structure précèdent les variations de fonction, ou inversement si les variations de structure sont le reflet des variations de fonction au cours de la différenciation. Cette question se rapproche de celle de la poule et de l'oeuf ; il est très probable qu'à partir d'au moins un certain point, cette interaction ait lieu à double sens, structure et fonction se régulant mutuellement progressivement au cours du temps (Rajapakse et Groudine (2011)).

Le fait que la topologie du génome ait plutôt l'effet de moduler que de déterminer l'expression des gènes a conduit à formuler un modèle dit *Fonction-Structure-Fonction*, dans lequel l'activité du génome<sup>1</sup> influence d'abord sa topologie, puis est à son tour modulée par cette topologie (Fig. A.13). Deux principes fondamentaux sous-tendent ce modèle, d'une part l'*auto-renforcement* qui résulte de l'interaction structure/fonction, qui a pour effet de stabiliser des états chromatiniens, et d'autre part l'*auto-propagation* de ces états au cours des divisions cellulaires (Cavalli et Misteli (2013)).

---

1. L'activité du génome étant encodée dans la structure chromatinienne à travers la mitose, il est légitime de se poser la question si ce modèle ne pourrait pas plutôt s'appeler *Structure-Fonction-Structure-Fonction*, la première occurrence du terme *Structure* se référant plutôt à la structure chromatinienne et la suivante à tous les niveaux structuraux, y compris la topologie du génome.

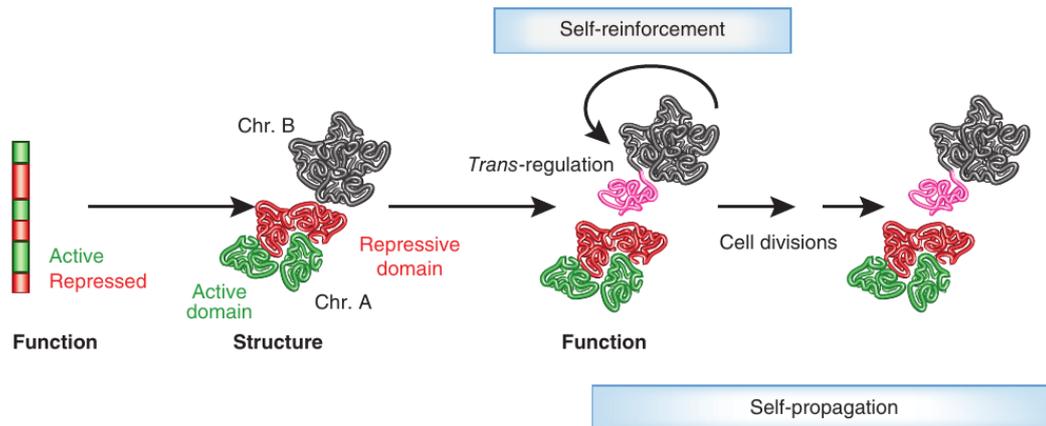


FIGURE A.13: Modèle décrivant les interactions entre structure et fonction (d'après Cavalli et Misteli (2013)). L'activité transcriptionnelle des régions génomiques, encodée par des modifications épigénétiques sur le chromosome, détermine la formation de domaines chromatiniens dans l'espace du noyau (vert et rouge). Ces domaines génèrent des micro-environnements capables de modifier l'activité de domaines se trouvant à proximité spatiale (indiqué par la transformation d'une portion de chromatine noire en rose, lorsqu'elle se trouve à proximité d'un domaine répressif (rouge)). Le statut fonctionnel du domaine et sa structure se renforcent mutuellement (*self-reinforcement*), par exemple par l'accumulation locale de facteurs répressifs. Ces modifications de statut transcriptionnel se propagent au cours des divisions cellulaires (*self-propagation*) par exemple par le biais des marques épigénétiques.

## Conclusion

Dans cette partie, nous avons montré que la chromatine est organisée dans le noyau. Son repliement est hiérarchique (nucléofilament, fibre de 30 nm, domaine chromatiniens et TC). Les différents niveaux sont inter-dépendants, et liés à la régulation de l'expression des gènes.

De plus le noyau contient de nombreux compartiments nucléaires associés à des fonctions précises. Le positionnement du génome au sein de l'espace nucléaire n'est pas aléatoire, et le positionnement de régions génomiques par rapport à des compartiments nucléaires a une importance fonctionnelle : certains compartiments constituent des micro-environnements actifs ou répressifs pour la transcription qui influencent l'état transcriptionnel des loci se trouvant à proximité. Ceci mène à considérer l'organisation nucléaire comme un troisième niveau de régulation de l'expression des gènes, qui s'ajoute aux niveaux de régulations génétique et épigénétiques, et qui intervient plutôt en modulant l'expression qu'en la déterminant<sup>1</sup>.

Nous avons volontairement laissé de côté les régions centromériques, afin de décrire plus spécifiquement leur organisation spatiale et leur rôle dans l'architecture nucléaire.

1. Nous proposons que ces trois niveaux soient associés à des informations qui pourraient être comparées par leur *persistance temporelle* et leur *susceptibilité* : une modification de la séquence d'ADN est en général plutôt rare, mais peut se propager à des échelles dépassant celle de l'organisme ; une modification épigénétique est beaucoup plus fréquente, se propage d'une cellule mère à une cellule fille, mais aussi (plus rarement) d'un individu à un autre ; et une modification de la topologie du génome est très fréquente mais ne se propage pas directement d'une cellule mère à une cellule fille (aucune preuve expérimentale de ce type de propagation n'existe à notre connaissance).

## A.2 Structure et fonction des régions centromériques

Le centromère est un domaine chromosomique spécialisé sur lequel s'assemble le kinétochore, complexe protéique couplant le fuseau mitotique au chromosome au cours de la mitose. Le centromère joue donc un rôle crucial dans la ségrégation des chromatides sœurs (Cheeseman et Desai (2008)). Sur le chromosome mitotique, le centromère correspond à la constriction primaire du chromosome. Les régions flanquantes du centromère, nommées régions péricentromériques, sont impliquées dans la cohésion des chromatides sœurs. Nous utilisons le terme de région centromérique (RC) pour désigner le domaine constitué du centromère et des régions péricentromériques.

Après avoir décrit les caractéristiques des RCs qui sont partagées entre les différentes espèces (§A.2.1), nous décrirons plus précisément la structure chromatidienne des RCs (§A.2.2) ainsi que les séquences d'ADN sous-jacentes chez l'homme (§A.2.3), puis exposerons les liens connus entre structure et fonction des RCs (§A.2.4).

### A.2.1 Généralités sur les régions centromériques

#### A.2.1.1 Séquences d'ADN

Les séquences des RCs sont très divergentes entre les espèces. Le centromère le plus simple est trouvé chez la levure bourgeonnante *Saccharomyces cerevisiae*, constitué d'un domaine de 125 pb, suffisant pour faciliter l'assemblage du kinétochore. Contrairement à ce centromère ponctuel, la plupart des eucaryotes (y compris la levure à fission *Schizosaccharomyces pombe*) possèdent des centromères régionaux<sup>1</sup> (Pluta *et al.* (1995)). Ces régions sont composées d'ADN répété en tandem ; la séquence et la taille de l'unité de répétition ainsi que le nombre de répétitions sont variables entre les espèces. On observe des tailles de centromères allant de 35-110 Kb chez *S. pombe* jusqu'à 0.3-5 Mb chez l'homme (voir Fig. A.14). On trouve parfois plusieurs familles de séquences sur les différents chromosomes, voire sur un même chromosome. Chez la souris, seules deux familles principales de répétitions ont été décrites : les satellites mineurs de 120 pb se trouvant au niveau des centromères, et les satellites majeurs de 234 pb au niveau des péricentromères. Ces séquences sont présentes sur tous les chromosomes à l'exception du chromosome Y (Pardue et Gall (1970)). Chez l'homme on trouve principalement les  $\alpha$ -satellites de 171 pb, ainsi que d'autres familles de séquences qui peuvent différer entre les chromosomes (voir §A.2.3).

Les connaissances dont nous disposons sur la séquence des RCs pour des espèces où elles sont très étendues sont limitées ; leur assemblage est en effet rendu très dif-

---

1. Une exception aux centromères ponctuels ou régionaux est la formation de kinétochores *holocentriques* tout le long des chromatides chez *Caenorhabditis elegans* (pour revue : Dernburg (2001)). .

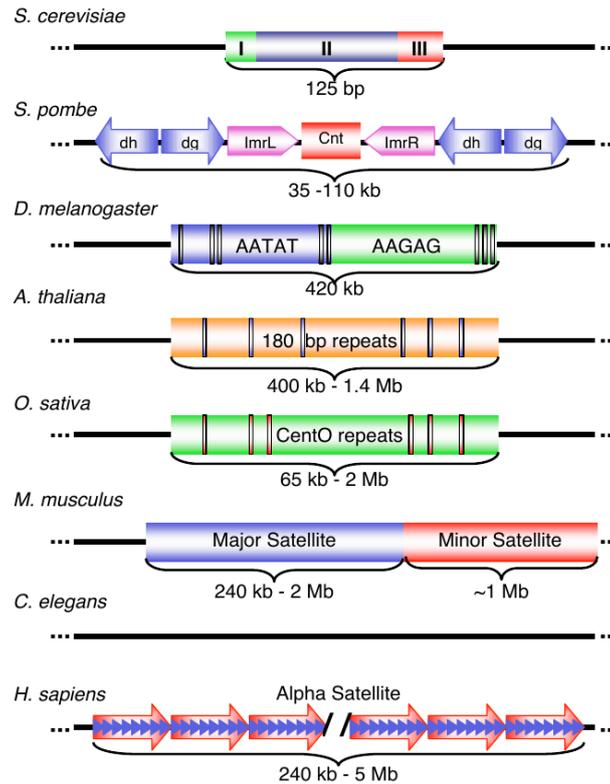


FIGURE A.14: Organisation des RCs entre les espèces (d'après Rudd (2005)). Les RCs de *S. cerevisiae* sont composées de trois domaines. Les RCs de *S. pombe* sont composées d'un noyau central unique (Cnt) flanquées de répétitions (ImrL, ImrR, dg, dh). Les RCs de la drosophile possèdent deux domaines satellite entrecoupés d'éléments transposables. Des répétitions de 180 pb (respectivement de CentO) entrecoupées d'éléments transposables constituent les RCs d'*arabidopsis* (respectivement du riz). Les RCs murines sont composées de répétitions de satellites majeurs et mineurs. Le centromère de *C. Elegans* est holocentrique. Les RCs humaines sont principalement composées de répétitions  $\alpha$ -satellites (la structure est détaillée au §A.2.3).

ficile par le degré élevé d'homogénéité de séquence entre les centaines de milliers de copies d'une séquence répétée donnée (Rudd et Willard (2004)). Les cartes des génomes présentent donc souvent un intervalle non assemblé sur chaque chromosome qui correspond à la RC, appelé *centromeric gap* en anglais. Néanmoins, de nombreuses expériences ont permis de rassembler des données sur les séquences des RCs humaines.

### A.2.1.2 Chromatine

Les RCs constituent, avec les télomères, l'hétérochromatine constitutive caractérisée par des marques d'histones répressives telles que H<sub>3</sub>K<sub>9</sub>me<sub>3</sub>, H<sub>3</sub>K<sub>27</sub>me<sub>3</sub> et H<sub>4</sub>K<sub>20</sub>me<sub>3</sub>, et par un fort taux de méthylation de l'ADN (Bestor (2005)) ainsi que par leur réplication tardive. Les centromères de toutes les espèces sont caractérisés par la présence du variant de l'histone H<sub>3</sub> CenpA (parfois nommé CenH3), très conservé et qui se trouve exclusivement au niveau de ces centromères. Cette dernière caractéristique est nécessaire et suffisante à la formation du centromère;

c'est pourquoi elle est considérée comme définissant le centromère, à défaut d'un consensus de séquences d'ADN.

## A.2.2 Structure chromatinienne des régions centromériques

### A.2.2.1 Centromères

Les centromères sont caractérisés par la présence du variant de l'histone H<sub>3</sub> CenpA. CenpA et H<sub>3</sub> diffèrent dans leur partie N-terminale, où sont apposées les modifications post-traductionnelles. La seule modification épigénétique décrite pour CenpA est la phosphorylation de la sérine 7 par Aurora B (Zeitlin *et al.* (2001)). CenpA possède aussi un domaine d'adressage spécifique nommé CATD (CenpA Targeting Domain), qui est nécessaire et suffisant pour son assemblage au niveau des centromères (Black *et al.* (2004)), ce notamment en interagissant spécifiquement avec la chaperone HJURP (Dunleavy *et al.* (2009)). Lors de la réplication, CenpA est diluée entre les brins néo-synthétisés et les brins matrices. Les histones H<sub>3,1</sub> et H<sub>3,3</sub> sont alors assemblées puis remplacées en début de phase G<sub>1</sub> par CenpA (Dunleavy *et al.* (2011)).

On observe une organisation particulière des marques épigénétiques au niveau des centromères : on trouve des domaines enrichis en la marque permissive H<sub>3</sub>K<sub>4</sub>me<sub>2</sub> en alternance sur le centromère avec des domaines enrichis en CenpA (Sullivan et Karpen (2004))(Fig. A.15a). La marque H<sub>3</sub>K<sub>4</sub>me<sub>2</sub> semble être directement impliquée dans l'assemblage de CenpA (Bergmann *et al.* (2011)), et son alternance avec des domaines CenpA pourrait jouer un rôle dans la structure secondaire du centromère (Fig. A.15b/c). La marque CenpA peut couvrir des régions de l'ordre de 2.5 Mb, mais cette taille varie beaucoup entre les chromosomes et les individus. Il semble que la taille du domaine CenpA soit proportionnelle à la taille du domaine couvert par les répétitions d'ordre supérieur (HOR : Higher-Order Repeat, voir plus bas) du chromosome (Sullivan *et al.* (2011)).

### A.2.2.2 Péricentromères

La chromatine péricentromérique possède des marques distinctes de la chromatine centromérique. En particulier, les marques CenpA et H<sub>3</sub>K<sub>4</sub>me<sub>2</sub> n'y sont pas détectables, et on observe un fort enrichissement dans la marque répressive H<sub>3</sub>K<sub>9</sub>me<sub>3</sub> (Sullivan et Karpen (2004), Fig. A.15a). Cette marque est impliquée dans le recrutement de la protéine HP1 (§A.1.1.1) ainsi que les protéines des complexes de cohésion.

## A.2.3 Séquences d'ADN centromérique humaines

### A.2.3.1 Séquences satellites

On trouve plusieurs familles de séquences satellites au niveau des RCs humaines. La famille majoritaire est constituée des séquences  $\alpha$ -satellites, répétées sur des

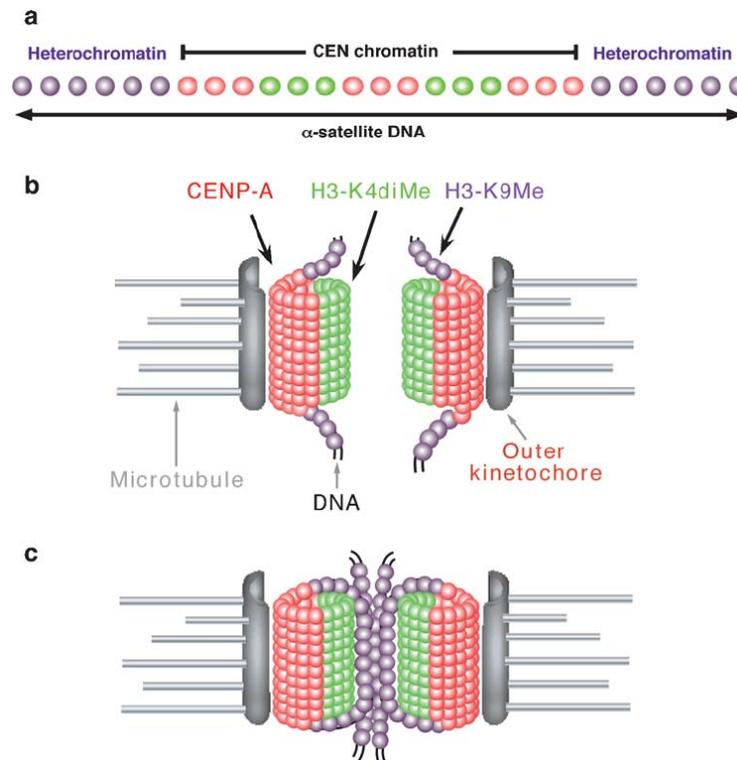


FIGURE A.15: Modèle de structure secondaire de la chromatine centromérique (d'après Schueler et Sullivan (2006)). L'association entre domaines enrichis en les mêmes marques épigénétiques pourrait contribuer à la formation de structures 3D distinctes entre les centromères et les péricentromères. (a) Structure linéaire de la chromatine au niveau des RCs. Au niveau des centromères, des domaines CenpA (rouge) et H<sub>3</sub>K<sub>4</sub>me<sub>2</sub> (vert) sont présents en alternance. Au niveau des péricentromères on trouve principalement la marque H<sub>3</sub>K<sub>9</sub>me<sub>3</sub> (violet). (b) L'alternance de domaines CenpA et de domaines H<sub>3</sub>K<sub>4</sub>me<sub>2</sub> pourrait former une structure 3D cylindrique observée dans les chromosomes métaphasiques. Les domaines CenpA seraient orientés du côté polaire, facilitant ainsi le recrutement des protéines du kinétochore externe, et les domaines H<sub>3</sub>K<sub>4</sub>me<sub>2</sub> orientés entre les chromatines soeurs (Blower *et al.* (2002)). (c) H<sub>3</sub>K<sub>9</sub>me<sub>3</sub> qui recrute des protéines de l'hétérochromatine telles que HP1 et des cohésines, est présente à la fois dans la zone du kinétochore interne entre les chromatides soeurs et dans les régions qui flanquent les centromères. Cette organisation pourrait d'une part orienter la chromatine CenpA du côté polaire du chromosome mitotique, et d'autre part promouvoir l'interaction d'HP1 avec elle-même, ainsi que la condensation et la cohésion appropriées de la chromatine.

longueurs pouvant atteindre plusieurs Mb (Manuelidis (1978)); à celles-ci s'ajoutent les  $\beta$ -/ $\gamma$ -satellites et les satellites I, II et III (Lee *et al.* (1997)).

Au sein de chaque RC, les unités répétées d' $\alpha$ -satellites peuvent être distingués en deux types : celles qui sont très divergentes et qui présentent peu d'homologies locales, appelés monomériques, et celles qui sont organisées en unités de répétitions multi-monomériques appelées HOR; elles se distinguent par leur arrangement en monomères d' $\alpha$ -satellites au sein de chaque unité de répétition. Les HORs sont répétées en tandem, et présentent une très forte homologie au sein d'un même ensemble de répétitions (typiquement 97-100%) (Alexandrov *et al.* (1993)); alors qu'il n'y a qu'environ 60-80% d'homologie en moyenne entre deux monomères. On trouve au niveau de chaque RC humaine une ou plusieurs HORs répétées sur des longueurs de l'ordre du Mb (Fig. A.16), et le variant d'histone CenpA est assemblé

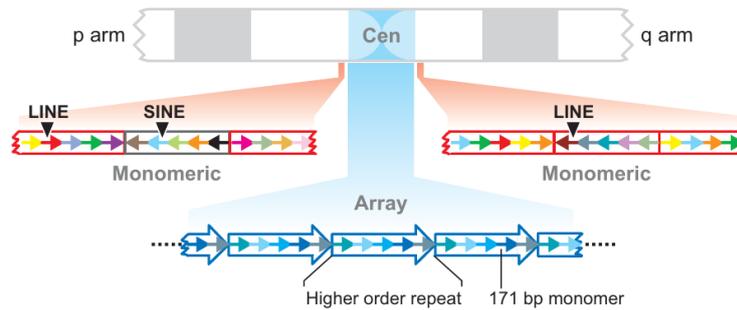


FIGURE A.16: Modèle d'organisation des séquences  $\alpha$ -satellites au niveau des RCs (d'après Schueler et Sullivan (2006)). Chaque petite flèche représente un monomère satellite simple. Au niveau des centromères, les  $\alpha$ -satellites sont organisés en HORs (grandes flèches bleues), qui sont des unités de répétitions composées de plusieurs monomères (5 dans le cas du schéma) divergents (la divergence est indiquée par des têtes de flèches de couleurs différentes). Au niveau des régions péricentromériques, les  $\alpha$ -satellites sont présents sous forme monomérique. On trouve des blocs de monomères de la même famille (indiqués par des boîtes rouges ou grises) qui contiennent parfois des rétrotransposons (LINEs ou SINEs). Des blocs de satellites adjacents peuvent avoir des orientations différentes. Les monomères sont très divergents entre eux, alors que les HORs sont quasiment identiques. D'autres séquences sont aussi présentes au niveau des régions péricentromériques, présentées dans la Table A.1.

sur l'une d'entre elles pour former le centromère (Alexandrov *et al.* (2001)).

### A.2.3.2 Différences entre les chromosomes

Satellite	Longueur	1	3	4	8	9	13	14	15	16	21	22	X	Y	autres
$\alpha^*$	171	+	+	+	+	+	+	+	+	+	+	+	+	+	+
$\beta$	68	+				+	+	+	+		+	+		+	
$\gamma$	220				+								+	+	
I	42		+	+			+	+	+		+	+			
II	5	+	+	+	+	+	+	+	+	+	+	+	+	+	+
III	5	++	+	+	+	++	+	+	+	++	+	+	+	++	+

TABLE A.1: Répartition des principales séquences satellites qui pourraient être présentes au niveau des régions péricentromériques sur les différents chromosomes humains. Ces données proviennent de Lee *et al.* (1997) et de données compilées de la littérature. Les séquences n'étant pas assemblées, il est impossible de connaître précisément la composition de chaque régions péricentromérique.

La longueur des séquences est exprimée en pb.

++ indique que les séquences sont très abondantes sur le chromosome. \*Les  $\alpha$ -satellites sont présents sous forme monomérique.

Chez l'homme, des variations dans la séquence des RCs sont observées entre les différents chromosomes. D'une part, les HORs se distinguent en général par la composition des monomères et le nombre de répétitions ; d'autre part, les autres familles de satellites ne sont pas réparties de manière homogène sur tous les chromosomes (voir Table A.1, et pour revue : Lee *et al.* (1997)). Notamment, les péricentromères des chromosomes 1, 9, 16 et Y possèdent de grands domaines enrichis en satellites III en position q12.

Ces différences peuvent permettre de distinguer les RCs portées par différents chromosomes dans des expériences de FISH (voir §A.4.1.2). Le caractère répété de ces séquences permet l'utilisation de sondes oligonucléotidiques. En effet, chaque sonde porte un (ou plusieurs) fluorophores et le grand nombre de répétitions permet de concentrer suffisamment de fluorophores pour détecter la région. L'incorporation de modifications LNA (Locked Nucleid Acid, Fig. A.17) dans un oligonucléotide permet d'augmenter son affinité pour sa séquence cible, et donc de le raccourcir. Ceci peut augmenter sa spécificité, certainement parce qu'un mésappariement déstabilise plus une séquence courte qu'une séquence longue (celui-ci représentant une fraction plus importante de l'énergie d'interaction d'une séquence courte).

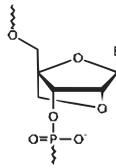


FIGURE A.17: Structure du LNA. Le LNA est un analogue d'acide nucléique contenant un pont méthylène entre l'hydroxyle en position 2 et l'atome de carbone 4 du sucre. Ceci lui confère une plus forte affinité pour l'ADN (Vester et Wengel (2004)).

Il existe cependant des RCs qui présentent beaucoup de similarités : c'est le cas des chromosomes 1/5/19, 13/21 et 14/22 (Hayden (2012)). Ces RCs sont difficilement distinguables dans des expériences de FISH et sont souvent détectées simultanément (Alcobia *et al.* (2000)). Au laboratoire, nous avons conçu des sondes oligonucléotidiques comportant des modifications LNA, qui permettent de détecter spécifiquement des RCs des ces chromosomes. La similitude des RCs de ces chromosomes reste inexpliquée ; il est possible que ces séquences aient une fonction précise et qu'une pression de sélection ait fait converger ces séquences. De manière alternative, sachant que ces séquences possèdent des mécanismes d'évolution particuliers (voir §A.2.3.3), il est possible que ces similitudes soient dues à des mécanismes d'homogénéisation de séquence, faisant par exemple intervenir la proximité spatiale de ces RCs<sup>1</sup>.

### A.2.3.3 Mécanismes d'évolution des séquences centromériques.

Le haut degré d'identité et le grand nombre de répétitions des HORs est très certainement maintenu par un processus d'évolution actif, conséquence du phénomène de dérive moléculaire (*molecular drive* en anglais). Il s'agit d'un processus qui se déroule à deux niveaux : les mutations sont homogénéisées au sein des membres d'une famille d'ADN répété, et cette famille est parallèlement stabilisée dans un groupe d'individus inter-féconds (Dover (1986)). Cela a pour conséquence une évolution concertée des monomères qui constituant une famille d'ADN satellite. Il existe plusieurs mécanismes de transferts non réciproques pouvant rendre compte

1. En effet, les chromosomes 13, 14, 21 et 22 portent les NORs et pourraient avoir tendance à se retrouver en contact au niveau des nucléoles.

du processus d'homogénéisation : les *crossing over* inégaux, la conversion génique, la réplication en cercle roulant (*rolling circle replication* en anglais) suivie de réinsertion, ainsi que des échanges médiés par des transposons (pour revue : Plohl *et al.* (2008)).

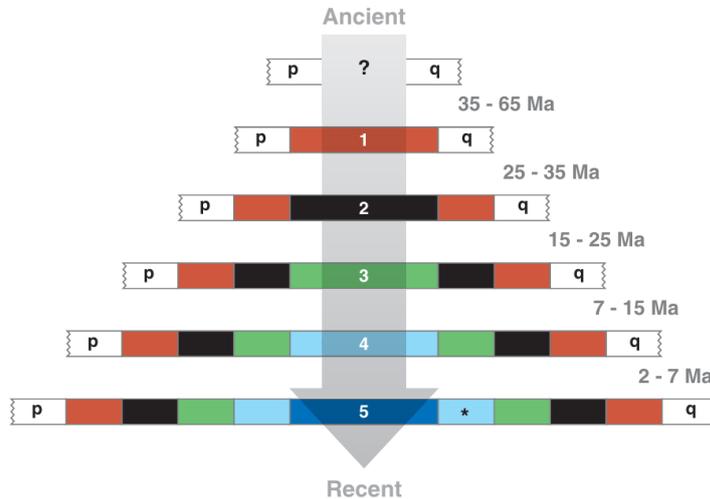


FIGURE A.18: Modèle d'évolution des RCs par expansion (d'après Schueler et Sullivan (2006)). Les additions successives (rectangles colorés) au centromère durant l'évolution des primates commence en haut du schéma avec un chromosome ancestral avant l'émergence des  $\alpha$ -satellites, et continue vers le bas jusqu'au centromère humain actuel. A chaque addition de nouveau bloc, les anciennes séquences centromériques sont poussées vers les extrémités du chromosome. Une astérisque indique la région d' $\alpha$ -satellites monomériques avec les sites de liaison pour CenpB. Le schéma est basé sur les données du centromère du chromosome X. Les dates approximatives en millions d'années (Ma : Millions of Years Ago) proviennent de l'arbre phylogénétique des primates et de la phylogénie des LINEs présents dans les RCs.

L'analyse phylogénétique des différents monomères le long des centromères permet d'expliquer l'organisation des séquences satellites au sein d'une RC. De manière générale, les  $\alpha$ -satellites monomériques sont plus anciens que les HORs, et on les retrouve chez la plupart des autres primates (Alexandrov *et al.* (2001)). De plus, la répartition des divergences et la distribution des transposons suggèrent que les domaines anciens ont été homogènes chez un ancêtre commun, mais que le processus d'homogénéisation s'est arrêté ; ce qui a permis l'accumulation de mutations et d'éléments transposables. Cela suggère que les anciennes répétitions d' $\alpha$ -satellites sont les restes des centromères de nos ancêtres primates, qui ont été actifs et homogènes puis obsolètes (Kazakov *et al.* (2003)). L'analyse récente de l'organisation des RCs des chromosomes X, 8 et 17 humains et la comparaison des séquences avec celles d'autres primates ont montré qu'il existe des blocs de satellites divergents, définis par une analyse phylogénétique, qui semblent s'être progressivement éloignés des séquences sur lesquelles s'assemble le centromère, et ce de manière assez symétrique vers les deux bras du chromosome. Cela suggère que les séquences des RCs évoluent par expansion depuis le centromère (voir Fig. A.18) et qu'il existe une pression sélective forte au niveau des séquences centromériques, certainement imposée par le fonctionnement du centromère, ainsi qu'une pression moins forte imposée par le fonctionnement du péricentromère au niveau des séquences péricentromériques

(Shepelev *et al.* (2009)).

## A.2.4 Lien entre structure et fonction

### A.2.4.1 Le paradoxe du centromère

La rôle joué par le centromère dans l'assemblage du kinétochore et la division cellulaire est très conservé au cours de l'évolution, alors que la séquence d'ADN est très divergente entre les espèces. Cela semble paradoxal car en général, la conservation de la fonction d'un locus au cours de l'évolution est corrélée avec la conservation de sa séquence. Cette singularité est connue sous le nom de *paradoxe du centromère*. En revanche deux caractéristiques sont conservées au cours de l'évolution : d'une part, on trouve un variant spécifique de l'histone H<sub>3</sub> (CenpA chez l'homme) au niveau de tous les centromères. D'autre part, le caractère répété et riche en A/T des séquences est présent chez quasiment tous les eucaryotes. Quelle est donc l'importance fonctionnelle de ces deux caractéristiques de structure ?

### A.2.4.2 Structure chromatinienne et fonction du centromère

CenpA est la seule protéine centromérique totalement conservée au cours de l'évolution. Des expériences de délétion ont montré que CenpA est nécessaire à l'assemblage du kinétochore et à la bonne ségrégation des chromatides sœurs au cours de la division cellulaire (Howman *et al.* (2000)). Mais il semble que la présence de CenpA ne soit pas suffisante à la formation d'un centromère fonctionnel. En effet, lorsque CenpA est sur-exprimée, elle se localise de manière ectopique sur de la chromatine non centromérique et recrute CenpC ainsi que d'autres protéines du kinétochore ; cependant, cela n'est pas suffisant pour créer des centromères ectopiques fonctionnels (Van Hooser *et al.* (2001)). Il a été proposé que la structure secondaire des régions enrichies en CenpA (Fig. A.15) est importante pour le recrutement du kinétochore et des complexes de cohésion des chromatides sœurs tout en assurant la séparation spatiale des domaines de cohésion et du kinétochore (Schueler et Sullivan (2006)).

Au niveau des péri-centromères, la marque H<sub>3</sub>K<sub>9</sub>me<sub>3</sub> interagit avec des complexes de cohésines, ce qui permet la cohésion des chromatides sœurs. Cette cohésion est un phénomène essentiel pour l'attachement correct des chromosomes aux microtubules (Amor *et al.* (2004)). De plus l'hétérochromatine péri-centromérique forme un compartiment distinct de l'hétérochromatine centromérique, et pourrait être impliquée dans la structure secondaire du centromère (Fig. A.15c). Chez l'homme, des cellules tumorales présentant une forte instabilité chromosomique ont souvent une diminution de la quantité de marque H<sub>3</sub>K<sub>9</sub>me<sub>3</sub> ; de plus, des cellules dans lesquelles la quantité de cette marque est diminuée par la surexpression de l'histone déméthylase JMJD2B présentent une forte instabilité chromosomique (Slee *et al.* (2012)). Cela suggère que l'hétérochromatine péri-centromérique joue un rôle dans la fonction du centromère.

### A.2.4.3 Séquence d'ADN et fonction du centromère

La conservation du caractère répété des séquences au cours de l'évolution pose la question de son rôle dans le fonctionnement du centromère. Des études portant sur des chromosomes humains anormaux, caractérisés par une relocalisation du centromère, montrent que de nouveaux centromères peuvent se former sur des régions dépourvues de séquences répétées, mais présentant néanmoins une séquence riche en A/T ; ces structures sont appelées *néocentromères* (du Sart *et al.* (1997)). De plus, les équidés possèdent des centromères dépourvus de séquences répétées qui sont fonctionnels et conservés depuis des millions d'années au niveau du chromosome 11 (Piras *et al.* (2010)). Les auteurs de cette étude suggèrent donc un mécanisme de formation de néocentromères par relocalisation dont ils ont pu observer les traces de chaque étape (Fig. A.19).

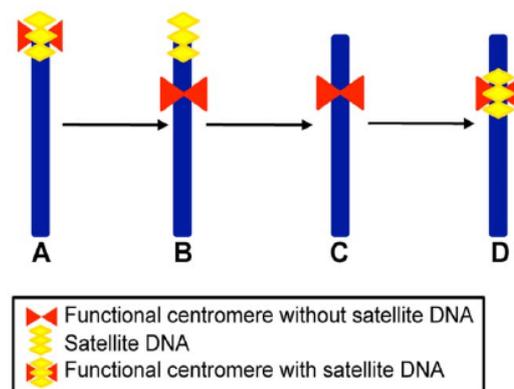


FIGURE A.19: Modèle de mécanisme de formation de néocentromères au cours de l'évolution en quatre étapes (d'après Piras *et al.* (2010)). (A) Le chromosome ancestral porte l'ADN satellite (jaune) au niveau du centromère (rouge). (B) Relocalisation du centromère au niveau d'une région ne contenant pas d'ADN satellite. Les séquences satellites sont toujours présentes au niveau de l'ancien locus du centromère. (C) Perte des séquences satellites au niveau du locus de l'ancien centromère. (D) Apparition de séquences répétées au niveau du néocentromère fonctionnel.

Ces observations montrent que le centromère peut être fonctionnel sans la présence d'ADN répété ; cette caractéristique n'est donc pas nécessaire. Il a été suggéré que des séquences satellites soient incorporées au niveau des sites de néocentromères, celles-ci lui conférant un avantage adaptatif, probablement en augmentant la précision de la ségrégation chromosomique (Marshall *et al.* (2008)). Des résultats de l'équipe d'Elena Giulotto présentés à la 19<sup>th</sup> *International Chromosome Conference* concernant le positionnement du centromère sur le chromosome 11 du cheval sont à l'appui de cette hypothèse. L'absence de séquences répétées au niveau de ce centromère a permis d'étudier précisément la position de CenpA sur le locus par des expériences de ChIP, qui montrent que la position du centromère varie entre les individus. Cela suggère qu'il existe une interaction entre CenpA et les séquences répétées, qui sont sélectionnées pour stabiliser la position du centromère. De plus cette équipe observe une perte des répétitions centromériques au niveau des sites des anciens centromères, ce qui suggère que les répétitions sont maintenues par une

pression sélective liée au fonctionnement du centromère.

Par ailleurs, contrairement aux régions monomériques, les HORs peuvent interagir directement avec des protéines impliquées dans l'assemblage du kinétochore et peuvent induire l'assemblage de centromères *de novo* (Harrington *et al.* (1997)). Les HORs sont enrichies en domaines de liaison de la protéine centromérique CenpB ; ces domaines d'une taille de 17 pb appelé CenpB-Box, sont suffisants pour établir un centromère fonctionnel dans des expériences de chromosomes artificiels (Ohzeki *et al.* (2002)). Ces expériences de chromosomes artificiels ont aussi montré que toutes les séquences ne sont pas équivalentes dans leur propension à établir, maintenir et transmettre le centromère, et que certaines HORs sont plus efficaces que des  $\alpha$ -satellites monomériques ou des séquences génomiques non  $\alpha$ -satellites (Harrington *et al.* (1997), Saffery *et al.* (2001)). Ces arguments s'opposent donc à une vision strictement épigénétique du centromère, dans laquelle toutes les séquences seraient équivalentes dans leur interaction avec le fonctionnement du centromère (Hayden (2012)).

#### A.2.4.4 Transcrits issus des régions centromériques

Il est possible que la fonction des séquences répétées soit associée aux transcrits qui en sont issus. En effet, des transcrits issus des RCs ont été observés dans de nombreux organismes, incluant la levure *S. pombe*, les plantes, la souris et l'homme. Ces transcrits s'accumulent dans différents contextes physiologiques, notamment au cours du cycle cellulaire, de la différenciation, du développement, en réponse à des stress, et dans des pathologies telles que le cancer (pour revue : Eymery *et al.* (2009)).

Chez la souris, ces transcrits semblent être un composant essentiel de la structure des RCs, et sont notamment impliqués dans l'association de la protéine HP1 à l'hétérochromatine péri-centromérique (Maison *et al.* (2002)). Cependant, les mécanismes dans lesquels ils interviennent sont encore peu compris. Leur accumulation forcée dans des cellules de souris mène à des problèmes de ségrégation des chromatides sœurs ce qui suggère un rôle de ces transcrits dans la fonction du centromère (Bouzinba-Segard *et al.* (2006)). Chez la levure *S. pombe*, des transcrits issus des régions péri-centromériques sont impliqués dans l'établissement et la maintenance de l'hétérochromatine centromérique via la machinerie d'ARN interférence (Grewal *et Elgin* (2007)). Un tel mécanisme n'a pas été mis en évidence chez les mammifères, mais plusieurs résultats expérimentaux suggèrent une implication des transcrits issus des RCs dans l'établissement et la maintenance de l'hétérochromatine centromérique (pour revue : Eymery *et al.* (2009)).

## Conclusion

Les régions centromériques sont caractérisées d'une part par la présence de séquences répétées, et d'autre part par la présence du variant d'histone CenpA au niveau du centromère et d'une structure chromatinienne condensée au niveau des

péricentromères. Ces caractéristiques sont liées au rôle du centromère dans la maintenance de l'information génétique. Cependant, ce lien avec la fonction ne permet pas d'expliquer la complexité des séquences humaines, et il est frappant de constater qu'elles sont beaucoup moins homogènes que les séquences murines ; à la fois au sein d'un chromosome et entre les chromosomes<sup>1</sup>. Nous allons voir que dans le noyau en interphase, les RCs montrent une organisation nucléaire particulière et possèdent des fonctions qui ne sont pas liées à la division cellulaire, notamment dans la régulation de l'expression des gènes. La compréhension de ces fonctions pourrait permettre de mieux comprendre la diversité des séquences centromériques.

---

1. La différence d'homogénéité entre homme et souris est peut-être à mettre en relation avec le fait que tous les chromosomes murins sont acrocentriques, sachant que les chromosomes humains acrocentriques présentent souvent des similarités de séquence.

## A.3 Organisation nucléaire des régions centromériques

Dans le noyau de cellules en interphase, les régions centromériques présentent une organisation particulière. D'une part elles se positionnent préférentiellement à la périphérie du noyau et des nucléoles, d'autre part elles ont tendance à s'associer entre elles. Dans cette partie, nous allons décrire ces deux aspects de l'organisation nucléaire des régions centromériques, puis nous exposerons les mécanismes moléculaires connus qui peuvent intervenir dans cette organisation.

### A.3.1 Positionnement nucléaire des régions centromériques

Depuis la découverte des séquences satellites dans les années 80, de nombreuses études ont abordé la question de la localisation spatiale des RCs humaines. Ces études ont montré d'une part que les RCs ont tendance à se localiser en périphérie du noyau et au bords des nucléoles; et d'autre part que leur localisation varie beaucoup au cours du cycle cellulaire, entre les types cellulaires, et en fonction de l'état physio-pathologique de la cellule (Haaf et Schmid (1991)).

Dans cette partie, nous allons décrire le positionnement des RCs par rapport aux deux points de repère les plus utilisés : la périphérie et les nucléoles, en distinguant à chaque fois les études qui prennent en compte l'ensemble des régions centromériques de celles qui les distinguent.

#### A.3.1.1 Positionnement par rapport à la périphérie du noyau

**Positionnement de l'ensemble des RCs.** L'ensemble des RCs peut être détecté simultanément dans des expériences d'immunocytochimie ciblant des protéines centromériques. Le sérum provenant de patient atteints du syndrome CREST<sup>1</sup> est souvent utilisé car il cible des protéines spécifiques des centromères. De manière générale, une grande proportion des RCs s'associe à la périphérie du noyau dans les cellules quiescentes; celle-ci peut atteindre plus de 95% dans une population de lymphocytes, dont 60% des noyaux ont l'ensemble de leurs RCs associées à la périphérie (Weierich *et al.* (2003), Solovei *et al.* (2004)). Au cours du cycle cellulaire, les RCs présentent une localisation plus interne en début de phase G<sub>1</sub>, et s'associent progressivement à la périphérie au cours de la phase G<sub>1</sub> et en début de phase S. De plus, le positionnement des centromères varie entre les types cellulaires (Solovei *et al.* (2004)). Ces résultats sont en contradiction avec d'autres tentatives de description quantitative du positionnement des centromères qui n'observent pas le même taux d'association à la périphérie ou les mêmes variations de positionnement au cours du cycle cellulaire. Ces contradictions peuvent s'expliquer par l'absence de prise en compte du cycle cellulaire, ou sa détermination trop approximative, car

---

1. Le syndrome CREST est une forme particulière de sclérodermie qui associe à des degrés divers : des Calcifications sous cutanées (C), un syndrome de Raynaud (R), des anomalies (Eosopagiennes (E), une Sclérodactylie (S) et des Télangiectasies (T).

seulement basée sur la ploïdie estimée par la quantité d'ADN (Bartholdi (1991), Weimer *et al.* (1992)). L'association à la périphérie semble être moins fréquente dans des cellules ES (Wiblin *et al.* (2005)), l'étude le montrant ne prend cependant pas en compte le cycle cellulaire et, les cellules ES se divisant en général plus rapidement que d'autres types cellulaires, il est possible que ces différences soient dues à ce facteur.

Toutes ces études quantifient la localisation des foyers marqués par le sérum CREST qui peuvent correspondre à l'agrégation de plusieurs centromères. Ces quantifications peuvent donc en être biaisées, puisqu'un foyer comprenant plusieurs centromères est compté comme un foyer ne comportant qu'un seul centromère. De plus, comme le nombre de foyers ainsi marqués varie au cours du cycle cellulaire et de la différenciation (voir §A.3.2.2), la comparaison entre différents stades du cycle ou différents types cellulaires est faussée, et les différences observées (ou absences de différences) pourraient être en partie du aux différences d'agrégation.

**Différences entre les RCs portées par différents chromosomes.** La fréquence d'association à la périphérie varie aussi entre les RCs portées par différents chromosomes (Carvalho *et al.* (2001), Table A.2). Les auteurs de ce travail établissent une corrélation entre la présence de bandes G sur le chromosome et l'association de la RC à la périphérie, suggérant que ces régions d'hétérochromatine facultative qui constituent les bandes G serviraient de sites d'ancrage à la périphérie du noyau, et influenceraient la position du centromère. Cette hypothèse est peut-être à mettre en relation avec le fait que les LADs sont principalement composées d'hétérochromatine facultative (§A.1.3.3), qui est enrichie dans les bandes G. Ces données sont en contradiction avec les résultats de Weierich *et al.* (2003), mesurés dans des lymphocytes non activés, qui suggèrent que les RCs portées par les chromosomes qu'ils ont pris en compte sont toutes localisées en périphérie du noyau ; même si en moyenne la localisation radiale de leur TCs varie.

Deux études récentes ont mesuré des relocalisations de RCs au cours du processus de différenciation, et ont montré que certaines RCs se relocalisent spécifiquement vers la périphérie du noyau (voir Table A.2).

### A.3.1.2 Positionnement par rapport aux nucléoles.

**Positionnement global des RCs.** Les RCs ont tendance à s'associer aux nucléoles. En marquant l'ensemble des centromères dans des expériences d'immunocytochimie, Wiblin *et al.* (2005) mesurent un taux d'association de 50% des centromères (ou agrégats de centromères) aux nucléoles, dans plusieurs types cellulaires y compris des fibroblastes. Ces mesures sont en contradiction avec les mesures de Solovei *et al.* (2004) qui observent 95% des centromères associés à la périphérie ; cette différence est certainement due au fait que les cellules utilisées par Wiblin *et al.* (2005) ne sont pas synchronisées. De manière intéressante, dans des types cellulaires différenciés, les centromères sont plus fréquemment associés soit à la périphérie du noyau soit aux nucléoles, en comparaison à des cellules ES (Wiblin *et al.* (2005),

Chromosomes	Localisation		Condition cellulaire		Référence
	Périphérie	Nucléoles	Cycle	Type	
1, 20	+/-	+/-	G <sub>0</sub> /G <sub>1</sub>	Lymphocytes B	Carvalho <i>et al.</i> (2001)
2, 4, 6, 7, 9, 10, 12	+	+/-			
15, 16, 17, 22	-	+			
18	+	+			
1, 11, 12, 15, 15, 17, 18, 20, X	+	ND	G <sub>0</sub>	Lymphocytes	Weierich <i>et al.</i> (2003)
8, 9, 18	9<8<18	ND	ND	Lymphocytes B	Salníková <i>et al.</i> (2000) <sup>a</sup>
	8<18≤9			Lymphocytes T	
	8<9<18			Moelle osseuse	
	8≤18<9			HL-60	
	8<18<9			U-937	
8	- ↔ +			Différentiation myéloïde (HL-60)	
9	+ ↔ +				
1, 5, 19 <sup>b</sup>	+ ↔ -	- ↔ +	ND	ES ↔ ? <sup>c</sup>	Bártová <i>et al.</i> (2008)
10, 15, 17	- ↔ +	ND			
12	- ↔ -	ND			
1, 3, 12, 17	- ↔ +	ND	ND	Myoblastes ↔ Myotubes	Rozwadowska <i>et al.</i> (2013)
7, 11	- ↔ -	ND			

TABLE A.2: Localisation nucléaire des centromères portés par des chromosomes spécifiques, dans différentes conditions. Dans le cas de l'étude de Carvalho *et al.* (2001), les informations de localisation proviennent d'une comparaison à un modèle aléatoire qui sert de référence; +/- indique qu'il n'y a pas de différence avec le modèle. Notons que le modèle utilisé dans cette étude est sujet à caution, car il utilise des paramètres provenant de mesures moyennées sur l'ensemble de la population, et ne prend donc pas en compte les variations inter-cellulaires de forme et de taille. Weierich *et al.* (2003) utilisent la position moyenne du contre-marquage de l'ADN comme référence. Dans les autres études, les données n'étant pas comparées à une référence, elles ne peuvent s'exprimer que de manière relative; soit entre les différents centromères (indiqué par les signes < pour une différence significative ou ≤ pour une différence non significative); soit au cours de la différenciation (indiqué par le signe ↔).

ND : non disponible

<sup>a</sup>Dans cette étude, le positionnement est mesuré par la distance normalisée au centre sur une projection 2D. <sup>b</sup>Les chromosomes 1, 5, et 19 ne sont pas distingués donc il s'agit d'une relocalisation moyenne entre les trois centromères; la relocalisation ne concerne pas forcément chacun des trois centromères.

<sup>c</sup>Dans cette étude, le type différencié n'est pas caractérisé, la différenciation des cellules ES est induite par l'acide rétinolique.

Bártová *et al.* (2008)). L'association des RCs aux nucléoles a aussi été détectée par des approches moléculaires, qui montrent que les séquences  $\alpha$ -satellites font partie des constituants majoritaires des NADs (Németh *et al.* (2010)).

**Variations entre les RCs portées par différents chromosomes.** Les centromères s'associent au nucléoles avec une fréquence qui varie entre les chromosomes (voir Table A.2, Carvalho *et al.* (2001)). Les centromères qui s'y associent le plus

fréquemment sont ceux des chromosomes portant les NORs, mais il existe aussi des différences entre les centromères des autres chromosomes. Certains centromères s'associent plus fréquemment aux nucléoles que ce qui serait attendu dans le cas d'une distribution aléatoire (Carvalho *et al.* (2001)). L'association aux nucléoles semble varier au cours de la différenciation ; Bártoová *et al.* (2008) observent une relocalisation des centromères à proximité des nucléoles au cours de la différenciation de cellules ES.

**Aspects fonctionnels.** L'implication fonctionnelle de ces associations n'est pas connue, mais le fait que les centromères de chromosomes ne portant pas les NORs s'associent aux nucléoles suggèrent que ces associations ont une importance fonctionnelle qui n'est pas liée à la biogenèse des ribosomes ; elle pourrait être liée au métabolisme du centromère. En effet, une étude a montré que des transcrits provenant des séquences  $\alpha$ -satellites s'accumulent dans le nucléole et sont impliquées dans le recrutement de protéines centromériques se trouvant aussi dans le nucléole (Wong *et al.* (2007)). Il est possible que le métabolisme du centromère fasse intervenir le nucléole. Par ailleurs l'association du chromosome X inactif est impliquée dans la maintenance de l'hétérochromatine (Zhang *et al.* (2007)) sur ce chromosome, il est possible que le nucléole joue aussi ce rôle pour l'hétérochromatine centromérique.

## Conclusion

Les RCs humaines ont tendance à s'associer à la périphérie du noyau et aux bords des nucléoles, avec des fréquences qui varient entre les chromosomes. Leur positionnement est dynamique et varie notamment au cours de la différenciation et du cycle cellulaire. La table A.2 montre que la majorité des études ne prennent pas en compte le cycle cellulaire qui influence pourtant le positionnement. De plus il est difficile d'exploiter les informations quantitatives provenant de ces études car les mesures sont normalisées différemment<sup>1</sup> et ne sont donc pas comparables d'une étude à l'autre. D'éventuelles variations de l'association des RCs aux nucléoles au cours du cycle cellulaire n'ont pas été mesurées à notre connaissance. La description du positionnement des RCs reste donc encore très partielle.

### A.3.2 Associations de régions centromériques

En 1908, le botaniste italien Pasquale Baccarini observait des structures denses dans le noyau en interphase de plantes, qu'il nomme *chromocentri* (chromocentres) (Baccarini (1908)). Ces structures ont été ensuite observées dans le noyau en interphase de nombreuses espèces par un simple marquage de l'ADN, y compris chez la drosophile, la souris ainsi que de nombreux mammifères (Solovei *et al.* (2013)) ; en revanche elles n'ont pas été observées chez l'homme. D'autres observations suggèrent que de telles structures pourraient exister chez l'homme ; mais elles ont été

---

1. voir §A.4.3.2 pour la description des différentes mesures de positionnement radial.

beaucoup moins étudiées que chez la souris. Nous allons donc nous appuyer sur les résultats connus chez la souris et les comparer aux observations faites chez l'homme.

### A.3.2.1 Le compartiment d'hétérochromatine péricentromérique murin

**Structure du compartiment d'hétérochromatine péricentromérique.** Le marquage de l'ADN par un intercalant tel que le DAPI permet de visualiser les RCs murines dans le noyau en interphase; elles correspondent à des foyers marqués plus intensément que leurs alentours, dont la frontière est nette, et qui sont nommés chromocentres. Ce contraste est certainement dû au fait que les séquences satellites murines sont beaucoup plus riches en A/T que le reste du génome, et que les intercalants comme le DAPI ont plus d'affinité pour les paires de bases A/T; de plus ces régions sont beaucoup plus condensées que le reste du génome (Fussner *et al.* (2012)) ce qui augmente la concentration locale d'intercalant. Des expériences de FISH utilisant des sondes dirigées contre les satellites majeurs et mineurs ont montré que ces foyers correspondent à l'association des péricentromères portés par plusieurs chromosomes, et que les centromères se localisent en périphérie de ces foyers (Fig. A.20, Guenatri *et al.* (2004)). Une étude a aussi décrit une réduction du nombre de signaux centromériques observés par immunocytochimie (10 signaux en moyenne au lieu de 40) dans des neurones (Martou et De Boni (2000)), suggérant que ce phénomène d'agrégation concerne aussi les centromères. Une autre étude a montré que les RCs ne s'associent pas aléatoirement dans le noyau; au contraire, il existe des associations préférentielles entre les RCs portées par différents chromosomes (Vadakkan *et al.* (2006)).

Des marques épigénétiques caractéristiques de l'hétérochromatine constitutive sont enrichies au niveau de ces foyers; on observe notamment  $H_3K_9me_3$ ,  $H_3K_{27}me_3$  (Peters *et al.* (2003)) et  $H_4K_{20}me_3$  (Kourmouli *et al.* (2004)), ainsi que l'absence des marques permissives  $H_3K_4me$  (Peters *et al.* (2003)), l'acétylation de  $H_4$  (Maison et Almouzni (2004)) et un fort taux de 5mC (Lewis *et al.* (1992)). Par conséquent on observe aussi les protéines se liant à la 5mC, MeCP2 (Lewis *et al.* (1992)) et MBD1-4 (Hendrich et Bird (1998)). Enfin d'autres protéines sont enrichies, telles que les protéines de l'hétérochromatine  $HP1_\alpha$ ,  $HP1_\beta$  (Taddei *et al.* (1999)) qui sont ciblées par  $H_3K_9me_3$ . HP1 agit comme médiateur induisant l'interaction d'autres protéines telles que Ki67, le suppresseur de tumeurs pRb, Suv39h, LBR et beaucoup d'autres (pour revue : Singh et Georgatos (2003)). Nous désignerons ces foyers correspondant à l'hétérochromatine péricentromérique, par le terme de compartiment d'hétérochromatine péricentromérique (PHC : Pericentromeric Heterochromatin Compartment).

**Variations au cours de la différenciation et du développement.** La taille et le nombre des chromocentres<sup>1</sup> varient entre les types cellulaires (Fig. A.21), ce qui indique que l'association des RCs est variable. Les différences les plus marquées

1. Ces observables reflètent le nombre d'associations des RCs : lorsque le nombre de chromocentres diminue, leur taille augmente ainsi que le nombre d'associations

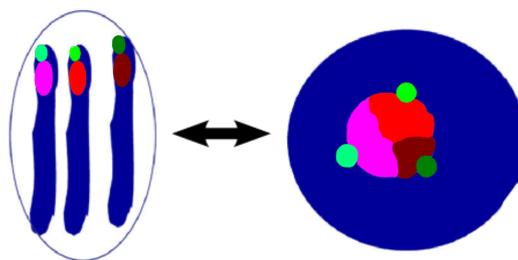


FIGURE A.20: Organisation spatiale des péricentromères (satellites majeurs, dégradé de rouges) et centromères (satellites mineurs, dégradé de verts) murins (d'après [Guenatri \*et al.\* \(2004\)](#)). Les péricentromères portés par plusieurs chromosomes s'associent pour former des foyers. Les centromères se localisent en périphérie de ces foyers.

sont observées entre les cellules sphériques d'une part (lymphocytes, macrophages, myotubes, neurones<sup>1</sup>) qui présentent peu de chromocentres mais de grande taille, et les cellules aplaties d'autre part (myoblastes, fibroblastes) dont les chromocentres sont plus nombreux et plus petits ([Mayer \*et al.\* \(2005\)](#)). Les cellules ES constituent une exception à cette corrélation; de plus les bords de leurs chromocentres sont moins nets que dans les types différenciés. Ces deux observations sont certainement à mettre en relation avec la plasticité de la chromatine dans ces cellules (§A.1.4.3). Un des modèles de différenciation les plus étudiés est la différenciation de myoblastes en myotubes, au cours de laquelle le PHC s'agrège ([Brero \*et al.\* \(2005\)](#)). Nous reparlerons de ce modèle au §A.3.3.2.

Par ailleurs, la maturation des ovocytes est accompagnée d'une réorganisation spectaculaire du PHC, au cours de laquelle le nombre de chromocentres est réduit de 8-9 à 3-4, et l'hétérochromatine péricentromérique se répartie en anneau autour du nucléole. Cette réorganisation est accompagnée de modifications épigénétiques et d'une extinction globale du génome, mais le lien entre ces deux événements n'est pas établi ([Bonnet-Garnier \*et al.\* \(2012\)](#)).

### A.3.2.2 Associations de régions centromériques chez l'homme

Le marquage de l'ADN par un colorant comme le DAPI ne permet pas d'observer des foyers de chromatine péricentromérique aussi nettement que chez la souris. Cette différence est certainement due au fait que les séquences satellites humaines sont moins riches en A/T que les séquences murines. D'autres approches faisant intervenir un marquage indirect des RCs ont permis de montrer que des RCs portées par différents chromosomes humains s'associent dans le noyau en interphase. La diversité des séquences portées par les RCs rend le marquage de l'ensemble des RCs difficile dans des expériences de FISH, en revanche il est possible de distinguer des RCs portés par des chromosomes spécifiques.

1. les neurones ne sont pas présentés dans l'étude de [Mayer \*et al.\* \(2005\)](#) mais suivent aussi cette observation ([Martou et De Boni \(2000\)](#))

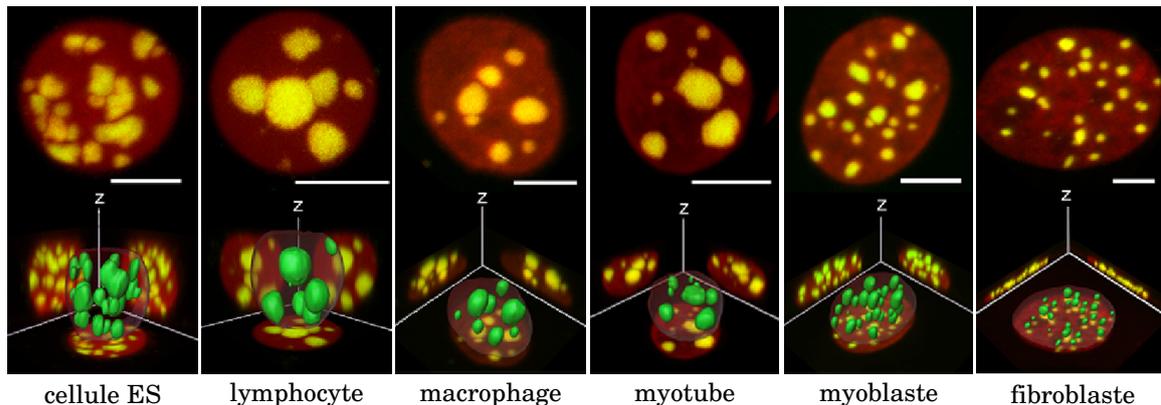


FIGURE A.21: Chromocentres murins observés par FISH 3D avec une sonde dirigée contre les satellites majeurs dans différents types cellulaires (d'après Mayer *et al.* (2005)). L'ADN (TO-PRO3) est coloré en rouge et le signal FISH apparaît en vert; la superposition des deux signaux se traduit par un marquage jaune au niveau des chromocentres. Pour chaque type cellulaire sont représentés : la projection maximale selon l'axe z (en haut) et une reconstruction 3D avec les projections maximales en intensité selon les axes xy, xz et yz (en bas). Barre d'échelle : 5  $\mu$ m.

**Associations de l'ensemble des centromères.** L'ensemble des centromères peut être détecté dans des expériences d'immunocytochimie; en général en utilisant un sérum ciblant plusieurs protéines centromériques provenant de patients atteints du syndrome de CREST, ou bien en utilisant un anticorps dirigé contre une protéine centromérique unique telle que CenpA. La détection nombre de foyers centromériques plus faible que le nombre de chromosomes de la cellule est en général interprétée comme l'association de plusieurs centromères. Plusieurs études ont montré que les associations mesurées de la sorte varient au cours du cycle cellulaire (Weimer *et al.* (1992), Solovei *et al.* (2004)), entre les types cellulaires (Alcobia *et al.* (2000), Solovei *et al.* (2004)) et au cours de différenciation (Beil *et al.* (2002)). Le nombre de foyers observés varie entre 13 et 46. L'effet observé le plus fort est une augmentation des associations dans des cellules quiescentes par rapport à des cellules en cycle, l'ampleur de ce phénomène variant entre les types cellulaires (Solovei *et al.* (2004)). Weimer *et al.* (1992) n'ont pas détecté de variation des associations au cours du cycle cellulaire, contrairement à Solovei *et al.* (2004) qui observent une association plus forte au cours de la phase S dans certains types cellulaire. Cette différence pourrait être due à la méthode de caractérisation du cycle cellulaire, qui est plus précise dans la deuxième étude.

En utilisant des anticorps dirigés contre des modifications d'histones, Zinner *et al.* (2006) ont observé que plusieurs centromères s'associent autour de foyers d'hétérochromatine caractérisés par la présence de la marque  $H_4K_{20}me_3$ , formant des structures en anneaux (Fig. A.22A). Ces structures ressemblent aux chromocentres observés chez la souris, mais les séquences d'ADN sous-jacentes n'ont pas été identifiées. Ces structures semblent aussi différer des chromocentres par leur forme en anneaux, mais il est possible que l'observation de cette morphologie particulière soit anecdotique, n'ayant pas été quantifiée, de plus il semble qu'elle ne soit pas systématique d'après les images montrées. La morphologie de ces foyers d'hétéro-

chromatine semble varier au cours du cycle cellulaire ; elle est plus agglomérée dans les cellules quiescentes (Zinner *et al.* (2006)).

Les RCs humaines ont donc tendance à s'associer dans le noyau en interphase, et deux types d'associations ont été identifié : des associations directes entre les centromères et des associations médiées par des foyers d'hétérochromatine.

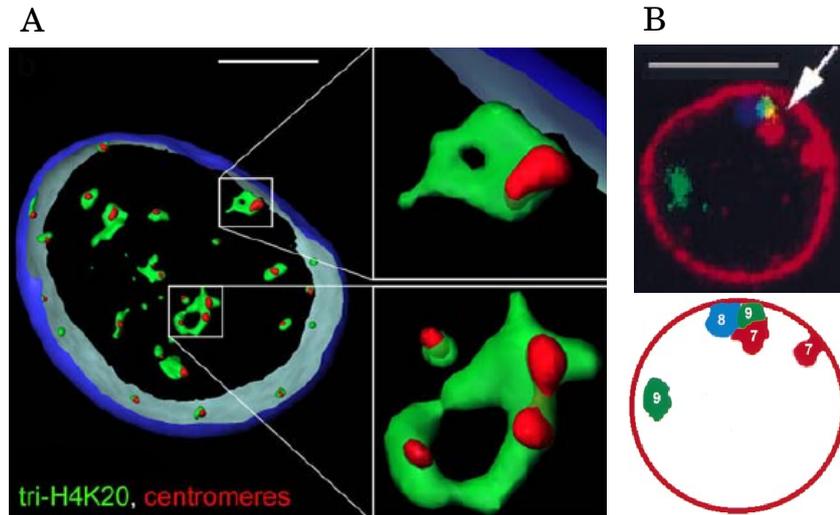


FIGURE A.22: Associations de régions centromériques chez l'homme. (A) Associations médiées par des foyers d'hétérochromatine (d'après Zinner *et al.* (2006)). Reconstitution 3D d'une image segmentée d'une expérience d'immunocytochimie sur un noyau de fibroblaste. L'hétérochromatine est détectée par un anticorps dirigé contre la marque H<sub>4</sub>K<sub>20</sub>me<sub>3</sub>, les centromères sont détectés par le sérum CREST. (B) Association de RCs portées par des chromosomes spécifiques (d'après Alcobia *et al.* (2000)). Plan médian d'une pile d'images d'une expérience FISH détectant des RCs spécifiques dans des lymphocytes. Les numéros des chromosomes correspondants sont indiqués sur le schéma en dessous de l'image. Barre d'échelle : 5  $\mu$ m.

**Associations de RCs spécifiques.** Des expériences de FISH détectant les RCs de chromosomes individuels ont permis d'étudier les associations entre RCs portées par des chromosomes spécifiques (Alcobia *et al.* (2000), Alcobia *et al.* (2003)). Les auteurs de ces études ont montré l'existence d'associations préférentielles de RCs (Fig. A.22B). Lorsque ces associations ont lieu à proximité des nucléoles, elles impliquent souvent, mais pas uniquement, des RCs des chromosomes qui portent les NORs. Ces associations préférentielles dépendent aussi du type cellulaire. Les différences observées sont plus importantes entre deux lignages (cellules hématopoïétiques et fibroblastes) qu'entre les types cellulaires d'un même lignage (Alcobia *et al.* (2000)) ; mais il y a aussi des variations héréditaires entre les différents stades de la différenciation des cellules hématopoïétiques (Alcobia *et al.* (2003)). Dans ces études, les foyers observés sont de tailles très variables<sup>1</sup>. Les associations étant mesurées par le contact entre les foyers observés, cette différence de taille peut mener à un biais ; les foyers plus larges ayant une plus forte chance d'être en contact avec d'autres RCs.

1. Cette variabilité de la taille peut s'expliquer par des différences d'abondance des séquences ciblées.

Des associations de centromères spécifiques ont aussi été détectées par des analyses de résultats de HiC<sup>1</sup> (Botta *et al.* (2010)). Comme les RCs ne sont pas présentes sur les cartes assemblées du génome, les auteurs de ce travail mesurent en fait les associations des régions assemblées qui se trouvent à proximité des RCs des chromosomes considérés.

**Le PHC humain.** Il existe actuellement un manque de clarté dans la définition du PHC chez l'homme. Le terme *chromocentre* est souvent utilisé dans la littérature (Bartholdi (1991), Beil *et al.* (2002), Alcobia *et al.* (2003), Solovei *et al.* (2004)), mais ce terme fait appel à une définition cytologique (foyers marqués intensément par un colorant de l'ADN) qui ne s'applique pas à l'homme. De plus, lorsque il est employé chez l'homme, le terme *chromocentre* se réfère soit aux associations de centromères (détectées par un marquage de CenpA par exemple), soit aux associations de RCs observées en FISH par Alcobia *et al.* (2000). Aucune de ces deux observables ne permet de comparer les structures sous-jacentes avec les chromocentres murins qui sont des associations de régions péri-centromériques :

- La détection des centromères par immunocytochimie ne prend en compte que les associations entre centromères, et pas d'éventuelles associations médiées par les régions péri-centromériques
- La FISH, avec les sondes utilisées par Alcobia *et al.* (2000), parce qu'elle ne détecte pas l'intégralité des RCs (centromères et péri-centromères), et donc d'éventuelles associations des péri-centromères chez l'homme ne peuvent être observées.

### A.3.2.3 Fonction du PHC.

**Régulation transcriptionnelle.** De nombreuses études conduites chez la souris ont montré une corrélation entre la répression transcriptionnelle d'un locus et son positionnement à proximité du PHC (pour revue : Jost *et al.* (2012)). Notamment, l'analyse du positionnement de gènes au cours de la différenciation de lymphocytes, a montré que l'activation de certains gènes corrèle avec un éloignement du locus par rapport au PHC (Delaire *et al.* (2004)). Inversement, la répression corrèle avec un rapprochement (Brown *et al.* (1997)). Il a été proposé que la répression soit liée à la compétition entre des protéines de l'hétérochromatine et des facteurs activant la transcription au niveau du locus situé à proximité du PHC (Brown *et al.* (1997)); le PHC isolant en quelque sorte le locus. La répression pourrait aussi être due à une propagation de l'état chromatinien en *trans*. Il est aussi possible que la répression soit due à un mécanisme indépendant du PHC, et que la relocalisation à proximité du PHC soit une conséquence de la répression.

---

1. voir A.4.1.1 pour la description du HiC.

**Autres types de régulations.** Plusieurs études ont montré que le PHC joue un rôle dans la recombinaison V(D)J<sup>1</sup> dans les lymphocytes B. Le mécanisme dit d'*exclusion allélique* permet une expression mono-allélique du gène, afin qu'un seul anticorps ne soit exprimé par cellule. Il a été montré qu'une des étapes de ce mécanisme implique une séquestration d'un des loci au niveau du PHC (Roldán *et al.* (2005)). Ce repositionnement est médié par la protéine ATM, et est accompagné de modifications chromatiniennes et de la répression transcriptionnelle (Hewitt *et al.* (2009)).

**Lien avec la différenciation.** Comme nous l'avons décrit précédemment, la morphologie du PHC murin varie beaucoup entre les types cellulaires. Il est difficile d'établir le rôle de la réorganisation du PHC au sein du processus de différenciation. Dans le modèle de différenciation de myoblastes en myotubes décrit précédemment, lorsque l'activité HDAC est inhibée les réorganisations du PHC sont bloquées en même temps que le processus de différenciation (Terranova *et al.* (2005)); mais cela ne démontre pas un lien de causalité entre ces deux phénomènes, qui pourraient tous deux être directement des conséquences indépendantes de l'inhibition de l'activité HDAC. Une étude a montré qu'en supprimant la protéine MeCP2, des cellules ES pouvaient se différencier en neurones sans que le PHC s'agrège, ce qui montre que cette réorganisation n'est pas une étape cruciale pour l'amorçage du processus de différenciation (Bertulat *et al.* (2012)). Les auteurs formulent l'hypothèse que la réorganisation du PHC pourrait avoir une importance dans la stabilisation et la maintenance de l'état de différenciation. Par ailleurs, dans un modèle de développement murin au stade 2 cellules, l'interférence avec la transcription des satellites majeurs induit l'arrêt du développement et perturbe la formation des chromocentres (Probst *et al.* (2010)). Ceci suggère que la régulation de la transcription des séquences satellite majeur joue un rôle dans le développement et la formation des chromocentres. Il n'est pas exclu que la formation des chromocentres soit aussi impliquée dans le développement. Notons qu'il est difficile d'évaluer la pertinence physiologique de ces modèles cellulaires cultivés *in vitro*, notamment pour des processus faisant intervenir des mécanismes épigénétiques qui sont certainement d'avantage susceptibles d'être sensibles à l'environnement de la cellule. En revanche, des différences d'agrégation du PHC murin ont été observées dans de nombreux tissus (Solovei *et al.* (2013)), ce qui suggère que ces réorganisations ne sont pas des artefacts de la culture *in vitro*.

**Rôle dans la topologie du génome.** Un travail de l'équipe de Wouter de Laat présenté au congrès *Dynamic Organization of Nuclear function 2012* a identifié des domaines chromosomiques s'associant au PHC murin par une adaptation de la technique de 4C. Ceci suggère que de la même manière que la lamina avec les LAD

---

1. La recombinaison V(D)J est un mécanisme de recombinaison de l'ADN qui permet de générer la diversité des récepteurs des lymphocytes, nécessaire à la reconnaissance de l'immense variété des antigènes.

et les nucléoles avec les NAD, le PHC pourrait servir de plate-forme d'ancrage de la chromatine.

L'agrégation des RCs portées par différents chromosomes pourrait être impliquée dans le positionnement des chromosomes, en imposant une contrainte physique, et participant ainsi à la mise en place de la topologie globale du génome. Cette hypothèse est souvent formulée dans le cadre du modèle d'auto-organisation (voir §A.1.5.1).

Notons que ces interactions ne peuvent être détectées par des approches moléculaires car les séquences ne sont pas présentes sur les cartes assemblées du génome (biais discuté en §A.4.1.1); elles sont donc souvent omises, sauf dans l'étude de *Botta et al.* (2010). L'étude de la structure interne des territoires chromosomiques (§A.1.1.4) montre que les domaines répressifs se positionnent préférentiellement à l'intérieur du TC et interagissent plutôt avec des domaines répressifs porté par le même bras chromosomique. Cependant, ces descriptions reposent sur des études moléculaires qui excluent les RCs, et ces dernières ne semblent pas suivre ce comportement puisqu'elles ont tendance à s'associer avec des RCs portées par d'autres chromosomes.

## Conclusion

Les régions péri-centromériques portées par plusieurs chromosomes s'agrègent dans le noyau en interphase pour former des foyers d'hétérochromatine. Ces foyers sont visibles par un simple marquage de l'ADN chez plusieurs espèces telles que la souris, mais n'ont jamais été observés chez l'homme à cause de la difficulté de détecter les régions péri-centromériques humaines. En revanche on observe une agrégation des centromères chez l'homme, ce qui suggère que des structures similaires pourraient exister. La fonction de ces structures a été très étudiée chez la souris; elles jouent notamment un rôle dans la régulation de la transcription, en fournissant un environnement répressif. Les variations morphologiques observées au cours de la différenciation suggèrent qu'elles pourraient aussi y jouer un rôle. Elles pourraient aussi être impliquées dans l'organisation globale du génome puisqu'elles résultent de l'association de RCs portées par plusieurs chromosomes, elles imposent ainsi une contrainte physique au génome.

### A.3.3 Mécanismes moléculaires impliqués dans l'organisation nucléaire des régions centromériques

#### A.3.3.1 Mécanismes liés au positionnement des RCs.

**Association à la périphérie.** Les mécanismes moléculaires induisant l'association des RCs à la périphérie ne sont pas connus. En revanche plusieurs observations réalisées chez la souris suggèrent que des protéines de l'enveloppe nucléaire pourraient être impliquées.

Les cellules à bâtonnet murines constituent un exemple frappant d'organisation nucléaire de la chromatine. Alors que, dans la majorité des cellules, l'euchromatine

se trouve en moyenne au centre du noyau et l'hétérochromatine en périphérie et aux bords des nucléoles (architecture dite *conventionnelle*), dans ces cellules l'euchromatine est principalement juxtaposée à la périphérie du noyau et l'hétérochromatine au centre formant un unique foyer (architecture dite *inversée*). Cette architecture est présente uniquement chez les mammifères nocturnes ; elle conférerait un avantage sélectif, en réduisant la perte de lumière dans la rétine (Solovei *et al.* (2009)). Une analyse par immunohistochimie conduite dans la même équipe a révélé que, dans toutes les espèces dont les cellules à bâtonnet présentent une architecture inversée, les protéines de l'EN lamine A/C et LBR ne sont pas exprimées (Solovei *et al.* (2013)). De plus, l'absence simultanée de ces deux protéines est nécessaire et suffisante à l'induction d'une architecture inversée (Fig. A.23) :

- Des expériences d'inactivation génique et d'extinction ciblée de ces gènes montrent que tous les types cellulaires qui n'expriment pas ces deux protéines présentent une architecture inversée deux semaines après la sortie du cycle cellulaire.
- La surexpression de LBR dans des cellules à bâtonnet suffit à rétablir une architecture conventionnelle.

En revanche, la surexpression de la lamine C dans des cellules à bâtonnet ne permet pas de rétablir une architecture conventionnelle. Ceci traduit certainement le fait que d'autres protéines interagissant à la fois avec la lamine A/C et la chromatine sont nécessaires pour retourner à une architecture conventionnelle, et que la surexpression de la lamine C ne suffit pas à faire exprimer leur gène par la cellule. Il s'agit certainement de protéines à domaine LEM, dont beaucoup sont fonctionnellement redondantes. Comme nous l'avons vu au §A.1.2.2, la lamina est composée de deux réseaux de protéines interagissant avec l'hétérochromatine et permettant de l'ancrer à la périphérie : d'une part les lamines B, interagissant avec LBR qui lie la chromatine, d'autre part les lamines A/C interagissant avec de nombreuses protéines qui se lient à la chromatine et qui sont partiellement redondantes. Les auteurs de ces études ont donc identifié la clef de voûte de chacun de ces deux réseaux<sup>1</sup> : LBR pour le réseau de type B et la lamine A/C pour le réseau de type A. L'inactivation simultanée de ces deux réseaux<sup>2</sup> est nécessaire et suffisante à l'induction du phénotype inversé.

Par ailleurs, un travail a montré que l'inhibition de l'activité HDAC par un traitement à la trichostatine A induit une relocalisation réversible des chromocentres à la périphérie de cellules murines<sup>3</sup> (Taddei *et al.* (2001)). Ce résultat est peut-être à mettre en relation avec le fait que la chromatine enrichie en marques H<sub>4</sub>K<sub>5</sub> interagit avec la protéine Nup83 à la périphérie du noyau (voir §A.1.3.2) : l'inhibition de

1. Le terme clef de voûte est introduit ici pour désigner l'élément du réseau sans lequel l'interaction entre la lamina et l'hétérochromatine péricentromérique spécifiquement ne peut se faire. Il est possible que ces protéines ne soient pas des clefs de voûte si l'on considérait l'interaction entre la lamina et une autre partie du génome.

2. Via l'inactivation de leurs clefs de voûte respectives.

3. Ce résultat est cependant controversé car il entre en contradiction avec une autre étude chez l'homme (Gilchrist *et al.* (2004)).

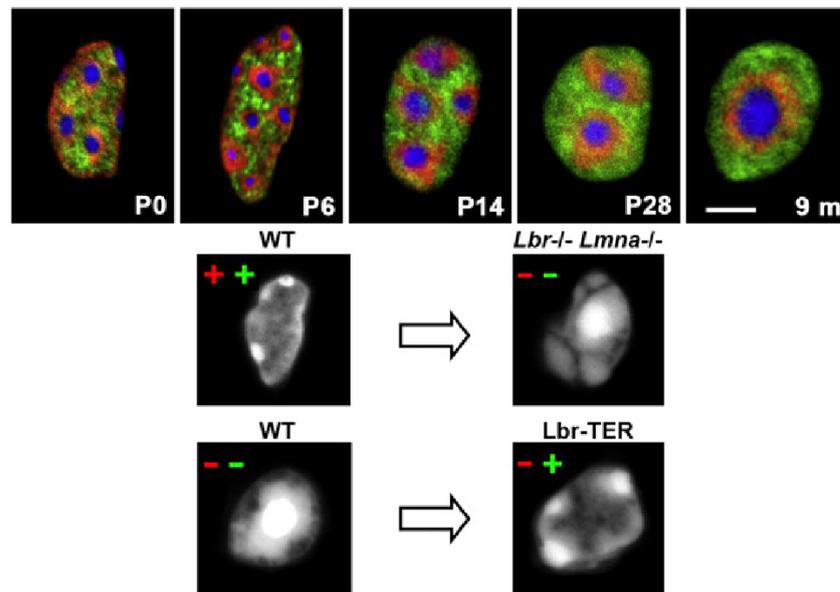


FIGURE A.23: Architecture inversée de la chromatine chez la souris. En haut : progression du phénotype inversé au cours du développement dans les cellules à bâtonnet de souris (d'après [Solovei et al. \(2009\)](#)). Images de noyaux marqués par FISH. En rouge : sonde dirigée contre les séquences L1 (classe majoritaire des LINEs, enrichis dans l'hétérochromatine facultative correspondant aux bandes G), en vert : sonde dirigée contre B1 (classe majoritaire des SINEs, enrichis dans l'euchromatine correspondant aux bandes R), en bleu : sonde dirigée contre les satellites majeurs (bandes C). En bas : expériences clés montrant le rôle de LBR et LMNA dans l'architecture inversée (d'après [Solovei et al. \(2013\)](#)). Images de noyaux de souris marqués au DAPI. Les signes + et - représentent le statut d'expression de la lamine A (rouge) et de LBR (vert). Ligne du haut : induction du phénotype inversé dans toutes les cellules post-mitotiques d'une souris *knock-out* pour les gènes *LMNA* et *LBR*. Ligne du bas : restauration du phénotype conventionnel dans les cellules à bâtonnet par expression transgénique de LBR.

l'activité HDAC induirait une sur-acétylation des RCs, ce qui augmenterait leur interaction avec Nup83 et entraînerait une relocalisation à la périphérie. Le taux d'acétylation des RCs est faible dans les cellules non traitées, donc ce mécanisme induit probablement la localisation périphérique d'autres régions du génome ; il est probable que cela ait pour conséquence d'éloigner les RCs de la périphérie par un effet indirect d'encombrement stérique au niveau des pores nucléaires.

**Associations aux nucléoles.** Dans le cas des chromosomes portant les NORs, l'association des RCs aux nucléoles est certainement liée au métabolisme des ribosomes, puisque les NORs se localisent à proximité des RCs sur le chromosome. Les associations préférentielles des autres RCs aux nucléoles pourraient être la conséquence du positionnement du TC, et facilitées par le fait que l'hétérochromatine a tendance à s'agréger avec elle-même (§A.3.2), mais elles pourraient aussi impliquer un mécanisme d'adressage des centromères aux nucléoles ([Németh et Längst \(2011\)](#)).

### A.3.3.2 Mécanismes liés à l'agrégation des régions centromériques.

Le PHC murin étant aisément détectable, de nombreuses études ont relié des variations induites d'expression de protéines ou des mutations génétiques avec des variations morphologiques du PHC, permettant ainsi de déduire l'implication directe ou indirecte de protéines dans la formation du PHC. Nous classons les protéines potentiellement impliquées en trois catégories : les protéines de l'EN, les protéines se liant à l'ADN, et les modificateurs chromatiniens (associés aux modifications qu'ils induisent).

**Protéines de l'EN.** Comme nous l'avons décrit précédemment, les deux réseaux de protéines liés aux lamines A/C et B sont impliqués dans l'ancrage du PHC à la périphérie ; et, lorsque les mécanismes d'ancrage de ces deux réseaux sont absents, le PHC n'est plus ancré à la périphérie et s'agrège en un seul foyer au centre de la cellule (architecture inversée, Fig. A.23). L'ancrage des RCs à la périphérie semble donc avoir un effet antagoniste à l'agrégation du PHC.

Le fait que la transition entre les architectures conventionnelles et inversées nécessite deux semaines d'arrêt du cycle cellulaire, donne une indication sur la cinétique du phénomène d'agrégation. Il est intéressant de constater que ces réorganisations ne nécessitent pas une division cellulaire, ce qui montre que l'hétérochromatine présente une certaine dynamique dans le noyau en interphase. L'ancrage du PHC à la périphérie est notamment médié par l'interaction entre LBR et des protéines liant le PHC telles que HP1 et MeCP2.

Par ailleurs, ce travail montre que le gène *Lmna* est impliqué dans la morphologie du PHC ; sachant que *Lmna* est aussi impliqué dans la morphologie du noyau (§A.1.2.2) et que son expression est régulée au cours de la différenciation (Solovei *et al.* (2013)) il est probable qu'il soit à l'origine des variations de morphologies observée dans la figure A.21.

**Protéines liant l'ADN.** Le PHC s'agrège au cours de la différenciation de myoblastes en myotubes (Brero *et al.* (2005)) ; ce modèle a donc été utilisé dans de nombreuses études du phénomène d'agrégation du PHC. L'agrégation est mesuré par la réduction du nombre de chromocentres<sup>1</sup>. Ce modèles a permis de montrer que les protéines HMGA1 et MeCP2 pourraient intervenir dans l'agrégation du PHC :

- La suppression d'HMGA1 augmente le nombre de chromocentres (donc diminue l'agrégation) (Brocher *et al.* (2010)), ce qui suggère que la protéine HMGA1 pourrait être directement impliquée dans l'agrégation des chromocentres grâce à la présence de plusieurs domaines *AT-hook*<sup>2</sup> de médier des interactions en *trans* (Vogel *et al.* (2011)).

1. Notons que dans ces études, la lignée cellulaire C2C12 généralement utilisée est aneuploïde, ce qui ne permet pas de faire des quantifications absolues du nombre de chromocentres ou de centromères ; les variations observées au cours de la différenciation restent cependant informatives.

2. Les domaines *AT-hook* sont des domaines de liaison à l'ADN, qui ont une forte affinité pour les régions riches en A/T telles que l'ADN satellite.

- Le niveau de la protéine MeCP2 augmente au cours de la différenciation de ces cellules, et son domaine de liaison à l'ADN méthylé est capable d'induire l'agrégation du PHC (Brero *et al.* (2005), Casas-Delucchi *et al.* (2012)). Des domaines *AT-hook*, récemment identifiés sur MeCP2, pourraient aussi être impliqués ; en effet, ils augmentent l'affinité de MeCP2 pour le PHC *in vivo* et ils induisent l'agrégation de la chromatine *in vitro* (Baker *et al.* (2013)).

De manière paradoxale, dans ce même modèle de différenciation, l'expression de la protéine HMGA1 est réduite au début du processus de différenciation ; de plus, la surexpression de cette protéine empêche l'agrégation des chromocentres, malgré une augmentation de l'expression de MeCP2 (Brocher *et al.* (2010)). Ceci suggère que HMGA1 pourrait avoir un effet bivalent sur l'agrégation du PHC : à trop faible ou à trop forte concentration, elle réduit l'agrégation. Deux mécanismes pourraient expliquer cette apparente contradiction :

- Si l'effet agrégeant d'HMGA1 est du au fait qu'elle possède plusieurs domaines *AT-hook* qui sont susceptibles d'interagir avec les RCs de plusieurs chromosomes, l'augmentation de la concentration d'HMGA1 pourrait avoir comme effet de saturer les sites de liaisons présents au niveau des RCs et de diminuer la probabilité qu'une même protéine HMGA1 interagisse avec des domaines situés sur des chromosomes différents, et donc diminuer l'effet agrégeant d'HMGA1.
- Il est possible que les domaines *AT-hook* de HMGA1 entrent en compétition avec ceux de MeCP2, réduisant ainsi l'affinité de MeCP2 pour le PHC, et l'empêchant d'agréger le PHC.

**Modifications de la chromatine.** Plusieurs travaux ont montré que des modificateurs chromatiniens étaient capables de modifier l'agrégation du PHC murin. L'inhibition conjointe des HMTs Prdm3 et Prdm16 responsables de la monométhylation de H<sub>3</sub>K<sub>9</sub><sup>1</sup> induit une déstructuration totale des chromocentres (Pinheiro *et al.* (2012)). Cela suggère que la méthylation de H<sub>3</sub>K<sub>9</sub> est impliquée dans l'agrégation des chromocentres.

De plus l'inhibition d'HDACs induit aussi une déstructuration partielle du PHC (Taddei *et al.* (2001)) et inhibe son agrégation au cours de la différenciation de myoblastes (Terranova *et al.* (2005)). MeCP2 interagissant avec des HDACs il est probable qu'elle soit impliquée dans le ciblage des HDACs au niveau du PHC. Sachant que l'acétylation de H<sub>3</sub>K<sub>9</sub> inhibe sa méthylation (Rea *et al.* (2000)), il est possible que le mécanisme d'agrégation inhibé par l'inactivation de Prdm3 et Prdm16 soit le même que celui avec lequel interfèrent les inhibiteurs d'HDAC.

**Mécanismes d'agrégation du PHC humain.** Tous les résultats précédents proviennent d'expériences conduites chez la souris ; le PHC humain étant difficile à détecter, les mécanismes d'agrégation ont été peu étudiés. Il est très probable

---

1. L'inhibition de chacune des deux HMTs séparément ne produit pas de phénotype suggérant qu'elles sont redondantes.

que ces mécanismes soient similaires à ceux décrits chez la souris. Dans des cellules HGPS, les centromères sont plus agrégés (Taimen *et al.* (2009)) ce qui suggère un rôle des protéines de membrane comme c'est le cas chez la souris.

## Conclusion

Chez de nombreuses espèces, y compris l'homme et la souris, les RCs portées par plusieurs chromosomes ont tendance à s'associer dans le noyau en interphase pour former des foyers d'hétérochromatine. Ces foyers se localisent en périphérie du noyau, aux bords des nucléoles, mais aussi dans le reste du nucléoplasme (il ne sont alors associés à aucun de ces deux compartiments). Très peu de travaux ont étudié le positionnement des RCs par rapport aux nucléoles et dans le nucléoplasme. Le PHC semble avoir une importance fonctionnelle, notamment dans la régulation de l'expression des gènes. Ce compartiment est difficilement observable chez l'homme à cause de la complexité des séquences d'ADN présentes au niveau des RCs, par conséquent, l'organisation spatiale des RCs humaines nécessite encore d'être précisée. Les mécanismes moléculaires responsables de cette organisation particulière commencent à être compris, et certains acteurs ont déjà été identifiés chez la souris. Cette organisation semble résulter d'un équilibre entre deux mécanismes antagonistes faisant intervenir :

- Des protéines de la membrane nucléaire, qui ont tendance à ancrer les RCs à la périphérie et à défavoriser les associations.
- Des protéines liant l'ADN et la chromatine qui ont tendance à promouvoir l'agrégation de l'hétérochromatine péricentromérique avec elle-même.

En revanche peu d'éléments concernant les mécanismes et la fonction du PHC humain sont disponibles, principalement à cause des difficultés liés sa détection, et du flou existant sur sa définition.

## A.4 Méthodes d'études de l'architecture nucléaire

Les deux grands types d'approches expérimentales permettant d'étudier l'organisation nucléaire sont les méthodes moléculaires et l'imagerie. Dans un premier temps nous allons présenter ces deux types, puis nous discuterons du caractère aléatoire de l'organisation du noyau et de quelle manière cela peut influencer l'approche expérimentale. Dans cette étude, nous avons choisi une approche faisant appel à l'imagerie (voir §A.5), nous nous concentrons par la suite sur les méthodes d'analyse des images provenant de la microscopie de fluorescence, en discutant des avantages et des biais de chaque méthode.

### A.4.1 Approches expérimentales

Dans cette section, nous allons présenter les principales méthodes utilisées à ce jour, et discuter de leurs avantages et limitations.

#### A.4.1.1 Méthodes moléculaires

**Étude des interactions entre régions génomiques.** La technologie de capture de conformation chromosomique (3C) et ses dérivés sont basées sur le *crosslink* de la chromatine par le formaldéhyde, qui crée un instantané des interactions entre toutes les paires de loci qui ont lieu dans l'espace du noyau. La chromatine est ensuite fragmentée (par exemple par digestion enzymatique) puis re-liguée intramoléculairement, ce qui génère des produits de ligation uniques au niveau des loci qui interagissent. Ces produits de ligation sont ensuite détectés par diverses techniques qui diffèrent par leur sensibilité et l'étendue des régions génomiques couvertes (pour revue : de Wit et de Laat (2012)). Il existe deux types de techniques dérivées : d'une part celles qui prennent une séquence de référence et détectent les régions avec lesquelles elle interagit (le 3C détecte les interactions entre deux loci, le 4C détecte les interactions entre un locus et l'ensemble du génome) et d'autres qui détectent les interactions possibles soit dans une région génomique définie d'environ quelques Mb (5C) soit sur l'ensemble du génome (HiC). La limite de profondeur de séquençage impose un compromis entre résolution des interactions détectées et taille de la région génomique couverte.

**Étude des interactions entre régions du génome et structures nucléaires.** Ces méthodes permettent de dresser des cartes d'interactions d'une protéine donnée avec la chromatine sur l'ensemble du génome. Dans le cas où la protéine en question est spécifiquement localisée dans un compartiment, on peut déduire de ces cartes l'interaction du génome avec ce compartiment. En revanche ceci n'est pas valable lorsque la protéine est partiellement soluble dans le nucléoplasme (Steenkel et Dekker (2010)).

**Immunoprécipitation de chromatine.** L'immunoprécipitation de chromatine est une technique dans laquelle une protéine d'intérêt est sélectivement immunoprécipitée à partir d'une préparation de chromatine, permettant de déterminer les séquences d'ADN qui lui sont associées. L'immunoprécipitation de chromatine dépend beaucoup du protocole de *crosslink* et nécessite de disposer d'un anticorps spécifique de la protéine d'intérêt (Collas (2009)).

**DamID.** Dans cette technique, une enzyme apposant une méthylation sur l'adénine de l'ADN (Dam : DNA adénine methyltransferase) est fusionnée à une protéine d'intérêt.

La méthylation de l'adénine est déposée de manière stable sur les séquences d'ADN se trouvant à proximité de la protéine de fusion dans des cellules vivantes, ce qui permet de marquer des interactions transitoires. Cette marque n'étant pas présente chez les métazoaires, l'empreinte déposée par l'enzyme peut être détectée spécifiquement (van Steensel *et al.* (2001)). Cette technique a notamment été utilisée pour déterminer les LADs en fusionnant la Dam à la lamine B (§A.1.3.3). Elle requiert l'expression d'une protéine de fusion et suppose que cette dernière se localise de la même manière que la protéine endogène. En revanche, contrairement à l'immunoprécipitation de chromatine, la technologie du DamID ne requière par de *crosslink* ni de détection indirecte par des anticorps.

**Avantages et limitations.** Ces techniques permettent d'obtenir des informations sur plusieurs loci en même temps voire même sur l'ensemble du génome, ce qui est un avantage considérable. Elles présentent deux limitations principales : d'une part elles fournissent en général des informations moyennes sur un grand nombre de cellules et d'autre part elles reposent sur l'existence de cartes assemblées du génome.

**Limites liée au moyennage.** Afin d'obtenir suffisamment de fragments pour couvrir l'ensemble de la carte du génome, ces techniques sont conduites sur un grand nombre de cellules et les résultats obtenus correspondent donc à une moyenne sur la population étudiée. Par conséquent cela ne permet pas d'étudier la variabilité inter-cellulaire et conduit parfois à étudier un comportement moyen virtuel qui n'existe finalement dans aucune cellule. Cela peut être une limitation très lourde pour l'étude de l'architecture spatiale, qui présente une forte variabilité inter-cellulaire due à sa nature dynamique et stochastique (§A.4.2.1). Deux types de méthodes ont été mise en oeuvre pour palier ce problème technique : d'une part la conduite de ces expériences sur cellule unique, récemment réalisée par l'équipe de Peter Fraser (Nagano *et al.* (2013)), d'autre part plusieurs équipes ont tenté d'extraire plusieurs conformations à partir des données de *multi-C* moyennées en les *déconvoluant*. Ce type de méthode repose sur l'introduction d'un modèle physique qui permet d'échantillonner l'ensemble des conformations possibles qui rendent compte des données expérimentales, et leur associe une probabilité (Kalhor *et al.* (2012)).

**Limites liées à l'assemblage du génome.** Ces techniques nécessitent de repérer les fragments d'ADN identifiés sur une carte assemblée du génome. D'une part cela suppose que la lignée cellulaire sur laquelle l'expérience est conduite possède le même génome que la carte de référence. D'autre part cela exclut les séquences qui ne sont pas présentes sur les cartes assemblées des génomes, telles que la majorité des séquences répétées en tandem (§A.2.3).

Une autre limitation s'ajoute à cela : il n'est en général pas possible de distinguer les allèles paternels et maternels d'un génome diploïde, leur comportement est donc moyenné.

#### A.4.1.2 Imagerie

Nous nous limitons ici à la microscopie de fluorescence, mais d'autres types de microscopie telle que la microscopie électronique à transmission ont été utilisés pour étudier l'organisation nucléaire.

**Techniques d'imagerie de fluorescence.** Ces techniques reposent sur le marquage indirect d'une région d'intérêt par le biais de molécules fluorescentes, afin d'étudier ses caractéristiques morphologique, ou ses relations spatiales avec d'autres structures nucléaires.

**L'immunocytochimie.** L'immunocytochimie est une technique cytologique qui permet de détecter des protéines d'intérêt par le biais d'un anticorps spécifique, qui peut soit être détecté directement s'il est marqué par une molécule fluorescente, soit indirectement ; il est alors révélé grâce à un anticorps secondaire fluorescent qui reconnaît spécifiquement l'anticorps primaire. Cette technique requiert la fixation des cellules, ce qui peut fortement altérer l'architecture 3D du noyau. Les protocoles qui utilisent une fixation chimique, par exemple avec du paraformaldéhyde, semblent préserver le mieux l'architecture 3D du noyau tout en permettant l'accessibilité aux épitopes (Heppenger *et al.* (2007)). Cependant, il est difficile de le vérifier car la fixation peut à la fois altérer la morphologie du noyau, celle de la chromatine, et la localisation des biomolécules. Cette technique nécessite aussi de disposer d'un anticorps spécifique de la protéine d'intérêt, et l'épitope reconnu doit être accessible dans le noyau fixé.

**L'hybridation de fluorescence *in situ*.** L'hybridation de fluorescence *in situ* (FISH : Fluorescent *In situ* Hybridization) est une technique cytologique qui utilise des sondes capables de s'hybrider sur une séquence d'ADN d'intérêt. Ces sondes peuvent être détectées directement ou indirectement, permettant de visualiser la région génomique d'intérêt. Cela suppose que la sonde soit bien spécifique de la région d'intérêt. Cette technique requiert la fixation de la cellule ainsi que la dénaturation de l'ADN, en général à haute température en présence de formamide, afin de permettre l'hybridation de la sonde ; ce qui peut altérer la structure 3D du noyau et notamment de la chromatine (Markaki *et al.* (2012)). Certaines sondes telles que

celles produites par des procédés enzymatiques, et notamment les peintures chromosomiques, doivent être utilisées avec de l'ADN Cot qui entre en compétition avec la sonde afin de masquer les séquences répétées, ce qui implique que seul un sous-ensemble des séquences cibles, celles présentes avec un faible nombre de répétitions, peut être détectés (Boyle *et al.* (2011)). La FISH peut aussi être utilisée pour détecter des ARN ; dans ce cas elle ne requiert pas de dénaturation, mais des précautions particulières pour éliminer les RNases. Dans certains cas cette technique peut être combinée avec un immuno-marquage afin de détecter simultanément des acides nucléiques et des protéines (Solovei *et Cremer* (2010)).

**Observation de cellules vivantes.** Des techniques alternatives permettent d'observer des compartiments dans des cellules vivantes. Par exemple, des loci pouvant être détectés par fluorescence peuvent être insérés dans le génome en utilisant le système opéron/répresseur Lac. Des protéines de fusion couplées à une protéine fluorescente telle que la GFP permettent de visualiser des compartiments nucléaires, sous réserve que la protéine de fusion se localise de la même manière que la protéine endogène. Ces techniques donnent accès à des informations dynamiques telles que la mobilité de régions génomiques ou la mobilité moléculaire<sup>1</sup>.

**Avantages et limitations des approches par imagerie.** Ces techniques ont globalement l'avantage de fournir des données sur cellules uniques, permettant de prendre en compte la variabilité inter-cellulaire. De plus elle peuvent permettre de corrélérer plusieurs informations spatiales ; il est par exemple possible avec ce genre de méthodes d'étudier l'association de deux loci, et de déterminer si cet évènement se produit dans un endroit particulier du noyau, ce qui est pour le moment impossible avec des méthodes moléculaires. En effet les méthodes moléculaires permettent dans certains cas de fournir ce type d'information, mais toujours à partir d'expériences distinctes ; ce qui ne permet pas de les corrélérer.

En revanche, contrairement aux approches moléculaires, les techniques faisant intervenir l'imagerie ne permettent pas d'étudier le comportement d'un grand nombre de loci dans la même expérience, à cause de la limitation du nombre de canaux de fluorescence, ni dans des expériences séparées car l'analyse est en général très difficile à automatiser. De plus ces techniques sont limitées dans leur résolution. D'une part à cause de la limite de résolution optique due au fait que la taille caractéristique des objets observés est de l'ordre de grandeur de la longueur d'onde des rayonnements utilisés en microscopie de fluorescence, qui diffractent donc sur ces objets<sup>2</sup>. D'autre part, ces techniques reposant toujours sur un marquage indirect, les marqueurs utilisés ont une certaine taille qui impose une limite. De plus, pour des questions de sensibilité, il est parfois nécessaire que plusieurs molécules fluorescentes soient présentes au niveau de la région d'intérêt pour qu'elle soit détectable ;

---

1. La mobilité moléculaire peut être étudiée par la technique de FRAP (Fluorescence Recovery After Photobleaching)

2. Des avancées récentes dans le domaine de la microscopie permettent de dépasser ces limites dans certains cas, et toujours sur cellules fixées (pour revue : Toomre *et Bewersdorf* (2010)).

ce qui limite encore plus la résolution.

## Conclusion

Les deux types d'approches présentés ont tous deux des avantages et des inconvénients :

- Les approches moléculaires permettent souvent d'étudier dans une même expérience un grand nombre de loci ou même l'ensemble du génome, mais sont en général conduites sur un grand nombre de noyaux, et fournissent ainsi des informations sur une conformation moyennée qui ne correspond à aucun noyau.
- Les approches d'imagerie au contraire ne permettent d'étudier qu'un nombre limité de loci, mais fournissent des informations à l'échelle de la cellule unique.

Ces deux familles d'approches sont donc complémentaires, et il est souvent intéressant de les combiner. Dans ce travail nous avons uniquement utilisé des approches d'imagerie de fluorescence, car ce sont les seules permettant d'étudier l'organisation nucléaire des RCs.

### A.4.2 Déterminisme dans l'organisation spatiale

#### A.4.2.1 Déterminisme et aléatoire

L'organisation du noyau est souvent décrite comme étant non-aléatoire (Nagano *et al.* (2013)), probabiliste (Misteli (2009)), ou même déterministe (Rajapakse *et al.* (2009)). Considérés hors contexte, ces termes peuvent paraître contradictoires. En réalité ces termes sont employés pour décrire le fait que malgré sa grande variabilité, la structure interne du noyau présente des régularités (dont certaines ont été décrites dans les parties précédentes de ce manuscrit). Ces régularités sont souvent de nature probabiliste, comme illustré par le fait que certaines mesures, telles que le positionnement radial d'un chromosome donné, suivent des lois de probabilités au sein d'une population de cellules fixées (§A.1.3.1). Cela signifie que la distribution du positionnement radial d'un chromosome sera conservée entre deux populations indépendantes de cellules comparables ; en revanche, la position du chromosome sera très probablement différente entre deux cellules de la même population. La structure du noyau n'est donc ni fixe, ni uniformément aléatoire ; nous proposons le terme de *déterminisme probabiliste* qui reflète à la fois la présence de régularités et un certain degré de stochasticité apparente. Ce terme se distingue donc à la fois du terme *aléatoire uniforme* et du terme de *déterminisme stricte*, qui sont en fait deux cas limites du déterminisme probabiliste.

#### A.4.2.2 Observation des phénomènes

Ces considérations sont importantes pour la conception d'expériences permettant d'observer un aspect en particulier de l'organisation spatiale du noyau, que nous désignons dans cette section par le terme de *phénomène*. Il existe deux grands types d'approches : des approches basées d'une part sur l'étude de cellules vivantes

(approche *dynamique*) et d'autre part sur l'étude de populations de cellules fixées (approche *statique*). Dans les deux cas, comme dans toute approche d'inférence statistique, il est crucial que l'échantillon de noyaux soit représentatif. Deux paramètres importants à prendre en compte sont :

- La variabilité du phénomène : si elle est faible, il se rapproche d'un phénomène déterministe, et ne nécessite pas d'étudier un échantillon de cellules trop important.
- Sa dynamique : plus il est transitoire, plus il faut augmenter la taille de l'échantillon dans une approche statique pour espérer l'observer, et plus il faut augmenter la résolution temporelle dans une approche dynamique.

**Approches dynamiques.** Ces approches demandent des développements particuliers pour introduire des marqueurs fluorescents compatibles avec la vie de la cellule. La photo-toxicité des rayonnements utilisés et le photo-blanchiment qu'ils induisent imposent souvent un compromis entre résolution spatiale, résolution temporelle, et durée totale d'acquisition. La résolution temporelle doit être adaptée à la dynamique du phénomène étudié ; par exemple la diffusion de molécules uniques peut être étudiée à l'échelle de la milliseconde, les mouvements de domaines chromatiniens à l'échelle de la minute, et le développement d'un organisme tel que *C. elegans* à l'échelle de l'heure.

Dans le cadre de l'étude de l'organisation nucléaire, les difficultés techniques et la durée de l'acquisition et de l'analyse d'images font que ces approches sont souvent conduites sur un nombre limité de noyaux. Ces approches sont parfois indispensables pour démontrer qu'un événement transitoire a nécessairement lieu dans chaque cellule. Par exemple dans l'étude de l'association transitoire des chromosomes X, lors de l'inactivation du chromosome X, une approche dynamique permettra de savoir si ce phénomène est à déterminisme stricte, c'est-à-dire s'il a lieu dans chaque cellule. A l'inverse, une approche statique permettra seulement de mesurer la proportion de cellules dans lesquelles cette association a lieu, et si cette proportion n'est pas égale à 1 on ne peut différencier un phénomène transitoire nécessaire d'un phénomène probabiliste.

**Approches statiques.** Ces approches reposent sur l'étude de populations de noyaux fixés. Le fait de fixer les noyaux permet dans certains cas une meilleure résolution pour les méthodes basées sur l'imagerie, et en général facilite les marquages et augmente la diversité des marquages et des combinaisons possibles. L'absence de la dimension temporelle réduit considérablement la quantité des données et permet en général d'étudier un plus grand nombre de noyaux.

Dans ces approches, un phénomène est étudié en réalisant des mesures sur des noyaux distincts provenant d'une population. Prenons comme exemple de phénomène la localisation radiale d'un TC. Dans chaque noyau, le TC sera localisé à une certaine position<sup>1</sup>. La mesure d'un nombre suffisamment élevé de noyaux per-

1. En termes statistique, la mesure du positionnement dans chaque noyau correspond à la réalisation d'une variable aléatoire associée à cette mesure.

met d'échantillonner l'ensemble des positions possibles pour le TC et ainsi d'inférer une loi de probabilité associée au positionnement du TC pour la population ; ceci permet donc de caractériser le positionnement du TC (est-il préférentiellement périphérique ? central ? etc.). Afin d'augmenter la puissance statistique (liée à la sensibilité de la mesure), et de rendre compte de l'ensemble des configurations possibles dans la population, la taille de l'échantillon de noyaux étudiée doit être augmentée avec la variabilité du phénomène étudié. De plus, il est important de déterminer les paramètres qui influencent le phénomène, afin de les prendre en compte pour homogénéiser la population et diminuer la part de variabilité due à ces paramètres. Ceci est essentiel pour permettre une meilleure reproductibilité des mesures entre différentes expériences.

Dans le cas où il est possible de définir plusieurs sous-populations ordonnées chronologiquement, il est possible d'inférer la dynamique d'un phénomène par une approche statique. Par exemple, en distinguant les cellules se trouvant dans différentes phases du cycle cellulaire, il est possible d'étudier la dynamique de phénomènes au cours du cycle cellulaire. La comparaison des différentes populations permet de retracer le cours du phénomène, à l'échelle de populations. Lors de la comparaison, il est important de prendre en compte tous les paramètres qui varient entre les populations : par exemple, la taille du noyau augmente au cours de la phase S et ce paramètre à lui seul peut expliquer des variations de positionnement au sein du noyau.

Les méthodes moléculaires fournissent en général des informations sur une conformation virtuelle résultant d'une moyenne sur une population de cellules. Elles seront donc utiles pour étudier des phénomènes de nature déterministe probabiliste (voir §A.4.1.1), mais contrairement aux approches sur cellules uniques, elle ne permettent pas de mesurer le caractère stochastique du phénomène par sa variabilité inter-cellulaire.

### A.4.3 Analyse spatiale des signaux de microscopie

L'analyse spatiale quantitative des signaux de microscopie est un des aspects limitants dans l'étude de l'organisation nucléaire. La difficulté d'appréhender la structure du noyau est probablement due au fait que sa morphologie présente toujours un certain degré de variabilité entre deux cellules de la même population, et qu'il n'y ait en général pas de points de repère pour décrire sa structure interne. Certains modèles cellulaires présentent néanmoins des points de repère qui ont permis d'établir des systèmes de coordonnées ; c'est le cas de la levure (Berger *et al.* (2008)) ainsi que des spermatozoïdes (Mudrak *et al.* (2012)). En l'absence de repères, une des mesures les plus utilisées pour caractériser la position d'un TC ou d'un locus est le positionnement radial qui prend comme point de repère la périphérie ou le centre du noyau (§A.4.3.2).

D'autres types d'analyses peuvent être réalisées en fonction de la question sous-

jacente, permettant d'analyser un *pattern*<sup>1</sup> de points dans son ensemble (§A.4.3.3), ou enfin de mesurer des relations spatiales entre plusieurs signaux (§A.4.3.4), telles que la colocalisation.

Un des aspects cruciaux est la normalisation de ces mesures afin de pouvoir comparer plusieurs noyaux et populations de noyaux en éliminant des facteurs de variabilité. Par exemple le positionnement radial est susceptible de dépendre de la taille et de la forme des noyaux. Afin de s'assurer qu'une différence observée entre deux populations de noyaux n'est pas uniquement due à des variations de taille et/ou de forme entre les deux populations, il est important d'utiliser une mesure normalisée par rapport à ces variations. Par ailleurs, certaines mesures font appel à la segmentation des signaux, qui est aussi une étape délicate.

Dans cette partie nous allons présenter succinctement la segmentation des signaux, puis les principaux types de mesures qui sont réalisés dans le cadre de l'étude de l'architecture du noyau.

#### A.4.3.1 Segmentation

La segmentation est une opération qui permet d'identifier des objets distincts à partir d'un signal en niveau de gris. Il existe de nombreuses méthodes qui sont souvent adaptées à un problème spécifique (voir [Russ \(2006\)](#)). Dans ce paragraphe nous allons simplement présenter les principales méthodes et leurs limites.

L'approche la plus utilisée pour segmenter un signal consiste à se baser sur un critère d'intensité, en définissant un seuil (*threshold* en anglais) sur les intensités de l'image au dessus duquel un voxel<sup>2</sup> est considéré comme appartenant au signal<sup>3</sup>. Les différents éléments (ou objets) du signal sont ensuite étiquetés par composante connexe<sup>4</sup>. Le seuil est parfois choisi manuellement, mais cela peut poser un problème d'objectivité et de reproductibilité. Alternativement, il existe des algorithmes qui permettent de calculer automatiquement ce seuil, tels que la méthode d'Otsu ([Otsu \(1975\)](#)).

Dans certains cas la présence de bruit de fond ou de défauts dans l'image, tels qu'une illumination non homogène, ne permet pas de définir un seuil unique pour l'ensemble de l'image. Dans ce cas il est possible de corriger certains défauts par filtrage, en réduisant le bruit<sup>5</sup> ou en augmentant le contraste local des objets<sup>6</sup> ([Andrey et al. \(2010\)](#), [Ram et al. \(2012\)](#)). Certains défauts du signal provenant du système d'acquisition peuvent être corrigés par leur connaissance *a priori*; par exemple la déconvolution permet de réduire l'étalement du signal dû à la diffraction.

---

1. Le terme anglais *pattern* peut être traduit par *motif* en français, mais nous préférons utiliser le terme anglais car il est plus couramment utilisé

2. Un voxel est un pixel en 3D, unité de discrétisation spatiale d'une image numérique.

3. De manière équivalente : en deçà duquel le voxel est considéré comme faisant partie du bruit de fond.

4. Cette opération consiste à regrouper les voxels en contact direct.

5. Par exemple grâce à un filtre Gaussien ou médian.

6. Par exemple en utilisant un filtre Laplacien de Gaussien ou un filtre en chapeau haut de forme (*top-hat* en anglais).

Le filtrage nécessite parfois une connaissance *a priori* de propriétés des objets que l'on cherche à détecter, notamment de leur taille, leur forme et s'il s'agit de foyers (*spots* en anglais). L'introduction de ces informations *a priori* permet en général d'augmenter l'efficacité du filtrage. Afin d'éliminer les faux-positifs, il est aussi possible de sélectionner les objets après la segmentation en définissant des critères de taille ou de forme par exemple (Andrey *et al.* (2010), Ram *et al.* (2012)).

L'approche par seuillage n'est pas adaptée pour séparer des foyers trop proches. Pour palier ce défaut, les algorithmes les plus utilisés sont les algorithmes de ligne de partage des eaux (*watershed* en anglais) (Soille (2003)). Brièvement, ces méthodes considèrent une image en niveaux de gris comme un relief topographique, et calculent la ligne de partage des eaux du relief par propagation depuis les extrémités des reliefs vers les lignes de partages. Ceci permet de partitionner une image en différents bassins versants, ou régions. Les régions qui correspondent au signal d'intérêt sont ensuite sélectionnées selon des critères souvent basés sur l'intensité, mais parfois aussi sur le contraste et la forme des régions (Andrey *et al.* (2010)). Ces méthodes produisent souvent de la sur-segmentation, c'est-à-dire que des régions sont trop divisées. Il existe plusieurs stratégies pour résoudre ce problème, telles qu'un filtrage adapté du signal avant d'appliquer l'algorithme, la sélection des marqueurs de source des bassins topologiques<sup>1</sup> (Ram *et al.* (2012)), ou la fusion des régions après l'étape de propagation. Un autre type d'approche adapté à la séparation de foyers de petite taille consiste à ajuster un modèle basé sur la connaissance *a priori* de la forme du signal (Thomann *et al.* (2002)).

En conclusion, la segmentation est une opération souvent difficile à mettre en oeuvre, qui doit être adaptée au signal. La procédure de segmentation dépend d'une part des objets que l'on cherche à détecter (forme, taille, etc.), et d'autre part du signal analysé, et notamment de la présence de bruit de fond et de défauts dans l'image. Nous avons vu qu'il est souvent nécessaire d'ajouter des informations *a priori* pour améliorer l'efficacité de la segmentation ; ceci a parfois pour conséquence de sélectionner une partie des objets, et cela n'est donc pas toujours adapté à des signaux constitués d'objets trop hétérogènes. De plus il n'est pas toujours pertinent de segmenter un signal, par exemple lorsqu'il est diffus et ne présente pas de régions distinctes de leurs alentours. Enfin, certaines procédures de segmentation sont difficiles à automatiser sur une population de noyaux et introduisent parfois un manque d'objectivité dû à l'intervention de l'utilisateur.

#### A.4.3.2 Mesures du positionnement radial

La mesure du positionnement radial fournit une information de localisation dans l'espace du noyau. Elle dépend du point de référence choisi, qui peut être soit le centre du noyau, soit son bord. Dans certains cas, il peut être pertinent de conduire ces analyses en 2D plutôt qu'en 3D : si les noyaux sont plats (comme chez les fibroblastes) les TCs s'étendent de la face basale jusqu'à la face apicale du noyau ; il

---

1. Dans ce cas, l'algorithme se nomme *region growing* ou *seeded watershed*, les graines (*seeds* en anglais) étant le point de départ de la propagation.

est donc raisonnable de considérer une analyse 2D (Bolzer *et al.* (2005)), mais dans ce cas une localisation préférentielle *périphérique* ne peut être interprétée comme une association préférentielle à la lamina, puisque celle-ci est aussi présente au niveau des faces basales et apicales du noyau. Dans tous les cas, ce type de mesures correspond à une projection de l'information 3D (ou 2D) dans un espace 1D, qui est nécessairement accompagnée d'une perte de l'information spatiale.

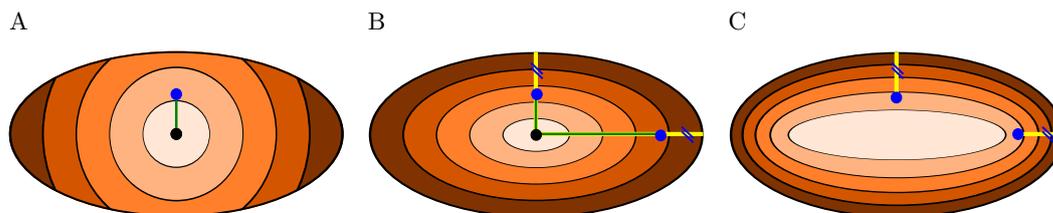


FIGURE A.24: Différentes mesures du positionnement radial. (A) Mesure à partir du centre du noyau : le positionnement du point bleu est mesuré par sa distance au centre (ligne verte). (B) Mesure à partir du centre normalisée : le positionnement d'un point bleu est mesuré par sa distance au centre (ligne verte) divisée par le rayon du noyau dans la direction du point (ligne verte + ligne jaune). Sur le schéma les deux lignes jaunes sont de même taille, donc les deux points bleus se situent à la même distance de la périphérie mais ne sont pas localisés dans la même couche. (C) Mesure à partir de la périphérie. Les couches sont de surfaces (ou volumes en 3D) égales, et deux points situés à la même distance de la périphérie se trouvent dans la même couche.

**Mesures reposant sur le centre du noyau.** Dans ce type de mesures, un point dans le noyau est localisé par le rapport entre sa distance au centre géométrique (ou au centre pondéré par les intensités de contre-marquage de l'ADN) du noyau (Fig. A.24A). Ces mesures ont l'avantage de ne pas être sensibles à la précision de la segmentation du noyau. Afin de normaliser cette mesure, cette distance absolue est souvent divisée par le rayon du noyau dans la direction du point<sup>1</sup>, et prend alors des valeurs entre 0 et 1 (Roix *et al.* (2003)). De plus, dans le cas de noyaux non sphériques, deux points se trouvant à la même distance euclidienne de la périphérie ou du centre peuvent ne pas se retrouver dans la même couche (Fig. A.24B). La pertinence de ce type de normalisation est donc discutable, notamment pour le cas de noyaux non sphériques. De plus cette mesure a tendance à sous-estimer les positions centrales, car deux couches d'épaisseur égale situées en périphérie et au centre du noyau ont des volumes très différents; donc la probabilité pour un point de se trouver en périphérie est plus élevée<sup>2</sup>.

Ces mesures ne sont pas adaptées à des objets qui ne sont pas assimilables à leurs centres, comme les TCs. Le groupe de Thomas Cremer a développé une mesure qui prend en compte l'ensemble de l'objet en pondérant la distance par l'intensité du signal fluorescent. Cette mesure repose sur la quantification du signal fluorescent dans différentes couches d'épaisseur égale<sup>3</sup>, définies comme un intervalle de distance

1. Notons que dans ce cas cette mesure devient sensible à la segmentation du noyau.

2. Considérons un noyau sphérique de rayon  $R$  : le volume de la sphère concentrique de rayon  $R/2$  représente 12.5% du volume du noyau.

3. Afin de limiter la sensibilité de la mesure au bruit de fond, celui-ci est réduit avant de quantifier le signal.

radiale (Cremer *et al.* (2001)). La distribution de signal dans les différentes couches est comparée à celle du contre-marquage de l'ADN ; les auteurs considèrent qu'une structure se localisant aléatoirement dans le noyau aura la même distribution que le contre-marquage de l'ADN - ce qui est une assertion discutable<sup>1</sup>. Cette mesure permet de déterminer une distribution d'intensité pour l'ensemble du signal, ainsi qu'une position préférentielle dans le noyau (Ronneberger *et al.* (2008)).

**Mesures reposant sur la distance à la périphérie du noyau.** Les mesures précédemment décrites ont l'inconvénient de se fonder sur le centre du noyau, qui n'a pas de sens biologique évident, et qui peut même dans certains cas se trouver en dehors du noyau. L'alternative consiste à repérer un point par sa distance aux bords du noyau. La distance euclidienne absolue à la périphérie<sup>2</sup> est une information qui peut avoir un sens biologique en soi, si la question sous-jacente est liée à l'association de l'objet d'intérêt à la périphérie du noyau. Afin d'utiliser cette mesure pour repérer la position du point d'intérêt dans l'ensemble du volume nucléaire, il est souvent utile de la normaliser afin de pouvoir la comparer à des mesures provenant de noyau différents.

Afin de normaliser la mesure, le noyau peut être découpé en couches successives de volumes (ou surfaces en 2D) égaux, depuis la périphérie vers le centre du noyau (Fig. A.24C). On repère alors un point par l'indice de la couche dans lequel il se trouve. On appelle parfois ce type de mesure *analyse en pelures d'oignon*. Dans le cas où les objets ne sont pas segmentés, il est possible de quantifier le signal fluorescent dans chaque couche<sup>3</sup> et de le comparer au contre-marquage de l'ADN ; c'est l'approche utilisée par le groupe de Wendy Bickmore (Croft *et al.* (1999)).

Dans le cas d'objets segmentés et assimilables à leur centre, une variante continue de cette mesure a été introduite, dénommée fraction de volume érodé (EVF : Eroded Volume Fraction) ; elle consiste à mesurer la fraction de volume qui se trouve entre un point d'intérêt et la périphérie du noyau (Ballester *et al.* (2008))<sup>4</sup>. L'EVF prend des valeurs qui varient entre 0 à la périphérie du noyau et 1 au centre du noyau. Cela correspond précisément à une comparaison à une distribution uniforme, donc l'EVF de points uniformément distribués dans un noyau est uniformément distribuée entre 0 et 1. Cela est valable quelque soit la forme et la taille des noyaux, donc cette mesure permet de normaliser par rapport aux variations de ces paramètres. L'intérêt de cette mesure par rapport à sa variante discontinue et qu'elle

---

1. Étant donné que l'hétérochromatine constitutive est souvent marquée plus intensément par le contre-marquage et ne se localise pas aléatoirement (voir : §A.3.2.1), le contre-marquage ne reflète pas forcément la localisation moyenne de la chromatine.

2. Sous-entendu la distance minimale entre le point considéré et le bord du noyau.

3. Il faut être prudent quant à l'influence du bruit de fond dans ce type d'approche, paramètre qui n'est à notre connaissance jamais commenté dans les études de ce groupe. Ce paramètre peut diminuer la sensibilité de la mesure, et éventuellement induire des différences artificielles entre deux expériences, s'il est globalement plus élevé dans l'une d'entre elles.

4. Notons que contrairement à ce que suggère son nom, cette mesure n'est pas calculée par érosions successives du noyau ce qui pourrait mener à des artefacts dus à l'échantillonnage spatial, mais par le calcul d'une carte de distance euclidienne.

évite d'avoir à choisir le nombre de couches (ou pelures), choix qui est en général arbitraire, et qui peut avoir une influence sur les résultats des tests statistiques. Une mesure équivalente faisant intervenir la comparaison de la distance du point d'intérêt à la périphérie à aux distances provenant d'un modèle de distribution aléatoire uniforme dans le même noyau a été proposée<sup>1</sup> (Jost *et al.* (2011)).

### A.4.3.3 Analyse de pattern de points

Ces méthodes permettent de caractériser dans son ensemble un pattern de distribution nucléaire de structures qui sont assimilables à des points, telles que des centromères ou des corps PML. Par exemple, elles permettent d'estimer si la distribution des objets est uniforme dans le noyau ou si elle est agrégée (*clustered* en anglais) (Fig. A.25A). Ces méthodes ont été développées et utilisées dans les domaines de l'écologie et de l'épidémiologie, et commencent à l'être dans le domaine de l'organisation nucléaire. Ces outils ont été adaptés au le domaine de l'organisation nucléaire (Andrey *et al.* (2010)); nous nous appuyerons sur ce travail pour la description de cette méthodologie, car nous le considérons comme le plus abouti dans ce domaine.

**Descripteurs de pattern de points.** Le terme *descripteur* est utilisé ici pour désigner une fonction mathématique qui permet de résumer l'information spatiale d'un pattern de points. En général, un descripteur permet de mettre en évidence un aspect particulier de l'information spatiale, tel que la présence d'agrégats. Plusieurs descripteurs ont été appliqués à l'étude de l'architecture nucléaire; par exemple la distribution de l'ensemble des inter-distances (quantifiée via les fonctions K et L) permettant une analyse multi-échelle de la présence d'agrégats (Beil *et al.* (2005)), ainsi que la distribution des distances au plus proche voisin (fonction G) et la fonction F qui mesure l'espace vide dans le noyau (Andrey *et al.* (2010), Fig. A.25B/C). Le fait de résumer l'information spatiale permet de comparer différents patterns de points entre eux. Cette méthodologie permet notamment de comparer un pattern observé à un modèle de distribution de points, simulé dans le même noyau. Le modèle le plus couramment utilisé est le modèle aléatoire uniforme, mais d'autres modèles ont été introduits; notamment un modèle prenant en compte la position radiale des points observés (Heride *et al.* (2010)). Le test de comparaison au modèle a pour hypothèse nulle que la distribution observée suit le modèle. Lorsque cette hypothèse est rejetée, le pattern observé ne suit pas le modèle. Dans le cas d'un modèle aléatoire uniforme, ce test permet de montrer que la distribution observée n'est pas aléatoire uniforme (en général ce résultat est désigné par *non-aléatoire*),

1. La mesure de Jost *et al.* (2011) est équivalente à l'EVF, car elle correspond au rang de la distance euclidienne à la périphérie du point observée au sein d'une distribution de distances à la périphérie de points tirés aléatoirement dans l'espace du noyau. Cette mesure correspond donc à une estimation de la probabilité de tirer un point plus proche de la périphérie que le point observé. Or cette probabilité converge vers la fraction de volume nucléaire plus proche de la périphérie que le point observé, puisque les points sont tirés aléatoirement dans l'espace du noyau. Cette fraction est exactement l'EVF.

et donc plus complexe ; ce qui suggère qu'il existe des mécanismes cellulaires responsables du positionnement non-aléatoire des objets observés. La déviation par rapport au modèle est mesurée par l'indice de distribution spatiale (SDI : Spatial Distribution Index) (Andrey *et al.* (2010)).

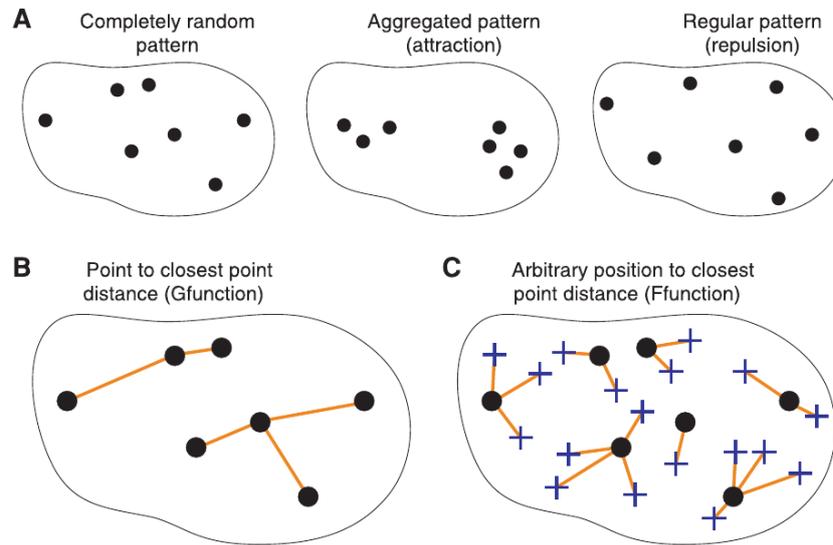


FIGURE A.25: Distributions spatiales et descripteurs (d'après Andrey *et al.* (2010)). (A) Différents type de distribution spatiale. Les points peuvent être distribués uniformément et indépendamment (pattern aléatoire uniforme), montrer une attraction mutuelle (pattern agrégé) ou une répulsion mutuelle (pattern régulier). (B) La fonction G est la distribution cumulative des distances entre chaque point et son plus proche voisin (lignes oranges). Cette fonction prend des valeurs faibles pour des patterns agrégés et élevées pour des patterns réguliers. (C) La fonction F est la distribution cumulative des distances entre des positions nucléaires arbitraires (croix bleues) et le point le plus proche (lignes oranges). Cette fonction prend des valeurs élevées lorsque le pattern est agrégé et faibles lorsque le pattern est régulier.

**Analyse à l'échelle d'une population.** Le SDI est une mesure normalisée (variant de 0 à 1) qui prend en compte les variations de forme et de taille des noyaux. Cette normalisation permet d'analyser les caractéristiques du positionnement de la structure d'intérêt à l'échelle de la population de noyaux. La distribution des SDIs d'un pattern de point conforme au modèle utilisé pour la simulation dans le noyau est uniforme, et une déviation de la distribution observée par rapport à une distribution uniforme signifie que les points ne sont pas distribués conformément au modèle. Il est parfois possible d'interpréter le sens de la déviation<sup>1</sup> (Andrey *et al.* (2010)).

L'apport de cette méthodologie est double : elle permet

- De tester la conformité à un modèle, et ceci à l'échelle de la population de noyaux.
- De normaliser les données à l'échelle d'une population afin de comparer plusieurs populations entre elles<sup>2</sup>.

1. Dans le cas des fonctions F et G une déviation peut correspondre soit à une attraction soit à une répulsion.

2. en utilisant le même modèle pour chaque population.

#### A.4.3.4 Analyse relative

Nous décrivons dans cette section une catégorie très générale d'analyse qui permet de mettre en évidence des relations spatiales entre des objets localisés dans le noyau et qui n'ont pas été traités dans les sections précédentes. Il est difficile de prendre en compte l'ensemble des problèmes qui existent en biologie, car un problème particulier nécessite souvent d'adapter ou de développer une mesure. A titre d'exemple, nous nous limiterons donc à deux problèmes qui sont souvent rencontrés : l'analyse de l'attraction/répulsion entre plusieurs objets segmentés et la co-localisation.

#### Mesure d'attraction/répulsion.

**Contact entre deux allèles.** Ce problème est très souvent rencontré en biologie ; par exemple il a été montré qu'un contact transitoire entre le centre d'activation des deux chromosomes X est nécessaire à l'activation d'un des deux chromosomes X (Bacher *et al.* (2006)). Si la question posée est la recherche d'un contact, il est possible de définir un seuil au dessous duquel les deux loci sont considérés comme étant en contact, puis d'analyser les événements *contact* et *absence de contact* à l'échelle de la population. Notons que la définition du seuil de contact n'est pas une question évidente ; il peut par exemple prendre en compte la taille caractéristique des loci, et leur mobilité à une certaine échelle temporelle. Il s'agit d'une approche *discrète*, par opposition à une approche *continue*, dans laquelle l'ensemble des distances est analysée, et qui permet de mesurer une attraction ou une répulsion. L'approche continue peut nécessiter une normalisation des distance afin de pouvoir comparer des noyaux de tailles différentes. Cette normalisation est en général préconisée lorsque la distribution des distances corrèle avec le volume des noyaux. L'approche la plus simple est de diviser par le rayon sphérique équivalent du noyau<sup>1</sup>. Cependant, cette approche n'est pas toujours adaptée car elle considère uniquement les variations de taille et non les variations de forme du noyau. Une alternative est d'utiliser la méthodologie décrite en §A.4.3.3, en utilisant comme descripteur la distance entre les deux loci, cela permet de déterminer une attraction/répulsion en prenant pour référence un modèle aléatoire. Les variations de forme et de taille entre les noyaux sont intrinsèquement prises en compte. Il est souvent utile de comparer les données concernant les loci d'intérêt à des contrôles négatifs, dans ce cas des loci présents sur le chromosome X et choisis aléatoirement. Comme nous l'avons vu au §A.4.2.2, ces approches statiques ne permettent de mettre en évidence qu'une attraction ou une répulsion ; une étude sur cellules vivantes est indispensable pour démontrer le caractère nécessaire du contact.

**Contact entre deux paires de loci.** Le problème se complexifie lorsqu'il s'agit de mesurer l'attraction/répulsion entre deux paires de loci A et B<sup>2</sup>, puis-

1. Soit  $V$  le volume du noyau et  $R$  le rayon sphérique équivalent :  $R = \sqrt[3]{\frac{3}{4\pi} \times V}$ .

2. Chaque paire correspondant aux deux allèles du même locus

qu'un allèle du locus A peut s'associer avec les deux allèles du locus B ; il y a donc quatre inter-distances à considérer. Comme pour le cas précédent, il est possible d'adopter une approche discrète ou une approche continue. Dans le cas de l'approche continue, un descripteur normalisé a été introduit, basé sur la comparaison entre les distances entre hétérologues et homologues (Ronneberger *et al.* (2008)). Cette normalisation suppose qu'en l'absence de répulsion ou d'attraction entre les hétérologues, les distances hétérologues devraient être à peu près équivalentes aux distances homologues, supposition qui pourrait être remise en question par le fait que les TCs homologues ont tendance à se repousser (Heride *et al.* (2010)). Comme pour le cas précédent, il est aussi possible d'exploiter la méthodologie décrite en §A.4.3.3, par exemple en utilisant comme descripteur la distance minimale entre les hétérologues si l'on cherche à mettre en évidence une attraction.

**Contact entre un plus grand nombre d'objets.** Des méthodes spécifiques ont été développées afin de déterminer si plusieurs objets, par exemple des gènes (G), ont tendance à s'associer à un nombre similaire ou supérieur d'un autre type d'objets, par exemple des corps nucléaires (C). Si le nombre d'objets de type G et C est grand, des méthodes permettent de quantifier l'association ou l'exclusion entre les objets des deux types. Elles sont par exemple implémentées pour le logiciel R (Baddeley *et Turner* (2005)), et ont été utilisées dans des études portant sur l'organisation nucléaire (Beil *et al.* (2005)). Si le nombre d'objet est faible, ces méthodes ne sont plus applicables. Un test a été développé dans le cadre de l'étude de l'association de gènes à des corps PML ; il est basé sur la comparaison de la distance entre chaque objet G et l'objet C le plus proche et les distances au plus proche voisin entre objets de type C (Shiels *et al.* (2001)). De manière alternative, un test de permutation peut être utilisé, dans lequel le type (G ou C) des objets est réattribué aléatoirement dans des simulations, en conservant le nombre d'objets de type G et de type C observés. L'association est caractérisée par un descripteur basé sur la distance entre chaque objet G et l'objet C le plus proche, et le descripteur de la distribution observée est comparé aux simulations (Ronneberger *et al.* (2008)). Enfin il est aussi possible d'utiliser la méthodologie décrite en §A.4.3.3, en utilisant un descripteur caractérisant l'association entre les objets G et C (par exemple la distribution de distance entre chaque objet B et l'objet C le plus proche), et en tirant aléatoirement les positions des objets de type G dans les simulations.

**Colocalisation.** La colocalisation est une mesure qui permet de déterminer si deux signaux différents sont localisés au même endroit du noyau. Cette approche n'est pas adaptée pour détecter des interactions moléculaires, car la résolution de la microscopie optique (même la super-résolution) est insuffisante pour identifier l'apposition physique de deux molécules via la comparaison de leur distribution dans des images de fluorescence. La colocalisation de deux marqueurs n'indique pas nécessairement qu'ils se localisent dans les mêmes structures, mais l'observation répétée de leur colocalisation au sein d'une cellule et entre plusieurs cellules augmente la confiance dans cette interprétation. Des analyses exploitant des techniques

à plus haute résolution telles que la microscopie électronique ou la FRET (Fluorescence Resonance Energy Transfer) peuvent être nécessaires pour confirmer une colocalisation à l'échelle moléculaire. Il existe deux grandes familles de mesures de colocalisation, selon si les signaux considérés sont segmentés ou non. Pour les explications qui suivent, nous considérerons la colocalisation de deux signaux nommés A et B.

**Colocalisation voxel-à-voxel.** Ces approches sont particulièrement utiles lorsqu'il n'est pas pertinent de segmenter les signaux, par exemple parce que leurs formes sont très hétérogènes. De plus elles peuvent dans certains cas s'avérer plus reproductibles que les approches basées sur la segmentation, car elles ne dépendent pas de celle-ci. Deux questions différentes peuvent être traitées par ces approches :

- Les marqueurs se localisent-ils au même endroit ? Il s'agit d'une information de co-occurrence spatiale ; dans ce cas la mesure va quantifier le recouvrement spatial.
- Existe-t-il une relation entre leurs intensités ? Il s'agit alors d'une information de corrélation, qui quantifie un lien entre les concentrations locales des objets marqués. Notons qu'il peut y avoir une co-occurrence en l'absence de corrélation.

De nombreuses mesures ont été développées ; il n'en existe pas qui permettent de répondre à toutes les situations, car elles sont sensibles à des effets différents. Le coefficient de corrélation de Pearson est une des mesures les plus utilisées ; il prend en compte à la fois co-occurrence et corrélation. Si les deux marqueurs présentent une co-occurrence mais pas de relation linéaire, ce coefficient sera plus faible. Plusieurs autres coefficients moins sensibles à cet effet ont été proposés (pour revue : [Bolte et Cordelières \(2006\)](#)). De manière générale, ces mesures caractérisent une colocalisation globale de l'image : dans le cas d'une forte corrélation ou anti-corrélation, il est possible de conclure à une colocalisation ou exclusion des marqueurs ; en revanche lorsque les valeurs de ces mesures sont faibles, elles ne permettent pas de faire la distinction entre une absence de relation spatiale et une colocalisation partielle.

**Colocalisation entre objets segmentés.** Lorsqu'il est pertinent de segmenter les deux signaux considérés, il existe différentes manières de mesurer la colocalisation. La mesure la plus intuitive repose sur la quantification du volume de recouvrement (*overlap* en anglais) entre les objets segmentés des deux signaux, rapporté au volume total de A ou de B ; ces coefficients appelés coefficients de Manders représentent donc la fraction de recouvrement. Cette mesure est normalisée et peut permettre de comparer des populations ; en revanche il est difficile d'en tirer une information absolue en cas de superposition partielle, car il est difficile de réaliser un test de comparaison à une référence interne. Certaines mesures ont été développées, permettant de faire la distinction entre l'exclusion totale des marqueurs, un recouvrement complet des marqueurs, et un recouvrement partiel qui peut être dû soit à une superposition parfaite de seulement une fraction des objets, soit à une superposition partiels des objets. Ces approches utilisent des critères de colocalisa-

tion basés sur le centre de chaque objet, et comptabilisent la fraction d'objets qui respectent le critère de colocalisation (Bolte et Cordelières (2006)). Elles sont en général très sensibles à la séparation des objets lors de la segmentation, ainsi qu'à l'hétérogénéité morphologique des objets.

## Conclusion

Les mesures que nous avons présentées permettent d'une part de déterminer des propriétés spatiales intrinsèques d'une population<sup>1</sup>, il s'agit alors d'une information *absolue*, et d'autre part de comparer plusieurs populations pour montrer une éventuelle différence<sup>2</sup>, dans ce cas il s'agit d'une information *relative*. Une information absolue s'obtient généralement par comparaison à une référence (souvent issue d'un modèle aléatoire). Par exemple, dans le cas de l'analyse d'un pattern de point, la comparaison à un modèle aléatoire uniforme permet de déterminer si le pattern est régulier, agrégé ou uniforme (voir §A.4.3.3).

Afin d'obtenir une mesure qui peut être comparée entre plusieurs noyaux ou plusieurs populations de noyaux, il est crucial qu'elle soit normalisée, car les noyaux sont souvent très variables en terme de forme, de taille, et d'intensité des marquages<sup>3</sup>. Notons que toutes les mesures ne sont pas sensibles aux variations de ces paramètres ; la normalisation doit donc être adaptée à chaque mesure. De manière générale, la normalisation s'effectue soit sur la base de paramètres déterminés sur l'ensemble de la population, soit au niveau de chaque noyau individuellement. La seconde manière de normaliser permet de prendre en compte les variations individuelles des noyaux. Cependant même dans ce cas, certains types de normalisation ne prennent pas en compte les variations de forme des noyaux ; c'est le cas par exemple de la normalisation de la distance radiale présentée dans la figure A.24B. Une manière de prendre en compte l'ensemble des variations morphologiques d'un noyau est de comparer la mesure à un modèle aléatoire évalué dans le même noyau, qui sert de référence ; la mesure normalisée correspond alors à la déviation par rapport au modèle. C'est l'approche utilisée par Andrey *et al.* (2010) avec les outils de statistique spatiale, ainsi que par Ballester *et al.* (2008) avec l'EVF, que nous utiliserons dans ce travail.

---

1. Par exemple si un objet a tendance à se localiser plutôt vers l'intérieur ou la périphérie du noyau dans le cas du positionnement radial.

2. par exemple comparer le positionnement de deux objets

3. Les variations en intensité peuvent être diminuées en uniformisant les conditions d'acquisition, mais ce n'est pas toujours le cas : il existe souvent une forte variation entre noyaux, indépendamment des conditions d'acquisition, qui peut être due par exemple à des variations stochastiques de l'expression de certains gènes.

## A.5 Projet de thèse

La question principale qui anime ce projet concerne l'implication des RCs dans l'organisation spatiale du génome. Fournir des éléments de réponse à cette question nécessite de donner une description précise de l'organisation nucléaire des RCs. La définition d'observables est nécessaire pour permettre l'étude des conséquences fonctionnelles d'inactivation (ou de surexpression) de protéines candidates. Plusieurs aspects de l'organisation spatiale des RCs humaines nécessitent d'être précisés :

- Comment varie le positionnement de RCs spécifiques, notamment vis-à-vis des nucléoles et de la périphérie du noyau, au cours du cycle cellulaire ?
- Existe-t-il des structures comparables aux chromocentres murins chez l'homme ?

Nous avons organisé notre travail en trois axes :

- La conception d'un outil d'analyse d'image permettant d'étudier différents aspects de l'organisation spatiale sur des grandes populations de cellules (§C.1).
- L'étude du positionnement de RCs spécifiques humaines vis-à-vis de la périphérie du noyau et des nucléoles (§C.2).
- Le développement de nouvelles approches pour l'étude de l'agrégation des RCs de manière globale, via l'utilisation d'un marqueur de l'ensemble des régions péri-centromériques humaines (§C.3).

Notre portera une attention particulière sur la définition des observables reproductibles permettant une description précise du positionnement et de l'agrégation des RCs. Étant donné que ces aspects de l'organisation nucléaire des RCs sont susceptibles de varier au cours du cycle cellulaire, nous le prendrons systématiquement en compte ; d'une part pour diminuer la variabilité de nos observables, mais aussi pour étudier leur dynamique au cours du cycle cellulaire.

# Matériels et Méthodes

---

## B.1 Culture Cellulaire

Dans ce travail, nous avons étudié deux types cellulaire humain : des cellules lymphoblastoïdes, et des cellules mésenchymateuses souches (hMES).

**Cellules lymphoblastoïdes.** Les lignées cellulaires lymphoblastoïdes humaines GM06990 et GM12878 sont cultivées sous agitation (200 rpm) dans des flasques Cellspin (Integra Biosciences) de 100 ml, contenant du milieu RPMI1640-GlutaMAX (Life Technologies) supplémenté avec 15% de SVF (Eurobio). Les cellules sont régulièrement diluées au  $1/2$  tous les deux jours, et gardées en culture 3 semaines au maximum.

Afin de permettre le marquage des cellules en phase S, le marquage de la répliation est obtenu en ajoutant du BrdU (Sigma-Aldrich) au milieu de culture à une concentration finale de 10  $\mu$ M pendant 30 min juste avant la fixation.

**Cellules hMES.** Les cellules hMES sont cultivée dans des flasques T75 traitées avec une solution 0.01% de gélatine (Sigma-Aldrich) contenant 10 ml de milieu  $\alpha$ -MEM complémenté avec 10% de SVF (Eurobio), dans une étuve humide à 37°C contenant 5% de CO<sub>2</sub>. Les cellules sont passées deux fois par semaine (diluées au  $1/5$ ). Afin d'être passées, les cellules sont lavée au PBS 1X puis traitée à la trypsine pendant 1 min à 37°C.

## B.2 Marquages *in situ*

### B.2.1 Préparation des échantillons

Les cellules sont ensemencées sur des lamelles 22×22 mm d'épaisseur moyenne de 0.17±0.01 mm (Thermo Fisher Scientific Menzel) nettoyées au préalable dans un plasma cleaner Femto v4 (Diener electronic GmbH).

**Cellules lymphoblastoïdes.** Les cellules sont centrifugées et resuspendues dans du PBS 1X à une concentration finale de  $2 \times 10^6$  cellules/ml. 300 $\mu$ l de la suspension sont déposés sur une lamelle traitée au préalable avec une solution de 0.01% poly-L-lysine (Sigma-Aldrich). Après 5 min d'incubation à température ambiante (TA) pour laisser les cellules adhérer, les cellules sont brièvement incubées dans une

solution de PSB 0.3X avant d'être fixées dans une solution de paraformaldéhyde (PFA) 4% dilué dans du PBS 0.3X pendant 10 min, puis lavées au PBS 1X (3 x 5 min).

**Cellules hMES.** Les lamelles sont traitées avec une solution 0.01% de gélatine (Sigma-Aldrich) durant 5 min. La gélatine est aspirée et 300  $\mu$ l d'une solution à 150000 cellules/ml dans du milieu de culture sont déposés sur une lamelle sans laisser sécher la gélatine. Afin de leur permettre d'adhérer, les cellules sont mises en incubateur pendant 1h puis 2 ml de milieu sont ajoutés sur la lamelle dans une plaque 6 puits. Après une nuit d'incubation, les cellules sont lavées au PBS 1X et fixées dans une solution de paraformaldéhyde 4% dilué dans du PBS 1X pendant 10 min, puis lavées à nouveau au PBS 1X (3 x 5 min).

**Stockage.** Les cellules fixées sont stockées à 4°C pendant au plus un mois. Le jour de l'expérience, afin de faciliter les manipulations, les lamelles sont collées sur des lames de verre 26x76 mm avec de la colle *rubber cement*.

### B.2.2 Immuno-FISH

Cette expérience permet de détecter simultanément des protéines et des séquences d'ADN. Il s'agit d'une expérience d'immunocytochimie suivie d'une expérience de FISH, intercalées d'une fixation.

#### B.2.2.1 Sondes

Les sondes oligonucléotidiques LNA proviennent de Eurogentec (Belgique). Les peintures chromosomiques proviennent de Metasystems (Allemagne). Les clones bactériens contenant les BACs proviennent de Invitrogen. Les préparations de BACs sont obtenues en utilisant le kit Quiagen large-construct. Les sondes à amorçage aléatoire (*random priming*) sont synthétisées en mélangeant 100 ng (dans 1  $\mu$ l) de BAC, 20  $\mu$ l de solution 2.5X de tampon et d'amorces du kit Bioprime (Invitrogen) (125 mM Tris Hcl pH 6.8, 12.5 mM MgCl<sub>2</sub>, 25mM  $\beta$ -mercaptoethanol, et 750 ng/ $\mu$ l d'octamères aléatoires) et 20  $\mu$ l d'eau bidistillée. Les échantillons sont dénaturés en chauffant à 100°C pendant 8 min, puis placés sur glace pendant 10 min. Après l'addition de 5  $\mu$ l de dNTP (dATP, dCTP, dGTP à 1 mM et dTTP à 0.65 mM), 3  $\mu$ l de dUTP modifié à 1mM and 1  $\mu$ l (40U) d'ADN polymérase fragment de Klenow (Invitrogen), les échantillons sont incubés sur la nuit à 37°C. Puis les tubes sont incubés à 74°C pendant 3min, et précipités par addition d'acétate de sodium et d'éthanol. Le culot séché est resuspendu dans 50  $\mu$ l de Tris 10 mM à pH 7.4, et la concentration est mesurée en utilisant un Nanodrop.

#### B.2.2.2 Protocole

**Immunocytochimie.** Les cellules sont perméabilisées dans une solution 0.1% Triton X-100/PBS 1X (PBST) pendant 5 min à TA, puis lavées deux fois dans du

PBS 1X. Après une étape de saturation dans une solution 1.5% *blocking reagent* (Roche Applied Science), l'immuno-marquage est effectué en incubant les anticorps primaires pendant 1 h à TA en chambre humide, en lavant 3 fois au PBST, en incubant les anticorps secondaires 45 min à TA en chambre humide dans le noir puis en lavant 3 fois au PBST.

Tous les anticorps sont dilués dans une solution de saturation contenant 1% de *blocking reagent* et 0.05% de Tween 20 dilués dans du SSC 2X à pH 7.0. Les anticorps primaires sont les suivants :

- CenpA (3-19,  $1/200$ , mouse monoclonal antibody, Abcam) pour la détection des centromères
- Nucleolin/C23 (H-250,  $1/400$ , rabbit polyclonal antibody, Santa Cruz) pour la détection des nucléoles
- Ki67 (S5,  $1/200$ , mouse monoclonal antibody, Millipore) pour la détection des cellules en cycle.

Les anticorps secondaires sont les suivants :

- FITC-conjugated goat anti-rabbit antibody (Jackson ImmunoResearch)
- FITC-conjugated goat anti-rat antibody (Jackson ImmunoResearch)
- Alexa488-conjugated goat anti-mouse antibody (Molecular Probes)
- Cy3-conjugated goat anti-rabbit antibody (Jackson ImmunoResearch)
- Cy3-conjugated goat anti-mouse antibody (Jackson ImmunoResearch)
- Cy5-conjugated goat anti-rabbit antibody (Jackson ImmunoResearch)
- Alexa647-conjugated goat anti-rat antibody (Jackson ImmunoResearch)
- DL649-conjugated donkey anti-mouse antibody (Jackson ImmunoResearch)
- Alexa750-conjugated goat anti-rabbit antibody (Molecular Probes).

Ils sont tous dilués  $1/200$  dans la solution de saturation.

**FISH.** Après une étape de post-fixation qui consiste à traiter les cellules avec 40  $\mu$ l d'une solution de PFA 2% PBS 1X pendant 10 min, puis rincées au PBST (3X 5 min), les cellules sont perméabilisées dans une solution de Triton X-100 0.5% PBS 1X, rincées 3 fois, traitée dans une solution d'HCl 0.1N pendant 2 min (précisément), rincées 2 fois, équilibrée dans une solution de SSC 2X, puis incubées dans une solution de 50% formamide désionisé diluée dans du SSC 2X à pH 5.6, pendant au moins 30 min avant l'hybridation. Pour la FISH, les solutions d'hybridation dépendent des sondes :

- les sondes oligonucléotidiques sont diluées à concentration finale de 0.1  $\mu$ M dans une solution d'hybridation contenant 50% de formamide deionisé, solution Denhart 1X, sulfate dextran 10% , SDS 0.1% et dilué dans un tampon SSC 2X à pH 5.6.
- pour les sondes BAC : 300ng de chaque sonde sont précipités à l'éthanol en présence d'un ratio 10:1 d'ADN Cot1 (Roche), puis resuspendues dans le tampon d'hybridation.
- dans le cas des peintures chromosomiques, les sondes oligonucléotidiques sont ajoutées directement dans le tampon commercial à concentration finale de

0.1 $\mu$ M.

45  $\mu$ l de la solution d'hybridation (10 $\mu$ l dans le cas des peintures chromosomiques) sont déposés sur la lamelle et recouverts d'une autre lamelle. Les lamelles sont chauffées 3 min à 85°C, puis doucement refroidies à 37°C à un taux de 1°C/s, dans un appareil de PCR *in situ* (MJ Research) humidifié. L'incubation à 37°C est arrêtée après 2 min pour les sondes oligonucléotidiques et prolongée en chambre humide sur la nuit pour les sondes BAC et les peintures chromosomiques. Dans le cas des hybridations prolongée, les deux lamelles sont scellées avec de la colle (*rubber cement*) afin d'éviter la dilution de la solution de sonde par l'atmosphère humide. Ensuite, la lamelle est lavée deux fois dans une solution de SSC 2X pH 6.3 à 63°C (72°C pour certaines sondes, voir C.1)

Pour la révélation des sondes non directement marquées, les préparations sont incubées dans une solution de saturation (BSA 4%, Tween20 0.05%, PBS 1X) pendant 30 min à 37°C en chambre humide afin de réduire les interactions non spécifiques. Ensuite, selon les combinaisons de marquage, les anticorps/protéines suivantes sont utilisées :

- Alexa488-conjugated streptavidin (1/200, Molecular Probes)
- Cy3-conjugated streptavidin (1/200, Caltag Laboratories)
- Cy5-conjugated streptavidin (1/200, Caltag Laboratories)
- FITC-conjugated sheep anti-digoxigenin (1/200, Roche)
- Rhodamine-conjugated sheep anti-digoxigenin (1/200, Roche)

Tous ces anticorps/protéines sont dilués dans une solution de saturation (BSA 4%, Tween20 0.05%, PBS 1X) La solution utilisée pour tous les lavages est 0.05% Tween20 SSC 2X pH 6.3. Les anticorps sont incubés pendant 20 min à 37°C en chambre humide.

**Détection du BrdU.** Pour la détection du BrdU, l'anticorps anti-BrdU (rat monoclonal antibody, 1/200, Roche) est ajouté directement dans la solution de révélation, et une autre incubation en présence d'un anticorps anti-rat (FITC-conjugated / Alexa647-conjugated, goat, Jackson ImmunoResearch) dans la solution de saturation à 20 min à 37°C en chambre humide.

**Marquage de l'ADN et montage.** L'ADN est marqué en incubant chaque lamelle dans un puits contenant 1 ml de solution de Hoechst 33342 (Thermo Scientific) à 4  $\mu$ M pendant 30 min. Après un bref rinçage au PBS 1X, les lamelles sont montées en enlevant la colle *rubber cement*, et en retournant les lamelles 22x22 mm sur une lamelle propre 26x76 mm sur laquelle une goutte de milieu de montage PPD8 a été déposée (Sigma-Aldrich, solution à 1 mg/ml dans un mélange de PBS 1X / glycerol 90% à pH 8 ajusté avec un tampon carbonate/bicarbonate). Les lamelles sont scellées avec du vernis, conservées à 4°C à court terme et à -20°C à long terme.

### B.2.2.3 Optimisations

Plusieurs étapes du protocole de FISH peuvent gêner la détection des épitopes, notamment le traitement à l'acide, et la dénaturation en présence de formamide. Ce protocole ne permet pas de détecter toutes les protéines, mais il a fonctionné pour les protéines CenpA, Ki67 et C23 que nous utilisons dans cette étude. Nous avons ajusté le protocole en mesurant 3 paramètres :

- les volumes des noyaux pour les cellules lymphoblastoïdes<sup>1</sup> : nous avons cherché à obtenir des distributions de volumes de noyaux reproductibles et proches de mesures réalisées sur cellules vivantes. Beaucoup de facteurs influencent ce paramètre, notamment : le protocole de fixation et le milieu de montage<sup>2</sup>. Nous avons aussi observé que l'utilisation de BSA pour saturer au lieu du *blocking reagent* lors du protocole d'immunocytochimie augmentait le volume des noyaux.
- l'intensité et la spécificité du marquage des protéines, qui est notamment réduite par le traitement HCl
- la localisation nucléaire des protéines : certaines protéines (telles que C23) ont parfois tendance à se délocaliser. Nous avons observé que le choc hypotonique réalisé juste avant la fixation préservait la localisation de cette protéine.

## B.3 Acquisition des images

Les acquisitions sont réalisées en utilisant un microscope inversé à épifluorescence de marque Zeiss (Axio Observer Z1), équipé d'un objectif à immersion à huile plan-apochromat 63x 1.4 NA, et des filtres suivants :

Fluorophore/Colorant	Excitation	Dichroïque	Émission	Source lumineuse
Hoechst	G365	FT395	BP445/50	LED : 365nm
FITC/Alexa488	BP470/40	FT495	BP525/50	LED : 470nm
Cy3/Rhodamine	BP546/10	FF555	BP583/22	HXP120 Metal Halide
Texas Red	BP580/23	FF605	BP628/32	HXP120 Metal Halide
Cy5	BP643/20	FF660	BP684/24	LED : 625nm
Cy7	vide	T760LP	ET810/90	LED : 625nm

L'huile d'immersion utilisée a pour indice 1.518 à 23°C. Les piles d'images contiennent en général 40 à 50 sections de 1344x1024 pixels (de taille 0.1  $\mu\text{m}$ ), acquises à 0.23  $\mu\text{m}$  d'interval avec une caméra CCD ORCA R2 ayant >70% d'efficacité quantique (Hamamatsu). Pour chaque section optique, les images sont acquises séquentiellement pour chaque canal de fluorescence.

Le décalage chromatique est corrigé en mesurant la fluorescence de billes Tetraspeck (Invitrogen). Ces billes nous ont aussi permis d'estimer une résolution maximale pour le système de 0.2  $\mu\text{m}$  en xy et 0.5  $\mu\text{m}$  en z.

1. En effet les noyaux de ces cellules semblent très déformables, et sont très sensibles au protocole de marquage.

2. Nous avons observé qu'un milieu durcissant pouvait réduire les volumes.

Chaque champ contient en général 10 à 20 noyaux. L'acquisition des noyaux se fait de manière aléatoire sur l'ensemble de la lamelle, afin d'obtenir un échantillon représentatif de la population de noyaux, sauf lorsque cela est précisé (par exemple pour augmenter le nombre de cellules dans une certaine phase du cycle cellulaire sans avoir à acquérir trop de champs de microscopie).

## B.4 Analyse statistique

### B.4.1 Test statistique

Le test de Kolmogorov-Smirnov à deux échantillons (K-S<sub>2</sub>) est utilisé pour la comparaison de deux distributions expérimentales. Le test de Kolmogorov-Smirnov à un échantillon (K-S<sub>1</sub>) est utilisé pour la comparaison d'une distribution expérimentale à une distribution théorique, sauf mention contraire la distribution théorique utilisée est la distribution uniforme. Ce test non paramétrique est utile pour comparer des distributions qui ne sont pas normales, car il est à la fois sensible aux différences de forme et de localisation entre deux distributions. Le seuil de signification (risque  $\alpha$ ) est fixé à 0.01.

### B.4.2 Distributions cumulatives

Les distributions sont en général représentées par des graphes de distribution cumulative, qui sont pratiques pour comparer plusieurs distributions. Ce type de graphe est associé à la statistique du test de Kolmogorov-Smirnov, qui quantifie l'écart maximal entre deux distributions cumulatives.

La courbe cumulative d'une distribution uniforme est une droite d'équation  $y = x$ .

CHAPITRE C

# Résultats

---

## Sommaire

---

<b>C.1 Développement d'outils pour l'analyse d'image quantitative</b>	<b>105</b>
C.1.1 TANGO : un outil générique pour l'analyse d'image 3D haut-débit . . . . .	105
C.1.2 Procédures de segmentations . . . . .	113
C.1.3 Mesures . . . . .	118
C.1.4 Analyse quantitative du cycle cellulaire . . . . .	121
C.1.5 Analyse du signal Ki67 . . . . .	122
C.1.6 Discussion et conclusion . . . . .	126
<b>C.2 Positionnement des régions centromériques humaines . . . . .</b>	<b>128</b>
C.2.1 Introduction . . . . .	129
C.2.2 Résultats . . . . .	131
C.2.3 Discussion . . . . .	140
C.2.4 Matériel supplémentaire . . . . .	143
<b>C.3 Agrégations des régions centromériques humaines . . . . .</b>	<b>148</b>
C.3.1 Résultats . . . . .	148
C.3.2 Discussion . . . . .	163

---

## C.1 Développement d'outils pour l'analyse d'image quantitative

Afin de pouvoir mener une étude quantitative de l'organisation nucléaire des régions centromériques, nous avons développé un logiciel de traitement et d'analyse d'image modulaire (§C.1.1), permettant la mise au point de procédures de segmentation adaptées aux signaux que nous avons étudié (§C.1.2), et le développement de mesures permettant de répondre aux questions que nous nous sommes posés (§C.1.3). Enfin nous avons développé une procédure d'analyse quantitative du cycle cellulaire (§C.1.4).

### C.1.1 TANGO : a generic tool for high-throughput 3D image analysis for studying nuclear organization

Dans cette partie, nous décrivons TANGO (Tools For Analysis of Nuclear Genome Organization), logiciel d'analyse d'image que nous avons développé.

Ce travail a fait l'objet d'une publication dans la revue *Bioinformatics* (Ollion *et al.* (2013)). La publication est insérée ci-après telle qu'elle a été soumise (à deux références près). Afin d'homogénéiser ce manuscrit et de simplifier sa lecture, l'article a été mis en page de la même manière que le reste du manuscrit, et les figures du matériel supplémentaire ont été insérées dans le texte, à l'exception des tableaux qui se trouvent en annexe.

**Résumé de l'article.** TANGO est un logiciel d'analyse d'image dédié à l'étude de l'architecture nucléaire. Il permet de réaliser une procédure d'analyse complète d'analyse d'image 3D provenant de microscopie de fluorescence en combinant deux environnements : ImageJ pour le traitement et l'analyse d'image et R pour l'analyse statistiques des résultats des mesures. Ces environnements sont tous deux très modulaires, libres, et très utilisés dans la communauté scientifique. Ainsi le potentiel de ces deux environnements est directement accessible lors de l'utilisation de TANGO.

Sa généricité est due a sa modularité, il contient en particulier des collections importantes :

- d'opérations de traitement 3D qui peuvent être combinées pour adapter des procédures de traitement à un signal donné
- de mesures 3D qui peuvent facilement être paramétrées.

Ces collections peuvent être enrichies grâce à un système de greffons.

Il est conçu pour permettre le traitement et l'analyse automatique de grande quantités d'images, tout en permettant une intervention manuelle ou semi-automatique de l'utilisateur qui est accélérée et facilitée par une interface d'exploration des données reflétant la structure hiérarchique des objets biologiques étudiés : les noyaux, les structures incluses dans les noyaux et les objets segmentés inclus dans les structures.

Enfin une des forces de ce logiciel est l'interfaçage à double sens avec le logiciel R :

- d'une part les résultats des mesures peuvent être importés directement dans R
- d'autre par des sous-populations d'objets peuvent être définies dans R et accessibles dans TANGO, ce qui permet par exemple de visualiser les images associées.

Toutes ces caractéristiques font de TANGO un outil performant et adapté à l'étude de l'organisation nucléaire par une approche d'imagerie de fluorescence haut-débit.

### C.1.1.1 Abstract

**Motivation :** The cell nucleus is a highly organized cellular organelle that contains the genetic material. The study of nuclear architecture has become a very important field of cellular biology. Extracting quantitative data from three-dimensional (3D) fluorescence imaging helps understand the functions of different nuclear compartments. However, such approaches are limited by the requirement for processing and analyzing large sets of images.

**Results :** Here we describe TANGO (Tools for Analysis of Nuclear Genome Organization), an image analysis tool dedicated to the study of nuclear architecture. TANGO is a coherent framework allowing biologists to perform the complete analysis process of 3D fluorescence images by combining two environments : ImageJ<sup>1</sup> (Schneider *et al.* (2012)) for image processing and quantitative analysis and R<sup>2</sup> (R (2013)) for statistical processing of measurement results. It includes an intuitive user interface providing the means to precisely build a segmentation procedure and set-up analyses, without possessing programming skills. TANGO is a versatile tool able to process large sets of images, allowing quantitative study of nuclear organization.

**Availability :** TANGO is composed of two programs : (1) an ImageJ plug-in ; (2) a package (rtango) for R. They are both free and open source, available<sup>3</sup> for Linux, Microsoft Windows and Macintosh OSX. Distribution is under the GPL v.2 licence.

### C.1.1.2 Introduction

The nucleus is a heterogeneous and compartmentalized organelle. The spatial localization of genes and nuclear compartments within the nucleus affects cellular functions, such as regulation of gene expression, DNA replication and repair (Austin *et Bellini* (2010)). The underlying mechanisms are still a matter of extensive investigation.

Nuclear organization can be studied using several imaging techniques. For example immunocytochemistry and Fluorescence *In Situ* Hybridization (FISH), allow the selective fluorescent labelling of protein-enriched compartments and chromosomal domains, respectively. Though recent microscopes facilitate the acquisition of large sets of 3D images, the automatic processing and analysis of those images remains a difficult task.

Several software programs have already been developed. CellProfiler (Kamentsky *et al.* (2011)), which is one of the most popular, only performs 2D analysis.

---

1. <http://imagej.nih.gov/ij/>  
2. <http://cran.r-project.org>  
3. <http://biophysique.mnhn.fr/tango>

FishFinder (Shirley *et al.* (2011)) and NEMO (Iannuccelli *et al.* (2010)), perform 3D analysis but have limited image processing and analysis capabilities.

To overcome these limitations, we have developed TANGO, an image analysis framework, which permits biologists to perform the complete analysis process of 3D fluorescence images : image processing, quantitative analysis and statistical processing. A potential application is the study of radial position variations of various structures (such as genes or centromeres) with respect to the nuclear membrane or to nucleoli, throughout the cell-cycle.

### C.1.1.3 Image Processing

**Data Organization and General Workflow.** The input data of TANGO consists of multichannel Z-stack images of microscopy fields, containing one or several nuclei. Each channel corresponds to a fluorescent marker and can be associated with a biological compartment of interest (further termed structure). All image files compatible with Bio-formats<sup>1</sup> (more than 100 formats), can be imported to TANGO. The processing workflow is composed of two successive steps (Fig. C.1) : segmentation of nuclei, followed by segmentation of structures within each nucleus (when needed). This results in a hierarchical organization of data : fields contain nuclei, that contain structures, that can contain segmented objects. This pre-defined workflow simplifies the workload : TANGO only requires specification of the number of channels, then the definition of associated structures. For performance and usability all images and parameters are stored in a MongoDB database, a highly powerful document-oriented system<sup>2</sup>.

**Segmentation/Processing Chains.** Segmentation, i.e. detection of nuclei and nuclear structures is achieved by applying a processing chain to the associated image. A processing chain is composed of three steps : pre-processing, segmentation and post-processing, as detailed in Figure C.2. Each step is a combination of one or several processing operations. TANGO contains rapid 3D implementations of more than thirty commonly used filters (such as top-hat, Laplacian of Gaussian or morphological filters) and several segmentation algorithms (such as seeded watershed or hysteresis thresholding) including those we have developed (listed in Table E.1). Moreover, programmers can easily add their own processing operations as ImageJ plug-ins. By taking advantage of this large collection, processing chains can be precisely adapted to the variety of structures to be analyzed, including the nucleus itself and structures composed of spot-like or larger objects. Although one must have a minimal knowledge in image processing to define a processing chain, no programming skills are required : for each step, users choose specific plug-ins and adjust their parameters (Fig. C.2). Robustness over field-to-field or cell-to-cell intensity variations can be increased by using methods to automatically compute thresholds instead of constant values, which is essential for automation of segmen-

---

1. <http://loci.wisc.edu/software/bio-formats>

2. <http://mongodb.org>

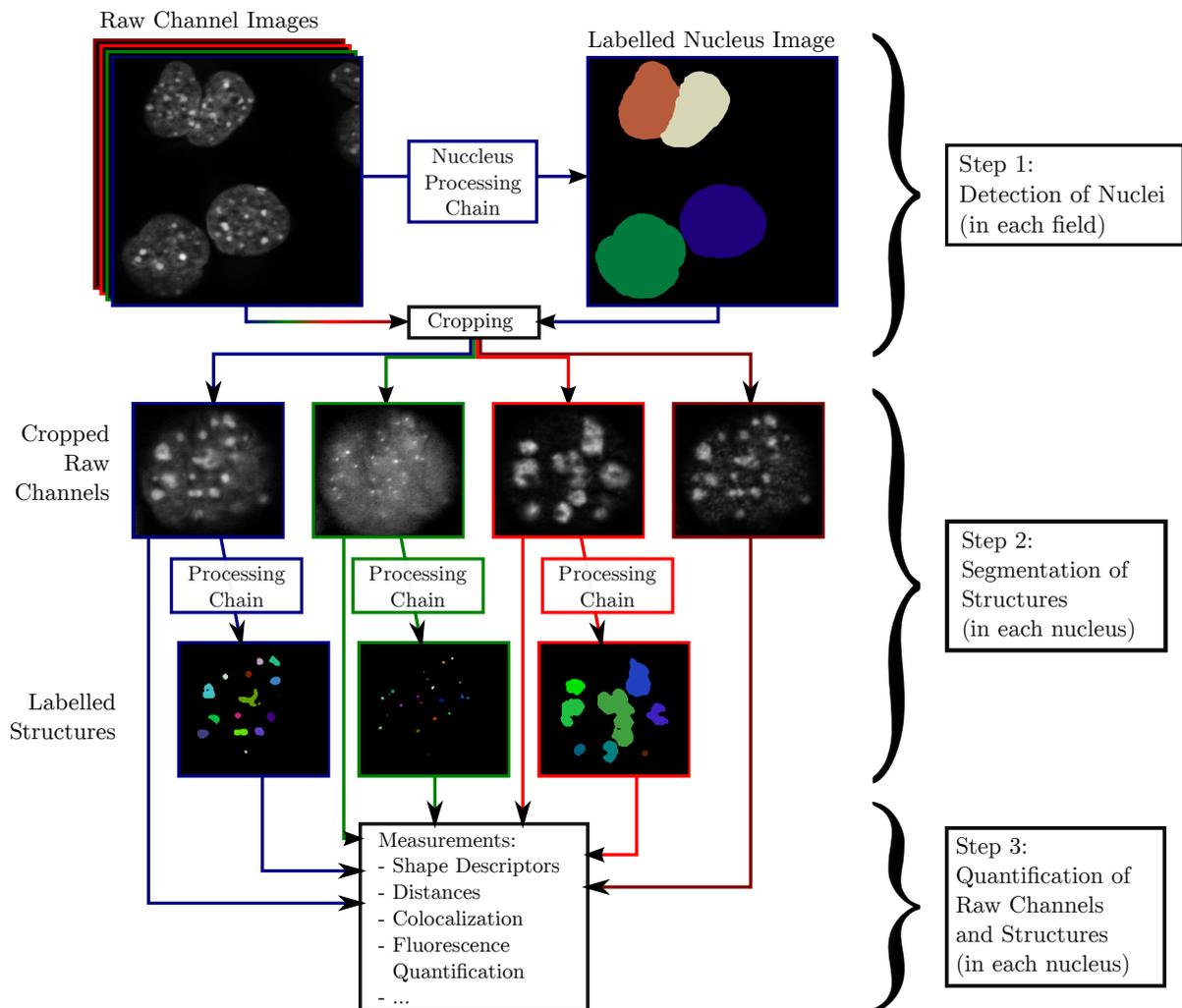


FIGURE C.1: Data processing workflow is composed of three main steps : (a) Segmentation of nuclei in each field, and cropping of channel images around the bounding box of each nucleus. (b) Segmentation of structures within each nucleus. (c) Quantitative image analysis in each nucleus, potentially using all raw channel images and labelled images.

tation procedures in a high-throughput context. Once defined, processing chains that have been validated for a specific type of structure can be re-used in different experiments as templates.

**Manual Intervention.** When automatic processing fails to correctly segment objects, a manual intervention is possible. Segmented objects can be merged, split, erased or created with semi-automatic procedures.

#### C.1.1.4 Data Analysis

**Data Browsing.** Figure 1 shows the data browsing tab that allows one to rapidly access the data (images and measurement results) related to each object or structure

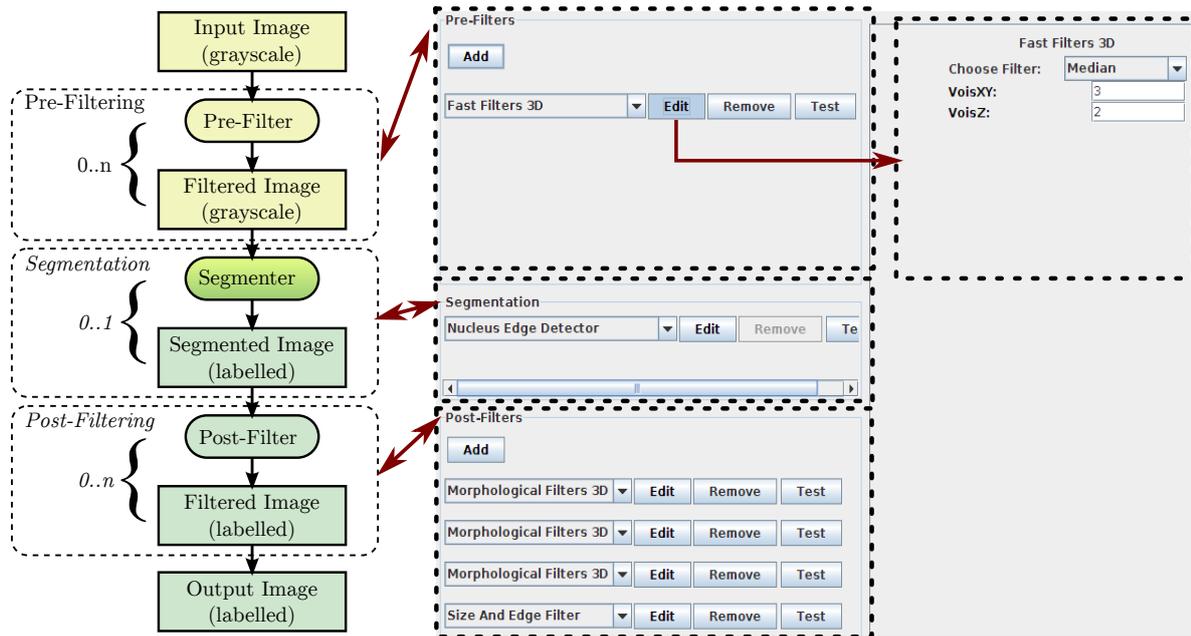


FIGURE C.2: Processing chains are composed of three steps : *pre-processing*, *segmentation* and *post-processing*. The *pre-processing* step is optional and can contain a sequence of several filters. The aim of this step is usually to correct common imaging artefacts, reduce noise (TANGO provides several filters, such as 3D-gaussian or 3D-median), and enhancing the signal to noise ratio (for instance, using 3D-top-hat filtering or 3D Laplacian of Gaussian). *Segmentation* is the process of identifying *objects* in the pre-processed image. Robust algorithms for segmentation of nuclei and of nuclear signals have been implemented, such as seed-based 3D watershed algorithm. The *post-processing* step is optional and can be composed of several filters. The aim of this step is to correct the segmented image. Classical operations are removing of small objects and morphological operations, such as closing. The right panel shows the processing chain editing interface in TANGO. For each of the three steps, processing operation can be chosen under operation of the corresponding category. If needed parameters can be adjusted (rightmost arrow). In order to help creation of processing chain, each operation can be tested individually on a selected nucleus (or on an opened image for nucleus segmentation).

contained in each nuclei. Raw and segmented images can be visualized as grayscale stacks, overlay stacks or 3D representations (Fig. C.4). Nuclei and objects can be annotated in order to define different sub-populations.

**Quantitative Image Analysis.** Quantitative image analysis is the process of computing numerical information describing each structure and the relationships between them. Voxel-based measurements quantify the staining pattern of a fluorescence signal within a segmented structure (e.g. mean intensity inside or outside objects, radial autocorrelation, layer measurements) or the co-localization of two signals. Segmentation-based measurements provide morphological descriptors of objects (e.g. : volume, elongation) or descriptors of spatial relations between several objects, including overlap, border-to-border distance between pairs of objects, or radial localization within the nucleus. TANGO includes a large collection of 3D measurements, some of which are rarely encountered in image analysis software (listed in Table E.2), and programmers can easily add their own plug-ins. Analysis results are stored in the database and can be exported to formatted text files

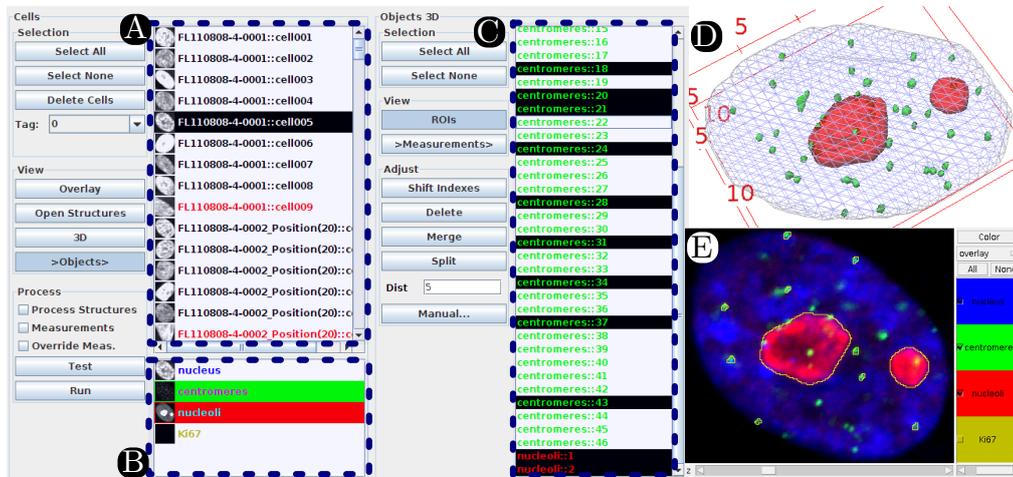


FIGURE C.3: The data browsing tab of the user interface provides access to the different levels of objects : (a) nuclei, (b) structures contained in the selected nuclei (in the present case : centromeres and nucleoli), (c) segmented objects contained in the selected structures. Different visualizations of a nucleus : (d) 3D, (e) overlay. Contours of selected objects in (c) can be interactively displayed on (e), which is useful to assess quality of segmentation.

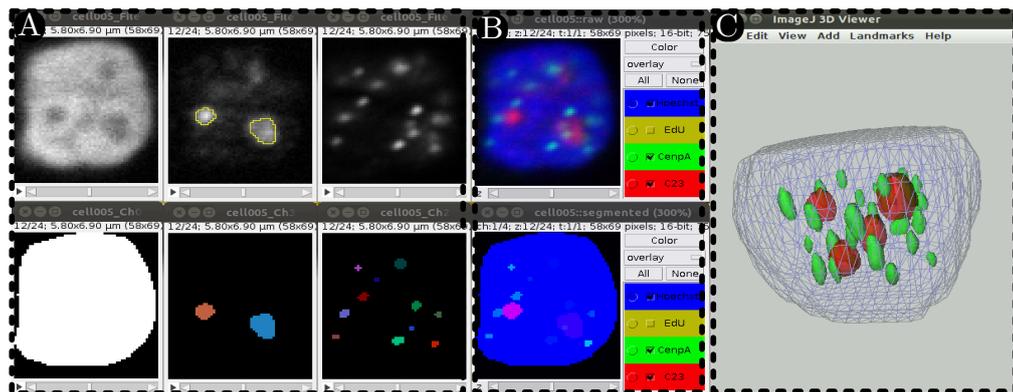


FIGURE C.4: Image Visualization. TANGO provides three modes of visualization for images : (a) standard ImageJ stack visualization, upper panel : grayscale images of different structures, lower panel : corresponding labelled images in which a color is assigned to each segmented object ; (b) overlay visualization of the images shown in (a) using the plug-in Image5d ; (c) 3D representation of segmented structures using the plug-in ImageJ 3D Viewer. To assess the quality of segmentation, contours of segmented objects can be displayed on any opened image (as in (a), middle upper image) and will be automatically updated when changing the Z-plane.

compatible with statistical processing software.

**Statistical Processing with R.** A part of TANGO is implemented as an r-package. Measurement results can be directly imported from the TANGO database to R as indexed dataframe objects, and thus can be processed using the wide panel of functions provided by R. When analyzing data, it can be very useful to visualize the images associated to a subset of the studied population (e.g. : cells in a specific stage of the cell-cycle, objects close to nuclear membrane). To do so, sub-populations of nuclei and objects can be defined in R and will be highlighted in TANGO. This

can be accomplished through the use of interactive plots (Fig. C.5). Finally, rtango is compatible with a flow-cytometry r-package (flowcore) that implements specific data visualization and analysis tools, thereby providing a complete solution for image-cytometry.

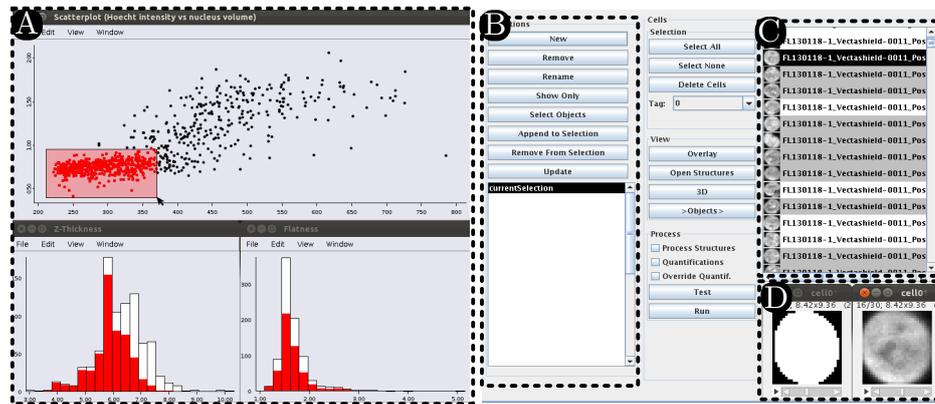


FIGURE C.5: Display and Analysis of sub-populations in R and TANGO. (a) Features from nuclei measured in TANGO can be imported in R using the r-package rtango, and plotted using the r-package Acinonyx (, package for interactive plots). In this case, each point of the scatter plot represents one nucleus. When a sub-population is selected, it is interactively reported on other plots. (b) Selections can be exported from R to TANGO, nuclei contained in the selection will be highlighted in the nucleus list (c) and the associated images can be visualized (d) or processed separately. Reciprocally, selections created or edited in TANGO are accessible within R, allowing one to analyse sub-populations of specific biological objects.

## Conclusion

TANGO is a versatile software which, thanks to a large collection of 3D processing operations and measurements as well as a convenient data visualization and analysis system, allows biologists to obtain quantitative information regarding the organization of the nucleus. We have successfully detected and analyzed various types of FISH (e.g. : BAC probes or chromosome territories) and immunocytochemistry (e.g. : centromeres or nucleoli) signals. TANGO is integrated into ImageJ and R, both of which are very modular open-source software widely used in the scientific community, allowing one to benefit from the rich potential of the two environments when using TANGO (Fig. C.6). In order to study large populations of nuclei, TANGO has been designed to automatically process substantial sets of images, while allowing semi-automatic intervention and visualization at several steps of the processing. Complementary information, documentation and a tutorial providing sample images, can be found on our website<sup>1</sup>.

1. There is also an integrated interactive detailed documentation within the interface.

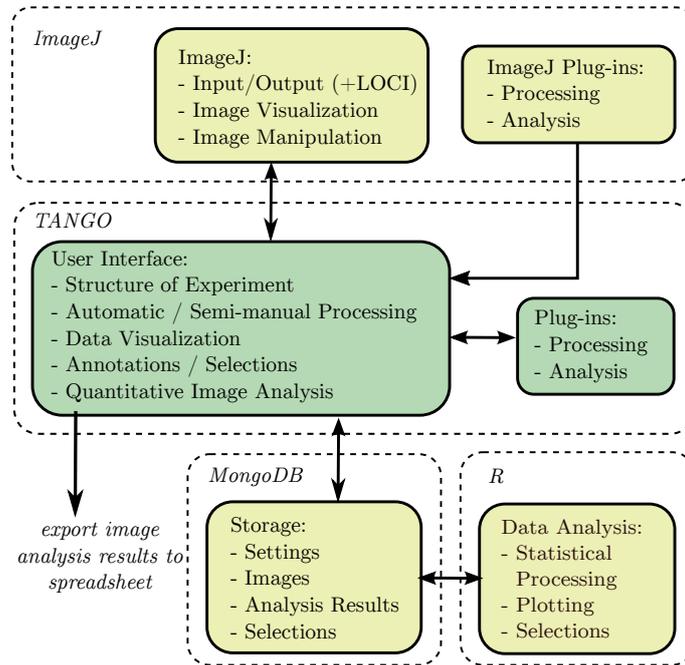


FIGURE C.6: TANGO is integrated inside two environments : ImageJ for image processing, analysis and visualization ; R for statistical analysis. All data are stored in a MongoDB database.

### Acknowledgement

The authors wish to thank Dr. Philippe Andrey for his helpful and valuable discussions on image processing, and advice on TANGO.

### C.1.2 Procédures de segmentations

Dans cette partie, nous décrivons les procédures de segmentation utilisées pour segmenter les différentes structures nucléaires que nous avons étudiées.

#### C.1.2.1 Segmentation des noyaux

**Problème posé.** La segmentation des noyaux présente en général deux défis techniques :

- définir précisément les bords des noyaux
- séparer les noyaux collés.

Dans notre système, les cellules sont en suspension etensemencées juste avant d'être fixées (voir §B.2), donc les noyaux sont assez bien séparés. Nous nous concentrerons donc sur le problème de définition précise des bords, les noyaux que la procédure n'a pas séparés seront éliminés à posteriori par inspection visuelle ou par un critère de volume. Dans notre système, nous utilisons un marquage de l'ADN au DAPI pour définir les noyaux. Ce marquage présente en général un très bon rapport signal sur bruit, les noyaux sont donc facilement distinguables du bruit de fond. La méthode utilisée en général est un simple seuillage, soit en utilisant une valeur fixe soit

en calculant un seuil par image, par exemple via la méthode d'Otsu (Otsu (1975)). Nous avons observé que les intensités moyennes des noyaux peuvent varier d'un noyau à l'autre, même au sein d'un champ de microscopie donné et donc que cette méthode est trop approximative et risque de biaiser les mesures telles que l'EVF, qui sont sensibles à la précision de la définition des bords du noyau. De plus nous avons observé qu'un seuil calculé par un algorithme automatique tel que la méthode d'Otsu dépend de la fraction de volume occupé par le signal (qui dépend notamment du nombre de noyaux présent sur le champ), ce qui rend cette méthode imprécise dans notre cas puisque le nombre de noyaux par champ est très variable.

**Principe de l'approche.** Nous avons donc développé une méthode qui détecte les contours de chaque noyau. Le contour est défini de manière objective en utilisant comme critère la variation maximale de l'intensité au niveau des bords du noyau. La détection des bords se fait par l'application d'un algorithme de *watershed*<sup>1</sup> (voir §A.4.3.1) sur l'image du gradient<sup>2</sup> : cet algorithme partitionne une image en de multiples régions dont les frontières sont localisées au niveau des maxima de l'intensité du gradient. Les régions correspondant à l'intérieur du noyau sont sélectionnées par un critère d'intensité puis fusionnées (voir Fig. C.7).

L'efficacité de cet algorithme est vérifiée par contrôle visuel et par la reproductibilité des distributions de volumes des noyaux entre des expériences indépendantes (voir §C.19).

### C.1.2.2 Segmentation des foyers

Nous utilisons le terme de *foyer*<sup>3</sup>, pour désigner un signal de forme particulière, caractérisé par :

- des bords nets (i.e. par un fort contraste local au niveau des bords)
- une intensité décroissante du point d'intensité maximal<sup>4</sup> jusqu'aux bords.

**Problème posé.** Dans le cas d'un marquage de l'ensemble des centromères par la protéine CenpA, on observe des foyers, mais nous avons observé que l'approche de détection des contours utilisée pour les noyaux ne s'applique pas car la taille de certains foyers est très faible par rapport à l'échantillonnage spatial, d'un rayon de 1 à 2 voxels. Pour ces foyers la procédure de détection des bords devient trop imprécise. Nous avons donc développé une autre approche qui permet d'une part de définir les contours de manière précise, et d'autre part de séparer les foyers très proches.

**Principe de l'approche.** Nous utilisons pour cette procédure la courbure de l'image, qui correspond à la valeur propre minimale de la transformée Hessienne. La

1. Algorithme de partage de ligne des eaux en français.

2. Le gradient d'une image correspond à sa dérivée dans un espace à 2 ou 3 dimensions. Les variations maximales d'une image se trouvent donc au niveau de l'intensité maximale du gradient.

3. *Spot* en anglais.

4. Ce point est nommé *centre* même s'il ne correspond pas exactement au centre géométrique de l'objet.

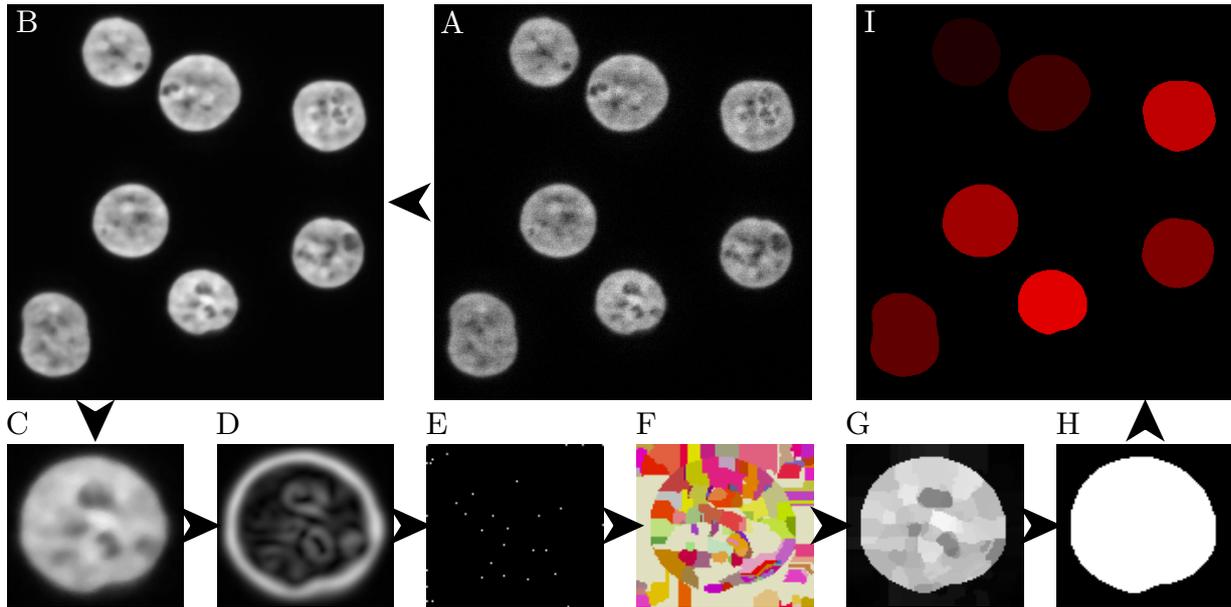


FIGURE C.7: Étapes de la procédure de segmentation des noyaux.

A-I : plans médians de la pile d'image correspondant à chaque étape de la procédure.

A : Image d'un champ de microscopie (l'image est ici découpée pour éviter qu'elle ne prenne trop de place sur le manuscrit).

B : Application d'un filtre median 3D  $2*2*2$  pour réduire le bruit.

Les algorithmes de *watershed* étant coûteux en temps de calcul dans le cas des images 3D de grande taille, l'algorithme procède en deux temps : (1) détection grossière des noyaux par seuillage en appliquant l'algorithme d'Otsu ; (2) ajustement de la segmentation autour de chaque noyau par l'algorithme de *watershed*.

C-H : L'ajustement des bords se fait au niveau de chaque noyau découpé.

C : Noyau découpé, une marge est ajoutée pour le cas où l'ajustement irait dans le sens d'une augmentation du volume du noyau.

D : Gradient 3D de l'image du noyau, avec un rayon de 2 voxels, prenant en compte l'anisotropie en Z.

E : Minima locaux du gradient qui servent de points de départ de la *watershed*. Ils ne sont soumis à aucun critère de sélection pour qu'ils soient présents à la fois dans le noyau et à l'extérieur du noyau.

F : L'application de l'algorithme de *watershed* partitionne l'image, on observe que les frontières des régions se trouvent au niveau des maxima de l'image du gradient.

G-H : Suppression des régions en dehors du noyau (procédure simplifiée à partir de [Andrey et al. \(2010\)](#)).

G : Afin de conserver uniquement les régions se trouvant au sein du noyau (donc de forte intensité), une image est créée dans laquelle chaque région est remplacée par son intensité moyenne.

H : L'algorithme de seuillage d'Otsu est appliquée à l'image G.

I : Masque des noyaux segmentés : les étapes C-H sont appliquées à chaque noyau, puis les noyaux sont recalés sur une image de la taille de l'image de départ.

courbure est très faible au niveau des centres des foyers, il s'agit d'un très bon critère pour les caractériser ([Soille \(2003\)](#)). Les foyers individuels détectés en calculant les minima régionaux de l'image de courbure, ce qui permet de séparer des foyers proches. En revanche la présence de bruit dans l'image produit des faux-positifs, qui sont éliminés en appliquant deux critères successifs : un critère portant sur la valeur de l'intensité au niveau du centre, et un critère portant sur la valeur de la courbure. Le fait de combiner ces deux critères diminue la sensibilité de l'algorithme à chacun des seuils, et permet de définir des seuils plus permissifs pour chacun des critères, tout en gardant une bonne sélectivité. Ceci permet d'automatiser facilement le calcul des seuils. Les bords des foyers sont déterminés par un seuil global sur l'image des

intensités filtrées (le filtrage est décrit plus bas). Ce seuil doit être calculé avec une bonne précision pour chaque noyau, car la définition des bords y est très sensible. Pour ce faire, une gaussienne est ajustée au niveau de chaque foyer, et ainsi un seuil d'intensité est déterminé par foyer. Le seuil correspond à la moyenne de l'ensemble des seuils des foyers individuels. L'ajustement de la gaussienne peut être faussé par la présence de foyers adjacents, pour éviter ce biais, l'image est segmentée par *watershed* à partir de chaque centre des foyers, et l'ajustement de chaque foyer se fait dans les voxels qui n'appartiennent pas à un autre foyer. Cette approche ressemble à celle développée par Ram *et al.* (2012), avec un ajustement supplémentaire des bords des foyers grâce au calcul d'un seuil très précis (voir Fig. §C.8 pour les détails techniques).

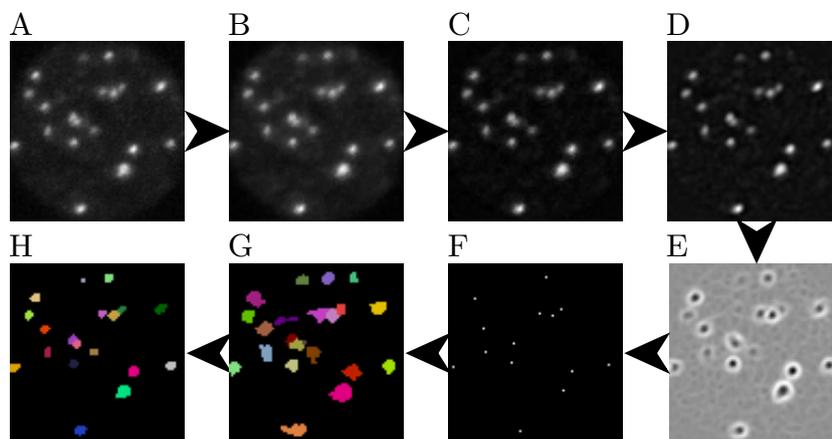


FIGURE C.8: Étapes de la procédure de segmentation des foyers. A-H : projections de 3 plans médian de la pile d'image correspondant à chaque étape de la procédure.

A : Image brute.

B-D : Étapes de filtrage.

B : Application d'un filtre median 3D avec un rayon de 1 voxel pour réduire le bruit.

C : Application d'un filtre top-Hat 3D avec un rayon de 5 voxels en XY et 3 voxels en Z pour augmenter le contraste local.

D : Application d'un filtre Laplacien de Gaussien 3D (voir 14) avec un rayon de 1 voxels pour augmenter le contraste local.

E-H : Étapes segmentation.

E : Image de courbure (valeur propre minimale de la transformée hessienne).

F : Image des centres des foyers. Les centres des foyers sont identifié selon deux critères : un critère de courbure et un critère d'intensité, parmi les minima régionaux de l'image de courbure sur une régions cubique de  $3*3*3$  voxels. Les seuils des critères sont calculés automatiquement pour chaque image. Le seuil d'intensité est calculé par l'algorithme d'Otsu et le seuil de courbure à partir d'un certain pourcentage des pixels les plus faibles de l'image de courbure.

G : Les foyers sont segmentés à partir de leurs centres, en appliquant une transformée *watershed* sur l'image de courbure à partir des centres des foyers. La segmentation est arrêtée à un seuil d'intensité très bas, correspondant à une estimation du bruit de fond, calculé automatiquement par l'algorithme de *sigma kappa clipping* (voir Tableau E.3.13), l'algorithme n'est pas sensible à ce seuil, il permet surtout d'accélérer la procédure.

H : Les foyers sont ajustés en calculant une intensité globale pour l'image : pour chaque foyer, on calcule un seuil en ajustant une gaussienne à partir du centre du foyer, en excluant les voxels se trouvant dans d'autres foyers. La moyenne des seuils est ensuite utilisée pour ajuster chaque foyer.

L'inspection visuelle montre que cette procédure produit environ 1% de faux positifs, qui-sont éliminés par un critère sur le volume, dont le seuil est placé à 0.02

$\mu\text{m}$ , et toujours vérifié par inspection visuelle avant d'être supprimés. Ces faux-positifs sont presque toujours présents dans les noyaux dont le bruit de fond est anormalement élevé.

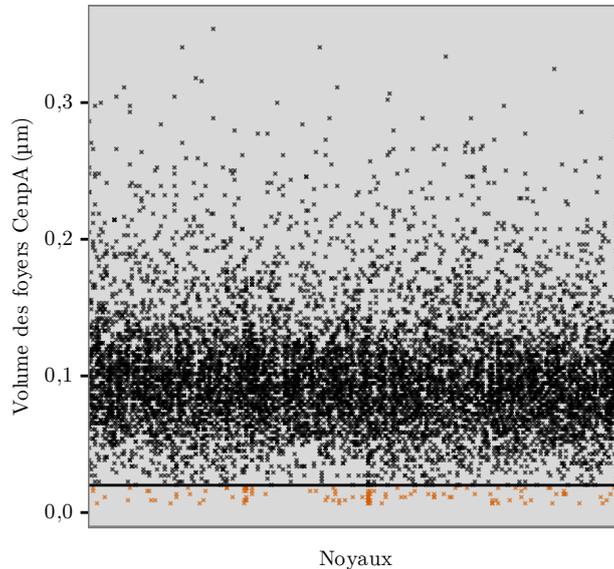


FIGURE C.9: Répartition des volumes des foyers CenpA par noyaux. Chaque ligne verticale représente un noyau, chaque croix un foyer CenpA. La ligne horizontale noire représente le seuil, les croix colorées sont sélectionnées dans TANGO pour l'inspection visuelle.

Cet algorithme est validé d'une part par inspection visuelle, d'autre part par la reproductibilité des mesures réalisée au §C.3. Notamment on observe que exactement 46 foyers CenpA sont détectés dans plus 75% des cellules en début de phase  $G_1$  (Fig. C.27A), ce résultat est en accord avec les observations de [Solovei \*et al.\* \(2004\)](#).

**Application à des foyers de plus grande taille.** Cette approche s'applique aussi aux foyers de plus grande taille tels que les foyers PCY (voir §C.3.1.2) ou les centromères spécifiques (voir §C.3.1.2). Le filtrage appliqué est légèrement différent :

- median 3D avec un rayon de 1 voxel pour réduire le bruit
- Laplacien de la gaussienne 3D avec un rayon de 2 voxels<sup>1</sup> pour augmenter le contraste local.

### C.1.2.3 Autres structures

Les nucléoles et les CT sont des régions qui ne présentent pas les caractéristiques de foyers, puisque l'intensité n'est pas décroissante depuis le centre des objets vers les bords. De plus les bords de ces structures sont parfois diffus. Les algorithmes

1. Prenant en compte l'anisotropie en Z.

précédemment décrits ne peuvent donc pas s'appliquer. Nous avons utilisé une approche par seuillage très simple, en utilisant la méthode d'Otsu, après une réduction du bruit par un filtre median 3D<sup>1</sup>. de rayon 2 voxels. L'efficacité de cette procédure est vérifiée par contrôle visuel et par la reproductibilité des distribution des volumes des nucléoles entre des expériences indépendantes (voir Fig. C.19).

## Conclusion

Nous avons développé plusieurs procédures de segmentation, en adaptant des algorithmes existants afin de segmenter les signaux que nous étudions, en essayant d'augmenter au maximum la précision de la segmentation lorsque les mesures ultérieures y sont sensibles (e.g. : précision du bord des noyaux, des foyers CenpA, etc.). Afin d'automatiser les procédures, nous avons pris grand soin de déterminer la sensibilité de l'algorithme à chaque paramètre, et d'éviter de se reposer sur des algorithmes de seuillage automatiques pour déterminer ces paramètres sensibles. Enfin nous avons essayé autant que possible d'optimiser les procédures afin d'accélérer les temps de traitement. La vérification de ces procédures repose principalement sur l'inspection visuelle d'une grande quantité d'images, et sur la reproductibilité des résultats qui les utilisent.

### C.1.3 Mesures

Afin de répondre aux questions que nous nous sommes posé, nous avons défini ou adapté des mesures. Dans cette partie nous décrivons ces mesures.

#### C.1.3.1 Mesure du positionnement dans le noyau

Nous mesurons le positionnement d'un point dans le noyau vis-à-vis d'une structure de référence via la mesure d'EVF (voir §A.4.3.2). La version développée par [Ballester et al. \(2008\)](#) prend pour référence les bords du noyau. Dans notre étude nous souhaitons aussi mesurer la position d'un point en prenant pour référence le bord des nucléoles.

**Généralisation de la mesure d'EVF.** Afin d'adapter cette mesure, nous l'avons généralisée en conservant ses 3 propriétés fondamentales :

- Deux points ayant la même EVF sont situés à la même distance de la structure de référence.
- L'EVF est normalisée et prend des valeurs entre 0 à proximité de la structure de référence, et 1 au point le plus éloigné de cette structure.
- La distribution d'EVF de points uniformément distribués dans le volume du noyau est uniforme, quelque soit la taille et la forme des noyaux. Cette mesure permet donc une normalisation par rapport aux variations de forme et de taille des noyaux.

---

1. Le filtre médian a la caractéristique de conserver les contrastes locaux tout en réduisant efficacement le bruit.

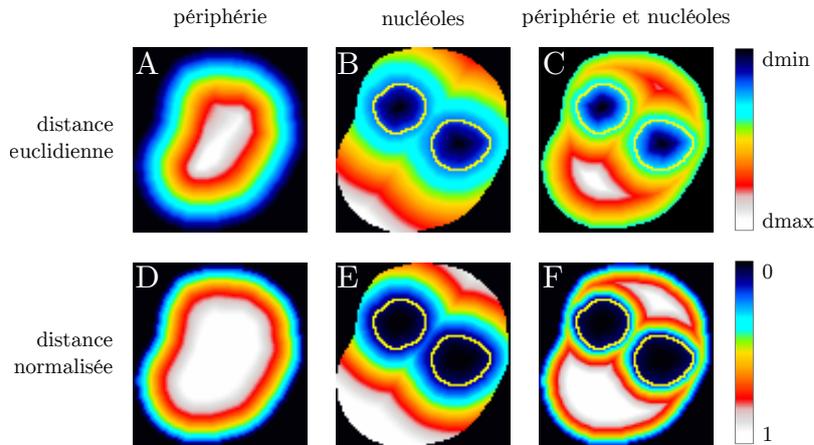


FIGURE C.10: Mesure de l'EVF depuis différentes structures de références. De la même manière que [Ballester et al. \(2008\)](#), notre estimation de l'EVF repose sur des cartes de distances euclidiennes qui associent à chaque voxel la distance minimale au bords de la structure de référence (ces cartes prennent en compte l'anisotropie en Z). Dans chaque cas elles permettent de calculer rapidement en tous les points du noyau la distance minimale à la structure de référence (A-C).

Les distances à l'intérieur des nucléoles sont comptabilisées en négatif et représentées en bleu foncé dans les images B et C. Ainsi deux points situés à une distance  $d$  du bord des nucléoles et présents à l'intérieur et à l'extérieur des nucléoles ne seront pas confondus, et le point situé à l'intérieur sera situé à une distance inférieure du centre des nucléoles que le point situé à l'extérieur.

La carte de distance est ensuite normalisée en remplaçant la distance de chaque voxel par son rang (lorsque plusieurs valeurs de distances sont égales, le rang utilisé correspond au rang médian) divisé par le nombre total de voxels. Nous obtenons ainsi une carte de distance normalisée, qui en chaque point indique la fraction de volume plus proche de la structure de référence, donc l'EVF (D-F).

L'EVF du centre d'un foyer (dont la précision est sub-voxélaire) est déterminée par une approximation bicubique sur la carte de distance normalisée (nous avons vérifié qu'il était équivalent de procéder ainsi ou de calculer directement le rang de la distance du point à la structure de référence).

Le contour jaune dans les images B, C, E et F correspond au contour des nucléoles. A et D : EVF depuis la périphérie du noyau. B et E : EVF depuis la périphérie des nucléoles. C et F : EVF depuis la périphérie du noyau et des nucléoles. Les couleurs indiquent les distances : le noir correspond aux distances les plus faibles, et elles augmentent jusqu'au blanc.

**Localisation par rapport à la périphérie.** Dans le cas où la structure de référence est la périphérie du noyau, nous observons un biais dans la mesure d'EVF par rapport à la périphérie. En effet, les foyers centromériques détectés ont une certaine taille<sup>1</sup>, et nous mesurons l'EVF au niveau du centre du foyer. Une zone en périphérie du noyau de largeur égale au rayon du foyer est donc inaccessible aux foyers. Lorsqu'un foyer est en contact avec la périphérie, son centre se trouve à une distance environ égale au rayon du foyer, et la mesure de l'EVF correspondante n'est donc pas nulle. Ceci a pour conséquence la mesure systématique d'une répulsion apparente aux faibles valeurs d'EVF, mais cette répulsion est artificielle. Cet effet est exacerbé par le fait que l'EVF par rapport à la périphérie varie très rapidement aux abords de la périphérie.

Afin de palier ce biais, nous avons modifié la définition de la mesure de l'EVF par rapport à la périphérie : pour chaque foyer, nous érodons le noyau d'une distance

1. Le rayon des foyers est en moyenne de  $0.3 \mu\text{m}$ , et dépend de la sonde utilisée.

égale au rayon moyen du foyer<sup>1</sup>. Ainsi lorsque le foyer se trouve en contact avec la périphérie, l'EVF mesurée est 0 (ou proche de 0). Cette modification supprime la répulsion apparente aux faibles valeurs d'EVF.

**Localisation par rapport aux nucléoles.** Afin de localiser un point par rapport aux bords des nucléoles, nous avons adapté la mesure en prenant comme référence les bords des nucléoles. Soit  $V_{nl}$  le volume total des nucléoles, et  $V_n$  le volume du noyau. Nous nommons *nucléoplasme* (np) l'espace nucléaire excluant celui des nucléoles. Nous avons observé que les centromères peuvent se localiser à l'intérieur des nucléoles, leur volume n'est donc pas exclu et nous avons distingué deux cas.

- Lorsque le point est localisé à l'extérieur d'un nucléole :  $EVF = \frac{V_{nl}+V}{V_n}$  avec  $V$  le volume nucléoplasmique plus proche des nucléoles que du point. L'EVF correspond ainsi à la probabilité de se trouver plus proche ou à l'intérieur d'un nucléole que le point.
- Lorsque le point est localisé à l'intérieur d'un nucléole :  $EVF = \frac{V}{V_n}$  avec  $V$  le volume nucléolaire qui est plus éloigné des bords du nucléole que le point (donc se trouvant plus à l'intérieur des nucléoles que le point). L'EVF correspond ainsi à la probabilité de se trouver plus à l'intérieur des nucléoles que le point.

Ainsi, l'EVF varie de 0 à l'intérieur des nucléoles jusqu'à 1, loin des nucléoles (voir Fig. C.10B et E).

**Localisation par rapport au nucléoplasme.** Afin de déterminer si un point se localise dans le nucléoplasme, c'est-à-dire s'il est éloigné à la fois de la périphérie du noyau et des nucléoles, nous avons défini une deuxième variante de l'EVF qui prend comme structure de référence à la fois la périphérie du noyau et les nucléoles. Soient :  $d_{nl}$  la distance au nucléole le plus proche d'un point,  $d_n$  la distance minimale d'un point à la périphérie du noyau, et  $d = \min(d_{nl}, d_n)$ . De la même manière que précédemment, nous distinguons deux cas :

- Lorsque le point est localisé à l'extérieur d'un nucléole :  $EVF = \frac{V_{nl}+V_1+V_2}{V_n}$  avec :  $V_1$  le volume nucléoplasmique se trouvant à une distance aux bords des nucléoles inférieure à  $d$ ,  $V_2$  le volume nucléoplasmique se trouvant à une distance à la périphérie du noyau inférieure à  $d$ .
- Lorsque le point est localisé à l'intérieur d'un nucléole :  $EVF = \frac{V_1}{V_n}$  avec :  $V_1$  le volume nucléolaire se trouvant à une distance du bords des nucléoles supérieure à  $d$ .

Ainsi, l'EVF varie de 0 à l'intérieur des nucléoles ou à proximité de la périphérie du noyau, jusqu'à 1 loin des nucléoles et de la périphérie, donc dans le nucléoplasme (voir Fig. C.10C et F).

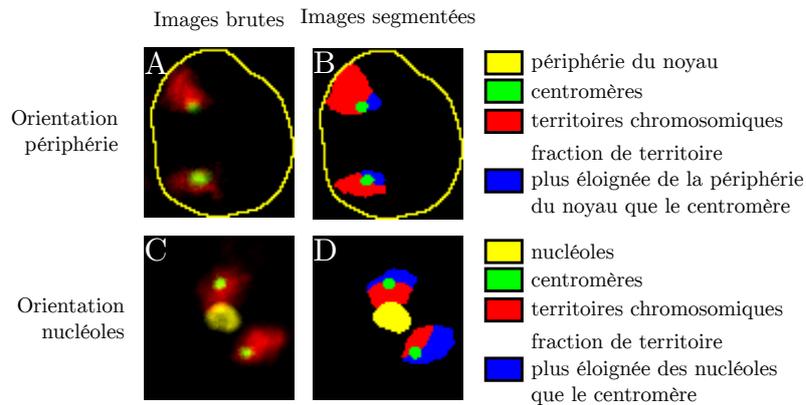


FIGURE C.11: Mesure d'orientation. A-D : projection maximales de 3 plan médian. A-B : orientation par rapport à la périphérie. C-D : orientation par rapport aux nucléoles. Cette mesure est estimée en 3D à l'aide d'une carte de distance euclidienne depuis la structure de référence.

### C.1.3.2 Mesure d'orientation

Afin de comparer le positionnement d'un centromère vis-à-vis d'une structure de référence à l'ensemble du chromosome (voir §C.2.2.6), nous avons développé une mesure d'orientation. Cette mesure estime la fraction de volume du territoire chromosomique qui est plus éloignée de la structure de référence que le centre du centromère. Lorsque cette mesure est :

- $> 0.5$ , le centromère est plus proche de la structure de référence que la majorité du TC
- $< 0.5$ , le centromère est plus éloignée de la structure de référence que la majorité du TC.

Cela correspond en quelque sorte à l'orientation du centromère vis-à-vis d'une structure de référence par rapport au TC. Nous avons réalisé cette mesure par rapport à deux structures de référence : la périphérie du noyau et les nucléoles (voir Fig. C.11).

### C.1.4 Analyse quantitative du cycle cellulaire

Afin de déterminer la phase du cycle cellulaire de chaque cellule, nous avons utilisé un marquage de la protéine Ki67 pour discriminer entre les cellules en cycle et les cellules quiescentes, et un marquage de la réplication par l'incorporation de BrdU pour identifier les cellules en phase S (Solovei *et al.* (2004)). Afin de quantifier la détermination de la phase du cycle, nous avons développé des mesures adaptées à l'analyse de chaque signal (Hoechst, Ki67 et BrdU). La figure C.12 montre les patterns caractéristiques obtenus pour chaque signal en fonction de la phase du cycle cellulaire.

L'intensité du contre-marquage de l'ADN sommée au sein de l'espace segmenté

1. Le rayon moyen du foyer est estimé par la moyenne des distances entre le centre du foyer et les voxels se trouvant sur le contour du foyer (un voxel se trouve sur le contour du foyer s'il est en contact direct avec au moins un autre voxel n'appartenant pas au foyer).

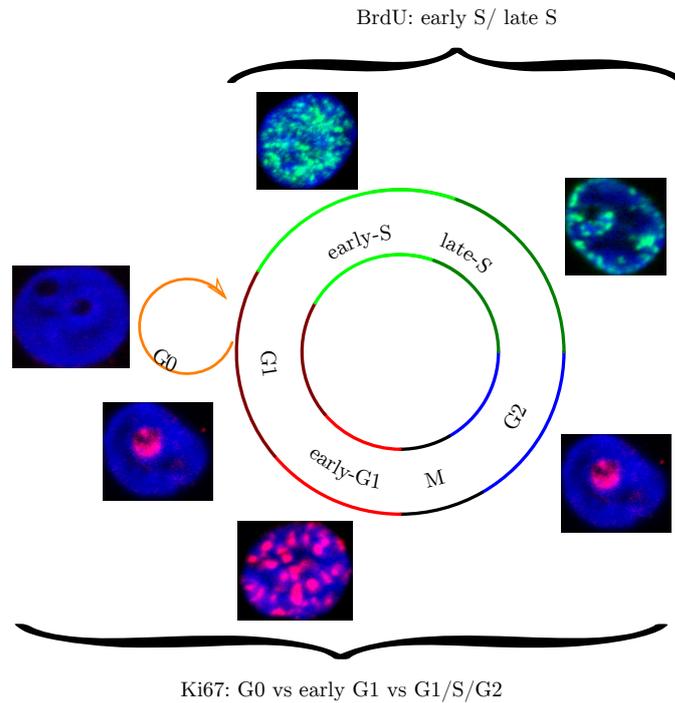


FIGURE C.12: Patterns caractéristiques des différents marqueurs du cycle cellulaire. Ki67 n'est pas exprimé dans les cellules quiescentes, présent au niveau des régions centromériques dans les cellules en début de phase  $G_1$  et dans les nucléoles durant le reste du cycle. Le signal BrdU est présent exclusivement dans les cellules en phase S, et présente une répartition spatiale différente au cours de la phase S : il est localisé préférentiellement en bordure de la périphérie du noyau et des nucléoles en fin de phase S, et inversement dans le reste du nucléoplasme en début de phase S.

du noyau et le volume du noyau représentent deux estimations de la quantité d'ADN dans le noyau. En représentant graphiquement les noyaux selon ces deux mesures (Fig. C.13A), on observe la présence de deux groupes de noyaux, qui représentent très certainement les cellules qui n'ont pas commencé la réplication de l'ADN (2N-2C) et les cellules qui ont terminé la réplication de leur ADN (2N-4C). Les noyaux situés entre ces deux groupes sont certainement les noyaux en phase S, dont la réplication n'est pas complètement terminée. Ce graphique, qui ne peut être utilisé pour déterminer avec précision des populations de cellules, est montré ici pour vérifier la cohérence de la procédure de détermination de la phase du cycle cellulaire décrite ci-après.

### C.1.5 Analyse du signal Ki67

**Détection des cellules quiescentes.** La quiescence est en général reconnue par l'absence de signal Ki67. Cependant, il est difficile de définir un seuil pour l'intensité moyenne du signal car sa distribution est un continuum au sein d'une population de noyaux. Nous avons donc combiné l'information fournie par l'intensité moyenne du signal avec une autre caractéristique du signal qui reflète son agrégation. L'agrégation

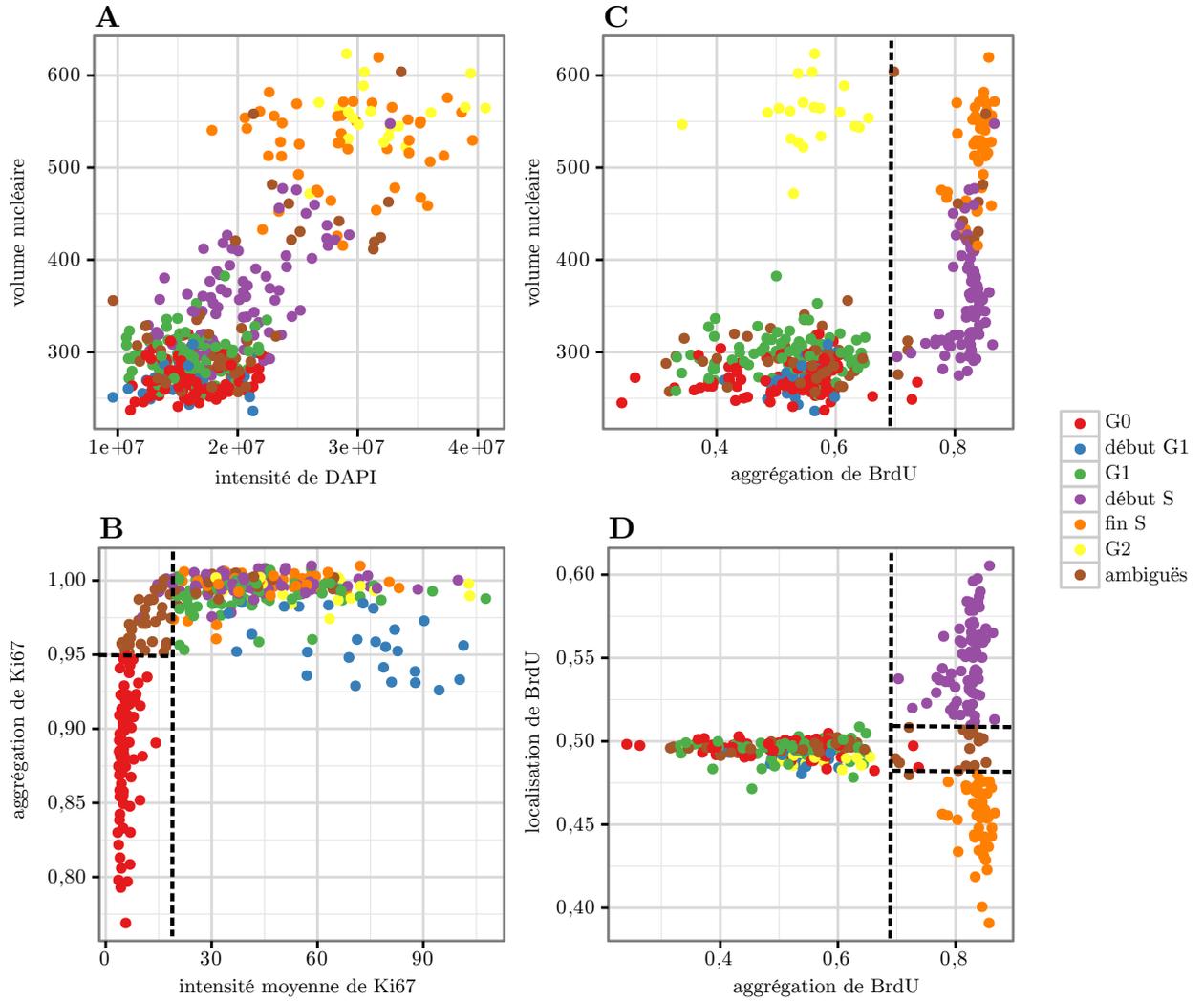


FIGURE C.13: Analyse quantitative du cycle cellulaire. Voir texte principal.

tion est estimée par la mesure d'auto-corrélation radiale 3D (RAC)<sup>1</sup> pour un rayon de  $0.2 \mu m$ .

Nous avons développé une approche graphique<sup>2</sup>, afin d'exploiter des points de repère définis au niveau de la population. La population est dispersée selon deux

1.  $RAC(R) = \frac{1}{N \times \sigma^2} \times \sum_{i,j \in N | d(i,j)=R} (S_i - \mu)(S_j - \mu)$ , avec  $R$  le rayon de l'auto-corrélation radiale,  $N$  le nombre de voxels dans le noyau,  $d(i, j)$  la distance entre les points d'indice  $i$  et  $j$ ,  $\sigma$  l'écart-type du signal,  $\mu$  la moyenne du signal. Afin de prendre en compte l'anisotropie en  $Z$ , l'image est ré-échantillonnée avant le calcul de la RAC. Le bruit de fond du signal est réduit par un filtre gaussien  $1*1*1$  avant la mesure afin de rendre les valeurs de la mesure plus reproductibles entre différentes expériences.

2. Ce type d'approche est souvent utilisée en cytométrie de flux.

axes, le premier correspond à l'intensité moyenne du signal Ki67, le second à l'agrégation du signal. Le pattern angulaire observé (Fig. C.13B) est très reproductible et peut être interprété par le fait que l'intensité moyenne du signal disperse bien les cellules en cycle, probablement parce que le niveau d'expression de la protéine Ki67 varie au sein du cycle cellulaire, mais ne disperse pas les cellules quiescentes, parce qu'elles n'expriment pas la protéine et donc le niveau du signal est proche de celui du bruit. D'autre part, l'agrégation disperse bien les cellules quiescentes mais pas les cellules en cycle. Cela est probablement dû au fait que lorsque Ki67 est exprimée, elle est très agrégée au sein des nucléoles dans les cellules en cycle et donc l'estimation de l'agrégation atteint une valeur limite proche de 1, alors que lorsqu'elle n'est pas exprimée, elle est progressivement éliminée et donc le niveau du signal est plus comparable à celui du bruit ce qui diminue l'estimation de l'agrégation. Ceci permet de définir graphiquement deux seuils : un seuil sur l'axe de l'agrégation définissant la limite supérieure pour les cellules quiescentes (rouge), et un seuil pour l'axe de l'intensité, définissant la limite inférieure pour les cellules en cycle (autres couleurs à l'exception du marron). Les cellules localisées entre ces deux seuils sont considérées comme ambiguës (marron) et éliminées des études. Cette méthode conduit à mesurer des proportions très reproductibles de cellules quiescentes et en cycle, entre différentes expériences conduites sur des cellules cultivées dans les mêmes conditions, même lorsque différents anticorps sont utilisés pour marquer Ki67 ou le BrdU. Les cellules quiescentes sont localisées dans le groupe 2N-2C du graphique C.13A, ce qui est cohérent.

**Détection des cellules en début de phase  $G_1$ .** Les noyaux ayant une forte intensité de signal Ki67 et une faible agrégation sont en général en début de phase  $G_1$  (bleu)<sup>1</sup>. Ce critère n'est pas assez sélectif pour les distinguer graphiquement des autres cellules en cycle, les cellules en début de phase  $G_1$  sont donc déterminées par inspection visuelle. Cela peut être fait de manière très rapide grâce à TANGO en n'inspectant uniquement les cellules potentiellement en début de phase  $G_1$  (qui vérifient le critère de faible agrégation et de forte intensité).

### C.1.5.1 Détection des cellules en phase S

La détermination des cellules en phase S est plus simple que celle des cellules quiescentes, car le signal BrdU est complètement absent des cellules qui ne sont pas en réplication, ces cellules sont facilement séparables des cellules en réplication. Nous avons observé que l'agrégation du signal BrdU permettait de définir un bon critère de séparation des cellules en réplication. Le graphe C.13C montre la dispersion des noyaux selon l'agrégation du signal BrdU et leur volume, et permet de définir trois populations : deux populations qui ont une faible agrégation de signal BrdU, dont l'une correspond aux cellules 2N-2C et l'autre aux cellules 2N-4C. Une troisième

---

1. L'agrégation des cellules en début de phase  $G_1$ ) est plus faible car le signal Ki67 est présent au niveau des régions centromériques et des corps pré-nucléolaire, et forme donc des foyers de plus petite tailles que les nucléoles matures (Solovei *et al.* (2004)).

population ayant une forte agrégation de signal BrdU et des valeurs de volume nucléaire compris entre les deux précédentes correspond aux cellules en réplication.

Afin de discriminer entre les cellules en début et fin de phase S, nous avons exploité les variations de la localisation nucléaire du signal BrdU entre ces deux phases. En effet, le début de la phase S correspond à la réplication de l'euchromatine qui est principalement localisée à distance de la périphérie du noyau et des bords des nucléoles. En revanche, la fin de la phase S correspond à la réplication de l'hétérochromatine qui est principalement localisée en périphérie du noyau et aux bords des nucléoles. La localisation de chaque voxel par rapport à la périphérie du noyau et aux bords des nucléoles est mesurée en utilisant la mesure d'EVF par rapport aux nucléoles et à la périphérie définie précédemment (voir §C.1.3.1, Fig. C.10F). La mesure de localisation du signal BrdU correspond à la localisation moyenne pondérée par l'intensité du signal BrdU. Ainsi, une valeur proche de 0.5 signifie que le signal est localisé aléatoirement entre les nucléoles et la périphérie du noyau, et une déviation de cette valeur signifie soit que le signal est en moyenne plus proche de ces deux structures de référence (pour des valeurs inférieures à 0.5) ou qu'il est plus éloigné des structures de référence (pour des valeurs supérieures à 0.5). Les population en début de phase S (violet) et en fin de phase S (orange) sont visibles sur la figure C.13D.

Dans les cellules qui ne sont pas en réplication, la distribution de cette mesure du positionnement du signal est très proche de 0.5, qui correspond à la répartition du bruit de fond qui est aléatoire, alors qu'elle est plus dispersée dans les cellules en réplication. De plus les volumes des noyaux en réplication sont compris entre ceux des populations 2N-2C et 2N-4C, et les volumes des noyaux en fin de réplication sont proches du volumes des cellules 2N-4C et supérieurs aux volumes des noyaux en début de réplication qui sont plus proches des volume des cellules 2N-2C, ce qui peut s'interpréter par le fait que plus la cellule avance dans la phase S, plus la quantité d'ADN répliqué (donc d'ADN total) augmente.

### C.1.5.2 Détection des cellules en phase $G_1$

Lorsque les cellules quiescentes et en cours de réplifications sont identifiées, les cellules en phase  $G_1$  (vert) et  $G_2$  (jaune) sont distinguées par la quantité d'ADN parmi les cellules restantes. Les cellules en phase  $G_2$  peuvent être visualisée sur le graphe C.13C, et correspondent au groupe ayant une faible valeur d'agrégation de BrdU et un volume nucléaire élevé.

## Conclusion

Nous avons défini une procédure d'analyse quantitative du cycle cellulaire, utilisant plusieurs marqueurs du cycle, permettant de distinguer au maximum les cellules en phase  $G_0$ , début  $G_1$ ,  $G_1$ , début S, fin S,  $G_2$ . La détermination du cycle est en général réalisée de manière qualitative en observant les patterns des différents marqueurs (Solovei *et al.* (2004)). Nous avons observé que notre approche quantitative, en plus d'éviter une procédure manuelle pénible, augmentait beaucoup la

reproductibilité des résultats en comparaison à la détermination manuelle du cycle. De plus cette procédure quantitative renforce l'objectivité de la détermination du cycle cellulaire.

### C.1.6 Discussion et conclusion

TANGO a été développé pour permettre de répondre à des questions très variées dans le domaine de l'organisation nucléaire. Le fait de restreindre son champ d'application à ce domaine a permis de définir une ontologie, associée à une architecture des données qui reflète les objets biologiques étudiés<sup>1</sup>, contrairement à des logiciels beaucoup plus génériques tels que ICY ou ImageJ, qui ne peuvent définir une telle architecture et qui prennent donc pour point de départ l'image sans lui donner un sens biologique. Le fait de définir une architecture fixe contraint le champ d'application au domaine de l'architecture nucléaire, mais ouvre beaucoup de possibilités au sein de ce domaine :

- l'architecture des données est reflétée dans l'interface graphique utilisateur du logiciel, ce qui permet à l'utilisateur de naviguer dans les données de manière très intuitive, et ainsi rapidement vérifier et corriger le résultats des procédures de segmentation.
- cette architecture est aussi reflétée dans l'interface avec le logiciel de traitement statistique R, ce qui permet un aller-retour entre les images visualisées dans TANGO et les données qui proviennent de ces images<sup>2</sup>. Ainsi TANGO s'intègre dans deux environnements : ImageJ pour le traitement et l'analyse d'image, et R pour le traitement statistique des données. Ces deux environnements contiennent chacun beaucoup de fonctionnalités, qui sont directement utilisables pour les données provenant de TANGO<sup>3</sup>.
- les étapes du traitement des images sont contraintes ( (1) segmentation des noyau, (2) découpage des noyaux, (3) segmentation des structures nucléaires et (4) mesures), ce qui facilite beaucoup l'utilisation sans pour autant limiter les possibilités grâce à la possibilité de paramétrer précisément chaque étape via des modules de traitement et de mesure. Un système de *plug-in* permet d'ajouter très simplement des modules, ce qui augmente encore les potentialités du logiciel.

Il s'agit d'aspects principalement d'ordre ergonomique, mais ils permettent de faciliter et d'accélérer le traitement des images et des données ce qui constitue une étape supplémentaire vers l'augmentation du *débit* dans le domaine de l'analyse d'image. Pour augmenter la vitesse du traitement, nous avons à la fois misé sur l'automatisation des procédures, mais aussi sur la possibilité d'une intervention

---

1. Le noyau et les structures nucléaires (incluses dans le noyau) qui peuvent être composées d'éléments multiples (objets)

2. Dans le sens TANGO  $\mapsto$  R, les données mesurées sont importées dans le logiciel sous un format directement exploitable, et dans le sens R  $\mapsto$  TANGO, les objets (noyaux ou éléments des structures nucléaires) peuvent être identifiées facilement dans TANGO.

3. Par exemple, R contient un module de traitement des données de cytométrie en flux, qui est directement exploitable pour les données de cytométrie en image provenant de TANGO.

manuelle très rapide permettant vérification et corrections éventuelles des erreurs des procédures automatiques. Cette intervention est accélérée par :

- les vignettes dans les listes d'objets<sup>1</sup> permettent de détecter rapidement des aberrations
- la possibilité de trier les listes d'objets (noyaux et objets contenus dans les structures) selon une mesure, ce qui permet par exemple d'accéder rapidement aux objets aberrants<sup>2</sup>
- la possibilité de générer des sélections d'objets via le logiciel R, qui sont facilement accessibles dans TANGO<sup>3</sup>.

Ce cadre de travail nous a permis de développer :

- des procédures de segmentation, adaptées aux différents signaux que nous avons étudiés (noyaux, centromères, nucléoles, etc.)
- des mesures originales permettant de répondre à des questions biologiques précises (orientation des centromères par rapport aux CTs, mesure d'attraction ou de répulsion par rapport à différentes structure nucléaires, etc.)
- une procédure d'analyse quantitative du cycle cellulaire à partir de différents marqueurs, permettant de distinguer plusieurs phases du cycle cellulaire de manière reproductible.

---

1. Qui correspondent aux projections maximales des images dans l'axe Z.

2. Par exemple les noyaux contenant plus de deux centromères, les noyaux coupés dans l'axe des Z, les nucléoles trop volumineux etc.

3. Par exemple une sélection peut porter sur plusieurs critères définis à partir d'une analyse graphique de la population complète

## C.2 Cell cycle-dependent positioning of individual centromeres in the interphase nucleus of human lymphoblastoid cell lines

Cette partie est présentée sous la forme d'un article qui sera soumis prochainement à la revue *Journal of Cell Science*. Par souci de présentation, il a été mis en page comme le reste du manuscrit. La partie matériels et méthode étant redondante avec celles autres résultats, nous ne l'avons pas incluse.

**Résumé des résultats principaux.** Plusieurs études ont montré que les RCs humaines ne se positionnent pas aléatoirement dans le noyau, et qu'elles ont tendance à s'associer avec la périphérie du noyau et les nucléoles. Cependant, la description du positionnement de RCs spécifiques nécessite encore d'être précisée.

Nous avons développé une méthode d'analyse du positionnement des RCs permettant de quantifier l'attraction ou la répulsion à n'importe quelle structure nucléaire. Cette méthode, combinée à l'utilisation de nouvelles sondes permettant de détecter des RCs portés par des chromosomes spécifiques, nous a permis d'étudier leur positionnement par rapport à la périphérie du noyau ainsi qu'aux nucléoles, en prenant en compte le cycle cellulaire. Nous observons que le positionnement des RCs présente des régularités à l'échelle de la population de cellules, qui sont reproductibles entre différentes expériences et deux lignées cellulaires lymphoblastoïdes :

- Les RCs de tous les chromosomes étudiés se positionnent préférentiellement soit aux bords des nucléoles soit en périphérie du noyau, comme le montre le fait que nous observons une forte répulsion pour le *nucléoplasme*<sup>1</sup>.
- La préférence pour la périphérie du noyau ou les nucléoles dépend de l'identité du chromosome, on observe notamment que les chromosomes qui portent les NORs se retrouvent plus souvent aux bords des nucléoles.
- Ce positionnement préférentiel dépend aussi du cycle cellulaire : de manière générale, les RCs s'associent plus souvent en périphérie dans les cellules quiescentes.
- Ce comportement semble spécifique des RCs : la comparaison à des loci choisis aléatoirement sur le chromosome montre un positionnement aléatoire vis-à-vis des nucléoles et une attraction pour le nucléoplasme.
- Par ailleurs, les RCs semblent orientées préférentiellement vers la périphérie du noyau ou les nucléoles par rapport au reste du TC.

---

1. Par *nucléoplasme* nous entendons ici le compartiment nucléaire interne qui exclu les nucléoles.

### C.2.1 Introduction

Large number of studies have now clearly established that genomes are not randomly organized within nuclei, and that the spatial relationships between chromatin domains and proteinaceous nuclear bodies are important for understanding nuclear functions such as DNA transcription, replication and repair as well as RNA metabolism (Meldi et Brickner (2011)). Moreover, the different spatial intranuclear organizations observed in different cell types led to the hypothesis that the topological organization of the genome in the interphase nucleus may play a role in the regulation of gene expression (Bickmore et van Steensel (2013)). Nevertheless, despite years of imaging studies and the recent introduction of genome wide molecular approaches (Rouquette et al. (2010), van Steensel et Dekker (2010)), the rules and mechanisms governing the structure and internal organization of the interphase nucleus still remain elusive, and the fundamental question of how this dynamic organization relates to nuclear function remains unanswered (Rajapakse et Groudine (2011)).

Satellite DNA sequences, which represent the main component of centromeric chromatin, are amongst the first genomic regions for which a non random nuclear distribution was evidenced. Since the pioneering work of Manuelidis and coworkers (Manuelidis (1984), Manuelidis (1985)), several groups have shown that centromeric regions are mobile structures that associate with the nuclear lamina as well as the nucleolus in both mouse and human cells, and that their distribution in the interphase nucleus is modified in relation to cell cycle as well as physiological or differentiation state (reviewed in Pluta et al. (1995), Wiblin et al. (2005)).

The genomic distribution of satellite DNA sequences largely extends outside the centromere to embrace regions surrounding centromeres, referred to as pericentromeric. Numerous studies performed in mouse have shown that pericentromeric regions undergo major reorganizations during cellular differentiation (Terranova et al. (2005), Almouzni et Probst (2011), Aguirre-Lavin et al. (2012)) and that these regions form heterochromatin foci that can induce transcriptional repression of juxtaposed genes (Brown et al. (1997)). Although these data clearly associate centromeric regions (which will be further abbreviated CR) with a peculiar intranuclear dynamics and to global gene regulation mechanisms, the functional relevance of the specific nuclear positioning of CR has still not been fully understood. CR of human chromosomes contain different types of satellite DNA. The presence of several thousands of near-identical sequences has until now thwarted standard sequence assembly (Hayden (2012)). As a result, they are not included in currently available genome reference assemblies and, consequently, molecular approaches involving alignments of numerous short sequences on assembled genomes do not provide any quantitative information about centromeric regions. Despite these difficulties, years of effort have led to some knowledge about the sequence composition of specific chromosomes. The most abundant satellite sequence, called  $\alpha$ -satellite, is present at each centromere with millions of copies.  $\alpha$ -satellite DNA is made of 171 bp related mono-

mers that display specific organizational patterns. The existence of so-called higher order repeats that differ slightly between different chromosomes enables the specific detection of centromeric regions from individual chromosomes by Fluorescence In Situ Hybridization (FISH), a feature that has been exploited in a few imaging studies. It is therefore known that CR from different chromosomes can associate with the nuclear border and the nucleoli with different frequencies (Salníková *et al.* (2000), Weierich *et al.* (2003)), and it has been suggested that such observations may be correlated with cytologically visible chromosomal features (Carvalho *et al.* (2001)).

Molecular and imaging experiments can be used to infer the rules that dictate genome structure and functioning inside the cell. One must keep in mind that these rules are probabilistic, not deterministic (de Wit *et de Laat* (2012)), meaning that the position of a given genomic region with respect to nuclear landmarks will likely be different between two cells, but that regularities will be observed at the population level. The questions of establishing if certain nuclear structures significantly associate with each other and if these associations have functional and/or organizational implications require therefore to be able to perform quantitative comparisons of spatial relationships between different cell populations. For this, it is essential to ensure that cell preparation methods preserve the native state existing in live cells as much as possible, but also to be able to analyze a sufficient amount of data. Fluorescent imaging methods, combined with either confocal or wide-field microscopy, have several advantages compared to genome-wide molecular approaches : they provide, at the single cell level, information regarding interactions between different genomic regions and different nuclear landmarks simultaneously. Moreover, genomic regions carried by homologous chromosomes can be distinguished. The main difficulties that are encountered when implementing imaging approaches stem from the challenges in generating data revealing the three dimensional internal organization of a large number of cells and analyzing these data in 3D, taking into account the inherent variabilities of the samples. Due to the lack of reference points in the nucleus, a lot of studies have focused on the radial distribution of nuclear structures, which are computed either by measuring distance to the nuclear border (Wiblin *et al.* (2005)) or by calculating a theoretical center for the nucleus (Weierich *et al.* (2003)). Measurements are often performed on maximal projections of 3D image stacks. Despite recent efforts that have been undertaken to implement 3D-imaging platforms Jost *et al.* (2011), there is still a need for methods to both rapidly acquire and automatically analyze data on the spatial distribution of nuclear structures in a statistically meaningful way.

In the present study, we used a combination of fluorescence in situ hybridization and immunochemical labeling to tag the CR from different human chromosomes together with nuclear proteins/structures, and high-throughput fluorescence imaging to determine their three-dimensional distribution pattern within the nucleus of two diploid human lymphoblastoid cell lines. Using a recently developed image analysis tool (Ollion *et al.* (2013)), we directly and quantitatively investigated the nuclear positioning of specific human centromeres, taking into account the presence of nu-

cleoli in the nuclear interior. We demonstrate differential attraction towards both the nuclear border and the nucleoli, the first one being enhanced in quiescent cells and the second one in cycling cells, and that the strength of these attraction differs between chromosomes. These features of CR are different from what is observed with non centromeric genomic loci. Moreover CR are also not randomly distributed within their CT. Potential mechanisms that may explain these observations are discussed.

## C.2.2 Results

### C.2.2.1 Development of a robust protocol for studying centromere nuclear positioning during the cell cycle

We chose to characterize the nuclear organization of centromeres from different chromosomes in human lymphoblastoid cell lines with normal diploid karyotypes, such as GM06990 and GM12878. Detection of the centromeric regions from specific chromosomes can be achieved by FISH, using oligonucleotide probes that were designed to target specific  $\alpha$ -satellite repeats (see supplementary material). As cell-cycle is known to dramatically alter the nuclear positioning of CR (Solovei *et al.* (2004)), we decided to use nuclear markers, such as Ki67 and BrdU, in order to determine unequivocally the cell-cycle stage at the single-cell level. This approach was preferred over synchronization which never leads to homogeneous populations. Moreover, it is difficult to ensure that there are no drug artifacts caused during cell synchronization that affect the experimental observations. Centromeres are known to accumulate at the nuclear periphery and close to nucleoli.

Thus, in a first series of experiments aimed at better characterizing the internal organization of centromeres, we decided to simultaneously visualize nuclei using Hoechst staining, centromeres from one specific chromosome pair by FISH using an oligonucleotide probe targeting  $\alpha$ -satellite DNA, nucleoli, using an antibody targeting the C23 protein, as well as the Ki67 staining patterns (see Fig. C.14A). We adapted an immuno-FISH protocol in order to obtain good signals for the different nuclear structures to be visualized (i.e. centromere, nucleoli, and cell cycle markers) and reproducible measurements. After image acquisition, segmentation of nuclei, nucleoli and CR was achieved in an efficient and straightforward way for over hundreds of nuclei, using robust segmentation algorithms that were purposely developed and integrated in TANGO (see §C.1.2). A quantitative analysis of the Ki67 signal provided a faithful identification of cycling and non-cycling cells (see §C.1.4). The best protocol, which was inspired by the one developed by the cremer's group (Solovei *et Cremer* (2010)), provided an excellent reproducibility for measurements of nuclear and nucleolar volumes in different experiments, including different batches of cells fixed on different days, provided cycling and quiescent cells were distinguished (Fig. C.19). Unexpectedly, we had to use specific buffers and mounting conditions for obtaining this reproducibility (see §B.2).

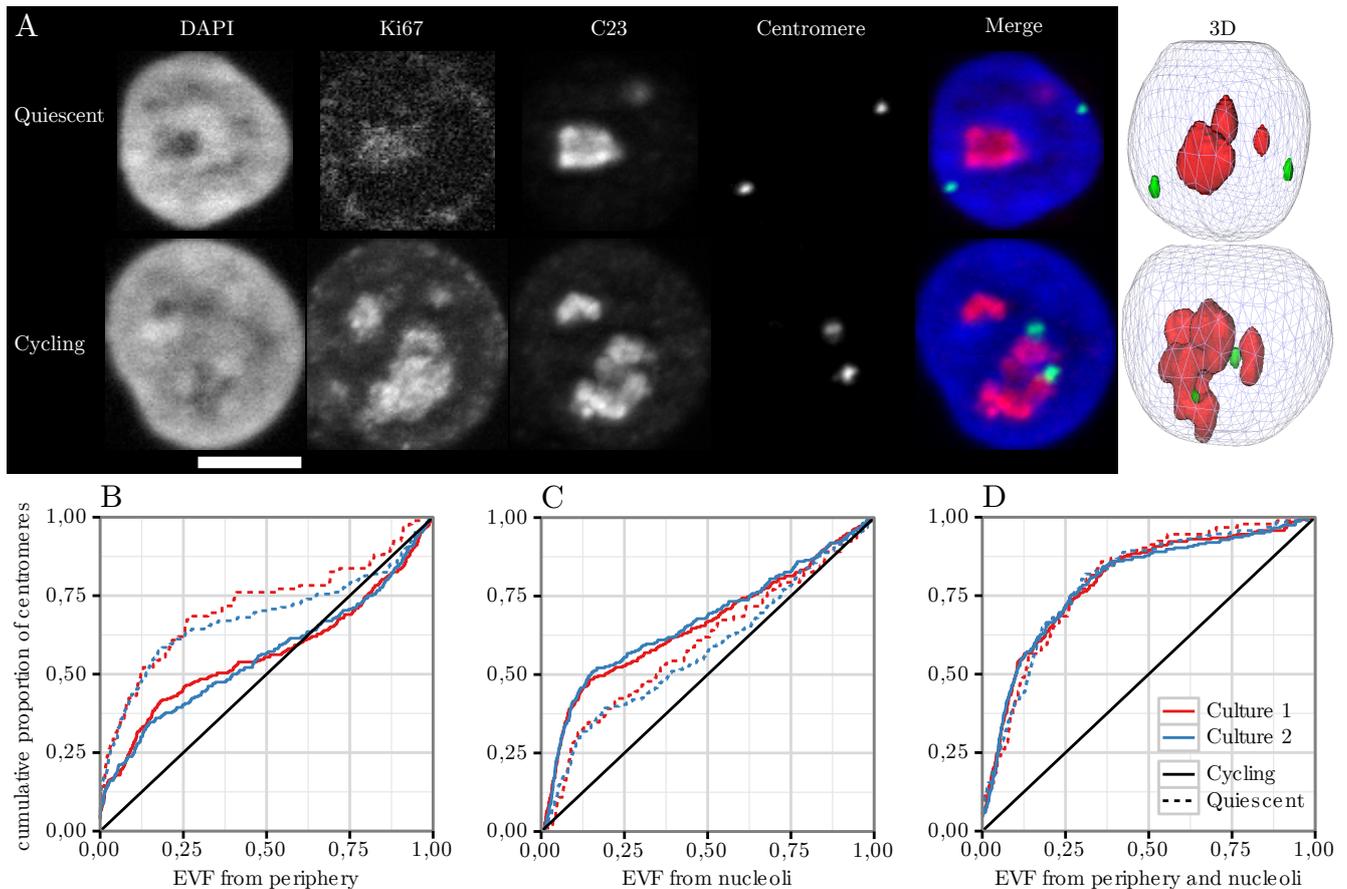


FIGURE C.14: Reproducibility of the nuclear positioning of the centromere of chromosome 18 between independent cultures. A. Typical images are shown for quiescent and cycling cells. Black and white images correspond to DAPI, Ki67, C23 and Chr18 centromere, as indicated. The merged image shows DAPI in blue, C23 in red and Chromosome 18 centromere in green. Scale bar represents 5  $\mu\text{m}$ . A three-dimensional reconstruction (by surface rendering) of nuclei is also shown at the far right end. Cumulative distributions of EVF to nuclear border (B), nucleolar border (C) and their combination (D) are shown for quiescent (dotted) and cycling (linear) cells. The red and blue curves were obtained from two different experiments, with fixation occurring at different times on two different batches of cells. Effectives (quiescent/cycling, centromeres) : 92/258 ; 180/390.

### C.2.2.2 Evaluation of the spatial distribution of centromeric signals

Our aim was to address the probabilistic positioning pattern of individual CR by measuring distances between these centromeres and specific nuclear landmarks in numerous nuclei. As lymphoblastoid cells have a spherical nuclei, it is essential to perform distance measurements in 3D. Such measurements were performed in large numbers of nuclei using our recently developed software, TANGO. In order to avoid biases due to variable sizes between different centromeric signals, the position of CR was estimated by the centroid of the FISH signals. Position of CR with respect to reference structures (in our case the nuclear periphery and the nucleolar periphery) can be described either with a discrete approach by defining a contact threshold and counting the proportion of centromeres in contact within the population (Carvalho *et al.* (2001)), or with a continuous approach by measuring the distance between the

centromere and the reference structure, and analyzing the distribution of distances. The discrete approach requires the setting of a contact threshold, which is difficult to define objectively and has a deep influence on the results. Therefore we chose a continuous approach. In order to normalize the positioning measurements against size and shape variations of nuclei, we used a recently developed measurement called eroded volume fraction (EVF) (Ballester *et al.* (2008)). The EVF of a point within the nucleus is defined as the fraction of nuclear volume lying between a considered point and the nuclear periphery (see §C.1.3.1 for a more thorough description of EVF calculation and properties). Figure C.14B displays the computed EVF, plotted as cumulative distributions (CD), obtained for the centromere of chromosome 18 from two different batches of GM12878 cells, after separation of the cycling and quiescent cells. It clearly shows that the CD are highly reproducible between different experiments and that there is a significant difference between the observed distributions in cycling and quiescent cells. In an attempt to better characterize the internal nuclear positioning of centromeres, we extended the concept of EVF to the study of the distribution of centromeres with respect to the nucleoli (Fig. C.14C, see §C.1.3.1). We also implemented another EVF measurement that takes into account both the nuclear border and the nucleoli, allowing to assess the attraction or repulsion towards the nucleoplasm (see §C.1.3.1) (Fig. C.14D). For these two measurements, CD were also highly reproducible and showed a significant difference between the observed distributions in cycling and quiescent cells.

### C.2.2.3 Interpretation of graphs

CD of EVF for loci that would be uniformly distributed within nuclei would be a straight line of equation  $x = y$  (indicated by the black line on graphs), reflecting a uniform distribution of EVF values. This is an inherent property of EVF (see §C.1.3.1). A positioning of the CD curve below this line can be translated by a decreased probability of being close to the reference structure (i.e. an apparent repulsion) while curves positioned above this line reflect an increased probability of being in proximity to the reference structure (i.e. and apparent attraction).

**EVF from nuclear border.** When EVF towards the nuclear border are considered, the distribution usually contain a small proportion of CR with null values, corresponding to CR in contact with the nuclear border according to our definition of the EVF from nuclear border (see §C.1.3.1). CD usually display a sharp increase at low EVF values and are located above the black line, followed by an inflexion of the curve, and an increase of the slope at high EVF values. This inflexion suggests that two populations of RC are present in cells : RC located close to the periphery and RC located in the interior of the nucleus. For CR from chromosome 18, the clear deviation of the CD from the uniform distribution (see §B.4 for details on statistical tests) provides evidence for an attraction of the CR towards the nuclear border. The attraction is stronger in quiescent cells, showing a relocation of CR from chromosome 18 towards the nuclear border in quiescent cells.

**EVF from nucleoli.** The CD of EVF to nucleoli display a different shape, with just two phases where the distributions seem to vary linearly. The definition of this EVF is such that points located on the border of the nuclei should have an EVF equal to the ratio of the nucleolar volume over the nuclear volume. This ratio can be calculated from measurements shown on figure C.19 to be approximately 0.1 in cycling cells and 0.15 in quiescent cells. We noticed that the two linear parts reach each other for positions with those EVF values. Therefore, the first part of the curve can be interpreted as reflecting the distribution of centromeres within nucleoli, and the second part their distribution outside the nucleoli. The CD clearly provide evidence for an attraction towards the nucleoli, which is stronger in cycling than in quiescent cells. Interestingly, we noticed that this attraction seemed to be correlated with the presence of a high slope (reflecting an important population of CR) for high EVF values (0.75 to 1) in the CD depicting the distribution toward the nuclear border, a fact that is not surprising knowing that nucleoli are in general located in the nuclear interior.

**EVF from nuclear border and nucleoli.** Finally the shape of the third version of EVF shows that centromeres are much more likely located close to the nuclear border or the nucleoli than expected under a uniform distribution, as there is a strong shift of every curves towards smaller EVF values. This can be interpreted as a repulsion of CR from the nucleoplasm as defined by the nuclear interior with the exclusion of nucleoli.

#### C.2.2.4 Differential nuclear positioning of individual centromeres

As a few reports have shown that different centromeres could interact with the nuclear border or with nucleoli with different frequencies, we decided to use the different measurements that we have implemented in order to compare a set of specific CR with each other. For this, we chose to focus on the GM06990 cell line, as this cell line provided the opportunity to compare more chromosomes than GM12878, and in particular to include acrocentric chromosomes that carry a nucleolar organizing region. The two probes that were designed to bind  $\alpha$ -satellite repeats from chromosomes 13 and 22 gave indeed the expected signals only on metaphases from GM06990. The difference between the two cell lines can be explained by the known polymorphism affecting the distribution and repeat number of these sequences (Lo *et al.* (1999)).

We performed experiments with probes targeting the CRs of chromosomes 1, 5, 7, 8, 10, 11, 13, 16, 17, 18, 22. We also included the analysis of some centromeres in the GM12878 cell line, to check if the distribution of individual centromeres would differ between two lymphoblastoid cells with different lineages. Figure C.15 displays the results obtained in the case of three centromeres (5, 17, and 18), which represent the 3 archetypal distributions we observed, for cycling and quiescent cells. The CD provide evidence for different distributions between individual CR that are conserved in both cell types, suggesting that the organizational patterns that we ob-

served reflect intrinsic properties of each centromere/chromosome that are common to EBV immortalized lymphoblastoid cell lines. For centromere 5, an attraction towards the nuclear border is observed in cycling cells, which is significantly enhanced in quiescent cells. This centromere is also attracted towards nucleoli in cycling but not at all in quiescent cells. CD for centromere 17 are shifted towards the opposite directions compared to those of centromere 18, revealing a strong attraction towards nucleoli and no attraction toward the nuclear border. The difference between cycling and quiescent cells is still present and significant for the 3 CR.

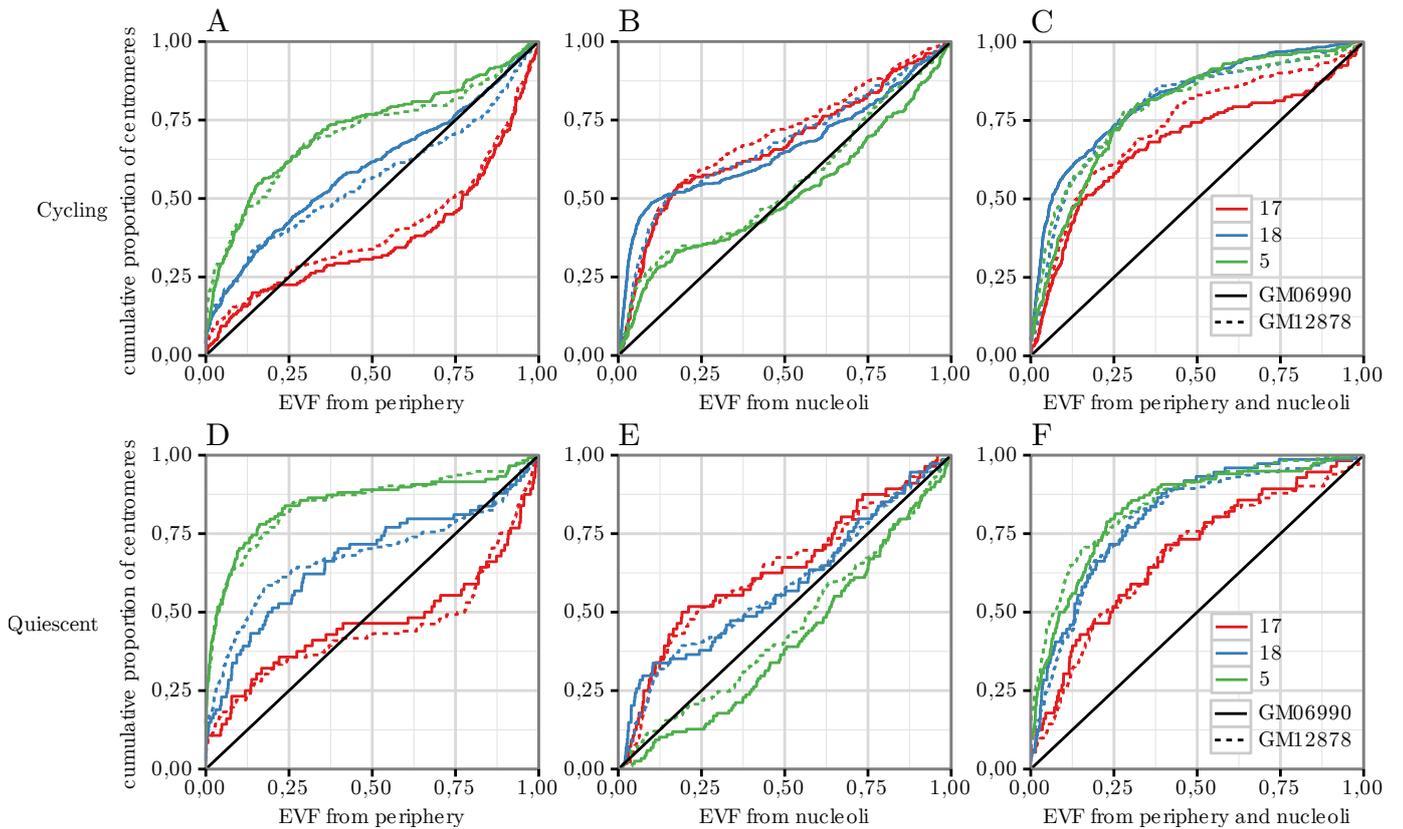


FIGURE C.15: Reproducibility of positioning pattern between GM12878 and GM06990. The three different cumulative distributions obtained with cycling GM06990 (linear) and GM12878 (dotted) cells are shown for three centromeres : 5 (green), 17 (red) and 18 (blue). Effectives (quiescent/cycling, number of centromeres) : GM06990 : 5 :118/246; 17 :56/160; 18 :150/236; GM12878 : 5 :124/206; 17 :132/272; 18 :92/258.

Additional data regarding centromeres 8 and 22 are provided in Figure C.20. The CR from the two acrocentric chromosomes that we could study (i.e. 13 and 22) displayed a much more pronounced attraction toward nucleoli than those from other chromosomes, as expected (Hernandez-Verdun (2006)). The attraction to the nucleoli was maximal in cycling cells, a situation where attraction toward the nuclear border seems to be abolished.

In order to provide an easy way of comparing every individual centromeres, we have plotted in figure C.16 the maximal deviations from the uniform distribution, a signed number which reflects if the cumulative distribution is above or below the

diagonal, for all the centromeres studied in the GM06990 cell line. The few cases where the maximal attraction and the maximal repulsion were within the same range of value are displayed in hatched bars. Although such a representation does not reflect all the information provided by the whole distribution, it contains its main information (except for the cases in which attraction and repulsion are of same magnitude) and it allows one to classify the chromosomes with respect to each others, and further demonstrates the existence of specific nuclear positioning of CR that is influenced by the proliferative state. The trend towards a significantly increased proximity of centromeres to the nuclear border in quiescent cells compared to cycling cells is true for all tested chromosomes, as well as the trend towards a significantly increased proximity of centromeres to nucleoli in cycling cells compared to quiescent cells. All CRs appear to associate to a certain extent with the nuclear border, except those from chromosomes 17, and from chromosome 22 in cycling cells. On the other hand, all CRs associate with the nucleoli in cycling cells except for CR from chromosomes 7 and 8, this association being highly enhanced for the two acrocentric chromosomes 13 and 22. Attraction toward nucleoli seems to be abolished in quiescent cells for CR from chromosomes 5, 7, 8, 10, 11. Interestingly, all CRs display a strong repulsion from the nucleoplasm, independent of the cell cycle. CD of EVF from nuclear border for CRs from chromosome 1 in cycling cells and chromosome 17 in quiescent cells display a strong inflexion which suggest that two population of CRs are present : a population attracted towards the nuclear border, and a population repulsed from the nuclear border ; the latter probably corresponds to CRs attracted towards the nucleoli, as suggested by the strong attraction towards the nucleoli and a strong repulsion from the nucleoplasm.

We wondered next if the distributions observed in cycling cells were in fact reflecting specific and different distributions during each phase of the cell cycle. As the configuration of our imaging apparatus allowed us to record up to 5 fluorescent signals, we were able to record a replication labeling pattern base on the incorporation of BrdU, in addition to the markers described above. An improvement of our cell cycle analysis strategy allowed us to distinguish  $G_1$ , early S and late S cells (see §C.1.4). Thus, we analyzed the positioning pattern of three centromeres (1, 7 and 11) by plotting the cumulative distributions for  $G_0$ ,  $G_1$ , early S and late S cells (Figure SC.21). The cumulative distributions observed for  $G_1$ , early S and late S were not superposable, but the differences that were observed between them for these three chromosomes were not significant and much lower than those observed between cycling and quiescent cells. Therefore, these results fully justify the choice of distinguishing only cycling and quiescent cells.

#### **C.2.2.5 Centromeres display a specific nuclear positioning compared to other genomic loci**

We further investigated how the nuclear organization of centromeric regions could be compared with that of other genomic regions. We decided first to submit two non-centromeric genomic regions from chromosome 1 to the same type of

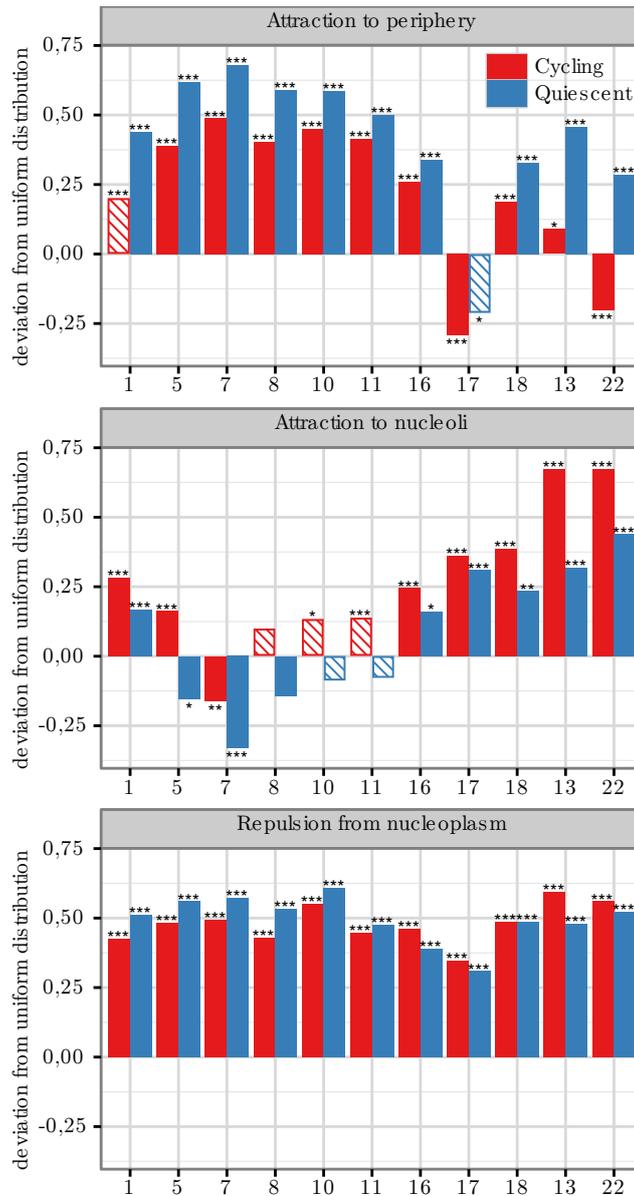


FIGURE C.16: Differential nuclear positioning of individual centromeres depending on the cell cycle stage in nuclei of lymphoblastoid cells. The graph indicates the maximal deviation from a uniform distribution of each centromere. Stars indicate that the p-value of the test comparing the experimental distribution to the uniform distribution. \* :p-value<0.01; \*\* :p-value<0.001; \*\*\* :p-value<0.0001. Hatched bars indicate that the maximal attraction and the maximal repulsion towards the reference structure are of comparable magnitude. Effectives (quiescent/cycling, number of centromeres) : 1 : 186/362; 5 :118/246; 7 :94/176; 8 :104/274; 10 :108/200; 11 :154/314; 16 :130/276; 17 : 56/160; 18 :150/236; 13 :120/336; 22 :128/298.

analysis as the one described in figure C.14. These regions, which were chosen on the long arm of chromosome 1, at distances from the centromere that correspond approximately to one third and two thirds of the full length of the arm, can be detected by FISH using probes synthesized from BACs clones (see §C.2.4 for genomic localization of probes). Figure C.17 shows the three cumulative distributions of

these two regions as well as those of the centromere from chromosome 1, for both cycling and quiescent cells. Statistical analysis showed that the distribution of the 1H158R locus was not different from a uniform distribution, and that 1L205G displayed a significant repulsion from nucleoli. Furthermore 1L205G display a strong attraction towards the nucleoplasm. These features make them different from what is observed for the centromere, i.e. a strong attraction to nucleoli, and a strong repulsion to the nucleoplasm, as described above. The two loci seem to differ slightly in their proximity to the nuclear border. One of them has a distribution that is similar to that observed for the centromere in cycling cells, but does not relocate to the nuclear border when cells get quiescent. The other one displays a more internal localization. Importantly, contrary to centromeres, there is no difference between cycling and quiescent cells for the intranuclear distribution of these two loci. These observations show that CR of chromosome 1 has a specific positioning pattern compared to at least two other genomic loci from the same chromosome.

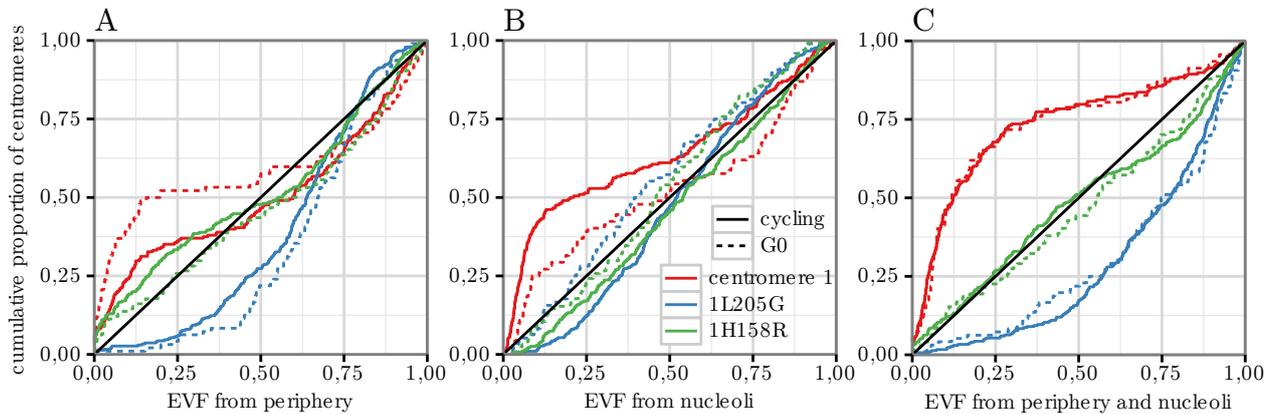


FIGURE C.17: Comparison of nuclear positioning of centromere and non centromeric regions for chromosome 1. The three different cumulative distributions obtained are shown for three chromosome 1 genomic loci : centromere (red), 1H158R (BAC, green), and 1L205G (BAC, blue). Data were split to distinguish quiescent (linear) and cycling (dotted) GM06990 cells. Effectives (quiescent/cycling, centromere number) : centro. 1 : 94/210; 1L205G : 96/302; 1H158R : 124/272.

### C.2.2.6 Centromeres display preferential orientations towards the nuclear border or the nucleoli within their CT depending on the nuclear localization of these CT

We developed a new approach that provides information regarding how a centromere is located with respect to the nuclear border or the nucleoli within its own chromosome territory (see Figure C.18 and §C.1.3.2). We have performed for chromosomes 1 and 7 experiments where nucleoli and Ki67 are detected by immunocytochemistry and both the chromosome territory and the centromere are detected by FISH. Then we selected among the nuclei those where the two alleles could be clearly distinguished. We decided to measure orientation after having defined chromosome subsets on the base of the contacts they make with the nuclear border or the nucleoli. More precisely, we distinguished three chromosomal populations-based

on a contact criterion : those where the CT makes contacts only with the nuclear border<sup>1</sup>, those where the CT makes contacts only with the nucleoli<sup>2</sup>, and those where the CT does not make any contact with either the nuclear border or the nucleoli. The number of nuclei where CT contacted both the nuclear border and the nucleoli was small, therefore these nuclei were not account in the analysis.

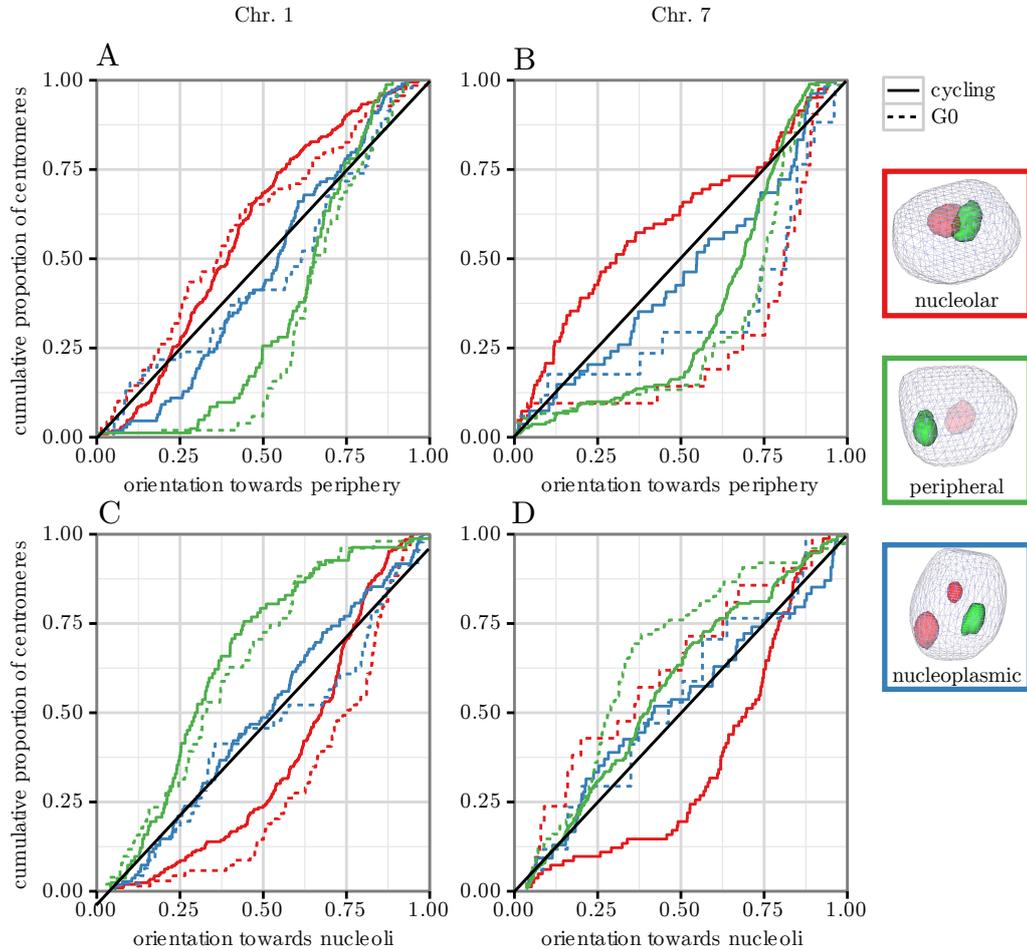


FIGURE C.18: Positioning of centromeres within chromosome territory for chromosomes 1 and 7. Each graph shows the cumulative distribution of computed orientations with respect to the nuclear border (A) and to the nucleoli (B), depending on cell-cycle and CT positioning (see legends on the graph). Effectives (quiescent/cycling, CT number) : 1 : nucleoli 69/209; nucleoplasm : 46/109; nuclear border : 51/82; 7 : nucleoli 21/82; nucleoplasm : 17/54; nuclear border : 75/191.

Figure C.18 displays the cumulative distributions of orientation obtained for cycling and quiescent cells for these three chromosomal subsets. For the centromere of chromosome 1, the red curve (CT associated to nucleoli) was significantly shifted towards low values for orientation towards the nuclear border and to high values for orientation toward the nucleolus, meaning that when the CT is close to the

1. CT were classified as being located next to the nuclear border when  $>10\%$  of its surface is in contact with the nuclear border.
2. CT were classified as being located next to nucleoli when more than  $5\%$  of its volume overlaps with a nucleoli.

nucleoli, the centromere tends to be closer to the nucleoli than the rest of the CT. Inversely, the green curve (CT associated to the nuclear border) was significantly shifted towards high values for orientation towards the nuclear border and to low values for orientation toward the nucleolus, meaning that when the CT is close to the nuclear border, the centromere tends to be closer to the nuclear border compared with the rest of the CT. The blue curve (CT that are neither close to the periphery nor to the nucleoli) does not significantly differ from a uniform distribution in both graphs, meaning that centromeres seem to be randomly oriented within CT for this population. This holds true for both quiescent and cycling cells, as there is no significant variation between cycling and quiescent cells for the cumulative distribution of orientation. For chromosome 7, we observed the same trends as for chromosome 1 for cycling cells, but the orientation in quiescent cells was different from that observed in the case of chromosome 1 for two CT configurations : for CT associated either with nucleoli only and CT lying in the nucleoplasm, a significant shift towards the nuclear border is observed in quiescent cells.

### C.2.3 Discussion

We have analyzed, using multicolor 3D-immunoFISH and 3D image analysis, the cell-cycle dependent spatial arrangement of individual CR in the nucleus of human lymphoblastoid cell lines. Despite the need for aggressive cell preparation methods that can have significant consequences on quantitative measurements, especially when FISH has to be used for the detection of specific centromeres, we were able to get reproducible measurements at the population level. The use of normalized measurements, such as the three EVF that we propose, and their display as cumulative distributions allowed us to infer, without the need for a distance threshold, that distinct centromeres had different probabilities of being in the vicinity of either the nuclear border or the nucleoli, and that these probabilities differed between cycling and quiescent cells. This organization is specific for centromeric regions and is not found for other genomic loci, which display a strong attraction toward the nucleoplasm.

Our results point to the cell cycle as a source of variability when studying nuclear organization. Reducing the influence of variable factors, by observing homogeneous populations of cells, can be crucial to ensure the reproducibility of the results. In the case of CR, distinguishing cycling and quiescent cells, thanks to a quantitative imaging approach, was sufficient. Current practice shows that the proportion of cycling and quiescent cells vary between cell lines, and for a single cell line depending on culture conditions (number of passages or different batches of serum). Such populations of cells will apparently display different association frequencies even if the nuclear distribution in each subpopulation of cells is the same.

Numerous studies have addressed the proximity of genomic loci to the nuclear lamina by implementing a so-called shell analysis, which sometimes included a normalization using the intensity of DNA counterstain. The present study emphasizes the importance of taking into account internal compartments such as the nucleoli.

The new EVF measurements that we introduce provide the opportunity to study how genomic loci are distributed with respect to the nucleoli as well as with the nucleoplasm. The differences that we observed between centromeric and non centromeric chromosomal loci emphasizes the interest of such descriptors which can be applied to any detectable genomic region and to any other nuclear structure provided it can be segmented.

Our results are in good agreement with previous imaging studies of CR. Using anticentromere antibodies, it has previously been shown that CR display an enhanced association with the nuclear border in quiescent cells while they localize more toward the nuclear interior in cycling cells (Solovei *et al.* (2004)). Our study shows that this property is true for each individual CR and that CR that are not close to the nuclear border are much more prone to be associated to the nucleoli than to remain in the nucleoplasm. We did not observe the enhanced attraction toward the nuclear border in S phase that was reported. This difference may be explained by the low number of nuclei that were included in this study, or by the fact that we observed individual centromeres by FISH instead of all centromeres using a CREST serum, an approach which may lead to some inaccuracy in counting aggregated centromeres. Differential associations of individual centromeres with the nucleoli and the lamina had been studied in G<sub>0</sub>/G<sub>1</sub> synchronized cells. If the global trends that were observed in this study are the same as in our observations, we provide clear additional evidence for increased association in G<sub>1</sub> versus G<sub>0</sub>, and our approach demonstrates that the underrepresentation of CR in the nucleoplasm is a property that is shared by all chromosomes. Molecular methods have also been implemented to study the positioning of DNA sequences with respect to the nuclear lamina or the nucleoli at the genome-wide level. Although CR are not analyzed in such approaches, increased association of pericentromeric regions with each structure have been reported, in accordance with our results (Németh *et al.* (2010), Zhao *et al.* (2006)). Nevertheless, it is important to note that such approaches rely only on the sequences that have been integrated into assembled genome maps. Therefore, they cannot be used to compare different CR in a quantitative way as the coverage of CR is only partial and differs between chromosomes.

Our results reveal that attractions of individual CR towards the nuclear periphery and towards the nucleoli vary in opposite directions. Interchromosomal differences may be explained by differences in the sequence composition of the different CR, which are known to vary regarding the type and number of repeated sequences, but also by differences in the global composition of the chromosome, which may be more or less prone to adopt a peripheral or internal localization. This idea points to the interest of studying CR and CT simultaneously. A previous study has addressed this question and concluded that CR localized at the periphery of CT (Weierich *et al.* (2003)). The orientation measurement that we introduced to get insights into how CR are positioned within their CT provides additional information, revealing that for chromosome 1, CR may be oriented either towards the periphery or toward the nucleoli depending on the localization of the CT. This suggests that two kinds of mechanisms may coexist in a same population of cells, one that anchors or attracts

centromeres to the lamina, another one that anchors or attracts centromeres to the nucleoli. The relative strength of these mechanisms would not be the same between different chromosomes, as we observed that chromosome 7 displayed in quiescent cells a preferential orientation towards the periphery even for CT associated to the nucleoli. This is consistent with the very poor attraction toward the nucleoli observed for CR from chromosome 7 in quiescent cells. These observations also suggest that CR may play a role in the non random nuclear organization of CT.

CR present original properties at the molecular level compared to other genomic loci. The repeated character of the DNA sequences leads to the presence of regularly spaced binding sites for DNA binding proteins and provides the opportunity to concentrate proteins or enzymatic activities at a specific genomic loci and therefore at spatially defined foci in the nucleus. Conversely, it also increases the probability to interact with a nuclear structure that is enriched in proteins that have some affinity for these sequences. Various proteins have been described to interact with satellite sequences, such as HMGA1 (Vogel *et al.* (2011)) and methyl-CpG binding proteins Brero *et al.* (2005), as well as postulated components of a nuclear matrix (Strissel (1996)) and nuclear intermediate filaments Tolstonog *et al.* (2000). Proteins from the nuclear lamina have been shown to interact with DNA Stierlé *et al.* (2003) or with chromatin (Shimi *et al.* (2008), Solovei *et al.* (2013)). In *Drosophila*, the nucleoplasmic homolog NLP has recently been shown to mediate anchoring of centromeres at the nucleolus (Padeken *et al.* (2013)). Such proteins may be actors of the specific nuclear positioning of CR by different mechanisms, for example by anchoring them to fixed or mobile nuclear components. One must keep in mind that centromeric regions from different chromosomes are also known to interact with each other (Solovei *et al.* (2004), Alcobia *et al.* (2003)), therefore a mechanism that drives the nuclear positioning of the CR from a single chromosome pair may also have influence on other chromosomes. For example, the presence of CR from non-acrocentric chromosomes close to nucleoli may result from associations of these CR with those from the acrocentric chromosomes, which are located at nucleoli when rDNA genes are expressed.

One important question that will have to be addressed regards the dynamics of CR during the cell cycle and the reorganizations that occur during cells division. Recent work has provided evidence for the existence of reorganizations of constitutive heterochromatin that proceed over more than two weeks in cells that have stopped proliferation (Solovei *et al.* (2013)). As it has been reported that nuclear organization remains stable after cells have started to proceed through G<sub>1</sub> phase (Walter *et al.* (2003)), it is very likely that kinetically disfavored interactions will become more frequent in quiescent cells. Such a mechanism may explain the enhanced interaction of CR with the nuclear border in quiescent cells.

Finally one may not exclude the existence of active relocalization mechanisms, as it has been reported that individuals or a small group of centromeres can occasionally move at rates of 7–10  $\mu\text{m}\cdot\text{h}^{-1}$  (Shelby *et al.* (1996)). Such questions will likely be addressed in the near future thanks to the use of sequence-specific DNA binding proteins that permit the visualization of CR in live cells (Ma *et al.* (2013)).

Understanding the mechanisms that drive the nuclear organization of CR should allow one to address the question whether the specific nuclear distribution of CR have some functional significance. One interesting hypothesis is that CR may control the global organization of genomes by constraining chromosome positioning. Mechanisms involving CR may thus explain recent observations in which proteins from the nuclear lamina as well as nuclear actin/myosins play a role in shaping the nuclear organization of CT (Malhas et Vaux (2009), Zuleger et al. (2013)). Another possibility is that association of CR to specific nuclear sub-compartments may facilitate some metabolic processes. The nucleolar border represents an interesting candidate as it has been shown that propagation of heterochromatin from the inactivated X chromosome is accompanied by transient associations of this chromosome with the nucleoli (Zhang et al. (2007)). Finally, CR may be involved in gene regulatory mechanisms. The association of genes with peripherally located heterochromatin has been suggested to contribute to their down-regulation (Dillon (2008)). Accumulation of pericentromeric heterochromatin at the periphery of the nucleus might provide a strategy for facilitating this process in quiescent cells. It will be interesting to check if the increased association of CR with the nuclear border observed during cell differentiation reflects global reorganization of heterochromatin in these cells or is just the consequence of a higher proportion of quiescent cells (Wiblin et al. (2005)).

In summary, the present study, which provides a good example of how accurate quantitative description of the spatial organization of the nucleus can be achieved thanks to the use of high throughput imaging of fixed cells, emphasizes the importance of differentiating cycling and quiescent cells and to take into account the presence of internal nuclear structures such as the nucleoli. The specific features of CR make them potential candidates as major drivers of genome organization in the nucleus, in addition to their role at mitosis. Our findings indicate that association of these regions with the nuclear lamina or the nucleoli differ between individual centromeres/chromosomes and that CR are not randomly distributed within CT, suggesting the existence of not yet characterized mechanisms that drive their peculiar organization. The methodologies described in this paper will be helpful to study anomalies of nuclear organization of CR that are observed in cancer cells as well as other diseases or in cases of infertility (Finch et al. (2008)). By providing the opportunity to reintegrate poorly sequenced repeat-rich chromosomal loci in nuclear organization studies, they should help uncover how the topology of the genome can impact the function of the living cell.

#### C.2.4 Supplementary Materials

**Probes.** Table C.1 depicts the sequences of the probes used for the detection of centromeric regions from specific chromosomes, as well as the washing temperature used for each of them and the fluorophore/hapten that is used for detection.

Chromosome	Sequence	Label	T(°C)
1	5' AaAcTgCtGCgTgAtGtGt <sup>3'</sup>	5' Cy3	70
5	5' AaAcTgCtCTgCgAtGtGt <sup>3'</sup>	3' dig	70
8	5' GaAaTPtTtCaGLALAPtT <sup>3'</sup>	5' Cy5	65
10	5' AaTtGgCcCtCtTtGaGc <sup>3'</sup>	3' dig	65
11	5' AgGgTtTcAgAgCtGcTc <sup>3'</sup>	5' Alx488	65
13	5' tAcCcAgCcAaAgGaGtT <sup>3'</sup>	5' Alx488	65
16	5' AaCaAgAcAaALTcGtTc <sup>3'</sup>	3' dig	65
17	5' tGaTtGaGtTgAaCtCaC <sup>3'</sup>	5' Cy5	65
18	5' GtGtGtCcTcAaCtAaAg <sup>3'</sup>	3' dig	65
22	5' GtGtGtCcTcAaCtAaAg <sup>3'</sup>	5' Cy5	72

TABLE C.1: Sequences of the probes used for the detection of centromeric regions from specific chromosomes. Small letters indicate Locked Nucleic Acids (LNA) modifications, dig is for digoxigenin, Alx488 is for AlexaFluor488. For each probe, the the washing temperature is indicated.

For each probe, we checked on metaphase that they hybridized to the expected chromosome and that both chromosomes were labeled within the studied cell line. An absence of signal was sometimes noticed on one or two chromosomes for the probes targeting chromosomes 13 and 22, which are due to a known polymorphism affecting the repeat number on those chromosomes. This is the reason why results regarding these chromosomes could only be obtained with the GM06990 cell line. The fluorescently labeled non centromeric probes were obtained thanks to a random priming kit (Bioprime, Invitrogen), using bacterial artificial chromosomes (BAC) as templates. The two BACs, which were obtained from Children's Hospital Oakland Research Institute, are the following ones (number refer to the positions of the first and last nucleotides in the human genome assembly, version GRCh37.p13). 1L205 : RP11-6B6 (205511288-205671057) 1H158 RP11-520H16 (158874920-159053664)

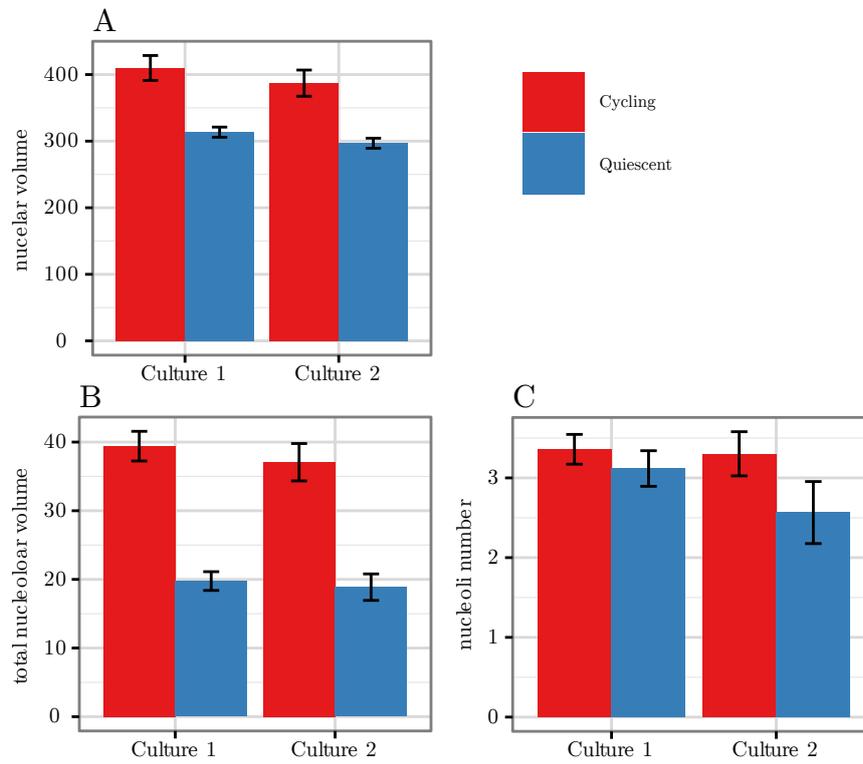


FIGURE C.19: Reproducibility of nuclear and nucleolar volume measurements. Cumulative distributions of nuclear volumes (A), nucleolar volumes (B), and nucleoli number (C) for cycling and quiescent cells. Mean values and standard deviations were computed for the experiments which results are shown in figure C.14. Effectives (quiescent/cycling, centromeres) : 92/258 ; 180/390.

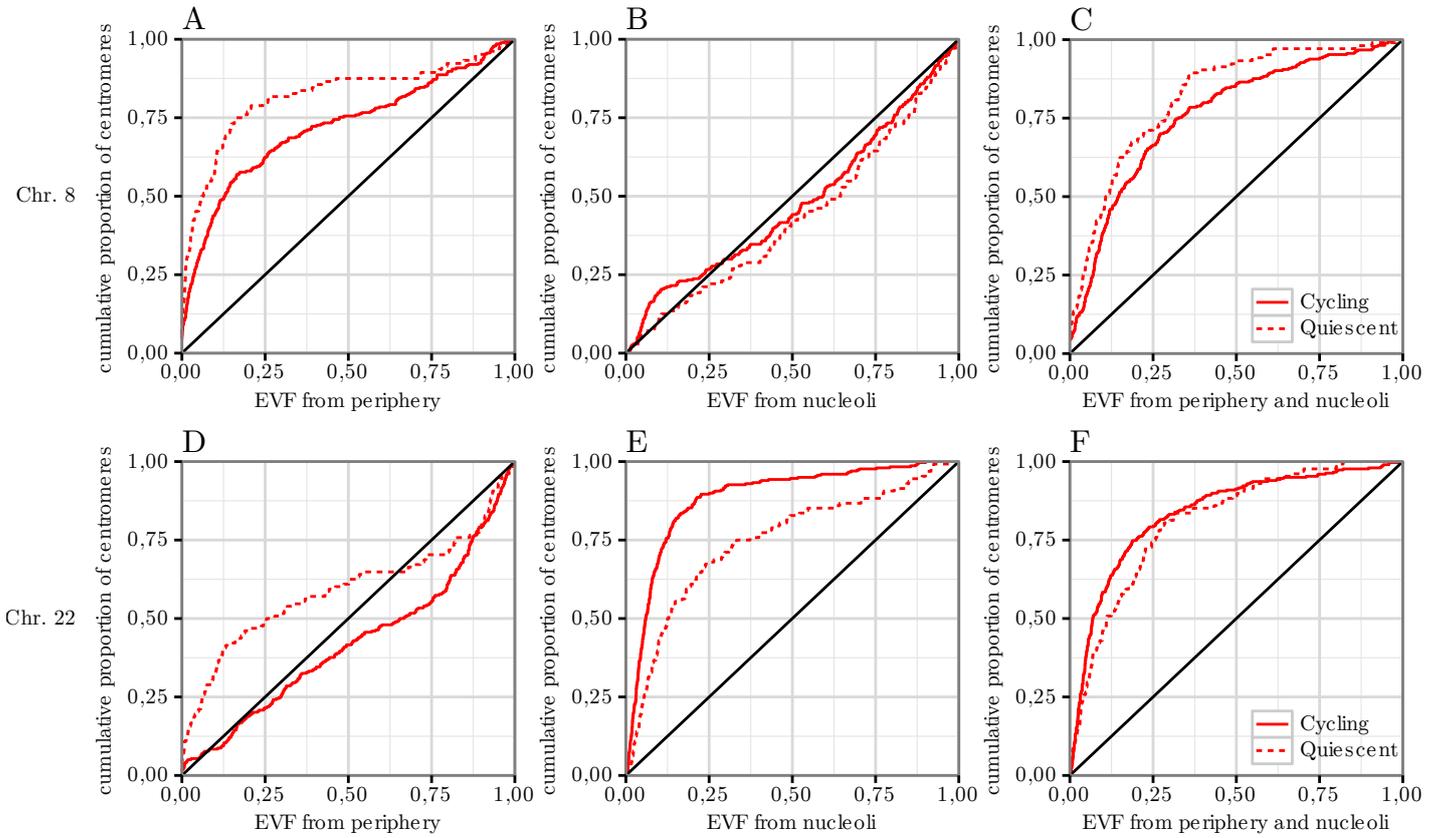


FIGURE C.20: Positioning pattern of centromere from chromosomes 8 and 22 in GM96990 cells. For each chromosome, the three different cumulative distributions are shown for cycling (linear) and quiescent (dotted) cells. Effectives (quiescent/cycling, centromere number) : 8 :104/274 ; 22 :128/298.

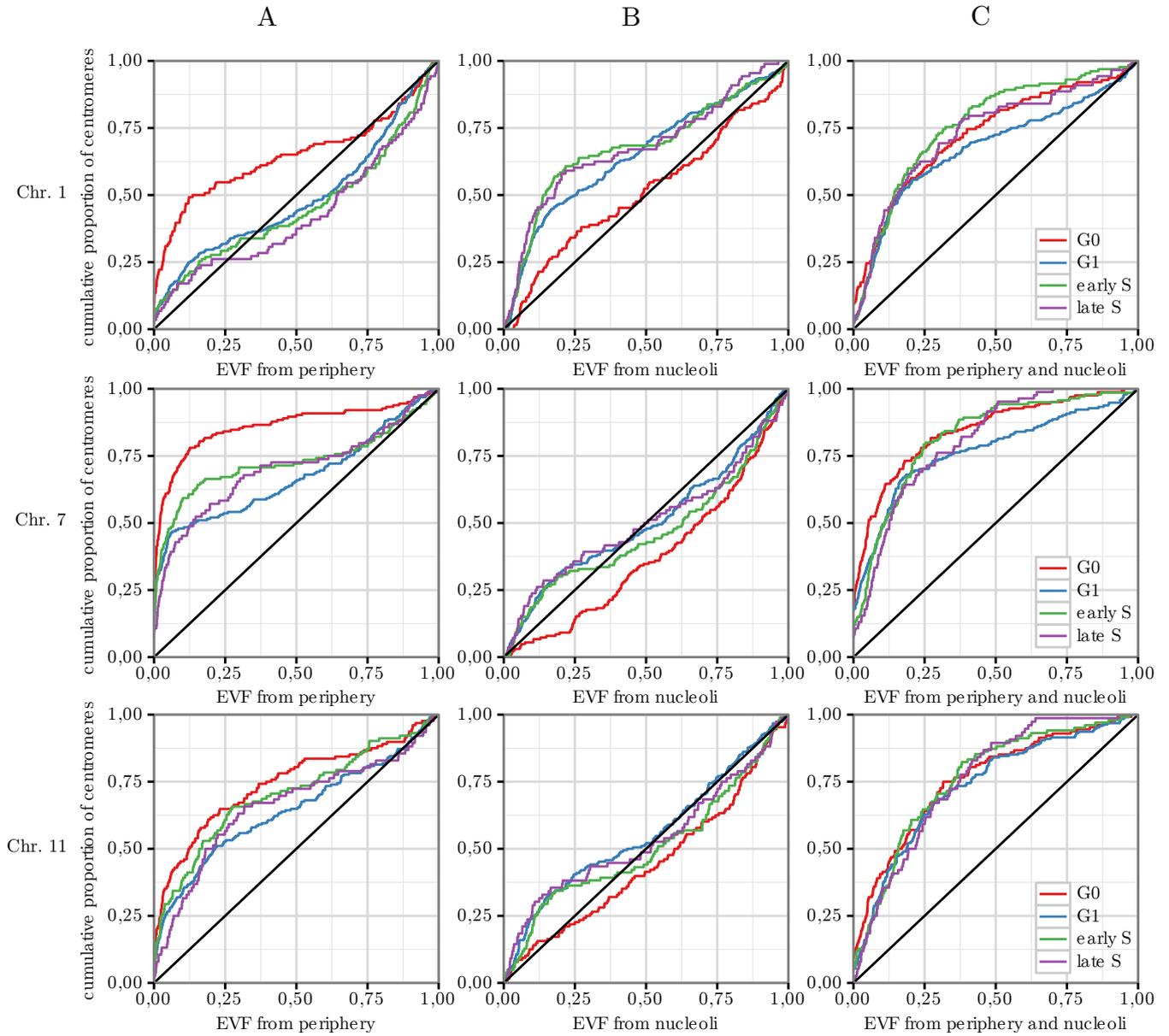


FIGURE C.21: Influence of the cell cycle on centromere nuclear positioning. Cumulative distributions of EVF computed from the nuclear border (A, above) and from the nucleoli (B, below) are shown for three centromeres, as indicated. In each case, data are split between quiescent cells (red), G1 cells (blue), and early (green) and late (purple) replicating cells. Data were obtained with GM12878 cells. Effectives (quiescent/G1/early S/late S, centromere number) : 1 : 63/120/44/65 ; 7 : 82/97/42/70 ; 11 : 64/94/38/51.

**Nuclear distribution of centromeres analyzed during S-phase.** The positioning pattern of three centromeres (1, 7 and 11) was analyzed by taking account the different phases of the cell cycle, using the strategy described in §C.1.4. G2 nuclei were excluded from this analysis because of insufficient numbers, which are likely due to elimination of these cells, that often display duplicated centromeric signals after the segmentation process. Results are shown on Figure SC.21.

### C.3 Agrégations des régions centromériques humaines

Comme décrit au §A.3.2.1, les régions péricentromériques portées par plusieurs chromosomes s'assemblent dans le noyau en interphase pour former des foyers d'hétérochromatine nommés chromocentres. Un marquage de l'ADN au DAPI permet de détecter ces structures chez de nombreuses espèces, incluant la souris. De telles structures n'ont jamais été observées chez l'homme, cependant plusieurs résultats expérimentaux suggèrent qu'elles pourraient exister. De nombreuses études ont mis en évidence que lorsque les centromères sont détectés par immunocytochimie avec des anticorps dirigés contre des protéines centromériques<sup>1</sup>, le nombre de foyers observés est souvent inférieur à 46 ; même dans le cas de cellules dont le caryotype est normal (Solovei *et al.* (2004)). Ces études en déduisent souvent que les foyers observés correspondent à des associations de centromères, et les nomment parfois *chromocentres*, par analogie avec les structures observées chez la souris (Bartholdi (1991), Beil *et al.* (2002), Alcobia *et al.* (2003), Solovei *et al.* (2004)). Nous pensons que cette dénomination n'est pas appropriée, car les structures observées par un marquage des protéines centromériques chez l'homme ne sont pas directement comparables aux structures observées chez la souris qui correspondent à des associations de régions péricentromériques (voir §A.3.2.1). Par ailleurs, des foyers d'hétérochromatine auxquels plusieurs foyers CenpA sont associés ont été observés par immunocytochimie en utilisant des anticorps dirigés contre H<sub>4</sub>K<sub>20</sub>me<sub>3</sub> (Zinner *et al.* (2006)). Il est possible que ces structures soient l'équivalent des chromocentres murins ; ceci supposerait que ces foyers correspondent à des séquences péricentromérique. Cela n'a jamais été montré probablement à cause de la difficulté de marquer l'ensemble des séquences péricentromériques humaines.

Dans cette partie nous allons étudier les associations de centromères et de régions péricentromériques en utilisant une nouvelle sonde qui permet la détection des régions péricentromériques humaines ; ce afin de déterminer si des structures similaires aux chromocentres murins existent chez l'homme. Afin de décrire le phénomène d'agrégation de RCs, nous commençons par définir des descripteurs permettant de caractériser les structures que nous observons. Ceci nous permet d'étudier quantitativement les variations du phénomène d'agrégation au cours du cycle cellulaire, la localisation sub-nucléaire des agrégats de RCs, et enfin de comparer plusieurs types cellulaires.

#### C.3.1 Résultats

##### C.3.1.1 Associations de centromères

**Observation des associations de centromères.** Dans toutes les lignées que nous avons utilisées, nous avons aussi détecté un nombre de foyers CenpA infé-

---

1. Telles que CenpA ou le sérum CREST.

rieur à 46 (voir C.22B); alors que d'une part le caryotype de nos lignées est normal<sup>1</sup>, et d'autre part nous avons développé un algorithme de segmentation performant pour séparer les foyers proches (voir §C.1.2.2). Nous observons que les noyaux qui comportent moins de centromères contiennent aussi des foyers plus volumineux et plus intenses. Afin de quantifier cette observation, nous avons mesuré la variabilité du signal CenpA entre les différents foyers de chaque noyau, et établi une corrélation négative entre cette variabilité et le nombre de foyers détectés par noyau (voir figure C.22C). Le coefficient de corrélation de Pearson estimé est de -0.798, et le test de corrélation de Pearson est très significatif<sup>2</sup>. Il semble évident de postuler que la taille et l'intensité<sup>3</sup> d'un foyer composé de plusieurs centromères augmente avec le nombre de centromères qui le composent, cette corrélation négative peut donc être interprétée par une agrégation des centromères.

Dans le reste de la partie, nous distinguons donc les termes *foyer CenpA* et *centromère*, le premier pouvant correspondre à l'association de plusieurs centromères.

**Attribution du nombre de centromères par foyer CenpA.** Afin de pouvoir étudier directement les centromères et non les foyers CenpA, nous avons élaboré un algorithme qui permet d'estimer le nombre de centromères formant chaque foyer CenpA. Comme nous l'avons montré, l'intensité du signal CenpA est liée au nombre de centromères; cet algorithme se base donc sur cette relation. L'approche la plus directe aurait été d'estimer l'intensité absolue correspondant à un centromère unique pour ensuite déduire directement le nombre de centromères par foyer, mais cette estimation est délicate; d'une part les centromères portés par différents chromosomes ne s'étendent pas sur des régions génomiques de même tailles donc le signal correspondant devrait être variable entre les centromères; et d'autre part on observe des variations dans l'intensité absolue du signal entre les différents noyaux. Nous avons donc conçu un algorithme qui se passe de cette estimation, et attribue un *poids* à chaque foyer CenpA correspondant à une estimation du nombre de centromères qui le composent, en supposant que la somme des intensités des centromères correspond à l'intensité du foyer CenpA. Son fonctionnement est décrit dans la figure C.22D.

Nous n'entendons pas obtenir un résultat exact par cette méthode, mais il s'agit d'une correction par pondération, en calculant la pondération la plus probable sans ajout d'informations supplémentaires, autres qu'un nombre total de centromères égal à 46. Le résultat de l'algorithme est sensible à la fois à la variabilité de taille des centromères, et à la qualité de la segmentation puisque l'intensité d'un foyer CenpA est calculée dans le volume segmenté de chaque foyer CenpA. Nous vérifierons la fiabilité de cette méthode par la reproductibilité des résultats qui l'utilisent.

---

1. Les cellules en culture pouvant facilement réorganiser leur caryotype, la ploïdie a été régulièrement vérifiée en comptant les chromosomes dans des étalements de métaphases.

2. p-value  $< 2.2 \times 10^{-16}$

3. Par intensité, nous entendons la somme de l'intensité des voxels se trouvant au sein du volume d'un foyer segmenté donné.

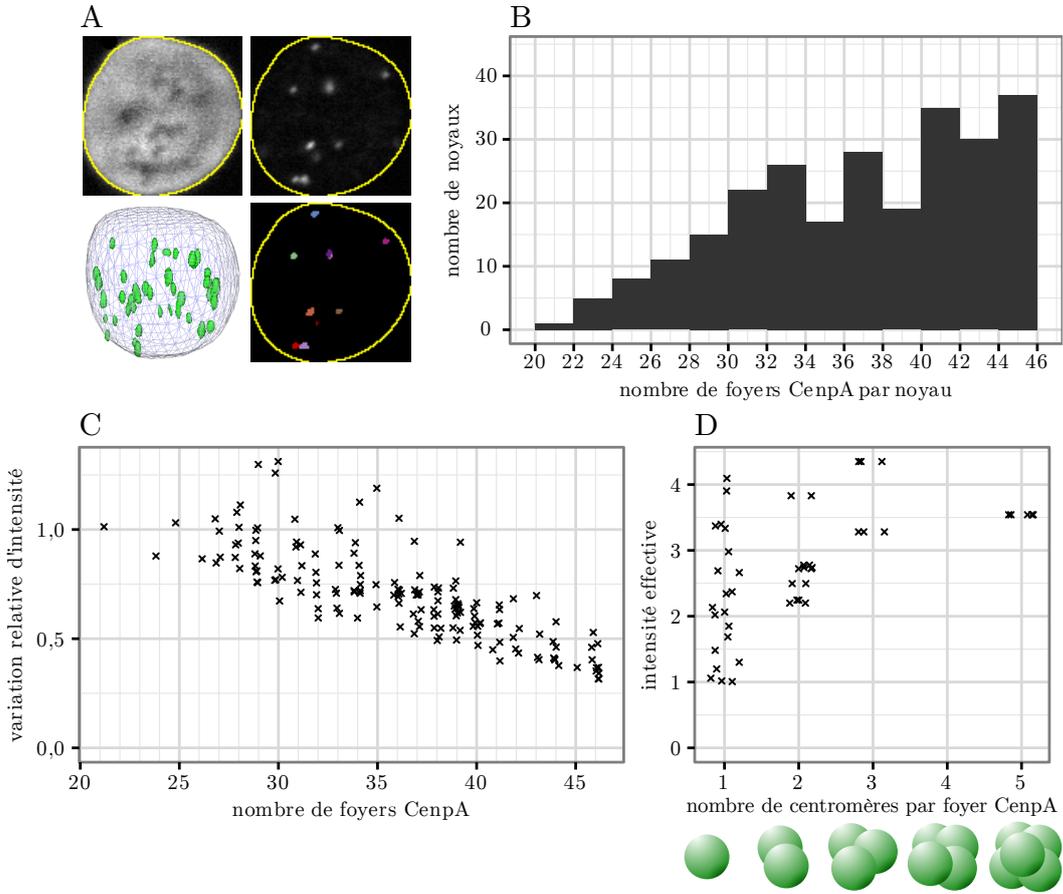


FIGURE C.22: Attribution du nombre de centromères par foyer CenpA. A : Expérience d'immunocytochimie sur des cellules lymphoblastoïdes GM12878. De haut en bas et de gauche à droite : marquage de l'ADN au DAPI (plan); marquage des centromères par l'anticorps anti-CenpA (plan); représentation 3D des centromères segmentés; masque des centromères segmentés (plan). Le contour du noyau segmenté est dessiné en jaune sur les images. Barre d'échelle : 5  $\mu\text{m}$ . B : Histogramme du nombre de foyers CenpA détectés par noyau. C : Corrélation entre la variation d'intensité du signal CenpA entre les différents foyers CenpA et leur nombre par noyaux. Pour chaque noyau, la variation d'intensité du signal CenpA est estimée par la mesure  $\frac{\sigma}{\mu}$  où  $\sigma$  est l'écart-type de la distribution d'intensité du signal CenpA intégré dans chaque foyer CenpA. Cette mesure est normalisée par la moyenne  $\mu$  des intensités afin de pouvoir comparer les noyaux entre eux, car il y a de grandes variations des valeurs absolues d'intensité d'un noyau à l'autre. Chaque croix représente un noyau. Seuls les noyaux ayant moins de 46 centromères ont été considérés. D : Exemple du résultat de l'algorithme d'attribution du nombre de centromères par foyer CenpA pour un noyau. Chaque croix représente un foyer CenpA. Le fonctionnement de l'algorithme est le suivant :

- Chaque foyer CenpA d'indice  $i$  est caractérisé par son intensité  $I_i$  et sa pondération  $N_i$  qui correspond au nombre de centromères attribués, initialisée à 1 au début de l'algorithme. Son intensité effective est définie par :  $I_i^{eff} = \frac{I_i}{N_i}$
- Le nombre de foyers CenpA est  $N^{CenpA}$ , le nombre total de centromères est donc  $N^{centro} = \sum_{i=1}^{N^{CenpA}} N_i$
- Il fonctionne itérativement : tant que  $N^{centro} < 46$ ;  $i^{max} = \text{argmax}_{i \in [1; N^{CenpA}]} (I_i^{eff})$ ;  $N_{i^{max}} = N_{i^{max}} + 1$
- En d'autres termes : tant qu'il y a moins de 46 centromères, on attribue un centromère supplémentaire au foyer CenpA dont l'intensité effective est la plus élevée.

### C.3.1.2 Associations de régions centromériques

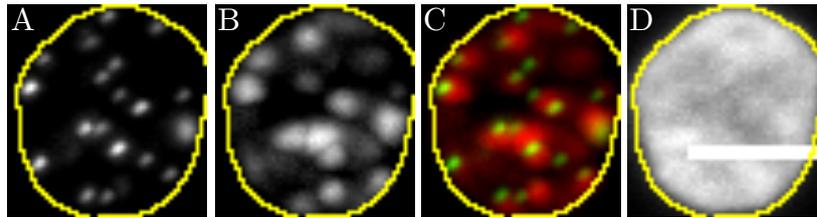


FIGURE C.23: Détection de foyers d'hétérochromatine par immunocytochimie dans les cellules GM12878. Les images A-D sont les projections maximales des 5 plans médians de la pile d'images, le contour du noyau segmenté est tracé en jaune. Barre d'échelle : 5  $\mu\text{m}$ . A : marquage des centromères par l'anticorps anti-CenpA. B : marquage de l'hétérochromatine par un anticorps anti- $\text{H}_3\text{K}_9\text{me}_3$ . C : superposition en fausses couleurs des signaux  $\text{H}_3\text{K}_9\text{me}_3$  et CenpA. D : marquage de l'ADN au DAPI.

**Observation des associations de régions péri-centromériques.** Des expériences d'immunocytochimie utilisant des anticorps dirigés contre  $\text{H}_3\text{K}_9\text{me}_3$  et CenpA montrent que  $\text{H}_3\text{K}_9\text{me}_3$  forme des foyers, auxquels plusieurs foyers CenpA sont associés dans les noyaux en interphase de cellules lymphoblastoïdes (Fig. C.23). Rappelant fortement l'organisation des RCs en chromocentres murins. Cependant, contrairement à la souris, l'anticorps dirigé contre  $\text{H}_3\text{K}_9\text{me}_3$  marque aussi des régions plus diffuses et éloignées du marquage CenpA, qui correspondent certainement à l'hétérochromatine non-péri-centromérique.

Afin de déterminer si ces structures sont similaires aux *chromocentres* murins, nous avons conçu une sonde permettant de détecter un grand nombre de régions péri-centromériques possibles. Les séquences les plus abondantes et présentes sur tous les chromosomes sont les  $\alpha$ -satellites, nous avons donc conçu une sonde ciblant une partie très conservée de ces séquences, que nous désignerons par PCY par la suite. Cette sonde de séquence *TTTGTGATGTGTG* est composée uniquement d'oligonucléotides LNA (voir Fig. A.17). Les régions marquées par PCY ont été analysées par FISH sur étalements de métaphases (voir figure C.24A-C). Cette sonde détecte une partie des régions péri-centromériques de chacun des chromosomes ; seuls deux chromosomes<sup>1</sup> sont plus faiblement marqués (voir figure C.24C). Cette sonde ne marque qu'un sous-ensemble des régions péri-centromériques, d'une part parce que les séquences  $\alpha$ -satellites forment une famille de séquences très divergentes, il est donc difficile de toutes les cibler avec la même sonde ; et d'autre part parce que les régions péri-centromériques ne sont pas composées uniquement de séquences  $\alpha$ -satellites (voir §A.2.3). Nous avons comparé le marquage de cette sonde avec celui de trois sondes oligonucléotidiques 50-mer couvrant la séquence consensus  $\alpha$ -satellite sur des étalements de métaphases et en 3D ; et nous avons observé que la sonde PCY inclut la plupart du marquage de ces trois sondes. C'est donc à notre connaissance le marqueur disponible permettant de détecter le plus grand nombre de répétitions péri-centromériques humaines.

Des expériences d'immuno-FISH, combinant PCY avec un marquage des centromères par un anticorps dirigé contre CenpA, montrent que dans le noyau en interphase, ces séquences forment des foyers à la périphérie desquels se localisent les

1. Il s'agit des chromosomes 17, qui ont été identifiés par FISH en utilisant une sonde spécifique.

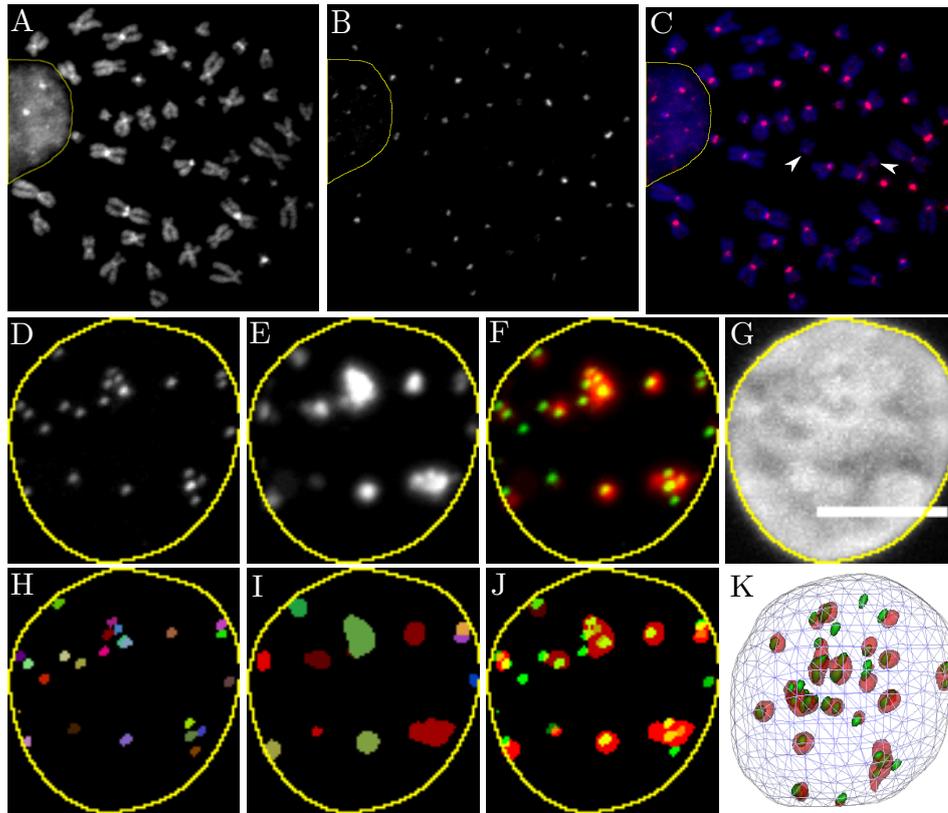


FIGURE C.24: Détection des séquences péracentromériques humaines par FISH. A-C : Expérience de FISH sur métaphases de cellules GM12878. L'objet séparé par un contour jaune est un noyau en interphase. A : Marquage de l'ADN au DAPI. B : Marquage des séquences  $\alpha$ -satellites par la sonde PCY. C : Superposition des canaux. Les pointes de flèche indiquent les deux chromosomes 17, très faiblement marqués par la sonde. D-K : Expérience d'immuno-FISH 3D sur cellules GM12878. Les images D-J sont les projections maximales des 5 plans médians de la pile d'images, le contour du noyau segmenté est tracé en jaune. Barre d'échelle : 5  $\mu$ m. D : Marquage des centromères par l'anticorps anti-CenpA. E : Marquage des régions péracentromériques par la sonde PCY. F : Superposition en fausses couleurs des signaux PCY et CenpA. G : Marquage de l'ADN au DAPI. H-I : masques de segmentation des images D-E respectivement ; chaque objet est identifié par une couleur différente sur l'image. J : Superposition des masques de segmentation H et I. K : Représentation surfacique 3D des objets segmentés des marquages PCY et centromériques.

foyers CenpA (voir figure C.24D-G). Le fait que plusieurs foyers CenpA se trouvent parfois associés à un même foyer PCY montre que ces foyers correspondent à l'association de RCs portées par plusieurs centromères. Ces résultats montrent que, comme chez la souris et de nombreux mammifères, l'hétérochromatine péracentromérique humaine forme un compartiment nucléaire caractérisé par l'association de régions péracentromériques et centromériques portées par plusieurs chromosomes. De telles structures n'avaient jamais été mises en évidence chez l'homme à notre connaissance.

Il est donc possible d'observer ces foyers par le biais d'un anticorps dirigé contre  $H_3K_9me_3$  ou avec la sonde PCY. Le premier marqueur n'est pas spécifique des régions péracentromériques, il est donc difficile de distinguer l'hétérochromatine péracentromérique du reste de l'hétérochromatine. En revanche, le marquage par la sonde PCY est spécifique, mais correspond à un sous-ensemble des régions péracen-

trométriques. Dans le reste de cette étude, nous avons choisi d'utiliser la sonde PCY malgré cette sous-estimation, car les foyers qu'elle marque sont strictement péricentromériques d'une part, et sont plus simples à segmenter automatiquement d'autre part car le marquage diffus H<sub>3</sub>K<sub>9</sub>me<sub>3</sub> gêne considérablement la segmentation.

Nous établissons par la suite plusieurs descripteurs permettant de caractériser les structures observées et de quantifier l'agrégation globale. Afin de valider ces descripteurs, nous étudions leur stabilité en comparant deux populations de noyaux provenant de cultures indépendantes, cultivées dans les mêmes conditions. Ces descripteurs permettront par la suite de comparer plusieurs populations de noyaux.

**Analyse de l'agrégation globale par noyau.** La façon la plus simple de quantifier l'agrégation globale des RCs d'un noyau est de compter le nombre d'agrégats présents dans chaque noyau. Ce nombre sera proche de 46 pour une agrégation faible, et proche de 1 pour une agrégation forte. Nous définissons deux niveaux d'agrégation :

- Une agrégation des centromères (décrite par la présence de plusieurs centromères dans un foyer CenpA).
- Une agrégation médiée par les RCs.

Afin de caractériser les objets du deuxième niveau d'agrégation, nous avons segmenté les foyers PCY (voir §C.1.2.2, et Fig. C.24H-K). L'association d'un foyer CenpA à un foyer PCY est prise en compte lorsque la distance entre les bords des deux objets segmentés est inférieure à un seuil de contact égal à la résolution optique dans la direction XY, soit 0.1  $\mu\text{m}$ . On observe qu'un faible nombre de foyers PCY n'est en contact avec aucun foyer CenpA, en moyenne  $1.2 \pm 1.6$ <sup>1</sup> par noyau<sup>2</sup>. L'observation des foyers correspondants montre qu'il ne s'agit en général pas d'un problème de segmentation (faux-positifs pour PCY ou faux-négatifs pour CenpA). On observe souvent un foyer CenpA à proximité mais pas en contact directe ce qui pourrait être expliqué par le fait que la sonde PCY ne marque pas l'ensemble de la région centromérique, et que dans certains cas la région marquée par PCY est éloignée du centromère dans l'espace du noyau. De plus on observe aussi un faible nombre de foyers CenpA qui ne sont pas en contact avec un foyer PCY, en moyenne  $4.5 \pm 2.3$  par noyau<sup>3</sup>. Cela peut s'expliquer par le fait que certains chromosomes sont plus faiblement marqués (voir Fig. C.24A-C) et l'intensité des foyers PCY correspondant est trop faible par rapport aux autres foyers pour être détectés par notre procédure d'analyse.

Le second niveau d'agrégation prend en compte à la fois l'agrégation de RCs médiées par PCY et les agrégats de centromères détectés par un marquage CenpA. Afin de prendre en compte les 46 centromères, nous définissons les agrégats par :

- Les foyers PCY qui sont en contact avec au moins un foyer CenpA.
- Les foyers CenpA isolés.

---

1. Cette notation correspond à  $\mu \pm \sigma$  avec  $\mu$  la moyenne et  $\sigma$  l'écart-type.  
 2. Cette valeur est stable entre les deux expériences comparées.  
 3. Cette valeur est stable entre les deux expériences comparées.

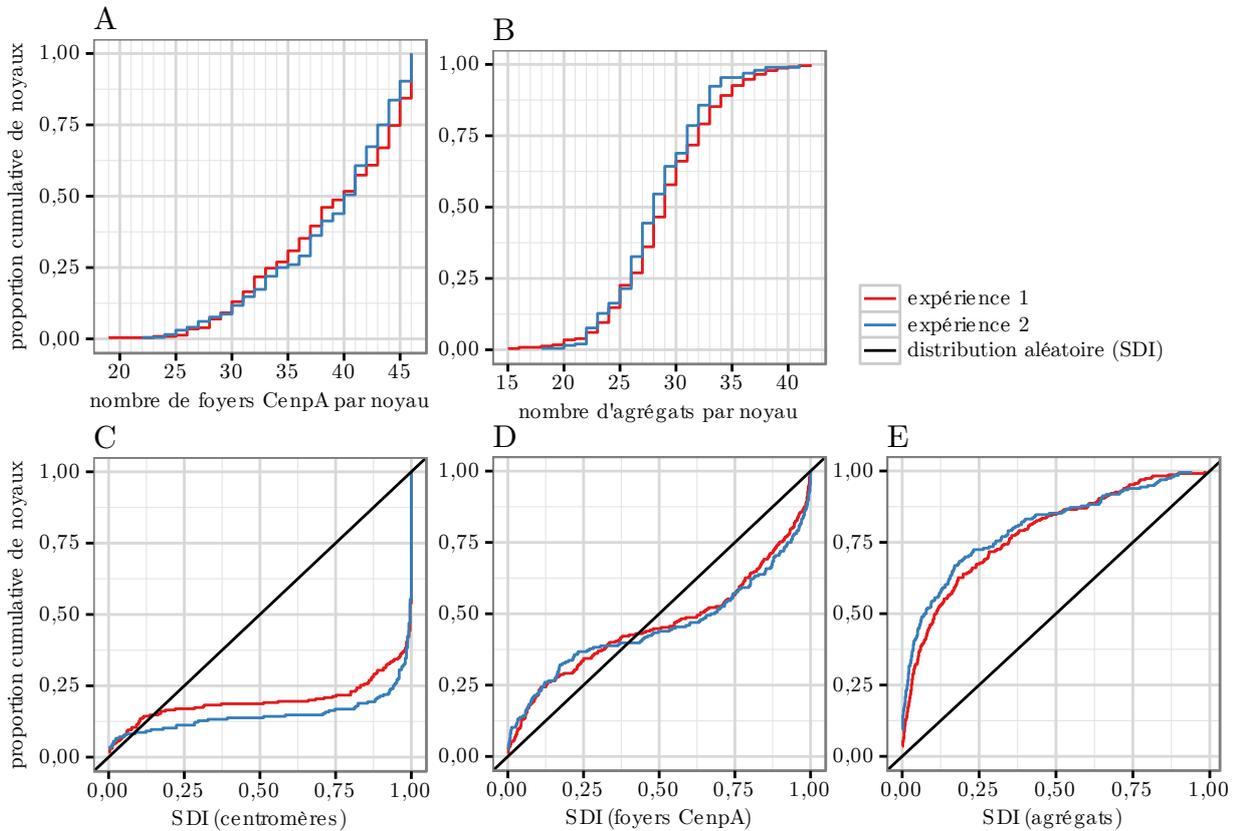


FIGURE C.25: Deux descriptions l'agrégation globale des RCs. A : Distribution du nombre de foyers CenpA par noyau. B : Distribution du nombre d'agrégats par noyau. Un agrégat est soit un foyer PCY en contact avec au moins un foyer CenpA, soit un foyer CenpA qui n'est pas en contact avec un foyer PCY. C-E : Analyse de l'agrégation des centromères (C), des foyers CenpA (D), ou des agrégats (E) faisant aux statistiques spatiales (voir §A.4.3.3). Pour chaque noyau, l'indice de distribution spatiale (SDI) est calculé via la fonction  $F$  qui quantifie l'espace vide dans le noyau ; les courbes représentent la distribution cumulative des SDI. Dans le cas des centromères, le pattern de points analysé est le pattern de points des foyers CenpA en répliquant les foyers CenpA du nombre de centromères qu'ils contiennent. Dans le cas des agrégats, le pattern de points analysé est le pattern des foyers PCY en contact avec au moins un foyer CenpA superposé au pattern de points des foyers CenpA en contact avec aucun foyer PCY. Les patterns sont comparés à un modèle de distribution spatiale aléatoire uniforme au sein du même noyau. Les droites noires représentent la distribution de SDI attendue sous le modèle. Une déviation du pattern observé vers la droite (des valeurs élevées de SDI) signifie que l'espace vide dans le noyau est plus important que sous le modèle, et donc que le pattern est agrégé (signe d'une attraction apparente) ; et inversement, une déviation vers la gauche signifie que le pattern est régulier (signe d'une répulsion apparente) (voir §A.4.3.3). Effectifs : expérience 1 : 196 noyaux, expérience 2 : 230 noyaux.

Ainsi la somme du nombre de centromères présents dans l'ensemble de ces agrégats est constante dans tous les noyaux et égale à 46. On observe que le nombre de foyers CenpA varie entre 20 et 46, avec une moyenne de  $38.7 \pm 5.9$ , et le nombre d'agrégats est en moyenne de  $28.3 \pm 3.9$  lorsqu'on prend en compte l'agrégation médiée par PCY. Les deux expériences présentent des distributions quasiment identiques pour les deux descripteurs<sup>1</sup>. Ces mesures sont très intuitives mais ont

1.  $p\text{-value} > 0.2$  avec le test de Welch à deux échantillons ( $W_2$ ), qui est utilisé ici à la place du test Kolmogorov-Smirnov de comparaison de deux échantillons ( $K\text{-}S_2$ ) car nous sommes dans le cas de distributions discrètes.

l'inconvénient de ne pas être des mesures absolues, car elle ne font pas appel à la comparaison à une référence.

Un autre type de méthodologie faisant appel aux statistiques spatiales permet de caractériser la distribution spatiale d'un pattern de points (voir §A.4.3.3). Nous utilisons ici la fonction  $F$  qui quantifie en fait l'espace vide dans le noyau, et permet de déduire si le pattern de points est régulier, uniforme ou agrégé, par comparaison à un modèle aléatoire uniforme. Nous avons appliqué cette méthodologie à la fois aux foyers CenpA et aux centromères. Dans le cas des centromères, on observe une déviation très forte vers les valeurs élevées de SDI, donc une forte agrégation de la population en comparaison au modèle (fig. C.25C). Dans le cas des foyers CenpA, on observe que la distribution des SDIs est en moyenne proche de celle du modèle, mais la population n'est pas pour autant conforme au modèle<sup>1</sup> et semble plutôt composée d'une sous-population agrégée et d'une sous-population régulière. En effet, la distribution des SDIs observée présente une forte inflexion, donc une déviation du modèle vers les faibles valeurs de SDIs ainsi qu'une déviation vers les valeurs élevées de SDIs (fig. C.25D). La déviation entre le descripteur calculé sur les centromères et celui sur les foyers CenpA est attendue puisqu'on prend en compte une agrégation supplémentaire des centromères via les foyers CenpA dans le deuxième cas. Enfin lorsqu'on prend en compte le deuxième niveau d'agrégation, on observe que l'ensemble de la population présente un pattern régulier, ce qui suggère une répulsion des agrégats (fig. C.25E). Nous obtenons les mêmes résultats avec la fonction  $G$  qui correspond à la distribution des distances au plus proche voisin (voir §A.4.3.3).

En conclusion, cette méthodologie permet de déduire que les RCs peuvent présenter des distributions agrégées ou régulières dans une grande proportion des noyaux; cela dépend beaucoup du niveau d'agrégation considéré. Dans les deux cas, et avec les fonction  $F$  et  $G$ , les différences entre les deux expériences ne sont pas significatives<sup>2</sup>.

**Caractérisation des structures : agrégation locale.** Le nombre de centromères qui forment le foyer est estimé par le nombre de foyers CenpA associés au foyer PCY, éventuellement pondéré par le nombre de centromères formant le foyer CenpA. Afin de caractériser les associations à l'échelle de la population, nous proposons deux types de descripteurs : d'une part le nombre moyen de foyers CenpA ou de centromères par foyer PCY (fig. C.26A-B) et d'autre part la distribution du nombre de foyers CenpA ou de centromères par foyer PCY par noyau (fig. C.26C-D). Le second descripteur permet d'avoir accès à la répartition de la composition (en terme de nombre de foyers CenpA ou de centromères associés) des foyers PCY par noyau à l'échelle de la population, ce qui donne une information plus fine que la moyenne par noyau tout en permettant d'évaluer la variabilité inter-noyaux. Aucune diffé-

1.  $p\text{-value} < 1 \times 10^{-7}$  avec le test de Kolmogorov-Smirnov de comparaison d'un échantillon à une distribution théorique uniforme (K-S<sub>1</sub>).

2.  $p\text{-value} > 0.2$  avec le test de K-S<sub>2</sub>.

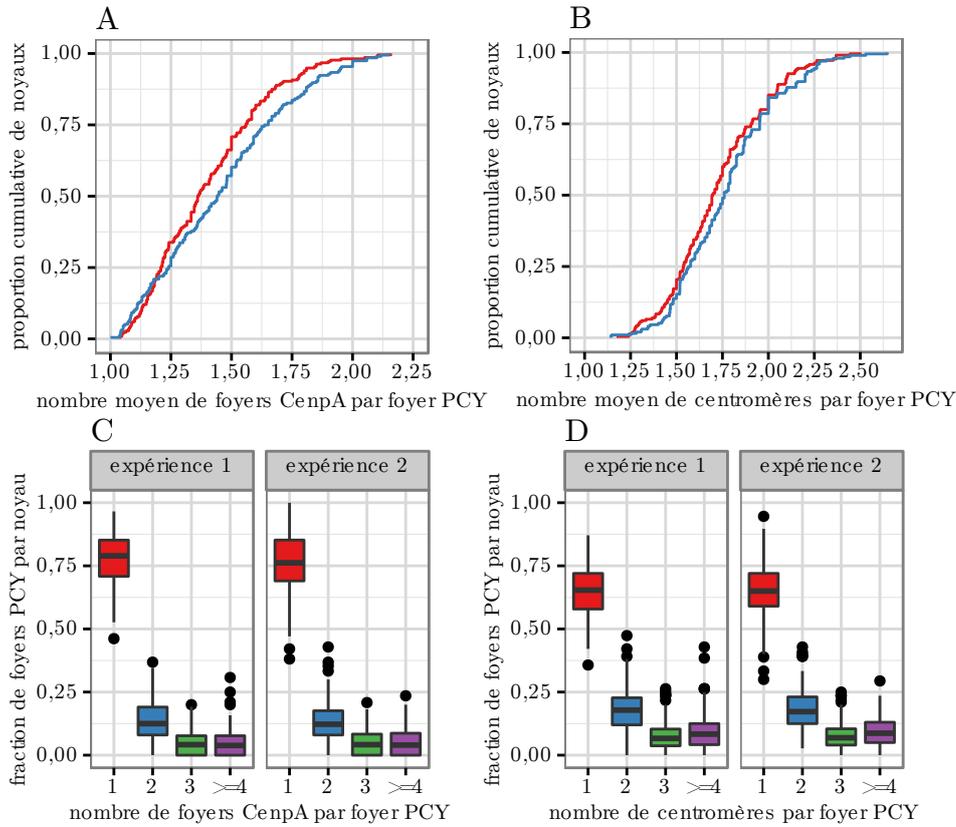


FIGURE C.26: Composition des foyers PCY. A-B : Distribution cumulative du nombre moyen de foyers CenpA (A) ou de centromères (B) par noyau. C-D : Répartition du nombre de foyers CenpA (C) ou de centromères (D) associés par noyau : chaque foyer PCY est classé selon son nombre de foyers CenpA (respectivement de centromères). Pour chaque catégorie de foyers PCY (1, 2, 3 ou  $\geq 4$ ), les diagrammes en boîtes\* représentent la distribution de la fraction de foyers PCY appartenant à la catégorie, par noyau.

\*La ligne centrale d'un diagramme en boîte représente la valeur médiane de la distribution des fractions, et les extrémités de la boîte représentent les 1<sup>ers</sup> et 3<sup>èmes</sup> quartiles de la distribution. Le segment supérieur s'étend jusqu'à la valeur maximale de la distribution se trouvant dans l'intervalle  $1.5 \times$  la distance inter-quartiles, et le segment inférieur jusqu'à la valeur minimale de la distribution se trouvant dans l'intervalle  $1.5 \times$  la distance inter-quartiles. Les segments indiquent à peu près l'intervalle de confiance à 95% d'un test de comparaison de médianes. Les noyaux sortant de cet intervalle représentés par des points.

Effectifs : expérience 1 : 196 noyaux, expérience 2 : 230 noyaux.

rence significative n'est observée entre les deux expériences contrôle<sup>1</sup>. Nous avons vérifié qu'aucun descripteur n'est sensible à une modification du seuil de contact.

### C.3.1.3 Dynamique des associations au cours du cycle cellulaire

Nous avons étudié la dynamique des structures que nous observons au cours du cycle cellulaire. Dans cette étude, le cycle a été quantifié en utilisant le marqueur Ki67, permettant de discriminer les phases G<sub>0</sub> et début de G<sub>1</sub> du reste du cycle cellulaire (voir §C.1.4). Dans nos conditions de culture, les cellules en début de G<sub>1</sub> ne représentent que 2% de la population totale; afin d'obtenir une population suffisante, nous avons sélectionné les champs de microscopie qui contenaient des

1. p-value  $\geq 0.2$  avec le test de K-S<sub>2</sub>.

cellules en début de  $G_1$  lors de l'acquisition. Les proportions de cellules dans les différentes phases du cycle ne représentent donc pas celles de la population, puisque l'acquisition n'est pas aléatoire.

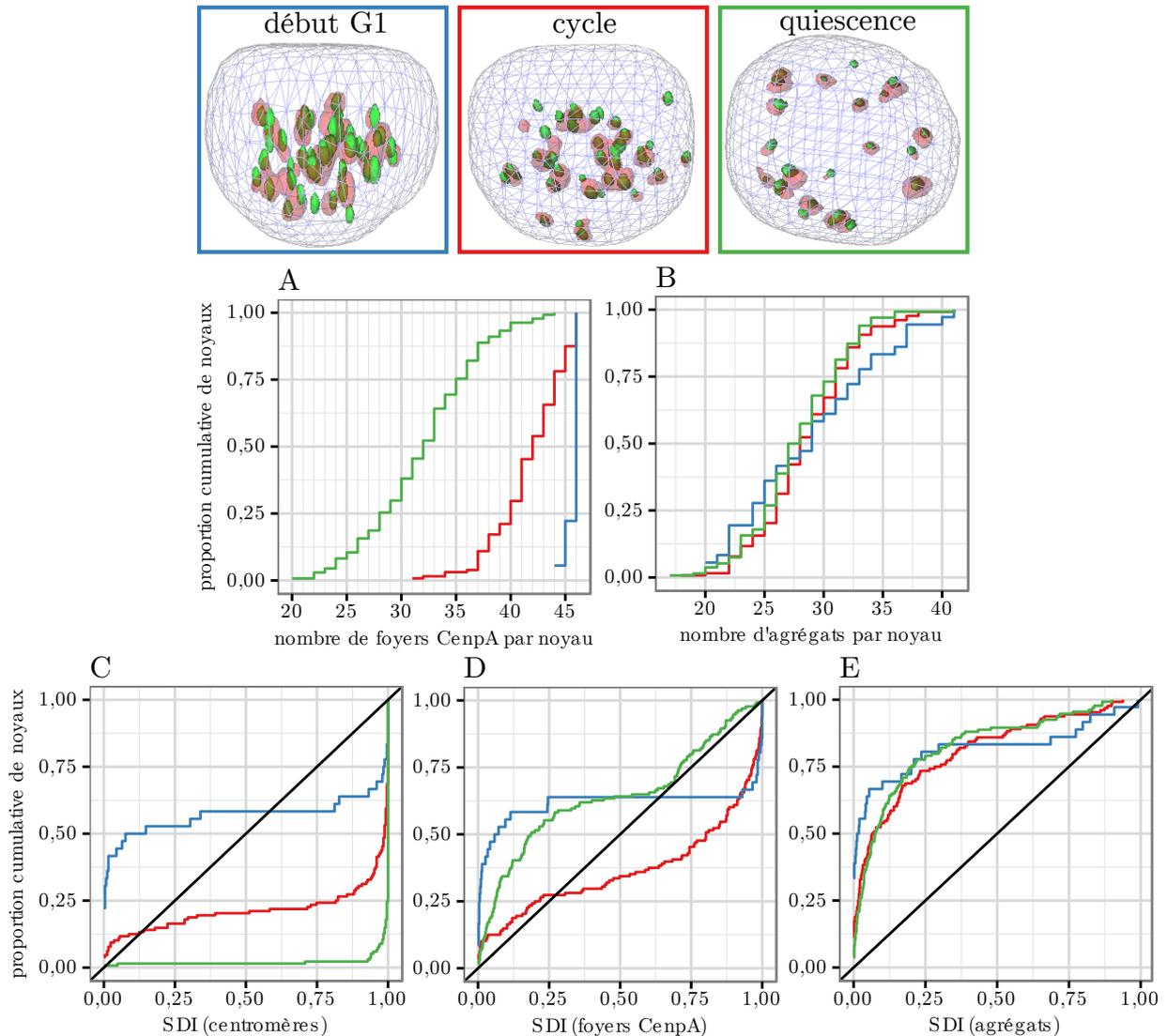


FIGURE C.27: Variation de l'agrégation globale des RCs au cours du cycle cellulaire. Les représentations 3D dans les cadres correspondent à trois noyaux typiques dans chacune des phases du cycle identifiées. Comme dans les figures précédentes, sont représentés les foyers PCY et les centromères. Les couleurs des cadres sont associées aux différentes phases du cycle dans les graphes A-E. A : Variation de la distribution du nombre de foyers CenpA par noyau au cours du cycle cellulaire. B : Variation de la distribution du nombre d'agrégats par noyau au cours du cycle cellulaire. C-E : Analyse des variations de de l'agrégation des centromères (C), des foyers CenpA (D) ou des agrégats (E) faisant appel aux statistiques spatiales, et utilisant la fonction F pour décrire le pattern (voir figure C.25 pour les détails).

Effectifs :  $G_0$  : 134 noyaux, début  $G_1$  : 36 noyaux, cycle : 128 noyaux.

**Dynamique de l'agrégation globale.** La figure C.27 montre les variations de l'agrégation globale entre les différentes phases du cycle. L'agrégation des cen-

centromères est maximale dans les cellules quiescentes, avec une valeur moyenne de  $31.9 \pm 4.0$  foyers, minimale dans les cellules en début de  $G_1$  avec une moyenne de  $45.7 \pm 0.56$ , et intermédiaire dans le reste du cycle avec une moyenne de  $41.7 \pm 4.0$ . Les distributions les plus proches (début de phase  $G_1$  et cycle) sont significativement différentes<sup>1</sup>. En revanche lorsqu'on considère les agrégats, aucune différence significative n'est observée<sup>2</sup> (voir fig. C.27A-B). Cela suggère que l'agrégation des centromères varie au cours du cycle cellulaire, alors que l'agrégation médiée par les RCs est invariante.

Ces données sont confirmées par l'analyse spatiale. Lorsqu'on considère les centromères, on observe une forte agrégation dans les cellules en cycle qui est encore augmentée dans les cellules quiescentes<sup>3</sup>. En revanche les cellules en début de  $G_1$  semblent composées de deux sous-populations, agrégée et régulière. Lorsque l'on considère les foyers CenpA, on observe que leur distribution est régulière dans les cellules quiescentes, agrégée dans les cellules en cycle, et composée de deux sous-populations agrégée et régulière dans les cellules en début de  $G_1$ <sup>4</sup>. Il y donc une répulsion apparente entre les foyers CenpA dans les cellules quiescentes. Cette répulsion est systématique et de plus forte amplitude lorsqu'on considère le deuxième niveau d'agrégation<sup>5</sup>, de manière invariante par rapport au cycle. Ces résultats sont aussi retrouvés avec la fonction G.

**Dynamique de l'agrégation locale.** Le nombre moyen de foyers CenpA par foyer PCY varie avec le cycle cellulaire ; il est notamment minimal dans les cellules quiescentes, ce qui est cohérent avec le fait que le nombre de foyers CenpA global est minimal dans les cellules quiescentes. De plus on observe qu'il n'y a quasiment aucun foyer PCY associé à plus de deux foyers CenpA dans les cellules quiescentes (figure C.28A-B). En revanche, aucune variation du nombre moyen de centromères par foyer PCY n'est observée au cours du cycle cellulaire<sup>6</sup>. On observe une faible augmentation de la proportion de foyers formés par deux centromères au détriment des foyers formés par un seul centromère dans les cellules quiescentes<sup>7</sup> ; et aucune variation des fractions de foyers formés par plus de deux centromères (figure C.28C-D). Nous observons donc que le nombre de centromères par foyer PCY ne varie pas au cours du cycle cellulaire alors que le nombre de foyers CenpA est variable. Ceci est cohérent avec les observations faites à l'échelle globale.

#### C.3.1.4 Localisation sub-nucléaire des agrégats de régions centromériques

- 
1. p-value  $< 2.2 \times 10^{-16}$  avec le test de  $W_2$ .
  2. p-value  $> 0.2$  entre les distributions les plus éloignées avec le test de  $W_2$ .
  3. p-value  $< 2.2 \times 10^{-16}$  avec le test de  $K-S_1$ .
  4. p-value  $< 2 \times 10^{-7}$  avec le test de  $K-S_1$  pour la distribution en début de  $G_1$  qui est la plus proche du modèle.
  5. p-value  $< 2.2 \times 10^{-16}$  avec le test de  $K-S_1$ .
  6. p-value  $> 0.7$  avec le test de  $K-S_2$  pour les deux distributions les plus éloignées.
  7. p-value  $< 1 \times 10^{-4}$  avec le test de  $W-S_2$ .

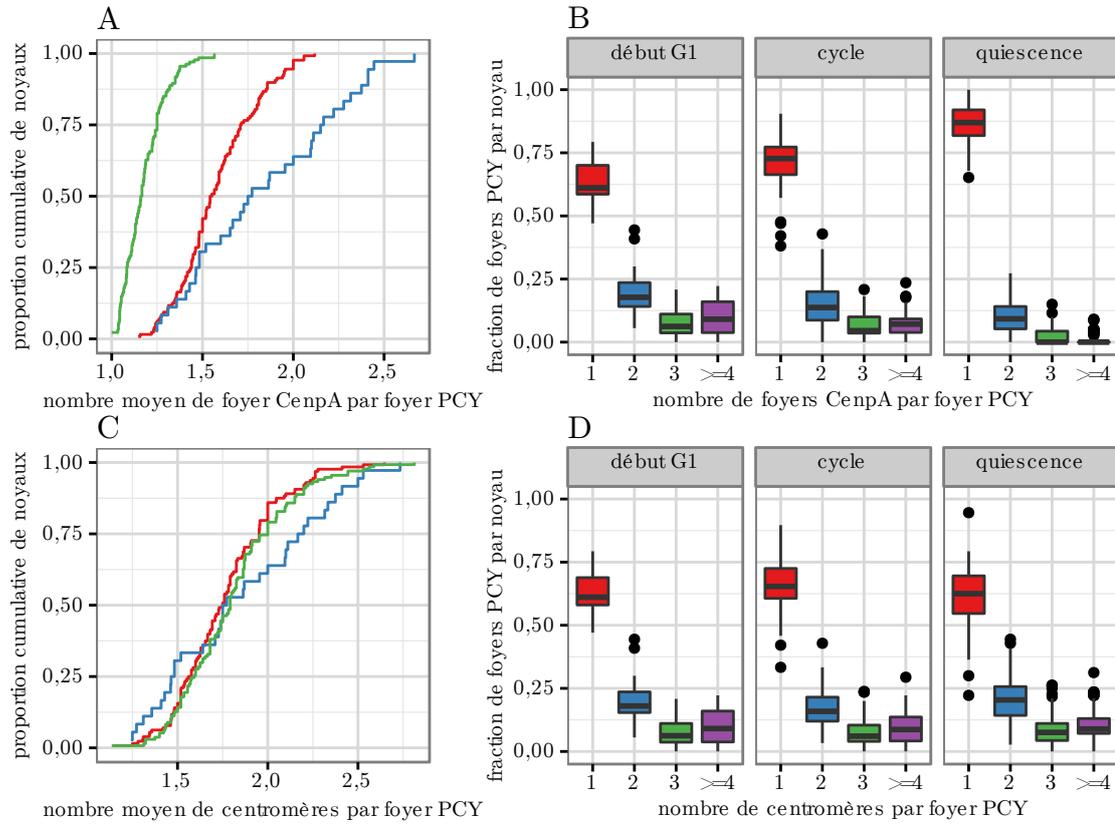


FIGURE C.28: Variation de la composition des foyers PCY au cours du cycle cellulaire. A et C : Variations de la distribution cumulative du nombre moyen de foyers CenpA par foyer PCY (A) ou de centromères (C) par noyau, au cours du cycle cellulaire. B et D : Variations de répartition du nombre de foyers CenpA (B) ou de centromères (D) associés à un foyer PCY par noyau.

Effectifs :  $G_0$  : 134 noyaux, début  $G_1$  : 36 noyaux, cycle : 128 noyaux.

**Différence d'agrégation en fonction de la localisation.** Afin de déterminer si l'agrégation des RCs diffère en fonction de la localisation sub-nucléaire, nous avons défini deux catégories de localisation pour les foyers PCY : d'une part les foyers *périphériques* dont la distance entre le bord du foyer et le bord du noyau est inférieure à  $0.1 \mu\text{m}$ , et d'autre part les foyers qui ne répondent pas à ce critère de localisation et qui sont donc dit *centraux*. La figure C.29A montre que la localisation varie au cours du cycle cellulaire, les foyers étant plus périphériques dans les cellules quiescentes (ce résultat est en accord avec des observations précédentes (Solovei *et al.* (2004))). Nous avons ensuite étudié la composition des foyers PCY en fonction de leur localisation et du cycle cellulaire. Les foyers PCY périphériques contiennent moins de centromères que les foyers centraux, à la fois dans les cellules quiescentes et en cycle<sup>1</sup> ; la population des foyers centraux est quant à elle significativement enrichie en foyers contenant plus de 4 centromères, et appauvrie en foyers ne contenant qu'un seul centromère, à la fois dans les cellules quiescentes et

1.  $p\text{-value} < 2.2 \times 10^{-16}$  avec le test de K-S<sub>2</sub> entre les distributions centrales et périphériques, à la fois dans les cellules quiescentes et en cycle.

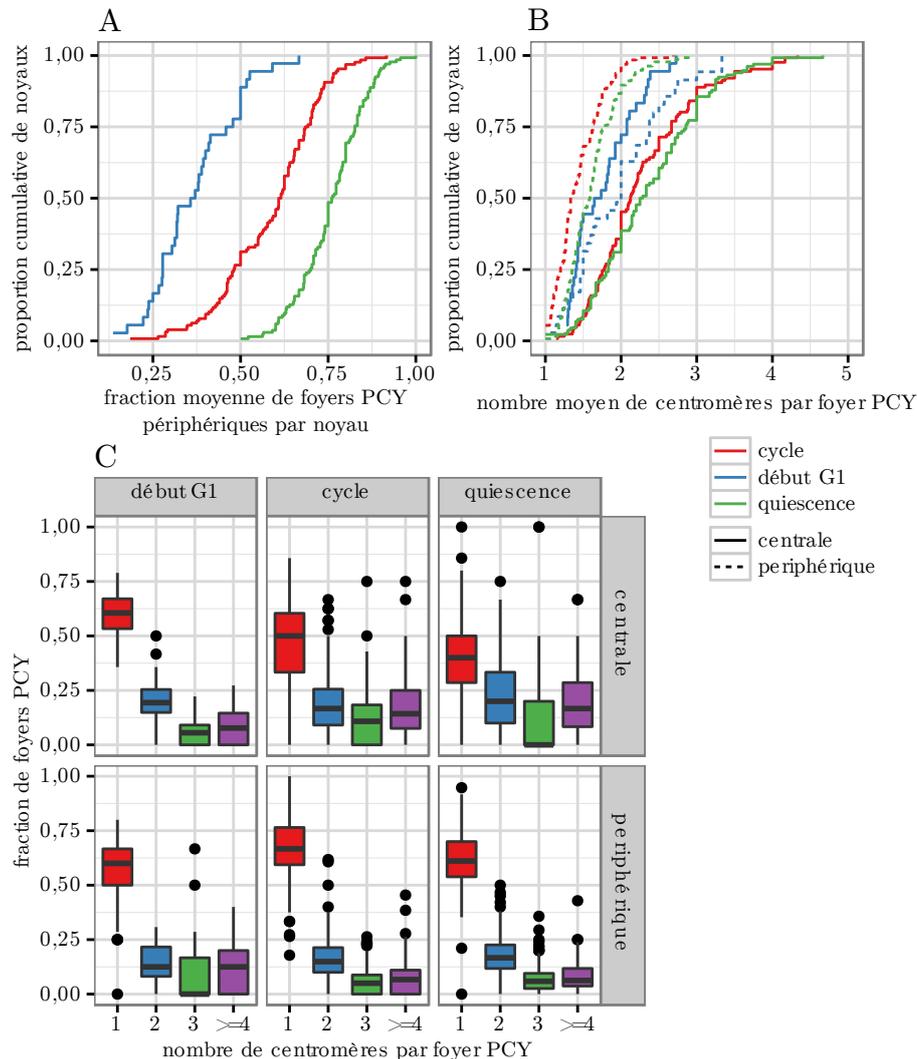


FIGURE C.29: Différences de composition des foyers PCY en fonction de la localisation sub-nucléaire. A : Distribution cumulative de la fraction de foyers PCY localisés en périphérie par noyaux en fonction du cycle cellulaire. B : Distribution cumulative de la moyenne du nombre de centromères associés à un foyer PCY par catégorie de localisation (centrale/périphérique), en fonction du cycle cellulaire. C : Répartition du nombre de centromères par foyer PCY en fonction du cycle cellulaire et de la localisation. Les fractions sont calculées au sein de chaque catégorie de localisation.

Effectifs :  $G_0$  : 134 noyaux, début  $G_1$  : 36 noyaux, cycle : 128 noyaux.

en cycle<sup>1</sup>. Cette observation ne s'applique pas aux cellules en début de phase  $G_1$ , probablement parce que les foyers sont moins localisés en périphérie du noyau et sont très mobiles durant cette phase du cycle (voir discussion).

**Orientation des centromères.** Afin de déterminer l'orientation d'un foyer CenpA par rapport au foyer PCY auquel il est associé, vis-à-vis de la périphérie du noyau, nous utilisons la mesure définie précédemment (voir §C.1.3.2). Lorsque cette mesure est proche de 0.5, il n'y a pas d'orientation préférentielle ; lorsqu'elle est comprise entre 0.5 et 0, le foyer CenpA est orienté vers le centre du noyau ; et

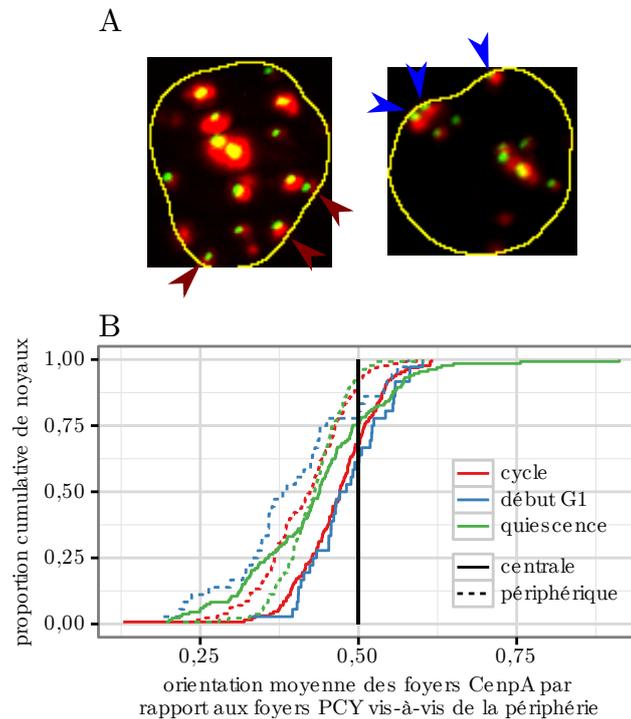


FIGURE C.30: Orientation des foyers CenpA par rapport aux foyers PCY et à la périphérie nucléaire. A : Images de noyaux (plans centraux) (vert : foyers CenpA, rouge : foyers PCY). Le contour du noyau au niveau du plan est indiqué par le trait jaune. Les couleurs des pointes de flèche indiquent l'orientation des foyers CenpA : bleues vers la périphérie, rouges vers le centre. B : Distribution cumulative de l'orientation moyenne des foyers CenpA en fonction de la localisation et du cycle cellulaire. Afin de ne pas biaiser la mesure, les foyers CenpA considérés comme périphériques sont soit directement en contact avec la périphérie, soit en contact avec un foyer PCY qui est en contact avec la périphérie. Le contact est évalué par le même critère que précédemment.

Effectifs :  $G_0$  : 134 noyaux, début  $G_1$  : 36 noyaux, cycle : 128 noyaux.

lorsqu'elle est comprise entre 0.5 et 1, il est orienté vers la périphérie du noyau. Dans les cellules en début de phase  $G_1$  ou en cycle, les foyers centraux ne sont pas orientés préférentiellement<sup>1</sup>. En revanche les foyers périphériques ou les foyers centraux des cellules quiescentes sont orientés vers le centre<sup>2</sup> (voir Fig. C.30). Nous avons vérifié la reproductibilité de cette mesure en comparant à une expérience indépendante (données non montrées).

### C.3.1.5 Comparaison avec un type cellulaire adhérent

Nous avons aussi observé des associations de foyers CenpA médiées par les foyers PCY dans des cellules mesenchymateuses souches humaines (hMES) (voir figure C.31). Dans ce paragraphe nous allons comparer les associations dans ces cellules aux associations dans les lymphoblastoïdes, en utilisant uniquement les descripteurs principaux.

Le nombre de foyers CenpA par foyer PCY varie de la même manière pour les

1.  $p\text{-value} > 0.7$  avec le test de Student à 1 échantillon en comparant à la valeur 0.5.  
 2.  $p\text{-value} < 2 \times 10^{-16}$  avec le test de Student à un échantillon en comparant à la valeur 0.5.

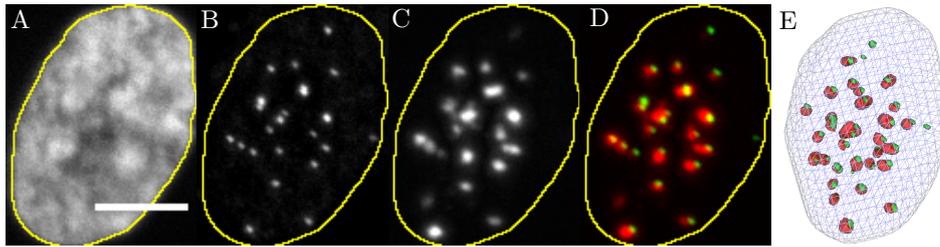


FIGURE C.31: Détection des séquences péri-centromériques humaines par immuno-FISH 3D sur cellules mésenchymateuses souches. Les images A-D sont les projections maximales des 5 plans médians de la pile d'images, le contour du noyau segmenté est tracé en jaune. Barre d'échelle : 5  $\mu\text{m}$ . A : Marquage de l'ADN au DAPI. B : Marquage des centromères par l'anticorps anti-CenpA. C : Marquage des régions péri-centromériques par la sonde PCY. D : Superposition en fausses couleurs des marquages des foyers PCY et des foyers CenpA. E : Représentation surfacique 3D des objets segmentés des marquages PCY et centromériques.

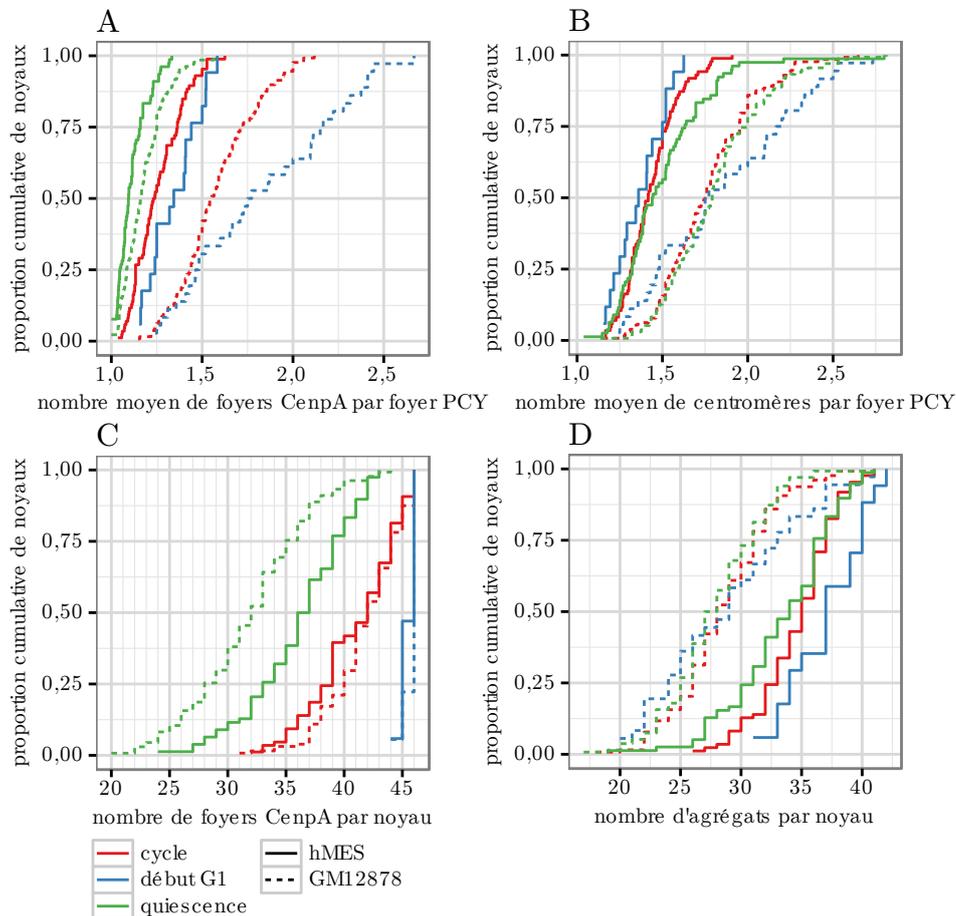


FIGURE C.32: Comparaison de l'agrégation des types cellulaires hMES et GM12878 au cours du cycle cellulaire. A : Distribution du nombre moyen de foyers CenpA par foyer PCY, par noyau. B : Distribution du nombre moyen de centromères par foyer PCY, par noyau. C : Distribution du nombre de foyers CenpA par noyau. D : Distribution du nombre d'agrégats par noyau. Effectifs (GM12878) :  $G_0$  : 134 noyaux, début  $G_1$  : 36 noyaux, cycle : 128 noyaux. Effectifs (hMES) :  $G_0$  : 78 noyaux, début  $G_1$  : 17 noyaux, cycle : 86 noyaux.

deux types cellulaires au cours du cycle, mais il est inférieur dans les cellules hMES<sup>1</sup> (Fig. C.32A). De plus, dans les deux types cellulaires, le nombre moyen d'agrégats ne varie pas au cours du cycle<sup>2</sup> et il est inférieur dans les cellules hMES (fig. C.32B). Comme dans les GM12878, l'agrégation globale des centromères dans les cellules hMES est minimale en début de G<sub>1</sub>, et forte en phase quiescente ; cependant dans une moindre mesure que pour les GM12878 (fig. C.32C). Enfin, comme dans les GM12878, le nombre d'agrégats ne varie pas au cours du cycle cellulaire dans les cellules hMES<sup>3</sup>, et l'agrégation est moins forte dans les cellules hMES que dans les cellules GM12878<sup>4</sup>(fig. C.32D).

En conclusion, la dynamique de l'agrégation des RCs au cours du cycle cellulaire est qualitativement la même dans les deux types cellulaires, mais les RCs sont moins agrégées dans les cellules hMES.

### C.3.2 Discussion

#### C.3.2.1 Deux niveaux d'agrégation des régions centromériques

En développant un marqueur permettant de détecter les régions péri-centromériques humaines sur la quasi-totalité des chromosomes nous avons pu mettre en évidence des associations de RCs, médiées par les régions péri-centromériques, qui ressemblent aux chromocentres murins. Nous définissons ainsi deux niveaux d'agrégation des RCs, d'une part l'agrégation des centromères, et d'autre part l'agrégation médiée par les régions péri-centromériques.

**Associations de centromères.** Les associations de centromères peuvent être observées par un simple marquage des centromères, par exemple avec un anticorps dirigé contre la protéine CenpA. Ces associations avaient déjà été décrites auparavant par l'observation d'un nombre de foyers inférieur à 46 dans des cellules diploïdes. L'observation d'un nombre réduit de foyers aurait pu être dû à un problème de détection, mais nous avons démontré que ça n'est pas le cas : en effet, lorsque l'on observe moins de foyers dans une cellule, la variation des intensités entre les foyers de la même cellule est plus importante, ce qui suggère que les foyers les plus intenses résultent de l'agrégation de plusieurs centromères. Nous avons utilisé cette propriété pour attribuer un nombre de centromères à chaque foyer CenpA, sachant que la cellule comporte 46 centromères en tout. Notre système d'observation ne permet pas de déterminer l'ultra-structure des foyers CenpA qui contiennent plusieurs centromères ; il est possible que les centromères se localisent à une distance inférieure à la résolution optique et ne soient donc pas distinguables ; mais l'on peut aussi imaginer que ces foyers résultent d'une structure chromatiniennne particulière, par exemple d'une interpénétration des fibres de chromatine de plusieurs

- 
1. p-value  $< 1 \times 10^{-4}$  avec le test de K-S<sub>2</sub> pour les distributions les plus proches.
  2. p-value  $> 0.2$  avec le test de K-S<sub>2</sub> pour les distributions les plus éloignées.
  3. p-value  $> 0.11$  avec le test de W<sub>2</sub> pour les distributions les plus éloignées.
  4. p-value  $< 2.2 \times 10^{-16}$  avec le test de W<sub>2</sub> pour les distributions les plus proches.

centromères. L'utilisation de la microscopie à super-résolution pourrait apporter des éléments de réponse à cette question.

**Associations médiées par les régions péricentromériques.** Les associations de centromères médiées par les régions péricentromériques n'avaient jamais été directement observées auparavant. Des associations de foyers centromériques médiées par des foyers d'hétérochromatine H<sub>4</sub>K<sub>20</sub>me<sub>3</sub> avaient déjà été décrites (Zinner *et al.* (2006)) sans que les séquences d'ADN sous-jacentes ne soient identifiées. Il est très probable qu'il s'agisse des mêmes structures. Cependant notre observable n'est pas parfaite car la sonde PCY ne marque pas l'ensemble des régions péricentromériques. Il faut donc garder à l'esprit que nous sous-estimons l'agrégation dans cette étude ; mais c'est à notre connaissance la manière la plus efficace de marquer les régions péricentromériques humaines. Ce marqueur pourrait éventuellement être amélioré en le combinant à d'autres sondes complémentaires dirigées contre les séquences centromériques humaines.

Ces structures ne sont pas une spécificité des lymphoblastoïdes puisqu'elles sont aussi présentes dans des cellules mésenchymateuses souches humaines, mais l'agrégation est moins forte dans ce dernier type cellulaire. Il nous semble important de mettre cette différence d'agrégation en relation avec le fait que les deux types cellulaires expriment la protéine LBR, mais que seules les cellules hMES expriment la lamine A (ces données proviennent d'expériences d'immunocytochimie que nous avons réalisées sur ces types cellulaires). Cette observation est cohérente avec le fait que l'expression des protéines de membrane, et notamment la lamine A, influence grandement la morphologie du PHC murin : il y a une corrélation entre l'expression de la lamine A et une faible agrégation du PHC (voir §A.3.3.1).

**Étude des mécanismes d'association des régions péricentromérique.** Cette observable ouvre la porte à l'étude des mécanismes moléculaires d'agrégation de l'hétérochromatine péricentromériques chez l'homme. Jusqu'ici, il était difficile de marquer l'ensemble des régions péricentromériques humaines ; ce qui rendait ces études délicates. Par exemple Agarwal *et al.* (2011) étudient l'effet de la surexpression de la protéine MeCP2 sur l'agrégation des RCs humaines, en observant uniquement les satellites III qui sont présents en abondance sur les RCs des chromosomes 1, 9 et 16, et en comptant le nombre de foyers distincts dans le noyau. L'observable n'est clairement pas adaptée puisque la probabilité d'observer un effet est diminuée par le fait qu'ils ne détectent qu'une partie des RCs de seulement 6 chromosomes sur 46.

Nous avons vu au §A.3.3.2 plusieurs protéines candidates qui pourraient être impliqués dans l'agrégation des RCs ; elles peuvent être des protéines de membrane, des protéines liant l'ADN ou des modifications épigénétiques (et les enzymes les régulant). Ces candidats proviennent d'expériences réalisées en général chez la souris, mais n'ont en général pas été testées chez l'homme. Il serait notamment intéressant de compléter l'étude de Agarwal *et al.* (2011) sur l'effet de l'inactivation la protéine MeCP2. Des mutations de cette protéine sont associées au syndrome neurologique

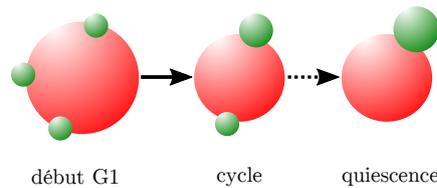


FIGURE C.33: Modèle décrivant les variations de l'agrégation au cours du cycle cellulaire. Les boules rouges représentent les foyers PCY et les boules vertes les foyers CenpA. En début de phase  $G_1$ , les centromères sont distincts, chaque boule verte représente un centromère. Au cours du cycle cellulaire (et dans une plus forte mesure dans les cellules quiescentes), certains centromères entrent en contact proche (indiqué par une boule verte plus grosse). Globalement, le nombre de centromères associés à un foyer PCY ne varie pas.

de Rett (*Amir et al. (1999)*), et ces études pourraient permettre de mieux comprendre les effets de ces mutations chez l'homme. Par ailleurs, il serait intéressant de voir s'il est possible de reproduire le phénotype inversé observé par *Solovei et al. (2013)* en l'absence des protéines LBR et lamine A. Pour cela il devrait être possible d'invalider LBR dans les cellules GM12878 (qui n'expriment pas la lamine A) par une approche d'inhibition utilisant des petits ARN en épingle à cheveux (shRNA). L'observation d'un phénotype inversé ne sera probablement possible qu'après plusieurs jours de quiescence, comme c'est le cas chez la souris ; pour cela il devrait être possible d'induire l'état de quiescence par déprivation de sérum au moins une semaine avant de réaliser l'observation du phénotype.

### C.3.2.2 Dynamique de l'agrégation au cours du cycle cellulaire

Nous avons montré que l'agrégation des centromères varie au cours du cycle cellulaire ; elle est notamment maximale dans les cellules quiescentes et quasi nulle dans les cellules en début de phase  $G_1$ . Ces résultats sont globalement en accord avec les mesures réalisées par *Solovei et al. (2004)*. Les auteurs de cette étude observent une agrégation légèrement plus forte dans les cellules quiescentes ; il est possible que la différence de mesure soit due au fait que les centromères sont détectés par un sérum (CREST) qui marque plusieurs protéines, et donc que les structures détectées soient de plus grande taille que les structures détectées par un anticorps dirigée uniquement contre la protéine CenpA. Il est aussi possible que cette différence soit due au faible effectif de leur étude. Nous observons pour les cellules en cycle une valeur intermédiaire entre les valeurs pour les cellules en phase  $G_1$  et les cellules en phase S de *Solovei et al. (2004)*, ce qui semble cohérent puisque nous ne distinguons pas la phase S, la population que nous étudions contient donc à la fois des cellules en phase S et des cellules en phase  $G_1$ .

Un des résultats les plus nouveaux de cette étude est que l'agrégation globale ne varie pas au cours du cycle cellulaire lorsqu'on prend en compte le deuxième niveau d'agrégation des RCs. Pour expliquer cette apparente contradiction avec le fait que l'agrégation des centromères varie au cours du cycle cellulaire, nous avons étudié la composition des foyers PCY. Nous avons montré que le nombre de centromères associés à un foyer PCY ne varie pas au cours du cycle cellulaire en moyenne ; en revanche le nombre de foyers CenpA associés est variable. Il est en effet maximal en

début de phase  $G_1$ , dans laquelle tous les centromères sont distincts, intermédiaire dans le reste du cycle cellulaire, et minimal dans les cellules quiescentes. Nous avons observé que la dynamique de l'agrégation des RCs est identique dans un autre type cellulaire, ce qui suggère que cette propriété est générale. Il semble donc qu'il y ait une réorganisation structurale des associations de RCs au cours du cycle cellulaire, les centromères pouvant entrer en contact proche (voir figure C.33).

### C.3.2.3 Agrégation et répulsion des régions centromériques.

L'analyse du pattern de localisation globale par une approche de statistiques spatiales permet de mettre en évidence que les centromères présentent un pattern très agrégé dans toutes les phases du cycle cellulaire par rapport à un modèle aléatoire, à l'exception des cellules en début de  $G_1$  qui semblent constituées de deux sous-populations présentant des pattern soit très réguliers soit très agrégés. Il est possible que la présence de deux sous-populations soit due au fait que la chromatine est très mobile en début de  $G_1$  (Walter *et al.* (2003)); les RCs sont donc très rapidement réorganisées et notre méthode de détermination de cette phase ne permet pas de distinguer une progression au cours de cette phase.

L'étude des patterns des foyers CenpA montre une répulsion apparente des foyers CenpA en  $G_0$  et une agrégation dans les cellules en cycle. Il est possible que cette différence soit due au fait que nous n'observons pas le même nombre de foyers dans ces deux phases du cycle. Cette hypothèse est confirmée par le fait qu'on observe une répulsion invariante au cours du cycle lorsqu'on considère le deuxième niveau d'agrégation (dont le nombre d'objets ne varie pas au cours du cycle). On peut donc conclure qu'il y a une forte agrégation des centromères, et une forte répulsion des agrégats de centromères. Il est possible que cette répulsion apparente soit la conséquence d'un encombrement stérique des territoires chromosomiques. Ces résultats peuvent être comparés à des expériences conduites chez le lapin, utilisant la même méthodologie et un marquage des centromères équivalent par le sérum CREST (Andrey *et al.* (2010)). Les auteurs de cette étude détectent une forte répulsion apparente des foyers de centromères. Cette différence avec nos résultats peut provenir de la différence de modèle mais aussi du fait que le cycle cellulaire n'est pas pris en compte dans cette étude, alors qu'il influence beaucoup les résultats des descripteurs dans notre étude.

### C.3.2.4 Lien avec le positionnement

Nous observons que l'agrégation dépend du positionnement des agrégats : les agrégats regroupant le plus grand nombre de centromères se localisent préférentiellement au centre du noyau. De plus nous observons que les centromères sont orientés préférentiellement vers le centre par rapport aux régions péricentromériques, lorsqu'ils se localisent en périphérie. En revanche nous observons que les centromères sont orientés aléatoirement lorsqu'ils se localisent au centre, à l'exception des cellules quiescentes. L'interprétation de ces observations sera discutée au §D.2.1.

### Conclusion

Dans cette partie, nous avons mis en évidence des structures correspondant à l'association de régions péri-centromériques portées par plusieurs chromosomes, qui ressemblent aux chromocentres murins. Cette observable ouvre la porte à des études mécanistiques sur l'agrégation de l'hétérochromatine péri-centromérique humaine, qui étaient surtout conduites chez la souris auparavant. Nous avons montré que les RCs peuvent être associées par le biais de leur centromères directement, ou par leurs péri-centromères. Si l'agrégation des RCs est globalement invariante au cours du cycle cellulaire, nous observons des réorganisations structurales avec une augmentation des associations directes de centromères dans les cellules quiescentes. Nous avons aussi mis en évidence des différences d'agrégation en fonction de la localisation sub-nucléaire, qui peuvent mettre sur la piste de mécanismes et qui seront discutées au §D.2.1.



# Discussion et Conclusion

---

## Sommaire

---

<b>D.1 Méthodologie d'étude de l'organisation nucléaire</b> . . . . .	<b>169</b>
D.1.1 Sources de variabilité inter-cellulaire . . . . .	169
D.1.2 Portée des analyses . . . . .	170
<b>D.2 Organisation nucléaire des régions centromériques humaines</b>	<b>171</b>
D.2.1 Liens entre positionnement et agrégation . . . . .	171
D.2.2 Dynamique de l'organisation spatiale des régions centromériques	172
<b>D.3 Conclusion</b> . . . . .	<b>176</b>

---

## D.1 Méthodologie d'étude de l'organisation nucléaire

Une partie de l'apport de ce travail réside dans la méthodologie mise en oeuvre, qui repose sur le développement :

- d'un outil d'analyse d'image générique et haut débit
- de procédures de segmentation
- de mesures originales
- d'une procédure d'analyse quantitative du cycle cellulaire.

Nous avons appliqué ces outils à l'étude de l'architecture nucléaire dans des populations de noyaux fixés. Ainsi, pour décrire les phénomènes auxquels nous nous sommes intéressés (le positionnement nucléaire d'une RC spécifique, ou l'agrégation de l'ensemble des RCs), nous avons cherché à mettre en évidence des régularités à l'échelle de la population de noyaux afin de prendre en compte la variabilité inter-cellulaire de ces phénomènes.

### D.1.1 Sources de variabilité inter-cellulaire

Il est essentiel d'identifier les sources de variabilités afin de les contrôler au maximum pour obtenir des mesures *reproductibles* entre des populations indépendantes de noyaux. La variabilité inter-cellulaire peut être due :

- Au caractère aléatoire du phénomène ; il est donc important d'augmenter la taille de l'échantillon pour bien décrire l'ensemble des configurations possibles.
- A la non-homogénéité de la population étudiée ; il est important de la réduire au maximum d'une part, et d'autre part de s'assurer que la composition des populations soit reproductible afin d'obtenir des mesures reproductibles. Par

exemple, nous avons mis en évidence que le cycle cellulaire influence les phénomènes étudiés ; une population asynchrone n'est donc pas homogène vis-à-vis de ces phénomènes. Le fait de prendre en compte le cycle cellulaire permet de réduire ce manque d'homogénéité. Dans le cas où le cycle cellulaire ne peut être pris en compte il, est important de s'assurer que la composition dans les différentes phases du cycle est bien reproductible entre deux populations.

- A la mesure en soi, par exemple à cause du manque de repère dans le noyau ou à la variabilité de forme et de taille des noyaux<sup>1</sup>. Pour réduire cette variabilité, il est important d'utiliser des mesures normalisées par rapport à ces variations au niveau de la cellule individuelle (telles que l'EVF) et non à partir de paramètres calculés sur la population globale<sup>2</sup>. On trouve parfois dans la littérature des mesures qui introduisent de la variabilité parce qu'elles ne sont pas adaptées<sup>3</sup>.
- A l'influence des protocoles expérimentaux de fixation et de marquage : il est important de déterminer les étapes susceptibles d'influencer le phénomène pour essayer de réduire leur impact, et à défaut, de réduire leur variabilité entre différentes expériences. Par exemple nous avons montré que le protocole de fixation des cellules lymphoblastoïdes pouvait modifier la taille des noyaux, ce qui est susceptible de faire varier les mesures de positionnement. Nous avons choisi un protocole de fixation qui permet d'obtenir de façon reproductible des noyaux dont la taille se rapproche de celle des cellules vivantes.

### D.1.2 Portée des analyses

Afin que les mesures soient informatives en soi et exploitables par la communauté scientifique, il est parfois utile de développer des mesures *absolues* (voir la conclusion du §A.4.3) qui résultent en général de la comparaison à une norme. Par exemple, en utilisant une mesure telle que l'EVF qui correspond à la comparaison à une distribution uniforme, nous avons pu décrire la répulsion ou l'attraction spécifique de RCs pour la périphérie du noyau, les nucléoles ou le reste du nucléoplasme. La table A.2 montre bien qu'il est très difficile d'obtenir ce genre d'informations à partir des données de la littérature : la majorité des mesures ne sont pas absolues, et uniquement comparables aux données de la même étude, ce qui limite leur portée.

---

1. Les variabilités de forme et de taille ne concernent pas toutes les mesures, il s'agit en général des mesures spatiales telles que le positionnement des RCs dans l'espace du noyau.

2. C'est par exemple ce que proposent [Carvalho et al. \(2001\)](#) : la référence à laquelle ils comparent leur mesure est calculée à partir d'une moyenne sur la population.

3. Par exemple [Mayer et al. \(2005\)](#) tentent de mesurer l'association ou la répulsion de TCs via l'angle entre les deux TCs et le centre du noyau. Cette mesure produit des distributions d'angles aléatoire uniformes et les auteurs concluent que l'organisation relative des TCs est aléatoire. Cette conclusion est en contradiction avec une analyse qui fait appel à une mesure beaucoup plus directe qui repose sur des distances entre les centres des TCs ([Heride et al. \(2010\)](#)) ; cette contradiction est certainement due au fait que la mesure n'est pas adaptée à la question.

## **D.2 Organisation nucléaire des régions centromériques humaines**

Les développements méthodologiques nous ont permis d'étudier deux aspects de l'organisation spatiale des RCs humaines : leur positionnement dans l'espace du noyau et leur agrégation.

Ces deux aspects sont inter-dépendants et influencés par le cycle cellulaire.

### **D.2.1 Liens entre positionnement et agrégation**

Dans la partie §C.2.2 nous avons mis en évidence que les RCs ont tendance à se positionner soit au niveau des nucléoles soit au niveau de la périphérie du noyau, et très peu dans le reste du nucléoplasme. Ces résultats sont à mettre en relation avec des observations de la littérature discutées au §A.1.3.4, qui montrent que des domaines associés à la périphérie du noyau peuvent être relocalisés au niveau de la périphérie du noyau ou des nucléoles d'une cellule mère à une cellule fille, de manière aléatoire. Ainsi, la périphérie du noyau et celle des nucléoles pourraient être considérées comme deux plate-formes d'ancrage de l'hétérochromatine péricentromérique, mais aussi d'autres régions hétérochromatiniennes telles que les NADs et les LADs.

Nous avons montré que la composition des agrégats varie entre ceux qui sont localisés en périphérie du noyau et ceux qui sont localisés au centre ; ces derniers étant composés d'un plus grand nombre de centromères.

#### **D.2.1.1 Agrégation à l'intérieur du noyau.**

L'observation des images indique que les agrégats composés de nombreuses RCs se trouvant au centre sont quasiment toujours localisés à proximité des nucléoles. Il aurait été intéressant de marquer les nucléoles dans cette étude afin de quantifier cette observation<sup>1</sup>. Puisque l'hétérochromatine a tendance à s'agréger avec elle-même (voir : §A.3.3.2, Gibcus et Dekker (2013)), cette observation suggère que les nucléoles pourraient jouer un rôle de plate-forme d'agrégation de ces structures. Afin de tester cette hypothèse, il devrait être possible de déstructurer les nucléoles et de mesurer l'agrégation des RCs au centre du noyau. Cependant la déstructuration des nucléoles pourrait aussi avoir des effets indirects ; il faudrait donc utiliser des moyens indépendants pour les déstructurer, par exemple en inhibant la transcription (Dundr et Misteli (2001)) ; ou en modifiant l'encombrement moléculaire du noyau (Hancock (2004)).

#### **D.2.1.2 Agrégation à la périphérie du noyau.**

Inversement, nous avons constaté qu'en périphérie les RCs sont moins agrégées, ce qui suggère que l'ancrage des RCs à la périphérie a une influence opposée.

---

1. Cela n'a pas été possible pour des raisons d'incompatibilités d'anticorps.

Il est possible que cela soit dû au fait que la surface de la périphérie du noyau est bien plus importante que celle des nucléoles<sup>1</sup>, ainsi les RCs ont moins de chances de se trouver en contact lorsqu'elles se localisent en périphérie du noyau.

De manière alternative, il est possible que ce soit les mécanismes d'ancrage à la périphérie eux-mêmes qui entrent en compétition avec l'agrégation. En effet, chez la souris, ils font intervenir des interactions entre des protéines de l'enveloppe et la marque H<sub>3</sub>K<sub>9</sub>me<sub>3</sub> présente au niveau des régions péricentromériques (voir §A.3.3.1). Ces mécanismes n'ont pas été identifiés chez l'homme mais il est probable qu'ils soient conservés. Par ailleurs nous avons vu que les RCs humaines peuvent interagir via leur péricentromères, ce qui pourrait donc entrer en compétition avec les mécanismes d'ancrage. Cette hypothèse de mécanisme d'ancrage est appuyée par le fait que nous observons une orientation préférentielle des foyers PCY vers la périphérie du noyau par rapport aux foyers CenpA, lorsque ces foyers sont situés en périphérie du noyau<sup>2</sup> (voir §C.3.1.4). Cette observation peut donner des indications sur les mécanismes d'ancrage des RCs à la périphérie : il est possible que cet ancrage fasse d'avantage intervenir les régions péricentromériques plutôt que le centromère directement. Il serait intéressant d'invalider certains acteurs hypothétiques de l'ancrage à la périphérie (LBR, Lmna, ou des modificateurs chromatiniens impliqués dans la méthylation de H<sub>3</sub>K<sub>9</sub> tels que Prdm3 et Prdm16, Suv39h, etc.) puis d'observer d'une part si cette orientation préférentielle est perdue, et d'autre part si l'agrégation est augmentée à la périphérie du noyau.

### D.2.2 Dynamique de l'organisation spatiale des régions centromériques

Les différences les plus flagrantes à la fois pour le positionnement et l'agrégation sont observées entre les cellules en cycle (indépendamment de la phase du cycle) et les cellules quiescentes. Dans les cellules quiescentes, nous observons pour toutes les RCs une localisation plus périphérique au détriment de l'association aux nucléoles, et une agrégation globale identique mais principalement expliquée par une agrégation directe des centromères (voir §C.3.2.1).

Ces variations nous permettent d'inférer un comportement dynamique des RCs<sup>3</sup>, mais il est important de prendre en compte que le fait que les états que nous comparons ne sont pas forcément homogènes (voir §D.2.2.1). Deux types de réorganisations peuvent être envisagés pour expliquer les variations entre cellules quiescentes et en cycle : des réorganisations non dirigées (voir §D.2.2.2) et des ré-

1. Nous mesurons qu'en moyenne les nucléoles occupent 10% du volume du noyau ; donc si l'on modélise le noyau par une sphère contenant un nucléole sphérique, le rapport des surfaces est d'environ 4,6.

2. De manière équivalente il s'agit d'une orientation préférentielle des foyers CenpA vers le centre du noyau par rapport aux foyers PCY.

3. Le comportement dynamique serait plus directement obtenu par des expériences sur cellules vivantes, mais de telles expériences sont pour le moment très délicates car il est alors difficile d'observer des RCs spécifiques ou des TCs spécifiques.

organisations dirigées (voir §D.2.2.3).

### D.2.2.1 Observation de l'état de quiescence.

Nous caractérisons l'état de quiescence par l'analyse du signal Ki67 (voir §C.1.5), et définissons ainsi une population que nous nommons *quiescente* mais qui n'est peut-être pas homogène. Dans la méthode même de caractérisation, cette population est dispersée selon la mesure d'agrégation du signal ; ce qui suggère que la protéine est présente à des niveaux variables au sein de la population. Nous interprétons cette dispersion par le fait que la protéine n'est plus exprimée, et progressivement éliminée car la population *quiescente* que nous observons est peut être composée de cellules qui sont sorties du cycle depuis des durées variables. Plusieurs observations provenant de la littérature suggèrent que l'état de quiescence est lent à atteindre :

- Le phénotype d'architecture inversé (voir §A.3.3.1) est atteint au bout de 2 semaines après la sortie du cycle cellulaire
- Dans le paragraphe portant sur le positionnement des régions centromériques (§A.3.1.1) nous avons pointé une contradiction entre deux études : [Carvalho et al. \(2001\)](#) étudient des cellules synchronisées en phase  $G_0/G_1$ <sup>1</sup> et observent des variations d'association à la périphérie entre les RCs portées par différents chromosomes, alors que [Weierich et al. \(2003\)](#) observent que toutes les RCs se localisent à la périphérie du noyau, dans des lymphocytes non activés qui sont probablement en arrêt de croissance depuis une plus longue période.

L'état de *quiescence* que nous observons est peut être en fait plus un état de *mise en quiescence* durant lequel de nombreuses réorganisations ont lieu. Il est aussi possible que la population *quiescente* que nous observons contienne des cellules en cours d'activation.

Les variations que nous observons entre cellules en cycle et cellules quiescentes pourraient être renforcées en réalisant une cinétique de quiescence<sup>2</sup>. Il serait aussi intéressant d'étudier l'activation de cellules quiescentes, afin de comparer les processus de mise en quiescence et d'activation.

### D.2.2.2 Réorganisations issues de processus non dirigés.

Il est possible que certaines réorganisations soit *thermodynamiquement* stables, mais *cinétiquement* défavorisées par la dynamique imposée par le cycle cellulaire. En effet, au cours du cycle cellulaire, les chromosomes sont condensés sous forme mitotique, puis décondensés en début de phase  $G_1$ , laquelle l'organisation du noyau s'établit et les chromosomes sont très mobiles ([Walter et al. \(2003\)](#)). Il y a donc un remaniement périodique de l'organisation des chromosomes, sur une période de la durée du cycle cellulaire (de l'ordre de 24h). Ces remaniements n'ont pas lieu dans les cellules quiescentes. De plus, en l'absence de signaux extérieurs, les chromosomes sont relativement peu mobiles dans le noyau (voir §A.1.4.1), ce qui permet de supposer que certains états cinétiquement défavorables mais thermodynamiquement

1. La synchronisation est réalisée en laissant les cellules 5 jours sans changer le milieu de culture.

2. L'état de quiescence peut être induit par déprivation de sérum par exemple.

stables peuvent être observés plus fréquemment dans les cellules quiescentes. Un exemple de remaniement qui semble suivre ce genre de mécanismes est l'architecture inversée présentée au §A.3.3.1, qui n'est observée que deux semaines après la sortie du cycle cellulaire. Plusieurs observations de notre étude pourraient relever de ce type de réorganisations.

Dans la figure C.33 nous proposons un modèle de réorganisation des associations de RCs au cours du cycle cellulaire, dans lequel les associations des RCs sont uniquement péricentromériques en début de phase  $G_1$ , essentiellement centromériques en  $G_0$ , et intermédiaires durant le reste du cycle cellulaire. Il est tentant de postuler qu'il s'agit d'un mécanisme séquentiel, les associations de régions péricentromériques précédant les associations de centromères ; par exemple parce que ces dernières résultent de réorganisations plus lentes de la chromatine. De telles réorganisations auraient plus souvent lieu dans des cellules quiescentes, dans lesquelles les centromères sont moins mobiles. Ce modèle pourrait être testé en réalisant une cinétique de quiescence et en quantifiant les différents types d'agrégations. De plus, s'il s'agit d'un mécanisme séquentiel, les associations directes de centromères devraient être perdues lorsque les associations de péricentromères sont inhibées. Cela pourrait être testé en inhibant la méthylation de  $H_3K_9$  (par l'inactivation de Prdm3 et Prdm16 par exemple).

Par ailleurs, nous observons que l'ensemble des RCs a tendance à occuper des positions plus périphériques dans les cellules quiescentes. Il est possible que l'ancrage de la chromatine à la périphérie soit un état stable mais défavorisé par la dynamique imposée par le cycle cellulaire.

### D.2.2.3 Réorganisations issues de processus dirigés.

Ce type de réorganisation a par exemple été observé dans un modèle cellulaire humain, dans lequel l'action d'un stress (en l'occurrence une déprivation de sérum) induit la relocalisation de certains TCs vers l'intérieur et d'autres vers la périphérie du noyau. Ces réorganisations sont détectables dès 15 min de stress, et dépendent de myosines nucléaires et d'ATP, suggérant qu'elles sont issues d'un processus dirigé actif (voir §A.1.4.3). Ce modèle de relocalisation est très intéressant car il semble très simple à mettre en oeuvre.

Les mécanismes de relocalisation semblent dépendre de moteurs tels que les myosines nucléaires, mais les régions génomiques avec lesquelles ils interagissent pour conférer aux chromosomes leur mobilité ne sont pas connues. Les RCs sont de bonnes candidates, car elles sont composées de séquences répétées ; et si une unité de répétition avait une forte affinité pour des composants de complexes moteurs, l'efficacité globale serait augmentée par la répétition de cette unité qui permettrait de multiplier les interactions sur une même région génomique.

Nous avons observé au §C.2.2.6 que dans les cellules en cycle, les RCs sont orientées vers le nucléole lorsque le TC est en contact avec le nucléole, et vers la périphérie lorsqu'il est en contact avec la périphérie. Il est possible que cette orientation préférentielle reflète un rôle des RCs dans la mobilité des TCs. Dans le

cas du chromosome 7, nous observons que l'orientation des RCs des TCs qui sont en contact avec les nucléoles ou dans le reste du nucléoplasme est modifiée par rapport aux cellules en cycle, et devient périphérique en comparaison avec le reste du TC. Ceci suggère que dans les cellules quiescentes, la relocalisation périphérique des RCs précède celle des TC. Nous observons cette variation pour le chromosome 7 dont la RC et le TC ont une très forte tendance à se localiser en périphérie du noyau, mais pas pour le chromosome 1 qui a tendance à se localiser plus au centre du noyau, même dans les cellules quiescentes. Cette hypothèse pourrait être approfondie en exploitant le modèle de relocalisation décrit au début de ce paragraphe. Dans ce modèle, certains chromosomes sont relocalisés vers la périphérie et d'autres vers le centre du noyau, en l'espace de temps courts. Il serait donc possible de réaliser une cinétique de déprivation de sérum, et observer d'éventuelles variations d'orientation de la RC qui précéderaient le mouvement du chromosome dans une direction donnée. Notre hypothèse serait renforcée si :

- Pour des chromosomes se relocalisant vers la périphérie, l'orientation de la RC des TCs qui ne sont pas en contact avec la périphérie devient périphérique.
- Pour des chromosomes se relocalisant vers les nucléoles, l'orientation de la RC des TCs qui ne sont pas en contact avec les nucléoles devient nucléolaire.

Par ailleurs une autre observation de cette étude indique qu'un mouvement actif pourrait être conféré aux RCs. Dans le cas des RCs localisées au centre du noyau, nous observons une variation d'orientation préférentielle des péricentromères (par rapport aux centromères) vers la périphérie du noyau. Dans les cellules quiescentes, les RCs ayant tendance à se relocaliser vers la périphérie, il est possible que cette variation d'orientation soit la conséquence d'une force<sup>1</sup> s'exerçant sur les péricentromères dans la direction de la périphérie du noyau. Cette hypothèse pourrait être testée en inhibant des candidats potentiels tels que des myosines nucléaires et en observant si cette variation d'orientation est perdue.

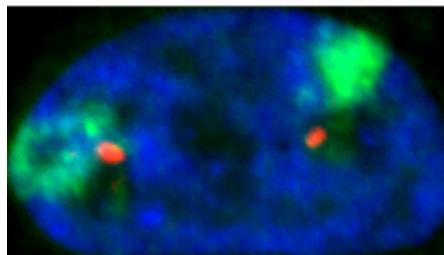


FIGURE D.1: Expérience de FISH 3D sur cellules hMES. Superposition en fausses couleurs du plan médian. Vert : TC du chromosome 1. Rouge : RCs du chromosome 1. Bleu : coloration de l'ADN au DAPI.

Enfin nous avons observé dans des cellules précurseurs mésenchymateuses que le centromère du chromosome 1 pouvait sortir de son TC. Lorsque le centromère se trouve en dehors du TC, nous l'avons toujours observé à proximité des nucléoles (voir Fig. D.1)<sup>2</sup>. Nous n'avons pas quantifié cet événement, mais il n'est pas rare.

1. Par exemple issue de myosines nucléaires

2. Les nucléoles sont ici observés par une coloration moins intense de l'ADN au DAPI.

Cette observation suggère que la mobilité des RCs pourrait aussi faire intervenir des boucles d'ADN.

### D.3 Conclusion

Dans ce travail, nous avons développé des observables permettant la description de l'organisation spatiale des RCs. Ces observables reposent à la fois sur des marqueurs (sondes spécifiques de RCs, sonde ciblant l'ensemble des régions péricentromériques) et sur des mesures (positionnement vis-à-vis de plusieurs structures nucléaires, orientation des centromères par rapport aux péricentromères ou aux TCs vis-à-vis d'une autre structure, description d'un pattern de points, etc.) Ces mesures ont été réalisées sur un grand nombre de noyaux de manière systématique grâce au développement d'un outil d'analyse quantitative d'image.

Nous avons décrit deux aspects de l'organisation spatiale des RCs humaines :

- Leur positionnement : nous avons montré qu'elles ont tendance à se positionner préférentiellement soit en périphérie du noyau, soit aux bords des nucléoles, et très peu dans le reste du nucléoplasme.
- Leur agrégation : nous avons montré que les régions péricentromériques portées par plusieurs chromosomes s'associent pour former des foyers d'hétérochromatine, comme chez la souris et de nombreux autres organismes.

Le positionnement et l'agrégation des RCs varient au cours du cycle cellulaire, et nous avons envisagé deux types de mécanismes pouvant intervenir dans cette dynamique : des réorganisations non dirigées et des mouvements dirigés. De plus ces deux aspects sont inter-dépendants, et nos observations nous mènent à formuler l'hypothèse que l'agrégation des RCs résulte d'un équilibre entre deux processus antagonistes : l'ancrage à la périphérie, et l'auto-agrégation de l'hétérochromatine qui pourrait avoir lieu principalement en périphérie des nucléoles.

Les observables que nous avons développées sont reproductibles entre des populations de noyaux indépendantes, et pourront donc permettre d'étudier les conséquences de traitements visant à disséquer les mécanismes moléculaires responsables de l'organisation particulière des RCs. Ces études pourront ensuite mener à la compréhension de la fonction des RCs humaines dans le noyau en interphase, mais peut être aussi de comprendre les contraintes fonctionnelles qui agissent sur l'évolution de leur séquences. Ces observables pourront enfin permettre de caractériser plus précisément les phénotypes de pathologies telles que les laminopathies ou le syndrome de Rett, dans lesquelles l'organisation spatiale du génome est perturbée, et ainsi de mieux comprendre les mécanismes à l'origine des perturbations.

## CHAPITRE E

# Annexes

TABLE E.1: Processing modules included in the first version of TANGO. Each module can contain several methods. See website for a more detailed description.

Module	Category	Short Description	Reference
Fast Filters 3D	Pre-filter	3D filters (mean, median, tophat, etc)	Table E.3.3
Histogram Transformation	Pre-filter	Operations on histogram (normalize, invert, scale, etc)	TANGO
Image Features	Pre-filter	Features from image (gradient scale, hessian)	Table E.3.11
Misc 3D filters	Pre-filter	3D filters (gaussian, bandpass, denoise, LoG3D)	Table E.3.8,7,15,14
Translate 3D	Pre-filter	Translation in x, y or z direction	Table E.3.11
Simple Segmenter	Segmenter	Simple Thresholding	Adapted from Table E.3.4
Hysteresis Segmenter	Segmenter	Segmentation using hysteresis thresholding (with two thresholds)	Adapted from Table E.3.11
Nucleus Edge Detector	Segmenter	Nucleus segmentation based on edge detection, robust over intensity variations	TANGO
Seeded Watershed 3D	Segmenter	Watershed 3D seeded with regional minima	TANGO
Spot Detector 3D	Segmenter	seeded watershed 3D run on the maximum eigen value of the hessian transform, dedicated to the segmentation of small spots	TANGO
Spot Segmenter	Segmenter	Segment spots using seeds, watershed and local thresholding	Table E.3.5
Erase Regions	Post-Filter	Erase region based on intensity criteria (to use after Seeded Watershed 3D)	TANGO

Continued on next page

TABLE E.1 – continued from previous page

Module	Category	Short Description	Reference
Fast Filters 3D	Post-Filter	3D filters (mean, median, tophat, etc)	Table E.3.3
Merge Regions	Post-Filter	Merge connected regions (to use after Seeded Watershed 3D)	TANGO
Morphological filters 3D	Post-Filter	3D Morphological filters performed individually on objects (opening, closing, 2D/3D fill holes) optimized for speed for large radii ( $O(\text{radius})$ )	TANGO
Size and Edge Filters	Post-Filter	Remove objects based on size or touching edges	TANGO
Watershed 2D	Post-Filter	2D watershed to separate object in 2D images	Table E.3.8
Re-label	Post-Filter	Labels connected objects	TANGO
AutoThreshold	Thresholder	ImageJ's autothreshold methods. Computes a threshold based on histogram	Table E.3.8
Kappa Sigma Clipping		Estimates the background value	Table E.3.13
Percentage of bright pixels	Thresholder	Computes the threshold value corresponding to a percentage of brightest pixels in the image	TANGO
Threshold Adjustment	Thresholder	Adjust threshold computed with another method (add $n*\text{sigma}$ , ...)	TANGO
Threshold operation	Thresholder	Operation between two thresholds computed with other methods (add, maximum, etc)	TANGO
Value	Thresholder	Constant value	TANGO

TABLE E.2: Quantitative image analysis modules included in the first version of TANGO. Each module can contain several measurements. See website for a more detailed description.

Module	Short Description	Reference
Distances	Center-center, center-border, border-border Distance between objects	TANGO
Grayscale spatial moments	3D Spatial Moments of signal	TANGO
JACOP colocalization	Colocalization based on intensity measurements	Table E.3.12
Measure Geometrical Advanced	Measurements of 3D surfaces using meshes	Table E.3.9
Measure Geometrical Simple	Geometrical measurements (volume, elongation, feret, etc)	Table E.3.1
Minimal distances	Minimal distances between objects of two structures	TANGO
Object Colocalization	Estimates overlap between segmented objects (percentage of overlap)	Table E.3.1
Shell Analysis	Normalized localization index in the nuclear space. Nucleus is divided into 3D layers of equal volume. Measures the index of the shell where the center of the object is located	TANGO
Spatial Statistics	Spatial Analysis of point pattern	as described in <i>Andrey et al. (2010)</i>
Eroded Volume Fraction	Continuous version of shell analysis, reference can be any combination of structures	adapted from <i>Ballester et al. (2008)</i>
Radial Autocorrelation 3D	3D version of radial autocorrelation. Image can be resampled to take anisotropy in account	TANGO
Signal quantification	Statistical descriptors of signal intensity within objects	Table E.3.1
Signal quantification Layer	Statistical descriptors of signal intensity within a defined layer of objects (inside or outside)	TANGO
Texture	Classical matrix texture descriptor (entropy, etc)	Table E.3.16
Texture 3D	3D radial version of the Texture module	TANGO

TABLE E.3: Third-party software used in TANGO

1. 3D Analysis : [http://imagejdocu.tudor.lu/doku.php?id=plugin:stacks:3d\\_roi\\_manager:start](http://imagejdocu.tudor.lu/doku.php?id=plugin:stacks:3d_roi_manager:start)
2. 3D Euclidean Distance Map : [http://www.optinav.com/Local\\_Thickness.htm](http://www.optinav.com/Local_Thickness.htm)
3. 3D Filters : [http://imagejdocu.tudor.lu/doku.php?id=plugin:filter:3d\\_filters:start](http://imagejdocu.tudor.lu/doku.php?id=plugin:filter:3d_filters:start)
4. 3D Object Counter : [http://imagejdocu.tudor.lu/doku.php?id=plugin:analysis:3d\\_object\\_counter:start](http://imagejdocu.tudor.lu/doku.php?id=plugin:analysis:3d_object_counter:start) (reference, see 12)
5. 3D Spots Segmentation : [http://imagejdocu.tudor.lu/doku.php?id=plugin:segmentation:3d\\_spots\\_segmentation:start](http://imagejdocu.tudor.lu/doku.php?id=plugin:segmentation:3d_spots_segmentation:start)
6. BoneJ : <http://bonej.org/> (used for ROI Z-interpolation during the manual nucleus segmentation procedure)
7. DropletFinder : [http://imagejdocu.tudor.lu/doku.php?id=plugin:analysis:droplet\\_counter:start](http://imagejdocu.tudor.lu/doku.php?id=plugin:analysis:droplet_counter:start)
8. ImageJ <http://rsbweb.nih.gov/ij/> Schneider CA, *et al.* (2012) NIH Image to ImageJ : 25 years of image analysis, *Nat. Methods*, **9**, 671-5
9. ImageJ 3D Viewer <http://3dviewer.neurofly.de/>. Schmid B., *et al.* (2010) A high-level 3D visualization API for Java and ImageJ, *BMC Bioinformatics*, **21**, 274
10. Image 5D <http://developer.imagej.net/plugins/image5d>
11. ImageScience : by Erik Meijering, <http://www.imagescience.org/meijering/software/featurej/>
12. JACOP : [http://imagejdocu.tudor.lu/doku.php?id=plugin:analysis:jacop\\_2.0:just\\_another\\_colocalization\\_plugin:start](http://imagejdocu.tudor.lu/doku.php?id=plugin:analysis:jacop_2.0:just_another_colocalization_plugin:start) Bolte S, *et al.* (2006) A guided tour into subcellular colocalization analysis in light microscopy, *J Microsc.*, **214**, 213-32
13. Kappa Sigma Clipping : <http://hdl.handle.net/1926/367> Gaetan Lehmann (2006) Kappa Sigma Clipping *The insight Journal*
14. LoG3D (Laplacian of Gaussian) : <http://bigwww.epfl.ch/sage/soft/LoG3D/> D. Sage, *et al.* (2005) Automatic Tracking of Individual Fluorescence Particles : Application to the Study of Chromosome Dynamics, *IEEE T. Image Proces.*, **14**, 1372-83
15. PureDenoise : <http://bigwww.epfl.ch/algorithms/denoise/>
16. Texture analysis : <http://rsbweb.nih.gov/ij/plugins/texture.html>

# Bibliographie

- ABNEY, J. R., CUTLER, B., FILLBACH, M. L., AXELROD, D. et SCALETTAR, B. a. (1997). Chromatin dynamics in interphase nuclei and its implications for nuclear structure. *The Journal of cell biology*, 137(7):1459–68. (Cité en page 46.)
- AGARWAL, N., BECKER, A., JOST, K. L., HAASE, S., THAKUR, B. K., BRERO, A., HARDT, T., KUDO, S., LEONHARDT, H. et CARDOSO, M. C. (2011). MeCP2 Rett mutations affect large scale chromatin organization. *Human molecular genetics*, 20(21):4187–95. (Cité en page 164.)
- AGARWAL, N., HARDT, T., BRERO, A., NOWAK, D., ROTHBAUER, U., BECKER, A., LEONHARDT, H. et CARDOSO, M. C. (2007). MeCP2 interacts with HP1 and modulates its heterochromatin association during myogenic differentiation. *Nucleic acids research*, 35(16):5402–8. (Cité en page 24.)
- AGUIRRE-LAVIN, T., ADENOT, P., BONNET-GARNIER, A., LEHMANN, G., FLEUROT, R., BOULESTEIX, C., DEBEY, P. et BEAUJEAN, N. (2012). 3D-FISH analysis of embryonic nuclei in mouse highlights several abrupt changes of nuclear organization during preimplantation development. *BMC developmental biology*, 12(1):30. (Cité en page 129.)
- ALCOBIA, I., DILÃO, R. et PARREIRA, L. (2000). Spatial associations of centromeres in the nuclei of hematopoietic cells : evidence for cell-type-specific organizational patterns. *Blood*, 95(5):1608–15. (Cité en pages 59, 71, 72 et 73.)
- ALCOBIA, I., QUINA, A. S., NEVES, H., CLODE, N. et PARREIRA, L. (2003). The spatial organization of centromeric heterochromatin during normal human lymphopoiesis : evidence for ontogenically determined spatial patterns. *Experimental Cell Research*, 290(2):358–369. (Cité en pages 72, 73, 142 et 148.)
- ALEXANDROV, I., KAZAKOV, A., TUMENEVA, I., SHEPELEV, V. et YUROV, Y. (2001). Alpha-satellite DNA of primates : old and new families. *Chromosoma*, 110(4):253–66. (Cité en pages 58 et 60.)
- ALEXANDROV, I., MEDVEDEV, L. I., MASHKOVA, T. D., KISSELEV, L. L., ROMANOVA, L. Y. et YUROV, Y. B. (1993). Definition of a new alpha satellite suprachromosomal family characterized by monomeric organization. *Nucleic acids research*, 21(9):2209–15. (Cité en page 57.)
- ALMOUZNI, G. et PROBST, A. (2011). Heterochromatin maintenance and establishment : lessons from the mouse pericentromere. *Nucleus*, 2(5):332–338. (Cité en page 129.)

- AMIR, R., den VEYVER, I. V. et WAN, M. (1999). Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. *Nature genetics*, 23(october):185–188. (Cité en page 165.)
- AMOR, D. J., KALITSIS, P., SUMER, H. et CHOO, K. H. A. (2004). Building the centromere : from foundation proteins to 3D organization. *Trends in cell biology*, 14(7):359–68. (Cité en page 61.)
- ANDREY, P., KIÊU, K., KRESS, C., LEHMANN, G., TIRICHINE, L., LIU, Z., BIOT, E., ADENOT, P.-G., HUE-BEAUVAIS, C., HOUBA-HÉRIN, N., DURANTHON, V., DEVINOY, E., BEAUJEAN, N., GAUDIN, V., MAURIN, Y. et DEBEY, P. (2010). Statistical Analysis of 3D Images Detects Regular Spatial Distributions of Centromeres and Chromocenters in Animal and Plant Nuclei. *PLoS Computational Biology*, 6(7):e1000853. (Cité en pages 88, 89, 92, 93, 97, 115, 166 et 179.)
- ANTEQUERA, F. et BIRD, a. (1993). Number of CpG islands and genes in human and mouse. *Proceedings of the National Academy of Sciences of the United States of America*, 90(24):11995–9. (Cité en page 23.)
- ASHBY, W. (1947). Principles of the self-organizing dynamic system. *The Journal of general psychology*, 37(2):125–8. (Cité en page 50.)
- AUSTIN, C. M. et BELLINI, M. (2010). The dynamic landscape of the cell nucleus. *Molecular reproduction and development*, 77(1):19–28. (Cité en pages 36 et 107.)
- BACCARINI, P. (1908). Sulle cinesi vegetative del *Cynomorium coccineum* L. *Nuovo giornale botanico ser.*, 15(2):189–203. (Cité en page 68.)
- BACHER, C. P., GUGGIARI, M., BRORS, B., AUGUI, S., CLERC, P., AVNER, P., EILS, R. et HEARD, E. (2006). Transient colocalization of X-inactivation centres accompanies the initiation of X inactivation. *Nature cell biology*, 8(3):293–9. (Cité en page 94.)
- BADDELEY, A. et TURNER, R. (2005). Spatstat : an R package for analyzing spatial point patterns. *Journal of statistical software*, 12(6). (Cité en page 95.)
- BAKER, S. A., CHEN, L., WILKINS, A. D., YU, P., LICHTARGE, O. et ZOGHBI, H. Y. (2013). An AT-hook domain in MeCP2 determines the clinical course of Rett syndrome and related disorders. *Cell*, 152(5):984–96. (Cité en page 79.)
- BALLESTER, M., KRESS, C., HUE-BEAUVAIS, C., KIÊU, K., LEHMANN, G., ADENOT, P. et DEVINOY, E. (2008). The nuclear localization of WAP and CSN genes is modified by lactogenic hormones in HC11 cells. *Journal of cellular biochemistry*, 105(1):262–70. (Cité en pages 91, 97, 118, 119, 133 et 179.)
- BANTIGNIES, F. et CAVALLI, G. (2011). Polycomb group proteins : repression in 3D. *Trends in genetics*, 27(11):454–64. (Cité en pages 35 et 43.)

- BANTIGNIES, F., GRIMAUD, C., LAVROV, S., GABUT, M. et CAVALLI, G. (2003). Inheritance of Polycomb-dependent chromosomal interactions in *Drosophila*. *Genes & development*, 17(19):2406–20. (Cité en page 42.)
- BANTIGNIES, F., ROURE, V., COMET, I., LEBLANC, B., SCHUETTENGRUBER, B., BONNET, J., TIXIER, V., MAS, A. et CAVALLI, G. (2011). Polycomb-Dependent Regulatory Contacts between Distant Hox Loci in *Drosophila*. *Cell*, 144(2):214–26. (Cité en page 42.)
- BARSKI, A., CUDDAPAH, S., CUI, K. et ROH, T. (2007). High-resolution profiling of histone methylations in the human genome. *Cell*, 129(4):823–37. (Cité en page 22.)
- BARTHOLDI, M. F. (1991). Nuclear distribution of centromeres during the cell cycle of human diploid fibroblasts. *Journal of cell science*, 99 ( Pt 2):255–63. (Cité en pages 66, 73 et 148.)
- BARTLETT, J., BLAGOJEVIC, J., CARTER, D., ESKIW, C., FROMAGET, M., JOB, C., SHAMSHER, M., TRINDADE, I. F., XU, M. et COOK, P. R. (2006). Specialized transcription factories. *Biochemical Society symposium*, 75(73):67–75. (Cité en page 35.)
- BÁRTOVÁ, E., GALIOVÁ, G., KREJČÍ, J., HARNICAROVÁ, A., STRASÁK, L. et KOZUBEK, S. (2008). Epigenome and chromatin structure in human embryonic stem cells undergoing differentiation. *Developmental dynamics*, 237(12):3690–702. (Cité en pages 67 et 68.)
- BEIL, M., DU, D., PASCHKE, S., SCHREINER, B., NOLTE, U., BRUEL, A. et IRINOPOULOU, T. (2002). Spatial distribution patterns of interphase centromeres during retinoic acid-induced differentiation of promyelocytic leukemia cells. *Cytometry*, 225:217–225. (Cité en pages 71, 73 et 148.)
- BEIL, M., FLEISCHER, F., PASCHKE, S. et SCHMIDT, V. (2005). Statistical analysis of the three-dimensional structure of centromeric heterochromatin in interphase nuclei. *Journal of microscopy*, 217(Pt 1):60–8. (Cité en pages 92 et 95.)
- BEISEL, C. et PARO, R. (2011). Silencing chromatin : comparing modes and mechanisms. *Nature reviews. Genetics*, 12(2):123–35. (Cité en pages 22 et 43.)
- BERGER, A. B., CABAL, G. G., FABRE, E., DUONG, T., BUC, H., NEHRBASS, U., GADAL, O. et ZIMMER, C. (2008). High-resolution statistical mapping reveals gene territories in live yeast. *Nature methods*, 5(12):1031–1037. (Cité en page 87.)
- BERGMANN, J. H., RODRÍGUEZ, M. G., MARTINS, N. M. C., KIMURA, H., KELLY, D. a., MASUMOTO, H., LARIONOV, V., JANSEN, L. E. T. et EARNSHAW, W. C. (2011). Epigenetic engineering shows H3K4me2 is required for HJURP targeting and CENP-A assembly on a synthetic human kinetochore. *The EMBO journal*, 30(2):328–40. (Cité en page 56.)

- BERTULAT, B., DE BONIS, M. L., DELLA RAGIONE, F., LEHMKUHL, A., MILDEN, M., STORM, C., JOST, K. L., SCALA, S., HENDRICH, B., D'ESPOSITO, M. et CARDOSO, M. C. (2012). MeCP2 dependent heterochromatin reorganization during neural differentiation of a novel Mecp2-deficient embryonic stem cell reporter line. *PLoS one*, 7(10):e47848. (Cité en page 74.)
- BESTOR, T. (2005). Transposons reanimated in mice. *Cell*, 122(3):320–2. (Cité en page 55.)
- BICKMORE, W. A. et van STEENSEL, B. (2013). Genome Architecture : Domain Organization of Interphase Chromosomes. *Cell*, 152(6):1270–1284. (Cité en pages 25 et 129.)
- BLACK, B. E., FOLTZ, D. R., CHAKRAVARTHY, S., LUGER, K., WOODS, V. L. et CLEVELAND, D. W. (2004). Structural determinants for generating centromeric chromatin. *Nature*, 430(6999):578–82. (Cité en page 56.)
- BLOWER, M. D., SULLIVAN, B. a. et KARPEN, G. H. (2002). Conserved organization of centromeric chromatin in flies and humans. *Developmental cell*, 2(3):319–30. (Cité en page 57.)
- BOISVERT, F.-M., van KONINGSBRUGGEN, S., NAVASCUÉS, J. et LAMOND, A. I. (2007). The multifunctional nucleolus. *Nature reviews. Molecular cell biology*, 8(7):574–85. (Cité en page 34.)
- BOLTE, S. et CORDELIÈRES, F. P. (2006). A guided tour into subcellular colocalization analysis in light microscopy. *Journal of microscopy*, 224(Pt 3):213–32. (Cité en pages 96 et 97.)
- BOLZER, A., KRETH, G., SOLOVEI, I., KOEHLER, D., SARACOGLU, K., FAUTH, C., MÜLLER, S., EILS, R., CREMER, C., SPEICHER, M. R., CREMER, T. et MU, S. (2005). Three-dimensional maps of all chromosomes in human male fibroblast nuclei and prometaphase rosettes. *PLoS biology*, 3(5):e157. (Cité en pages 37 et 90.)
- BOND, C. S. et FOX, A. H. (2009). Paraspeckles : nuclear bodies built on long noncoding RNA. *The Journal of cell biology*, 186(5):637–44. (Cité en pages 36 et 44.)
- BONNET-GARNIER, A., FEUERSTEIN, P., CHEBROUT, M., FLEUROT, R., JAN, H.-U., DEBEY, P. et BEAUJEAN, N. (2012). Genome organization and epigenetic marks in mouse germinal vesicle oocytes. *The International journal of developmental biology*, 56(10-12):877–87. (Cité en page 70.)
- BORNFLETH, H., EDELMANN, P., ZINK, D., CREMER, T. et CREMER, C. (1999). Quantitative Motion Analysis of Subchromosomal Foci in Living Cells Using Four-Dimensional Microscopy. *Biophysical Journal*, 77(5):2871–2886. (Cité en page 47.)

- BOTTA, M., HAIDER, S., LEUNG, I. X. Y., LIO, P. et MOZZICONACCI, J. (2010). Intra- and inter-chromosomal interactions correlate with CTCF binding genome wide. *Molecular systems biology*, 6(426):426. (Cité en pages 73 et 75.)
- BOULON, S., WESTMAN, B. J., HUTTEN, S., BOISVERT, F.-M. et LAMOND, A. I. (2010). The nucleolus under stress. *Molecular cell*, 40(2):216–27. (Cité en page 34.)
- BOUZINBA-SEGARD, H., GUAIS, A. et FRANCASTEL, C. (2006). Accumulation of small murine minor satellite transcripts leads to impaired centromeric architecture and function. *Proceedings of the National Academy of Sciences of the United States of America*, 103(23):8709–14. (Cité en page 63.)
- BOYLE, S., GILCHRIST, S., BRIDGER, J. M., MAHY, N. L., ELLIS, J. A. et BICKMORE, W. A. (2001). The spatial organization of human chromosomes within the nuclei of normal and emerin-mutant cells. *Human molecular genetics*, 10(3):211–220. (Cité en page 37.)
- BOYLE, S., RODESCH, M. J., HALVENSLEBEN, H. a., JEDDELOH, J. a. et BICKMORE, W. a. (2011). Fluorescence in situ hybridization with high-complexity repeat-free oligonucleotide probes generated by massively parallel synthesis. *Chromosome research*, pages 901–909. (Cité en pages 28 et 84.)
- BRERO, A., EASWARAN, H. P., NOWAK, D., GRUNEWALD, I., CREMER, T., LEONHARDT, H. et CARDOSO, M. C. (2005). Methyl CpG-binding proteins induce large-scale chromatin reorganization during terminal differentiation. *The Journal of cell biology*, 169(5):733–43. (Cité en pages 70, 78, 79 et 142.)
- BROCHER, J., VOGEL, B. et HOCK, R. (2010). HMGA1 down-regulation is crucial for chromatin composition and a gene expression profile permitting myogenic differentiation. *BMC cell biology*, 11:64. (Cité en pages 78 et 79.)
- BROERS, J. L. V., PEETERS, E. a. G., KUIJPERS, H. J. H., ENDERT, J., BOUTEN, C. V. C., OOMENS, C. W. J., BAAIJENS, F. P. T. et RAMAEKERS, F. C. S. (2004). Decreased mechanical stiffness in LMNA-/- cells is caused by defective nucleocytoskeletal integrity : implications for the development of laminopathies. *Human molecular genetics*, 13(21):2567–80. (Cité en pages 31 et 33.)
- BROWN, C. R., KENNEDY, C. J., DELMAR, V. a., FORBES, D. J. et SILVER, P. a. (2008). Global histone acetylation induces functional genomic reorganization at mammalian nuclear pore complexes. *Genes & development*, 22(5):627–39. (Cité en page 40.)
- BROWN, K. E., GUEST, S. S., SMALE, S. T., HAHM, K., MERKENSCHLAGER, M. et FISHER, a. G. (1997). Association of transcriptionally silent genes with Ikaros complexes at centromeric heterochromatin. *Cell*, 91(6):845–54. (Cité en pages 73 et 129.)

- BROWN, S. (1966). Heterochromatin. *Science*, 151(709):417–25. (Cité en page 18.)
- CAMPOS, E. et REINBERG, D. (2009). Histones : annotating chromatin. *Annual review of genetics*, 43:559–99. (Cité en page 22.)
- CARON, H., van SCHAIK, B., van der MEE, M., BAAS, F., RIGGINS, G., van SLUIS, P., HERMUS, M. C., van ASPEREN, R., BOON, K., VOÛTE, P. a., HEISTERKAMP, S., van KAMPEN, A. et VERSTEEG, R. (2001). The human transcriptome map : clustering of highly expressed genes in chromosomal domains. *Science*, 291(5507): 1289–92. (Cité en pages 25 et 26.)
- CARVALHO, C., PEREIRA, H. M., FERREIRA, J., PINA, C., MENDONÇA, D., ROSA, a. C. et CARMO-FONSECA, M. (2001). Chromosomal G-dark bands determine the spatial organization of centromeric heterochromatin in the nucleus. *Molecular biology of the cell*, 12(11):3563–72. (Cité en pages 66, 67, 68, 130, 132, 170 et 173.)
- CASAS-DELUCCHI, C. S., BECKER, A., BOLIUS, J. J. et CARDOSO, M. C. (2012). Targeted manipulation of heterochromatin rescues MeCP2 Rett mutants and re-establishes higher order chromatin organization. *Nucleic acids research*, 40(22): e176. (Cité en page 79.)
- CAUDRON-HERGER, M. et RIPPE, K. (2012). Nuclear architecture by RNA. *Current opinion in genetics & development*, 22(2):179–87. (Cité en page 36.)
- CAVALLI, G. et MISTELI, T. (2013). Functional implications of genome topology. *Nature Structural & Molecular Biology*, 20(3):290–299. (Cité en pages 30, 52 et 53.)
- CHAKALOVA, L. et FRASER, P. (2010). Organization of transcription. *Cold Spring Harbor perspectives in biology*, 2(9):a000729. (Cité en page 35.)
- CHEESEMAN, I. M. et DESAI, A. (2008). Molecular architecture of the kinetochore-microtubule interface. *Nature reviews. Molecular cell biology*, 9(1):33–46. (Cité en page 54.)
- CHUANG, C.-H., CARPENTER, A. E., FUCHSOVA, B., JOHNSON, T., LANEROLLE, P. D., BELMONT, A. S. et de LANEROLLE, P. (2006). Long-range directional movement of an interphase chromosome site. *Current biology : CB*, 16(8):825–31. (Cité en page 47.)
- CHUBB, J. R., BOYLE, S., PERRY, P. et BICKMORE, W. a. (2002). Chromatin motion is constrained by association with nuclear compartments in human cells. *Current biology : CB*, 12(6):439–45. (Cité en page 47.)
- COLLAS, P. (2009). The state-of-the-art of chromatin immunoprecipitation. *Chromatin Immunoprecipitation Assays*, 567:1–25. (Cité en page 82.)
- COOK, P. R. (2002). Predicting three-dimensional genome structure from transcriptional activity. *Nature genetics*, 32(3):347–52. (Cité en page 35.)

- CREMER, M., von HASE, J., VOLM, T. et BRERO, A. (2001). Non-random radial higher-order chromatin arrangements in nuclei of diploid human cells. *Chromosome research*, 9(7):541–67. (Cité en page 91.)
- CREMER, T. et CREMER, M. (2010). Chromosome territories. *Cold Spring Harbor perspectives in biology*, 2(3):a003889. (Cité en page 27.)
- CROFT, J. a., BRIDGER, J. M., BOYLE, S., PERRY, P., TEAGUE, P. et BICKMORE, W. A. (1999). Differences in the localization and morphology of chromosomes in the human nucleus. *The Journal of cell biology*, 145(6):1119–31. (Cité en pages 37 et 91.)
- DAHL, K., RIBEIRO, A. et LAMMERDING, J. (2008). Nuclear shape, mechanics, and mechanotransduction. *Circulation research*, 102(11):1307–1318. (Cité en page 33.)
- de WIT, E. et de LAAT, W. (2012). A decade of 3C technologies : insights into nuclear organization. *Genes and development*, 26(1):11–24. (Cité en pages 81 et 130.)
- DECHAT, T., PFLEGHAAR, K., SENGUPTA, K., SHIMI, T., SHUMAKER, D. K., SOLIMANDO, L. et GOLDMAN, R. D. (2008). Nuclear lamins : major factors in the structural organization and function of the nucleus and chromatin. *Genes & development*, (312):832–853. (Cité en pages 32 et 39.)
- DELAIRE, S., HUANG, Y. H., CHAN, S. W. et ROBEY, E. a. (2004). Dynamic repositioning of CD4 and CD8 genes during T cell development. *The Journal of experimental medicine*, 200(11):1427–35. (Cité en page 73.)
- DENEGRI, M., MORALLI, D., ROCCHI, M., BIGGIOGERA, M., RAIMONDI, E., COBIANCHI, F., CARLI, L. D., RIVA, S. et BIAMONTI, G. (2002). Human Chromosomes 9, 12, and 15 Contain the Nucleation Sites of Stress-Induced Nuclear Bodies. *Molecular biology of the cell*, 13(June):2069–2079. (Cité en page 44.)
- DERNBURG, a. F. (2001). Here, there, and everywhere : kinetochore function on holocentric chromosomes. *The Journal of cell biology*, 153(6):F33–8. (Cité en page 54.)
- DILLON, N. (2008). The impact of gene location in the nucleus on transcriptional regulation. *Developmental cell*, 15(2):182–6. (Cité en page 143.)
- DIXON, J. R., SELVARAJ, S., YUE, F., KIM, A., LI, Y., SHEN, Y., HU, M., LIU, J. S. et REN, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, 485(7398):376–80. (Cité en pages 26 et 27.)
- DOVER, G. (1986). Molecular drive in multigene families : how biological novelties arise, spread and are assimilated. *Trends in genetics*, 275(June):159–65. (Cité en page 59.)

- du SART, D., CANCELLA, M., EARLE, E., MAO, J., SAFFERY, R., TANTON, K. M., KALITSIS, P., MARTYN, J., BARRY, A. E. et CHOO, K. H. A. (1997). A functional neo-centromere formed through activation of a latent human centromere and consisting of non-alpha-satellite DNA. *Nature genetics*, 16(2):144–53. (Cit  en page 62.)
- DUNDR, M. (2012). Nuclear bodies : multifunctional companions of the genome. *Current opinion in cell biology*, 24(3):415–22. (Cit  en pages 32 et 36.)
- DUNDR, M. et MISTELI, T. (2001). Functional architecture in the cell nucleus. *Biochem. J*, 1783(11):2041–3. (Cit  en pages 50, 52 et 171.)
- DUNHAM, I., BIRNEY, E., LAJOIE, B. et SANYAL, A. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature*, 489(7414):57–74. (Cit  en page 24.)
- DUNLEAVY, E. M., ALMOUZNI, G. et KARPEN, G. H. (2011). H3.3 is deposited at centromeres in S phase as a placeholder for newly assembled CENP-A in G phase. *Nucleus*, 2(2):146–57. (Cit  en page 56.)
- DUNLEAVY, E. M., ROCHE, D., TAGAMI, H., LACOSTE, N., RAY-GALLET, D., NAKAMURA, Y., DAIGO, Y., NAKATANI, Y. et ALMOUZNI-PETTINOTTI, G. (2009). HJURP is a cell-cycle-dependent maintenance and deposition factor of CENP-A at centromeres. *Cell*, 137(3):485–97. (Cit  en page 56.)
- EDELMANN, P., BORNFLETH, H., ZINK, D., CREMER, T. et CREMER, C. (2001). Morphology and dynamics of chromosome territories in living cells. *Biochimica et biophysica acta*, 1551(1):M29–39. (Cit  en page 47.)
- EFRONI, S., DUTTAGUPTA, R., CHENG, J., DEGHANI, H., HOEPPNER, D. J., DASH, C., BAZETT-JONES, D. P., LE GRICE, S., MCKAY, R. D. G., BUETOW, K. H., GINGERAS, T. R., MISTELI, T. et MESHORER, E. (2008). Global transcription in pluripotent embryonic stem cells. *Cell stem cell*, 2(5):437–47. (Cit  en page 48.)
- ERNST, J., KHERADPOUR, P. et MIKKELSEN, T. (2011). Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature*, 473(7345):43–49. (Cit  en page 24.)
- EYMERY, A., CALLANAN, M. et VOURC’H, C. (2009). The secret message of heterochromatin : new insights into the mechanisms and function of centromeric and pericentric repeat sequence transcription. *The International journal of developmental biology*, 53(2-3):259–68. (Cit  en page 63.)
- FELSENFELD, G. et GROUDINE, M. (2003). Controlling the double helix. *Nature*, 421(6921):444–8. (Cit  en page 22.)
- FILION, G. J., van BEMMEL, J. G., BRAUNSCHWEIG, U., TALHOUT, W., KIND, J., WARD, L. D., BRUGMAN, W., de CASTRO, I. J., KERKHOVEN, R. M., BUSSEMAKER, H. J. et van STEENSEL, B. (2010). Systematic protein location mapping

- reveals five principal chromatin types in *Drosophila* cells. *Cell*, 143(2):212–24. (Cité en page 24.)
- FINCH, K. a., FONSEKA, G., IOANNOU, D., HICKSON, N., BARCLAY, Z., CHATZIMETIOU, K., MANTZOURATOU, A., HANDYSIDE, A., DELHANTY, J. et GRIFFIN, D. K. (2008). Nuclear organisation in totipotent human nuclei and its relationship to chromosomal abnormality. *Journal of cell science*, 121(Pt 5):655–63. (Cité en page 143.)
- FINLAN, L. E., SPROUL, D., THOMSON, I., BOYLE, S., KERR, E., PERRY, P., YLSTRA, B., CHUBB, J. R. et BICKMORE, W. a. (2008). Recruitment to the nuclear periphery can alter expression of genes in human cells. *PLoS genetics*, 4(3):e1000039. (Cité en page 41.)
- FLEMMING, W. (1965). Contributions to the knowledge of the cell and its vital processes. *The Journal of cell biology*. (Cité en page 18.)
- FUSSNER, E., STRAUSS, M., DJURIC, U., LI, R., AHMED, K., HART, M., ELLIS, J. et BAZETT-JONES, D. P. (2012). Open and closed domains in the mouse genome are configured as 10 nm chromatin fibres. *EMBO reports*, 13(11):992–6. (Cité en pages 25 et 69.)
- GEYER, P. K., VITALINI, M. W. et WALLRATH, L. L. (2011). Nuclear organization : taking a position on gene expression. *Current opinion in cell biology*, 23(3):354–9. (Cité en page 41.)
- GIBBUS, J. H. et DEKKER, J. (2013). The hierarchy of the 3D genome. *Molecular cell*, 49(5):773–82. (Cité en pages 26, 27, 30 et 171.)
- GILCHRIST, S., GILBERT, N., PERRY, P. et BICKMORE, W. a. (2004). Nuclear organization of centromeric domains is not perturbed by inhibition of histone deacetylases. *Chromosome research*, 12(5):505–16. (Cité en page 76.)
- GOETZE, S., LEEUW, W. D., GIROMUS, O., INDEMANS, M. H. G., KOSTER, J., ONDREJ, V., VERSTEEG, R. et DRIEL, R. V. (2007). The three-dimensional structure of human interphase chromosomes is related to the transcriptome map. *Molecular and cellular biology*, 27(12):4475–87. (Cité en page 27.)
- GREWAL, S. I. S. et ELGIN, S. C. R. (2007). Transcription and RNA interference in the formation of heterochromatin. *Nature*, 447(7143):399–406. (Cité en page 63.)
- GUARDA, A., BOLOGNESE, F., BONAPACE, I. M. et BADARACCO, G. (2009). Interaction between the inner nuclear membrane lamin B receptor and the heterochromatic methyl binding protein, MeCP2. *Experimental cell research*, 315(11):1895–903. (Cité en page 40.)
- GUELEN, L., PAGIE, L., BRASSET, E., MEULEMAN, W., FAZA, M. B., TALHOUT, W., EUSSEN, B. H., de KLEIN, A., WESSELS, L., de LAAT, W. et van STEENSEL,

- B. (2008). Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature*, 453(7197):948–51. (Cité en page 39.)
- GUENATRI, M., BAILLY, D., MAISON, C. et ALMOUZNI, G. (2004). Mouse centric and pericentric satellite repeats form distinct functional heterochromatin. *The Journal of cell biology*, 166(4):493–505. (Cité en pages 69 et 70.)
- GUETG, C. et SANTORO, R. (2012). Formation of nuclear heterochromatin : The nucleolar point of view. *Epigenetics*, (August):811–814. (Cité en page 42.)
- HAAF, T. et SCHMID, M. (1991). Chromosome topology in mammalian interphase nuclei. *Experimental cell research*, 192(2):325–32. (Cité en page 65.)
- HABERMANN, F. a., CREMER, M., WALTER, J., KRETH, G., von HASE, J., BAUER, K., WIENBERG, J., CREMER, C., CREMER, T. et SOLOVEI, I. (2001). Arrangements of macro- and microchromosomes in chicken cells. *Chromosome research*, 9(7):569–84. (Cité en page 29.)
- HANCOCK, R. (2004). A role for macromolecular crowding effects in the assembly and function of compartments in the nucleus. *Journal of structural biology*, 146(3): 281–90. (Cité en page 171.)
- HARRINGTON, J., BOKKELEN, G. V., MAYS, R. W., GUSTASHAW, K. et WILLARD, H. F. (1997). Formation of de novo centromeres and construction of first-generation human artificial microchromosomes. *Nature genetics*, 15(4):345–55. (Cité en page 63.)
- HAYDEN, K. E. (2012). Human centromere genomics : now it's personal. *Chromosome research*, 20(5):621–33. (Cité en pages 59, 63 et 129.)
- HEBBES, T., THORNE, A. et CRANE-ROBINSON, C. (1988). A direct link between core histone acetylation and transcriptionally active chromatin. *The EMBO journal*, 7(5):1395–1402. (Cité en page 22.)
- HEITZ, E. (1928). Das Heterochromatin der Moose. *Jahrb. wiss. Bot. J*, 69:762–818. (Cité en page 18.)
- HENDERSON, A., WARBURTON, D. et ATWOOD, K. (1972). Location of ribosomal DNA in the human chromosome complement. *Proceedings of the National Academy of Sciences*, 69(11):3394–3398. (Cité en page 33.)
- HENDRICH, B. et BIRD, a. (1998). Identification and characterization of a family of mammalian methyl-CpG binding proteins. *Molecular and cellular biology*, 18(11):6538–47. (Cité en page 69.)
- HEPPERGER, C., OTTEN, S., von HASE, J. et DIETZEL, S. (2007). Preservation of large-scale chromatin structure in FISH experiments. *Chromosoma*, 116(2):117–33. (Cité en page 83.)

- HERIDE, C., RICOUL, M., KIËU, K., von HASE, J., GUILLEMOT, V., CREMER, C., DUBRANA, K. et SABATIER, L. (2010). Distance between homologous chromosomes results from chromosome positioning constraints. *Journal of cell science*, 123(Pt 23):4063–75. (Cité en pages 92, 95 et 170.)
- HERNANDEZ-VERDUN, D. (2006). Nucleolus : from structure to dynamics. *Histochemistry and cell biology*, 125(1-2):127–37. (Cité en pages 33 et 135.)
- HEWITT, S. L., YIN, B., JI, Y., CHAUMEIL, J., MARSZALEK, K., TENTHOREY, J., SALVAGIOTTO, G., STEINEL, N., RAMSEY, L. B., GHYSDAEL, J., FARRAR, M. A., SLECKMAN, B. P., SCHATZ, D. G., BUSSLINGER, M., BASSING, C. H. et SKOK, J. A. (2009). RAG-1 and ATM coordinate monoallelic recombination and nuclear positioning of immunoglobulin loci. *Nature immunology*, 10(6):655–64. (Cité en page 74.)
- HO, C. Y. et LAMMERDING, J. (2012). Lamins at a glance. *Journal of cell science*, 125(Pt 9):2087–93. (Cité en page 33.)
- HOROWITZ, R. a., AGARD, D. a., SEDAT, J. W. et WOODCOCK, C. L. (1994). The three-dimensional architecture of chromatin in situ : electron tomography reveals fibers composed of a continuously variable zig-zag nucleosomal ribbon. *The Journal of cell biology*, 125(1):1–10. (Cité en page 25.)
- HOU, C. et CORCES, V. G. V. (2012). Throwing Transcription for a Loop : Expression of the genome in the 3D Nucleus. *Chromosoma*, 121(2):107–116. (Cité en page 27.)
- HOWMAN, E. V., FOWLER, K. J., NEWSON, a. J., REDWARD, S., MACDONALD, a. C., KALITSIS, P. et CHOO, K. H. A. (2000). Early disruption of centromeric chromatin organization in centromere protein A (Cenpa) null mice. *Proceedings of the National Academy of Sciences of the United States of America*, 97(3):1148–53. (Cité en page 61.)
- HÜBNER, M. R., ECKERSLEY-MASLIN, M. a. et SPECTOR, D. L. (2012). Chromatin organization and transcriptional regulation. *Current opinion in genetics & development*, 23(2):89–95. (Cité en page 20.)
- HÜBNER, M. R. et SPECTOR, D. L. (2010). Chromatin dynamics. *Annual review of biophysics*, 39:471–89. (Cité en page 49.)
- IANNUCELLI, E., MOMPART, F., GELLIN, J., LAHBIB-MANSAIS, Y., YERLE, M. et BOUDIER, T. (2010). NEMO : a tool for analyzing gene and chromosome territory distributions from 3D-FISH experiments. *Bioinformatics (Oxford, England)*, 26(5):696–7. (Cité en page 108.)
- JACKSON, D. et HASSAN, A. (1993). Visualization of focal sites of transcription within human nuclei. *The EMBO journal*, 1(3):1059–1065. (Cité en page 35.)

- JANICKI, S. M., TSUKAMOTO, T., SALGHETTI, S. E., TANSEY, W. P., SACHIDANANDAM, R., PRASANTH, K. V., RIED, T., SHAV-TAL, Y., BERTRAND, E., SINGER, R. H. et SPECTOR, D. L. (2004). From silencing to gene expression : real-time analysis in single cells. *Cell*, 116(5):683–98. (Cité en page 45.)
- JENUWEIN, T. et ALLIS, C. (2001). Translating the histone code. *Science*, 293(5532): 1074–80. (Cité en page 23.)
- JOST, K. L., BERTULAT, B. et CARDOSO, M. C. (2012). Heterochromatin and gene positioning : inside, outside, any side? *Chromosoma*, 121(6):555–63. (Cité en page 73.)
- JOST, K. L., HAASE, S., SMEETS, D., SCHRODE, N., SCHMIEDEL, J. M., BERTULAT, B., HERZEL, H., CREMER, M. et CARDOSO, M. C. (2011). 3D-Image analysis platform monitoring relocation of pluripotency genes during reprogramming. *Nucleic acids research*, 39(17):e113. (Cité en pages 92 et 130.)
- KAISER, T., INTINE, R. et DUNDR, M. (2008). De novo formation of a subnuclear body. *Science*, 322(December):1713–1717. (Cité en page 50.)
- KALHOR, R., TJONG, H., JAYATHILAKA, N., ALBER, F. et CHEN, L. (2012). Genome architectures revealed by tethered chromosome conformation capture and population-based modeling. *Nature biotechnology*, 30(1):90–8. (Cité en pages 28, 29 et 82.)
- KAMENTSKY, L., JONES, T. R., FRASER, A., BRAY, M.-A., LOGAN, D. J., MADDEN, K. L., LJOSA, V., RUEDEN, C., ELICEIRI, K. W. et CARPENTER, A. E. (2011). Improved structure, function and compatibility for CellProfiler : modular high-throughput image analysis software. *Bioinformatics (Oxford, England)*, 27(8): 1179–80. (Cité en page 107.)
- KAZAKOV, A. E., SHEPELEV, V. a., TUMENEVA, I. G., ALEXANDROV, A. a., YUROV, Y. B. et ALEXANDROV, I. a. (2003). Interspersed repeats are found predominantly in the “old” alpha-satellite families. *Genomics*, 82(6):619–627. (Cité en page 60.)
- KHOCHBIN, S. (2001). Histone H1 diversity : bridging regulatory signals to linker histone function. *Gene*, 271(1):1–12. (Cité en page 21.)
- KIND, J., PAGIE, L., ORTABOZKOYUN, H., BOYLE, S., de VRIES, S. S., JANSSEN, H., AMENDOLA, M., NOLEN, L. D., BICKMORE, W. a. et van STEENSEL, B. (2013). Single-cell dynamics of genome-nuclear lamina interactions. *Cell*, 153(1):178–92. (Cité en pages 44 et 45.)
- KONINGSBRUGGEN, S. V., GIERLIN, M., MARTIN, D., BARTON, G. J., ARIYUREK, Y., DUNNEN, J. T. D. et LAMOND, A. I. (2010). High-Resolution Whole-Genome Sequencing Reveals That Specific Chromatin Domains from Most Human Chromosomes Associate with Nucleoli. *Molecular biology of the cell*, 21:3735–3748. (Cité en pages 42 et 44.)

- KOSAK, S. T., SKOK, J. a., MEDINA, K. L., RIBLET, R., LE BEAU, M. M., FISHER, A. G. et SINGH, H. (2002). Subnuclear compartmentalization of immunoglobulin loci during lymphocyte development. *Science*, 296(5565):158–62. (Cité en page 39.)
- KOURMOULI, N., JEPPESEN, P., MAHADEVHAIH, S., BURGOYNE, P., WU, R., GILBERT, D. M., BONGIORNI, S., PRANTERA, G., FANTI, L., PIMPINELLI, S., SHI, W., FUNDELE, R. et SINGH, P. B. (2004). Heterochromatin and tri-methylated lysine 20 of histone H4 in animals. *Journal of cell science*, 117(Pt 12):2491–501. (Cité en page 69.)
- KRUIHOF, M., CHIEN, F.-T., ROUTH, A., LOGIE, C., RHODES, D. et van NOORT, J. (2009). Single-molecule force spectroscopy reveals a highly compliant helical folding for the 30-nm chromatin fiber. *Nature structural & molecular biology*, 16(5):534–40. (Cité en page 25.)
- KUMARAN, R. I. et SPECTOR, D. L. (2008). A genetic locus targeted to the nuclear periphery in living cells maintains its transcriptional competence. *The Journal of cell biology*, 180(1):51–65. (Cité en page 41.)
- LEE, C., WEVRICK, R., FISHER, R. B., FERGUSON-SMITH, M. a. et LIN, C. C. (1997). Human centromeric DNAs. *Human genetics*, 100(3-4):291–304. (Cité en pages 57 et 58.)
- LEE, D. C., WELTON, K. L., SMITH, E. D. et KENNEDY, B. K. (2009). A-type nuclear lamins act as transcriptional repressors when targeted to promoters. *Experimental cell research*, 315(6):996–1007. (Cité en page 41.)
- LEVI, V., RUAN, Q., PLUTZ, M., BELMONT, A. S. et GRATTON, E. (2005). Chromatin dynamics in interphase cells revealed by tracking in a two-photon excitation microscope. *Biophysical journal*, 89(6):4275–85. (Cité en pages 46 et 47.)
- LEWIS, J., MEEHAN, R. et HENZEL, W. (1992). Purification, sequence, and cellular localization of a novel chromosomal protein that binds to methylated DNA. *Cell*, 69(6):905–14. (Cité en page 69.)
- LIEBERMAN-AIDEN, E., van BERKUM, N. L., WILLIAMS, L., IMAKAEV, M., RAGOCZY, T., TELLING, A., AMIT, I., LAJOIE, B. R., SABO, P. J., DORSCHNER, M. O., SANDSTROM, R., BERNSTEIN, B., BENDER, M. a., GROUDINE, M., GNIRKE, A., STAMATOYANNOPOULOS, J., MIRNY, L. a., LANDER, E. S., DEKKER, J. et van BERKUM, N. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science (New York, N.Y.)*, 326(5950):289–93. (Cité en pages 28 et 30.)
- LO, a. W., LIAO, G. C., ROCCHI, M. et CHOO, K. H. (1999). Extreme reduction of chromosome-specific alpha-satellite array is unusually common in human chromosome 21. *Genome research*, 9(10):895–908. (Cité en page 134.)

- LUGER, K., MÄDER, A. et RICHMOND, R. (1997). Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature*, 7:251–260. (Cité en pages 20 et 21.)
- MA, H., REYES-GUTIERREZ, P. et PEDERSON, T. (2013). Visualization of repetitive DNA sequences in human chromosomes with transcription activator-like effectors. *Proceedings of the National Academy of Sciences of the United States of America*, pages 1–6. (Cité en page 142.)
- MA, H., SAMARABANDU, J., DEVDHAR, R. S., ACHARYA, R., CHENG, P. C., MENG, C. et BEREZNEY, R. (1998). Spatial and temporal dynamics of DNA replication sites in mammalian cells. *The Journal of cell biology*, 143(6):1415–25. (Cité en page 27.)
- MAISON, C. et ALMOUZNI, G. (2004). HP1 and the dynamics of heterochromatin maintenance. *Nature reviews. Molecular cell biology*, 5(4):296–304. (Cité en page 69.)
- MAISON, C., BAILLY, D., PETERS, A. H. F. M., QUIVY, J.-P., ROCHE, D., TADDEI, A., LACHNER, M., JENUWEIN, T. et ALMOUZNI, G. (2002). Higher-order structure in pericentric heterochromatin involves a distinct pattern of histone modification and an RNA component. *Nature genetics*, 30(3):329–34. (Cité en page 63.)
- MALHAS, A. et VAUX, D. (2009). Transcription factor sequestration by nuclear envelope components. *Cell Cycle*, (April):959–964. (Cité en page 143.)
- MANUELIDIS, L. (1978). Chromosomal localization of complex and simple repeated human DNAs. *Chromosoma*, 32:23–32. (Cité en page 57.)
- MANUELIDIS, L. (1984). Different central nervous system cell types display distinct and nonrandom arrangements of satellite DNA sequences. *Proceedings of the National Academy of Sciences of the United States of America*, 81(10):3123–7. (Cité en page 129.)
- MANUELIDIS, L. (1985). Indications of Centromere Movement during Interphase and Differentiation. *Annals of the New York Academy of Sciences*, pages 205–221. (Cité en page 129.)
- MAO, Y. S., ZHANG, B. et SPECTOR, D. L. (2011). Biogenesis and function of nuclear bodies. *Trends in genetics*, 27(8):295–306. (Cité en pages 31, 32 et 36.)
- MARGALIT, A., BRACHNER, A. et GOTZMANN, J. (2007). Barrier-to-autointegration factor—a BAFfling little protein. *Trends in cell biology*, 17(4):202–8. (Cité en page 40.)
- MARKAKI, Y., SMEETS, D., FIEDLER, S., SCHMID, V. J., SCHERMELLEH, L., CREMER, T. et CREMER, M. (2012). The potential of 3D-FISH and super-resolution structured illumination microscopy for studies of 3D nuclear architecture.

- BioEssays : news and reviews in molecular, cellular and developmental biology*, 34(5):412–26. (Cité en page 83.)
- MARKO, J. et POIRIER, M. (2003). Micromechanics of chromatin and chromosomes. *Biochemistry and cell biology*, 81:209–20. (Cité en page 48.)
- MARSHALL, O., CHUEH, A., WONG, L. et CHOO, K. H. A. (2008). Neocentromeres : new insights into centromere structure, disease development, and karyotype evolution. *The American Journal of Human Genetics*, 82(February):261–282. (Cité en page 62.)
- MARTOU, G. et DE BONI, U. (2000). Nuclear topology of murine, cerebellar Purkinje neurons : changes as a function of development. *Experimental cell research*, 256(1):131–9. (Cité en pages 69 et 70.)
- MATTOU, A. et MESHORER, E. (2010). Chromatin plasticity and genome organization in pluripotent embryonic stem cells. *Current opinion in cell biology*, 22(3):334–41. (Cité en page 49.)
- MAYER, R., BRERO, A., SCHROEDER, T., CREMER, T. et DIETZEL, S. (2005). Common themes and cell type specific variations of higher order chromatin arrangements in the mouse. *BMC cell biology*, 22:1–22. (Cité en pages 70, 71 et 170.)
- MCSTAY, B. et GRUMMT, I. (2008). The epigenetics of rRNA genes : from molecular to chromosome biology. *Annual review of cell and developmental biology*, 24:131–57. (Cité en page 34.)
- MEABURN, K. J., CABUY, E., BONNE, G., LEVY, N., MORRIS, G. E., NOVELLI, G., KILL, I. R. et BRIDGER, J. M. (2007). Primary laminopathy fibroblasts display altered genome organization and apoptosis. *Aging cell*, 6(2):139–53. (Cité en page 39.)
- MEABURN, K. J. et MISTELI, T. (2007). Chromosome territories. *Nature*, 445(January). (Cité en page 29.)
- MEHTA, I., AMIRA, M., HARVEY, A. et BRIDGER, J. M. (2010). Rapid chromosome territory relocation by nuclear motor activity in response to serum removal in primary human fibroblasts. *Genome biology*, 11:1–17. (Cité en page 48.)
- MEHTA, I. S., ESKIW, C. H., ARICAN, H. D., KILL, I. R. et BRIDGER, J. M. (2011). Farnesyltransferase inhibitor treatment restores chromosome territory positions and active chromosome dynamics in Hutchinson-Gilford progeria syndrome cells. *Genome biology*, 12(8):R74. (Cité en pages 39 et 48.)
- MÉJAT, A. et MISTELI, T. (2010). LINC complexes in health and disease. *Nucleus (Austin, Tex.)*, 1(1):40–52. (Cité en pages 31 et 32.)
- MELDI, L. et BRICKNER, J. H. (2011). Compartmentalization of the nucleus. *Trends in cell biology*, 21(12):701–8. (Cité en page 129.)

- MESHORER, E. et MISTELI, T. (2006). Chromatin in pluripotent embryonic stem cells and differentiation. *Nature reviews. Molecular cell biology*, 7(7):540–6. (Cité en page 49.)
- MESHORER, E., YELLAJOSHULA, D. et GEORGE, E. (2006). Hyperdynamic plasticity of chromatin proteins in pluripotent embryonic stem cells. *Developmental cell*, 10:105–116. (Cité en page 48.)
- MEULEMAN, W., PERIC-HUPKES, D., KIND, J., BEAUDRY, J.-B., PAGIE, L., KELIS, M., REINDERS, M., WESSELS, L. et van STEENSEL, B. (2013). Constitutive nuclear lamina-genome interactions are highly conserved and associated with A/T-rich sequence. *Genome research*, 23(2):270–80. (Cité en page 40.)
- MILON, B. C., CHENG, H., TSELEBROVSKY, M. V., LAVROV, S. a., NENASHEVA, V. V., MIKHALEVA, E. a., SHEVELYOV, Y. Y. et NURMINSKY, D. I. (2012). Role of histone deacetylases in gene regulation at nuclear lamina. *PloS one*, 7(11):e49692. (Cité en page 41.)
- MISTELI, T. (2007). Beyond the sequence : cellular organization of genome function. *Cell*, 128(4):787–800. (Cité en page 51.)
- MISTELI, T. (2009). Self-organization in the genome. *Proceedings of the National Academy of Sciences of the United States of America*, 106(17):6885–6. (Cité en pages 51 et 85.)
- MOHN, F., WEBER, M., REBHAN, M., ROLOFF, T. C., RICHTER, J., STADLER, M. B., BIBEL, M. et SCHÜBELER, D. (2008). Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors. *Molecular cell*, 30(6):755–66. (Cité en page 23.)
- MOREY, C., KRESS, C. et BICKMORE, W. A. (2009). Lack of bystander activation shows that localization exterior to chromosome territories is not sufficient to up-regulate gene expression. *Genome research*, 19:1184–1194. (Cité en page 30.)
- MUDRAK, O. S., B NAZAROV, I., JONES, E. L. et ZALENSKY, A. O. (2012). Positioning of chromosomes in human spermatozoa is determined by ordered centromere arrangement. *PloS one*, 7(12):e52944. (Cité en page 87.)
- NAGANO, T., LUBLING, Y., STEVENS, T. J., SCHOENFELDER, S., YAFFE, E., DEAN, W., LAUE, E. D., TANAY, A. et FRASER, P. (2013). Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature*, 502(7469):59–64. (Cité en pages 82 et 85.)
- NÉMETH, A., CONESA, A., SANTOYO-LOPEZ, J., MEDINA, I., MONTANER, D., PÉTERFIA, B., SOLOVEI, I., CREMER, T., DOPAZO, J. et LÄNGST, G. (2010). Initial genomics of the human nucleolus. *PLoS genetics*, 6(3):e1000889. (Cité en pages 42, 44, 67 et 141.)

- NÉMETH, A. et LÄNGST, G. (2011). Genome organization in and around the nucleolus. *Trends in genetics*, 27(4):149–56. (Cité en pages 33, 42 et 77.)
- NIZAMI, Z., DERYUSHEVA, S. et GALL, J. G. (2010). The Cajal body and histone locus body. *Cold Spring Harbor perspectives in biology*, 2(7):a000653. (Cité en pages 36 et 44.)
- NORA, E. P., LAJOIE, B. R., SCHULZ, E. G., GIORGETTI, L., OKAMOTO, I., SERVANT, N., PIOLOT, T., van BERKUM, N. L., MEISIG, J., SEDAT, J., GRIBNAU, J., BARILLOT, E., BLÜTHGEN, N., DEKKER, J. et HEARD, E. (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature*, 485(7398):381–5. (Cité en page 26.)
- OHZEKI, J.-i., NAKANO, M., OKADA, T. et MASUMOTO, H. (2002). CENP-B box is required for de novo centromere chromatin assembly on human alphoid DNA. *The Journal of cell biology*, 159(5):765–75. (Cité en page 63.)
- OLINS, A. et OLINS, D. (1974). Spheroid chromatin units ( $\nu$  bodies). *Science*, 183(8):330–332. (Cité en page 20.)
- OLINS, D. E. et OLINS, A. L. (2003). Chromatin history : our view from the bridge. *Nature reviews. Molecular cell biology*, 4(10):809–14. (Cité en page 20.)
- OLLION, J., COCHENNEC, J., LOLL, F., ESCUDÉ, C. et BOUDIER, T. (2013). TANGO : A Generic Tool for High-throughput 3D Image Analysis for Studying Nuclear Organization. *Bioinformatics*, pages 1–2. (Cité en pages 106 et 130.)
- OSBORNE, C. S., CHAKALOVA, L., BROWN, K. E., CARTER, D., HORTON, A., DEBRAND, E., GOYENECHEA, B., MITCHELL, J. a., LOPES, S., REIK, W. et FRASER, P. (2004). Active genes dynamically colocalize to shared sites of ongoing transcription. *Nature genetics*, 36(10):1065–71. (Cité en page 35.)
- OTSU, N. (1975). A Threshold Selection Method from Gray-Level Histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1):62–66. (Cité en pages 88 et 114.)
- PADEKEN, J., MENDIBURO, M. J., CHLAMYDAS, S., SCHWARZ, H.-J., KREMMER, E. et HEUN, P. (2013). The nucleoplasmin homolog NLP mediates centromere clustering and anchoring to the nucleolus. *Molecular cell*, 50(2):236–49. (Cité en page 142.)
- PARADA, L. a., MCQUEEN, P. G. et MISTELI, T. (2004). Tissue-specific spatial organization of genomes. *Genome biology*, 5(7):R44. (Cité en page 37.)
- PARDUE, M. L. et GALL, J. G. (1970). Chromosomal localization of mouse satellite DNA. *Science*, 168(3937):1356–8. (Cité en page 54.)

- PAULER, F., SLOANE, M. et HUANG, R. (2009). H3K27me3 forms BLOCs over silent genes and intergenic regions and specifies a histone banding pattern on a mouse autosomal chromosome. *Genome research*, 19:221–233. (Cité en page 26.)
- PEDERSON, T. (2010). The Nucleolus. *Cold Spring Harbor Perspectives in Biology*, 3(3):a000638–a000638. (Cité en page 34.)
- PERIC-HUPKES, D., MEULEMAN, W., PAGIE, L., BRUGGEMAN, S. W. M., SOLOVEL, I., BRUGMAN, W., GRÄF, S., FLICEK, P., KERKHOVEN, R. M., van LOHUIZEN, M., REINDERS, M., WESSELS, L. et van STEENSEL, B. (2010). Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation. *Molecular cell*, 38(4):603–13. (Cité en page 41.)
- PETERS, A. H. F. M., KUBICEK, S., MECHTLER, K., O’SULLIVAN, R. J., DERIJCK, A. a. H. a., PEREZ-BURGOS, L., KOHLMAIER, A., OPRAVIL, S., TACHIBANA, M., SHINKAI, Y., MARTENS, J. H. A., JENUWEIN, T., BIOCENTER, T. V. et KAWARACHO, S. (2003). Partitioning and plasticity of repressive histone methylation states in mammalian chromatin. *Molecular cell*, 12(6):1577–89. (Cité en page 69.)
- PINHEIRO, I., MARGUERON, R., SHUKEIR, N., EISOLD, M., FRITZSCH, C., RICHTER, F. M., MITTLER, G., GENOUD, C., GOYAMA, S., KUROKAWA, M., SON, J., REINBERG, D., LACHNER, M. et JENUWEIN, T. (2012). Prdm3 and Prdm16 are H3K9me1 methyltransferases required for mammalian heterochromatin integrity. *Cell*, 150(5):948–60. (Cité en page 79.)
- PIRAS, F. M., NERGADZE, S. G., MAGNANI, E., BERTONI, L., ATTOLINI, C., KHORRAULI, L., RAIMONDI, E. et GIULOTTO, E. (2010). Uncoupling of satellite DNA and centromeric function in the genus *Equus*. *PLoS genetics*, 6(2):e1000845. (Cité en page 62.)
- PLOHL, M., LUCHETTI, A., MESTROVIĆ, N. et MANTOVANI, B. (2008). Satellite DNAs between selfishness and functionality : structure, genomics and evolution of tandem repeats in centromeric (hetero)chromatin. *Gene*, 409(1-2):72–82. (Cité en page 60.)
- PLUTA, A., MACKAY, A. et AINSZTEIN, A. (1995). The centromere : hub of chromosomal activities. *Science*, 270(7):1591–4. (Cité en pages 54 et 129.)
- POLLOCK, C. et HUANG, S. (2010). The Perinucleolar Compartment. *Cold Spring Harbor Perspectives in Biology*, 2(2):a000679–a000679. (Cité en page 33.)
- POMBO, a., JACKSON, D. a., HOLLINSHEAD, M., WANG, Z., ROEDER, R. G. et COOK, P. R. (1999). Regional specialization in human nuclei : visualization of discrete sites of transcription by RNA polymerase III. *The EMBO journal*, 18(8):2241–53. (Cité en page 35.)
- PROBST, A. V., OKAMOTO, I., CASANOVA, M., EL MARJOU, F., LE BACCON, P. et ALMOUZNI, G. (2010). A strand-specific burst in transcription of pericentric

- satellites is required for chromocenter formation and early mouse development. *Developmental cell*, 19(4):625–38. (Cité en page 74.)
- R, C. T. (2013). R : A language and environment for statistical computing. *R Foundation for Statistical Computing*. (Cité en page 107.)
- RAJAPAKSE, I. et GROUDINE, M. (2011). On emerging nuclear order. *The Journal of cell biology*, 192(5):711–21. (Cité en pages 50, 52 et 129.)
- RAJAPAKSE, I., PERLMAN, M. D., SCALZO, D., KOOPERBERG, C., GROUDINE, M. et KOSAK, S. T. (2009). The emergence of lineage-specific chromosomal topologies from coordinate gene regulation. *Proceedings of the National Academy of Sciences of the United States of America*, 106(16):6679–84. (Cité en pages 52 et 85.)
- RAM, S., RODRÍGUEZ, J. et BOSCO, G. (2012). Segmentation and detection of fluorescent 3D spots. *Cytometry Part A*, (5). (Cité en pages 88, 89 et 116.)
- RAZIN, S. V., GAVRILOV, a. a., PICHUGIN, a., LIPINSKI, M., IAROVAIA, O. V. et VASSETZKY, Y. S. (2011). Transcription factories in the context of the nuclear and genome organization. *Nucleic acids research*, 39(21):9085–92. (Cité en page 35.)
- REA, S., EISENHABER, F., O’CARROLL, D., STRAHL, B. D., SUN, Z. W., SCHMID, M., OPRAVIL, S., MECHTLER, K., PONTING, C. P., ALLIS, C. D. et JENUWEIN, T. (2000). Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature*, 406(6796):593–9. (Cité en page 79.)
- REDDY, K. L., ZULLO, J. M., BERTOLINO, E. et SINGH, H. (2008). Transcriptional repression mediated by repositioning of genes to the nuclear lamina. *Nature*, 452(7184):243–7. (Cité en page 41.)
- ROIX, J. J., MCQUEEN, P. G., MUNSON, P. J., PARADA, L. A. et MISTELI, T. (2003). Spatial proximity of translocation-prone gene loci in human lymphomas. *Nature genetics*, 34(3):287–291. (Cité en pages 39 et 90.)
- ROLDÁN, E., FUXA, M., CHONG, W., MARTINEZ, D., NOVATCHKOVA, M., BUSSLINGER, M., SKOK, J. A. et ROLDA, E. (2005). Locus ‘decontraction’ and centromeric recruitment contribute to allelic exclusion of the immunoglobulin heavy-chain gene. *Nature immunology*, 6(1):31–41. (Cité en page 74.)
- RONNEBERGER, O., BADDELEY, D., SCHEIPL, F., VERVEER, P. J., BURKHARDT, H., CREMER, C., FAHRMEIR, L., CREMER, T. et JOFFE, B. (2008). *Spatial quantitative analysis of fluorescently labeled nuclear structures : problems, methods, pitfalls*, volume 16. (Cité en pages 91 et 95.)
- ROUQUETTE, J., CREMER, C., CREMER, T. et FAKAN, S. (2010). *Functional nuclear architecture studied by microscopy : present and future.*, volume 282. Elsevier Inc., 1 édition. (Cité en page 129.)

- ROZWADOWSKA, N., KOLANOWSKI, T., WILAND, E., SIATKOWSKI, M., PAWLAK, P., MALCHER, A., MIETKIEWSKI, T., OLSZEWSKA, M. et KURPISZ, M. (2013). Characterisation of Nuclear Architectural Alterations during In Vitro Differentiation of Human Stem Cells of Myogenic Origin. *PloS one*, 8(9):e73231. (Cité en page 67.)
- RUDD, M. (2005). *Organization, evolution and function of alpha satellite DNA at human centromeres*. Thèse de doctorat. (Cité en page 55.)
- RUDD, M. et WILLARD, H. (2004). Analysis of the centromeric regions of the human genome assembly. *Trends in genetics*, 20(11):524–9. (Cité en page 55.)
- RUSS, J. (2006). *The image processing handbook*. (Cité en page 88.)
- SAFFERY, R., WONG, L. H., IRVINE, D. V., BATEMAN, M. a., GRIFFITHS, B., CUTTS, S. M., CANCELLA, M. R., CENDRON, a. C., STAFFORD, a. J. et CHOO, K. H. A. (2001). Construction of neocentromere-based human minichromosomes by telomere-associated chromosomal truncation. *Proceedings of the National Academy of Sciences of the United States of America*, 98(10):5705–10. (Cité en page 63.)
- SALNÍKOVÁ, M., KOZUBEK, S. et LUKÁŠOVÁ, E. (2000). Spatial arrangement of genes, centromeres and chromosomes in human blood cell nuclei and its changes during the cell cycle, differentiation and after irradiation. *Chromosome Research*, 8:487–99. (Cité en pages 67 et 130.)
- SARMA, K. et REINBERG, D. (2005). Histone variants meet their match. *Nature reviews. Molecular cell biology*, 6(2):139–49. (Cité en page 21.)
- SASAI, N. et DEFOSSEZ, P.-A. (2009). Many paths to one goal? The proteins that recognize methylated DNA in eukaryotes. *The International journal of developmental biology*, 53(2-3):323–34. (Cité en page 24.)
- SCAFFIDI, P. et MISTELI, T. (2006). Lamin A-dependent nuclear defects in human aging. *Science*, 1059(2006). (Cité en page 39.)
- SCHIRMER, E. et FOISNER, R. (2007). Proteins that associate with lamins : many faces, many functions. *Experimental cell research*, 313(10):2167–79. (Cité en pages 33, 34 et 41.)
- SCHNEIDER, C., RASBAND, W. et ELICEIRI, K. (2012). NIH Image to ImageJ : 25 years of image analysis. *Nature methods*. (Cité en page 107.)
- SCHOENFELDER, S., CLAY, I. et FRASER, P. (2010a). The transcriptional interactome : gene expression in 3D. *Current opinion in genetics & development*, 20(2):127–33. (Cité en page 43.)

- SCHOENFELDER, S., SEXTON, T., CHAKALOVA, L., COPE, N. F., HORTON, A., ANDREWS, S., KURUKUTI, S., MITCHELL, J. a., UMLAUF, D., DIMITROVA, D. S., ESKIW, C. H., LUO, Y., WEI, C.-l., RUAN, Y., BIEKER, J. J. et FRASER, P. (2010b). Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells. *Nature Genetics*, 42(1):53–61. (Cité en page 43.)
- SCHUELER, M. et SULLIVAN, B. (2006). Structural and functional dynamics of human centromeric chromatin. *Annual review of human genetics*, 7:301–13. (Cité en pages 57, 58, 60 et 61.)
- SEXTON, T., YAFFE, E., KENIGSBERG, E., BANTIGNIES, F., LEBLANC, B., HOICHMAN, M., PARRINELLO, H., TANAY, A. et CAVALLI, G. (2012). Three-dimensional folding and functional organization principles of the Drosophila genome. *Cell*, 148(3):458–72. (Cité en pages 26 et 28.)
- SHELBY, R. D., HAHN, K. M. et SULLIVAN, K. F. (1996). Dynamic elastic behavior of alpha-satellite DNA domains visualized in situ in living human cells. *The Journal of cell biology*, 135(3):545–57. (Cité en page 142.)
- SHEPELEV, V. a., ALEXANDROV, A. a., YUROV, Y. B. et ALEXANDROV, I. a. (2009). The evolutionary origin of man can be traced in the layers of defunct ancestral alpha satellites flanking the active centromeres of human chromosomes. *PLoS genetics*, 5(9):e1000641. (Cité en page 61.)
- SHIELS, C., ISLAM, S. a., VATCHEVA, R., SASIENI, P., STERNBERG, M. J., FREEMONT, P. S. et SHEER, D. (2001). PML bodies associate specifically with the MHC gene cluster in interphase nuclei. *Journal of cell science*, 114(Pt 20):3705–16. (Cité en page 95.)
- SHIMI, T., PFLEGHAAR, K., KOJIMA, S.-i., PACK, C.-g., SOLOVEI, I., GOLDMAN, A. E., ADAM, S. A., SHUMAKER, D. K., KINJO, M., CREMER, T. et GOLDMAN, R. D. (2008). The A-and B-type nuclear lamin networks : microdomains involved in chromatin organization and transcription. *Genes & development*, pages 3409–3421. (Cité en pages 32, 33 et 142.)
- SHIRLEY, J. W., TY, S., TAKEBAYASHI, S.-i., LIU, X. et GILBERT, D. M. (2011). FISH Finder : a high-throughput tool for analyzing FISH images. *Bioinformatics (Oxford, England)*, 27(7):933–8. (Cité en page 108.)
- SINGH, P. B. et GEORGATOS, S. D. (2003). HP1 : facts, open questions, and speculation. *Journal of structural biology*, 140(1-3):10–6. (Cité en page 69.)
- SLEE, R. B., STEINER, C. M., HERBERT, B.-S., VANCE, G. H., HICKEY, R. J., SCHWARZ, T., CHRISTAN, S., RADOVICH, M., SCHNEIDER, B. P., SCHINDELHAUER, D. et GRIMES, B. R. (2012). Cancer-associated alteration of pericentromeric heterochromatin may contribute to chromosome instability. *Oncogene*, 31(27):3244–53. (Cité en page 61.)

- SMIGOVÁ, J., JUDA, P., CMARKO, D. et RAŠKA, I. (2011). Fine structure of the "PcG body" in human U-2 OS cells established by correlative light-electron microscopy. *Nucleus (Austin, Tex.)*, 2(3):219–28. (Cité en page 35.)
- SOILLE, P. (2003). *Morphological image analysis : principles and applications*. Springer-Verlag. (Cité en pages 89 et 115.)
- SOLOVEI, I. et CREMER, M. (2010). 3D-FISH on cultured cells combined with immunostaining. *Methods in molecular biology*, 659(Chapter 3):117–126. (Cité en pages 84 et 131.)
- SOLOVEI, I., KREYSING, M., LANCTÔT, C., KÖSEM, S., PEICHL, L., CREMER, T., GUCK, J. et JOFFE, B. (2009). Nuclear architecture of rod photoreceptor cells adapts to vision in mammalian evolution. *Cell*, 137(2):356–68. (Cité en pages 76 et 77.)
- SOLOVEI, I., SCHERMELLEH, L., DÜRING, K., ENGELHARDT, A., STEIN, S., CREMER, C. et CREMER, T. (2004). Differences in centromere positioning of cycling and postmitotic human cell types. *Chromosoma*, 112(8):410–23. (Cité en pages 65, 66, 71, 73, 117, 121, 124, 125, 131, 141, 142, 148, 159 et 165.)
- SOLOVEI, I., WANG, A. S. S., THANISCH, K., SCHMIDT, C. S. S., KREBS, S., ZWERTGER, M., COHEN, T. V. V., DEVYS, D., FOISNER, R., PEICHL, L., HERRMANN, H., BLUM, H., ENGELKAMP, D., STEWART, C. L. L., LEONHARDT, H. et JOFFE, B. (2013). LBR and Lamin A/C Sequentially Tether Peripheral Heterochromatin and Inversely Regulate Differentiation. *Cell*, 152(3):584–598. (Cité en pages 68, 74, 76, 77, 78, 142 et 165.)
- SOMECH, R., SHAKLAI, S., GELLER, O., AMARIGLIO, N., SIMON, A. J., REHAVI, G. et GAL-YAM, E. N. (2005). The nuclear-envelope protein and transcriptional repressor LAP2 beta interacts with HDAC3 at the nuclear periphery, and induces histone H4 deacetylation. *Journal of cell science*, 118(Pt 17):4017–25. (Cité en page 41.)
- SOUTOGLOU, E. et MISTELI, T. (2007). Mobility and immobility of chromatin in transcription and genome stability. *Current opinion in genetics & development*, 17(5):435–42. (Cité en page 46.)
- STEENSEL, B. V. et DEKKER, J. (2010). Genomics tools for unraveling chromosome architecture. *Nature biotechnology*, 28(10):1089–1095. (Cité en page 81.)
- STIERLÉ, V., COUPRIE, J., OSTLUND, C., KRIMM, I., ZINN-JUSTIN, S., HOSSENLOPP, P., WORMAN, H. J., COURVALIN, J.-C. et DUBAND-GOULET, I. (2003). The carboxyl-terminal region common to lamins A and C contains a DNA binding domain. *Biochemistry*, 42(17):4819–28. (Cité en pages 40 et 142.)
- STRISSEL, P. L. (1996). Scaffold attachment regions in centromere-associated DNA. *105(2):122–133*. (Cité en page 142.)

- SULLIVAN, B. a. et KARPEN, G. H. (2004). Centromeric chromatin exhibits a histone modification pattern that is distinct from both euchromatin and heterochromatin. *Nature structural & molecular biology*, 11(11):1076–83. (Cité en page 56.)
- SULLIVAN, L. L., BOIVIN, C. D., MRAVINAC, B., SONG, I. Y. et SULLIVAN, B. a. (2011). Genomic size of CENP-A domain is proportional to total alpha satellite array size at human centromeres and expands in cancer cells. *Chromosome research*, 19(4):457–70. (Cité en page 56.)
- TADDEI, A., MAISON, C., ROCHE et ALMOUZNI, G. (2001). Reversible disruption of pericentric heterochromatin and centromere function by inhibiting deacetylases. *Nature cell biology*, 3(2):114–20. (Cité en pages 76 et 79.)
- TADDEI, A., ROCHE, D., SIBARITA, J. B., TURNER, B. M. et ALMOUZNI, G. (1999). Duplication and maintenance of heterochromatin domains. *The Journal of cell biology*, 147(6):1153–66. (Cité en page 69.)
- TAIMEN, P., PFLEGHAAR, K., SHIMI, T., MÖLLER, D., BEN-HARUSH, K., ERDOS, M. R., ADAM, S. a., HERRMANN, H., MEDALIA, O., COLLINS, F. S., GOLDMAN, A. E. et GOLDMAN, R. D. (2009). A progeria mutation reveals functions for lamin A in nuclear assembly, architecture, and chromosome organization. *Proceedings of the National Academy of Sciences of the United States of America*, 106(49):20788–93. (Cité en page 80.)
- TANIURA, H., GLASS, C. et GERACE, L. (1995). A chromatin binding site in the tail domain of nuclear lamins that interacts with core histones. *The Journal of cell biology*, 131(1):33–44. (Cité en page 40.)
- TERRANOVA, R., SAUER, S., MERKENSCHLAGER, M. et FISHER, A. G. (2005). The reorganisation of constitutive heterochromatin in differentiating muscle requires HDAC activity. *Experimental cell research*, 310(2):344–56. (Cité en pages 74, 79 et 129.)
- THOMANN, D., RINES, D. R., SORGER, P. K. et DANUSER, G. (2002). Automatic fluorescent tag detection in 3D with super-resolution : application to the analysis of chromosome movement. *Journal of microscopy*, 208(Pt 1):49–64. (Cité en page 89.)
- THOMSON, I., GILCHRIST, S., BICKMORE, W. A. et CHUBB, J. (2004). The radial positioning of chromatin is not inherited through mitosis but is established de novo in early G1. *Current biology*, 14:166–172. (Cité en page 49.)
- TIWARI, V. K., COPE, L., MCGARVEY, K. M., OHM, J. E. et BAYLIN, S. B. (2008). A novel 6C assay uncovers Polycomb-mediated higher order chromatin conformations. *Genome research*, 18(7):1171–9. (Cité en page 42.)
- TOLSTONOG, G. V., WANG, X., SHOEMAN, R. et TRAUB, P. (2000). Intermediate filaments reconstituted from vimentin, desmin, and glial fibrillary acidic protein

- selectively bind repetitive and mobile DNA sequences from a mixture of mouse genomic DNA fragments. *DNA and cell biology*, 19(11):647–77. (Cité en page 142.)
- TOOMRE, D. et BEWERSDORF, J. (2010). A new wave of cellular imaging. *Annual review of cell and developmental biology*, 26:285–314. (Cité en page 84.)
- TOWBIN, B. D., GONZÁLEZ-AGUILERA, C., SACK, R., GAIDATZIS, D., KALCK, V., MEISTER, P., ASKJAER, P. et GASSER, S. M. (2012). Step-wise methylation of histone H3K9 positions heterochromatin at the nuclear periphery. *Cell*, 150(5):934–47. (Cité en page 41.)
- VADAKKAN, K. I., LI, B. et DE BONI, U. (2006). Trend towards varying combinatorial centromere association in morphologically identical clusters in Purkinje neurons. *Cell & chromosome*, 5(1):1. (Cité en page 69.)
- VAN HOOSER, A. A., OUSPENSKI, I. I., GREGSON, H. C., STARR, D. A., YEN, T. J., GOLDBERG, M. L., YOKOMORI, K., EARNSHAW, W. C., SULLIVAN, K. F. et BRINKLEY, B. R. (2001). Specification of kinetochore-forming chromatin by the histone H3 variant CENP-A. *Journal of cell science*, 114(Pt 19):3529–42. (Cité en page 61.)
- van STEENSEL, B. et DEKKER, J. (2010). Genomics tools for unraveling chromosome architecture. *Nature Biotechnology*, 28(10):1089–1095. (Cité en page 129.)
- van STEENSEL, B., DELROW, J. et HENIKOFF, S. (2001). Chromatin profiling using targeted DNA adenine methyltransferase. *Nature genetics*, 27(3):304–8. (Cité en page 82.)
- VESTER, B. et WENGEL, J. (2004). LNA (locked nucleic acid) : high-affinity targeting of complementary RNA and DNA. *Biochemistry*, 43(42). (Cité en page 59.)
- VIRÉ, E., BRENNER, C., DEPLUS, R., BLANCHON, L., FRAGA, M., DIDELOT, C., MOREY, L., VAN EYNDE, A., BERNARD, D., VANDERWINDEN, J.-M., BOLLEN, M., ESTELLER, M., DI CROCE, L., de LAUNOIT, Y. et FUKS, F. (2006). The Polycomb group protein EZH2 directly controls DNA methylation. *Nature*, 439(7078):871–4. (Cité en page 24.)
- VOGEL, B., LÖSCHBERGER, A., SAUER, M. et HOCK, R. (2011). Cross-linking of DNA through HMGA1 suggests a DNA scaffold. *Nucleic acids research*, 39(16):7124–33. (Cité en pages 78 et 142.)
- VOO, K. S., CARLONE, D. L., JACOBSEN, B. M. et SKALNIK, D. G. (2000). a mammalian transcriptional activator that binds unmethylated CpG motifs and shares a CXXC domain with DNA methyltransferase, human trithorax, and methyl-CpG. *Molecular and cellular biology and cellular biology*, 20(6):2108–21. (Cité en page 23.)

- WALTER, J., SCHERMELLEH, L., CREMER, M., TASHIRO, S. et CREMER, T. (2003). Chromosome order in HeLa cells changes during mitosis and early G1, but is stably maintained during subsequent interphase stages. *The Journal of cell biology*, 160(5):685–97. (Cité en pages 49, 142, 166 et 173.)
- WANG, J., SHIELS, C., SASIENI, P., WU, P. J., ISLAM, S. a., FREEMONT, P. S. et SHEER, D. (2004). Promyelocytic leukemia nuclear bodies associate with transcriptionally active genomic regions. *The Journal of cell biology*, 164(4):515–26. (Cité en page 44.)
- WATT, F. et MOLLOY, P. (1988). Cytosine methylation prevents binding to DNA of a HeLa cell transcription factor required for optimal expression of the adenovirus major late promoter. *Genes & development*, 2(9):1136–1143. (Cité en page 23.)
- WEIERICH, C., BRERO, A., STEIN, S., von HASE, J., CREMER, C., CREMER, T. et SOLOVEI, I. (2003). Three-dimensional arrangements of centromeres and telomeres in nuclei of human and murine lymphocytes. *Chromosome research*, 11(5):485–502. (Cité en pages 65, 66, 67, 130, 141 et 173.)
- WEIMER, R., HAAF, T., KRÜGER, J., POOT, M. et 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. *Human genetics*, 88(6):673–82. (Cité en pages 66 et 71.)
- WIBLIN, A. E., CUI, W., CLARK, a. J. et BICKMORE, W. a. (2005). Distinctive nuclear organisation of centromeres and regions involved in pluripotency in human embryonic stem cells. *Journal of cell science*, 118(Pt 17):3861–8. (Cité en pages 66, 129, 130 et 143.)
- WILSON, K. L. et FOISNER, R. (2010). Lamin-binding Proteins. *Cold Spring Harbor perspectives in biology*, 2(4):a000554. (Cité en page 33.)
- WONG, L. H., BRETtingham-MOORE, K. H., CHAN, L., QUACH, J. M., ANDERSON, M. a., NORTHROP, E. L., HANNAN, R., SAFFERY, R., SHAW, M. L., WILLIAMS, E., CHOO, K. H. A. et ACCESS, O. (2007). Centromere RNA is a key component for the assembly of nucleoproteins at the nucleolus and centromere. *Genome research*, 17(8):1146–60. (Cité en page 68.)
- WOODCOCK, C. L. (1994). Chromatin fibers observed in situ in frozen hydrated sections. Native fiber diameter is not correlated with nucleosome repeat length. *The Journal of cell biology*, 125(1):11–9. (Cité en page 25.)
- YE, Q. et WORMAN, H. J. (1996). Interaction between an Integral Protein of the Nuclear Envelope Inner Membrane and Human Chromodomain Proteins Homologous to Drosophila HP1. *Journal of Biological Chemistry*, 271(25):14653–14656. (Cité en page 40.)

- YUSUFZAI, T. M., TAGAMI, H., NAKATANI, Y. et FELSENFELD, G. (2004). CTCF tethers an insulator to subnuclear sites, suggesting shared insulator mechanisms across species. *Molecular cell*, 13(2):291–8. (Cité en page 42.)
- ZEITLIN, S. G., SHELBY, R. D. et SULLIVAN, K. F. (2001). CENP-A is phosphorylated by Aurora B kinase and plays an unexpected role in completion of cytokinesis. *The Journal of cell biology*, 155(7):1147–57. (Cité en page 56.)
- ZHANG, L.-F., HUYNH, K. D. et LEE, J. T. (2007). Perinucleolar targeting of the inactive X during S phase : evidence for a role in the maintenance of silencing. *Cell*, 129(4):693–706. (Cité en pages 42, 68 et 143.)
- ZHAO, Z., TAVOOSIDANA, G., SJÖLINDER, M., GÖNDÖR, A., MARIANO, P., WANG, S., KANDURI, C., LEZCANO, M., SANDHU, K. S., SINGH, U., PANT, V., TIWARI, V., KURUKUTI, S. et OHLSSON, R. (2006). Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. *Nature genetics*, 38(11):1341–7. (Cité en page 141.)
- ZINK, D., AMARAL, M. D., ENGLMANN, A., LANG, S., CLARKE, L. a., RUDOLPH, C., ALT, F., LUTHER, K., BRAZ, C., SADONI, N., ROSENECKER, J. et SCHINDELHAUER, D. (2004). Transcription-dependent spatial arrangements of CFTR and adjacent genes in human cell nuclei. *The Journal of cell biology*, 166(6):815–25. (Cité en pages 39 et 45.)
- ZINNER, R., ALBIEZ, H., WALTER, J., PETERS, A. H. F. M., CREMER, T. et CREMER, M. (2006). Histone lysine methylation patterns in human cell types are arranged in distinct three-dimensional nuclear zones. *Histochemistry and cell biology*, 125(1-2):3–19. (Cité en pages 71, 72, 148 et 164.)
- ZORN, C., CREMER, C., CREMER, T. et ZIMMER, J. (1979). Unscheduled DNA synthesis after partial UV irradiation of the cell nucleus : distribution in interphase and metaphase. *Experimental cell research*, 124:111–119. (Cité en pages 28 et 29.)
- ZULEGER, N., BOYLE, S., KELLY, D. a., DE LAS HERAS, J. I., LAZOU, V., KORFALI, N., BATRAKOU, D. G., RANGLES, K. N., MORRIS, G. E., HARRISON, D. J., BICKMORE, W. a. et SCHIRMER, E. C. (2013). Specific nuclear envelope transmembrane proteins can promote the location of chromosomes to and from the nuclear periphery. *Genome biology*, 14(2):R14. (Cité en pages 38 et 143.)
- ZULEGER, N., ROBSON, M. et SCHIRMER, E. (2011). The nuclear envelope as a chromatin organizer. *Nucleus*, (October):339–349. (Cité en page 38.)
- ZULLO, J. M., DEMARCO, I. a., PIQUÉ-REGI, R., GAFFNEY, D. J., EPSTEIN, C. B., SPOONER, C. J., LUPERCHIO, T. R., BERNSTEIN, B. E., PRITCHARD, J. K., REDDY, K. L. et SINGH, H. (2012). DNA sequence-dependent compartmentalization and silencing of chromatin at the nuclear lamina. *Cell*, 149(7):1474–87. (Cité en page 40.)

