



**HAL**  
open science

# Etude de la transcription des séquences satellites du génomme humain

Angéline Eymery

► **To cite this version:**

Angéline Eymery. Etude de la transcription des séquences satellites du génome humain. Biologie cellulaire. Université Joseph-Fourier - Grenoble I, 2008. Français. NNT: . tel-00368271

**HAL Id: tel-00368271**

**<https://theses.hal.science/tel-00368271>**

Submitted on 15 Mar 2009

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Université Joseph Fourier- Grenoble 1  
Chimie et Sciences du Vivant  
(Arrêtés ministériels du 5 juillet 1984 et du 30 mars 1992)

# **THESE**

Pour obtenir le titre de

**Docteur de l'Université Joseph Fourier**  
**Discipline Biologie**

Présentée et soutenue publiquement par

**Angéline EYMERY**

Le 17 Décembre 2008

## **Etude de la transcription des séquences satellites du génome humain**

### **Jury**

Pr Stefan Nonchev (Président du jury)  
Dr Pierre-Antoine Defossez (Rapporteur)  
Dr Patrick Lomonte (Rapporteur)  
Dr Marie-Geneviève Mattei (Examinatrice)  
Dr Chantal Vaury (Examinatrice)  
Pr Claire Vourc'h (Directrice de thèse)

Thèse préparée au sein du laboratoire Dyogen (**DY**namique de l'**O**rganisation du **GEN**ome)  
Institut Albert Bonniot  
Centre de recherche Inserm U823 - Université Joseph Fourier Grenoble 1

## Remerciements...

*Je remercie le Professeur Stefan Nonchev d'avoir accepté la présidence de ma thèse. C'est toi qui as été mon premier professeur de biologie cellulaire à l'université et cette présidence est donc plus qu'évidente. Merci également pour ton soutien tout au long de mon parcours universitaire.*

*Je remercie les Dr Pierre-Antoine Defossez et Patrick Lomonte d'avoir accepté d'être les rapporteurs de ma thèse. Je remercie également les Dr Marie-Geneviève Mattei et Chantal Vauray d'avoir accepté d'évaluer ce travail.*

*J'exprime toute ma reconnaissance à Claire Vourc'h qui m'a accueillie dans son laboratoire, il y a maintenant plus de cinq ans lors de mon premier stage. Depuis, tu m'as permis de m'épanouir sur ce sujet que nous affectionnons tant. Merci pour la liberté et la confiance que tu m'as accordée pendant cette thèse.*

*Je remercie également les membres de l'équipe, les « Dyogen Girls » Super Cathy, Caro, Sab, Edwige, Véro, Elisabeth et la petite dernière, Lydia.*

*Un grand merci à tous mes collaborateurs sans qui cette thèse ne serait pas ce qu'elle est. Merci à Saadi Khochbin et Eric Gilson, Sophie Rousseaux et Mary Callanan sans oublier Michelle El Atifi, pour votre aide, votre patience et votre bonne humeur. Merci également à la Fondation pour la Recherche Médicale et à la Région Rhône Alpes qui ont financé ces travaux de thèse, respectivement pour un et trois ans.*

*Je remercie mes collègues et amis de l'Institut Albert Bonniot : ceux du début, Charrette et Eric, et ceux de la fin, Seb, « Faycal », Junior, Johnny, Aurélia, Leila, Alexandra et Sevgi. « Last but not least » un grand merci à Sophie, ma « sœur jumelle », pour son soutien moral et physique.*

*Enfin, Merci à ceux qui de près ou de loin ont favorisé le bon déroulement de cette thèse et en particulier Hélène et Dr Manu.*

# Table des matières

<b>Table des matières.....</b>	<b>- 3 -</b>
<b>Avant propos.....</b>	<b>- 7 -</b>
<b>Introduction.....</b>	<b>- 8 -</b>
<b>Chapitre I Génétique &amp; Epigénétique.....</b>	<b>- 9 -</b>
I.1 Génétique.....	- 9 -
I.2 Epigénétique .....	- 11 -
I.2.A La méthylation de l'ADN.....	- 13 -
1. Définition.....	- 13 -
2. Les ADN méthyltransférases .....	- 14 -
a) La famille DNMT1 .....	- 15 -
b) La famille DNMT3 .....	- 15 -
c) La famille DNMT2.....	- 17 -
3. Fonctions de la méthylation de l'ADN .....	- 18 -
I.2.B Le nucléosome.....	- 19 -
1. Les modifications d'histones .....	- 20 -
a) L'acétylation .....	- 21 -
b) La méthylation.....	- 23 -
c) Fonctions des modifications d'histones .....	- 25 -
2. Les variants d'histones .....	- 27 -
a) H2A .....	- 27 -
b) H2B .....	- 28 -
c) H3.....	- 29 -
3. Les complexes de remodelage .....	- 29 -
a) Les différentes familles .....	- 30 -
b) Régulation des complexes de remodelage .....	- 31 -
I.2.C Les ARN non codants .....	- 32 -
1. Xist et l'inactivation du chromosome X.....	- 32 -
2. L'empreinte parentale.....	- 35 -
3. L'interférence par ARN.....	- 36 -
a) PTGS.....	- 37 -
b) TGS.....	- 37 -
<b>Chapitre II Centromères &amp; Péracentromères.....</b>	<b>- 40 -</b>
II.1 Euchromatine et hétérochromatine .....	- 42 -
II.1.A Les différences.....	- 42 -

1.	Caractéristiques épigénétiques .....	- 42 -
2.	Transcription .....	- 44 -
II.1.B	Hétérochromatine et génomes .....	- 45 -
1.	Comparaison des génomes .....	- 46 -
2.	Les différentes classes d'ADNnc du génome humain.....	- 47 -
a)	Les séquences répétées en tandem.....	- 48 -
b)	Les éléments répétés dispersés.....	- 52 -
II.2	Centromères et péricentromères .....	- 55 -
II.2.A	Les centromères .....	- 56 -
1.	Les séquences génomiques.....	- 56 -
2.	Les modifications épigénétiques.....	- 57 -
3.	Les protéines centromériques.....	- 59 -
II.2.B	Les péricentromères.....	- 60 -
1.	Les séquences génomiques.....	- 61 -
2.	Epigénétique et HP1.....	- 62 -
3.	Cohésion des chromatides sœurs .....	- 64 -
II.2.C	Les CT et PCT dans différentes espèces.....	- 66 -
1.	Saccharomyces cerevisiae .....	- 66 -
a)	Le centromère .....	- 66 -
b)	Le péricentromère .....	- 67 -
2.	Saccharomyces pombe.....	- 68 -
a)	Le centromère .....	- 68 -
b)	Le péricentromère .....	- 69 -
3.	Drosophila melanogaster .....	- 70 -
a)	Le centromère .....	- 71 -
b)	Le péricentromère .....	- 72 -
4.	Mus musculus .....	- 73 -
a)	Le centromère .....	- 73 -
b)	Le péricentromère.....	- 74 -
5.	L'Homme.....	- 74 -
a)	Le centromère .....	- 74 -
b)	Le péricentromère.....	- 75 -

**Chapitre III Hétérochromatine & Transcription .....- 78 -**

III.1	Hétérochromatine et ARN non codant .....	- 78 -
III.1.A	L'hétérochromatine péricentromérique de <i>S.pombe</i> .....	- 79 -
III.1.B	L'hétérochromatine péricentromérique murine .....	- 81 -
III.2	Transcription des séquences centromériques et péricentromériques chez les mammifères....	- 82 -
III.2.A	Conditions d'expression des séquences répétées.....	- 82 -
1.	La réponse cellulaire au stress .....	- 83 -
2.	Développement et différenciation cellulaire.....	- 84 -
3.	Cycle cellulaire.....	- 86 -
4.	Sénescence.....	- 87 -

III.2.B Mécanismes impliqués dans la transcription des CT et des PCT .....	- 88 -
1. Facteurs de transcription spécifiques.....	- 88 -
2. Remodelage de la chromatine .....	- 89 -
a) Méthylation de l'ADN .....	- 89 -
b) Modifications d'histones .....	- 90 -
3. Régulation post-transcriptionnelle.....	- 92 -
<b>Résultats .....</b>	<b>- 94 -</b>
<b>OBJECTIFS.....</b>	<b>- 95 -</b>
<b>Chapitre I Identification de nouvelles cibles chromosomiques d'HSF1 .....</b>	<b>- 97 -</b>
I.1 Introduction : la réponse cellulaire au stress .....	- 97 -
I.1.A Le facteur de transcription HSF1 .....	- 98 -
1. Structure de la protéine .....	- 98 -
2. Régulation de l'activation d'HSF1.....	- 99 -
I.1.B Les granules de stress.....	- 101 -
I.2 Matériel et Méthodes .....	- 103 -
1. Culture cellulaire et choc thermique.....	- 103 -
2. Immunofluorescence .....	- 104 -
3. Hybridation in situ fluorescente pour la détection d'ARN et d'ADN (FISH ARN et FISH ADN) -	105 -
a) FISH ADN.....	- 105 -
b) FISH ARN avec oligonucléotide.....	- 106 -
4. Reconstitution de granules sur chromosomes métaphasiques.....	- 106 -
a) Préparation de chromosomes métaphasiques .....	- 106 -
b) Reconstitution de granules.....	- 107 -
5. Western blot .....	- 107 -
a) Extraction de protéines.....	- 107 -
b) Western Blot.....	- 107 -
6. Gel retard .....	- 108 -
7. Microscopie et analyse d'image.....	- 108 -
I.3 Résultats.....	- 109 -
I.3.A Identification de granules de stress secondaires.....	- 109 -
I.3.B HSF1 se lie aux satellites 2 et satellites 3.....	- 113 -
I.3.C La formation des granules de stress secondaires est dépendante de la quantité d'HSF1.....	- 116 -
I.3.D Les granules de stress secondaires sont des sites de transcription.....	- 118 -
I.4 Conclusion .....	- 120 -
<b>Chapitre II Transcription des séquences satellites du génome humain .....</b>	<b>- 123 -</b>
II.1 Introduction .....	- 123 -
II.2 Matériel et Méthodes .....	- 125 -
1. Culture cellulaire et traitements .....	- 125 -
2. Extractions d'ARN.....	- 126 -
3. RepChip .....	- 126 -

4.	Immunoprécipitation d'ADN méthylé (MeDIP) .....	- 128 -
5.	Immunofluorescence .....	- 128 -
6.	Western blot .....	- 128 -
7.	RT PCR .....	- 128 -
II.3	Résultats .....	- 129 -
1.	Conception des oligonucléotides .....	- 129 -
2.	La réponse au stress thermique: un modèle de validation pour la RepChip .....	- 132 -
3.	Impact du remodelage de la chromatine sur la transcription des séquences satellites.....	- 139 -
4.	Impact de Dicer sur l'expression des séquences satellites.....	- 143 -
5.	Expression des séquences satellites in vivo, dans des tissus normaux et tumoraux.....	- 145 -
II.4	Conclusion .....	- 150 -
II.5	Résultats complémentaires .....	- 153 -
1.	Rep'Array.....	- 153 -
2.	Irradiation et différenciation cellulaire .....	- 154 -
a)	Irradiation .....	- 154 -
b)	Différenciation cellulaire .....	- 155 -
3.	Etude de l'expression des séquences répétées au cours d'un stress physiopathologique ....	- 155 -
II.6	Perspectives .....	- 160 -
<b>Chapitre III Le syndrome ICF .....</b>		<b>- 162 -</b>
III.1	Introduction .....	- 162 -
III.2	Matériel et méthodes .....	- 163 -
1.	Culture cellulaire et choc thermique.....	- 163 -
2.	Immunofluorescence, et hybridation in situ fluorescente .....	- 164 -
3.	Western Blot .....	- 164 -
4.	Reconstitution de granules sur chromosomes métaphasiques.....	- 164 -
5.	RepChip .....	- 165 -
III.3	Résultats .....	- 165 -
1.	Absence de granule de stress dans les cellules ICF .....	- 165 -
2.	La réponse au stress est conservée.....	- 167 -
3.	La déméthylation de l'ADN n'empêche pas la formation des granules de stress .....	- 168 -
III.4	Discussion .....	- 171 -
III.5	Perspectives .....	- 173 -
<b>Conclusion .....</b>		<b>- 174 -</b>
<b>Références .....</b>		<b>- 179 -</b>

# Avant propos

Dans nos cellules, le support de l'information génétique n'est autre que l'ADN. Selon le Dogme Central de la génétique, l'ADN est transcrit en ARN qui est lui même traduit en protéine. Les protéines assurent des fonctions indispensables au développement et à la survie des organismes. Ces parties d'ADN codant pour des protéines, communément appelées gènes, ne représentent qu'une minuscule portion du génome humain : à peine 2% ! Les 98% restant du génome ont souvent été qualifiés de « déchets génomiques » à cause de l'absence apparente d'utilité de cet ADN non codant (absence de transcription, donc de traduction et de protéine) et du caractère souvent répétitif de sa séquence en base (contrairement à la séquence unique et originale des gènes).

De l'ADN inutile parasiterait ainsi notre ADN? Peut être pas...

L'objectif de ma thèse a été d'évaluer le potentiel transcriptionnel de séquences particulières de l'ADN non codant du génome humain : les séquences satellites centromériques et péricentromériques. J'ai également cherché à mettre en évidence les mécanismes impliqués dans cette transcription et plus particulièrement à déterminer l'impact de modifications épigénétiques sur celle-ci.

Qu'est ce que l'épigénétique ? Quelles sont les caractéristiques de cet ADN non codant ? Que sait-on à l'heure actuelle de la transcription de ces séquences et de leurs éventuelles fonctions? C'est ce que nous allons découvrir dans le chapitre introductif. Les résultats que j'ai obtenus au cours de ma thèse seront présentés dans un second chapitre, sous la forme de trois articles en cours de préparation. En conclusion, je présenterai brièvement les perspectives concernant les potentielles fonctions des ARN issus de l'ADN non codant.



# Introduction

Chapitre I	Génétique & Epigénétique
Chapitre II	Centromères & Péricentromères
Chapitre III	Hétérochromatine & Transcription

# Chapitre I Génétique & Epigénétique

Le terme « gène » est devenu un concept majeur en biologie. Depuis l'apparition du mot, au début du XX<sup>ème</sup> siècle, sa définition n'a cessé d'évoluer au rythme de l'accroissement des connaissances. A travers un court historique, nous allons traverser le siècle précédent, à la découverte de la génétique avant d'aborder l'épigénétique dans une seconde partie.

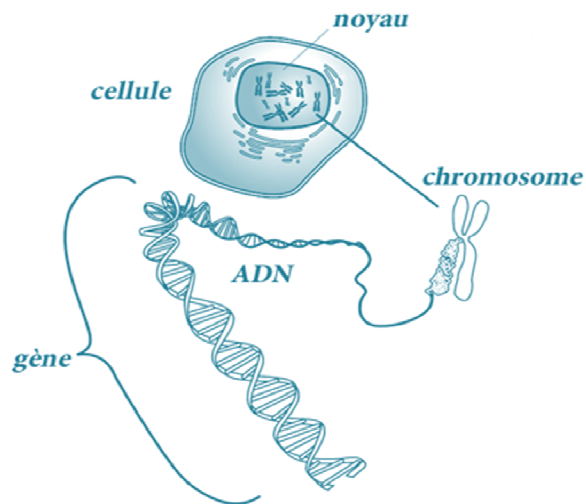
## I.1 Génétique

L'origine étymologique du terme « génétique » est genno, « donner naissance » en grec. Les débuts de la génétique remontent à la dernière moitié du XIX<sup>ème</sup> siècle, sous l'impulsion des 2 pionniers que sont Charles Darwin et Gregor Mendel.

L'hypothèse de la pangenèse, proposée par Darwin en 1859, repose sur l'idée selon laquelle chaque cellule peut produire un petit germe, appelé gemmule, qui représente le support de l'hérédité. Mendel montre quant à lui en 1866 que les caractères héréditaires sont gouvernés par une double commande et qu'une seule des deux est transmise aux descendants par chaque parent. Quelques années plus tard, en 1879, Strasburger et Flemming en étudiant la division des cellules végétales découvrent des filaments présentant une grande affinité pour des colorants basiques : les chromosomes. Dès le début du XX<sup>ème</sup> siècle, les chromosomes sont décrits comme les supports physiques de l'hérédité. En effet, en 1902 Sutton montre que le nombre de chromosomes est réduit de moitié au cours de la méiose afin que suite à la fécondation le zygote (la cellule œuf) soit constitué d'un nombre correct de chromosomes.

C'est en 1909 qu'est introduite, par Wilhelm Johannsen, la notion de gène. C'est un dérivé de pangene, nom donné par Hugo de Vries pour les entités impliquées dans la pangenèse de Darwin, et qui signifie étymologiquement « l'origine de tout » (en grec, pan=tout ; genos=origine). En 1910, grâce aux travaux de Morgan portant sur la ségrégation des mutations chez la drosophile, le gène est alors décrit comme un fragment de chromosome, localisé sur un endroit précis de celui-ci, un locus. En 1941, les travaux de Beadle et Tatum chez *Neurospora*, montrent que des mutations des

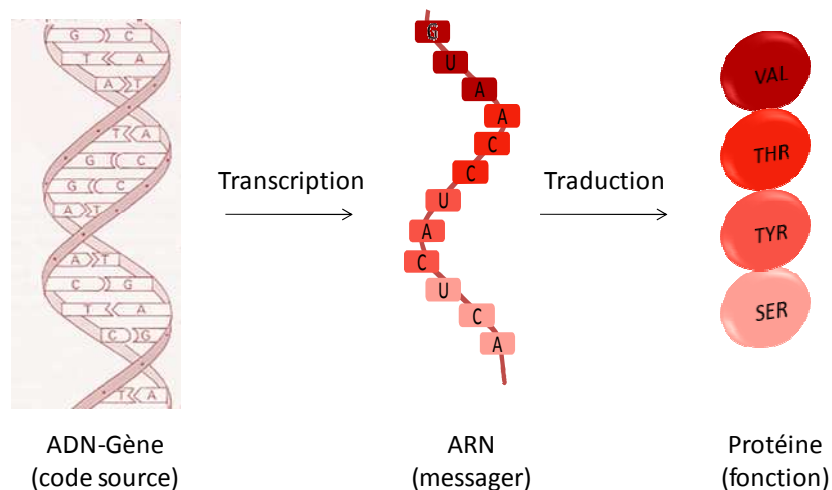
gènes induites par des rayons X sont associées à des déficiences dans des voies métaboliques mettant en cause des protéines spécifiques appelées enzymes. Ainsi les gènes, portés par les chromosomes, dirigent la synthèse des protéines qui contrôlent des fonctions métaboliques.



**Figure 1 : Le gène.** Le noyau des cellules contient des chromosomes. Chaque chromosome est constitué d'ADN. Un gène est un fragment de l'ADN, localisé à un endroit précis, un locus, sur le chromosome.

Il faut attendre 1944 pour qu'entre en scène l'ADN (Acide Désoxyribonucléique) (Figure 1). Avery, MacLeod et McCarty montrent en effet que l'ADN, connu pour être ubiquitaire dans tous les organismes, est la base chimique de la transmission d'information chez les bactéries. L'intérêt pour l'ADN s'accroît alors et c'est en 1953 que Watson et Crick proposent un modèle d'appariement des bases pour la réplication de la double hélice d'ADN. A la fin des années 50 il apparaît clairement que les gènes sont nécessaires à la production de protéines. Cependant il reste encore à savoir comment l'information contenue dans l'ADN peut être traduite en protéines. Le lien manquant est établi notamment par Jacob et Monod à travers des séries d'expériences d'invasion de bactéries par des bactériophages. Ces expériences mettent en évidence une nouvelle molécule, l'ARN (Acide Ribonucléique) dont la composition ressemble très fortement à l'ADN. L'intermédiaire entre l'ADN et les protéines est ainsi un ARN qu'ils nomment « ARN messager ». Finalement, Crick résume les données accumulées et propose alors en 1958 le « Dogme Central » selon lequel un gène est transcrit en un messager, l'ARN, qui est lui-même traduit en protéine (Figure 2). Dans ce modèle les fonctions cellulaires sont assurées par les protéines et l'ARN n'est qu'un intermédiaire. Mais des exceptions font alors leur apparition: certains gènes ne codent pas pour des protéines mais pour des ARN fonctionnels tels que les ARNr ou les ARNt.

Ainsi, dans la première moitié du XX<sup>ème</sup> siècle, la génétique subit un formidable essor et semble apporter de nombreux éléments de réponse quant à la grande énigme de l'hérédité. Cependant, les généticiens de l'époque se posent encore quelques questions. Ils cherchent en effet à comprendre pourquoi les variations génétiques sont très rarement corrélées à des changements phénotypiques (c'est-à-dire à des modifications de caractères observables telles que la couleur des yeux) et surtout pourquoi des différences phénotypiques ne sont pas forcément liées à des modifications génétiques. Une nouvelle discipline apparaît alors : l'épigénétique.



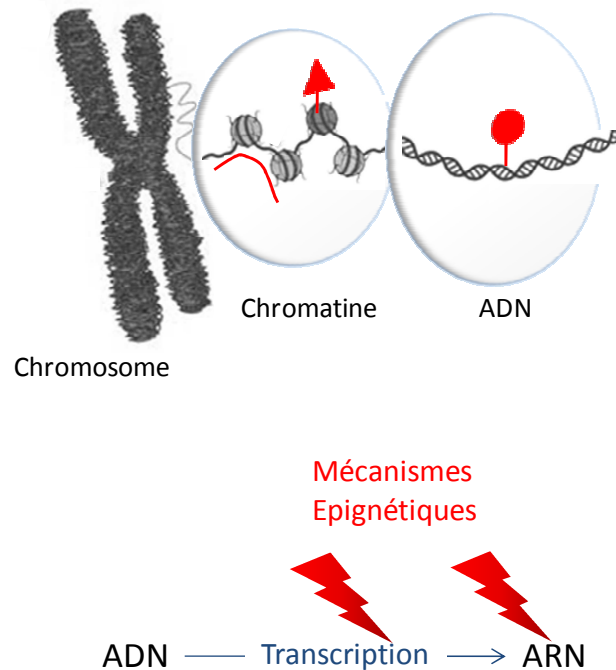
**Figure 2 : Le Dogme Central : un gène donne une protéine qui assure une fonction.** L'ADN est composé de 4 bases azotées (A, T, C, G) qui s'apparient entre elles pour former la double hélice. Le gène, qui est un fragment d'ADN, va être transcrit pour former l'ARN. Chaque triplet de bases (codon) de l'ARN correspond à un acide aminé. La traduction de l'ARN aboutit à la production d'une chaîne d'acides aminés appelée protéine. C'est cette protéine qui permet d'assurer une fonction particulière et l'ARN n'est qu'un simple intermédiaire.

## I.2 Epigénétique

*« L'épigénétique a toujours été l'ensemble de ces choses bizarres et merveilleuses que la génétique ne sait pas expliquer. » Denise Barlow*

En 1892, Weissmann suggère que la spécialisation des cellules conduit à la perte de certains éléments. Une décennie plus tard, Spemann élabore l'hypothèse selon laquelle ces fragments d'informations ne sont pas perdus mais simplement mis sous silence... Il restait encore à savoir comment ! L'épigénétique allait apporter une solution.

Épigénétique est un terme introduit par Conrad Waddington en 1942 afin de nommer la branche de la biologie qui étudie les relations de cause à effet entre les gènes et leurs produits, conduisant à l'apparition des phénotypes. Littéralement, le terme épigénétique, par l'utilisation du préfixe « épi », sous entend que la génétique n'explique pas tout et qu'il est donc nécessaire de chercher des explications au-delà, en addition du code génétique.



**Figure 3 : Les mécanismes épigénétiques.** Dans le noyau, l'ADN n'est pas nu. Il s'enroule autour de protéines, appelées histones, qui s'assemblent pour former le nucléosome (bille). Cette association entre les nucléosomes et l'ADN est nommée chromatine. La chromatine est le siège de modifications dites épigénétiques, car elles influencent la transcription des gènes, sans modifier le code génétique de l'ADN. Il existe trois acteurs permettant la mise en place de modifications épigénétiques : la méthylation de l'ADN (rond rouge), les modifications du nucléosome (triangle rouge), et les ARN non codants (trait rouge).

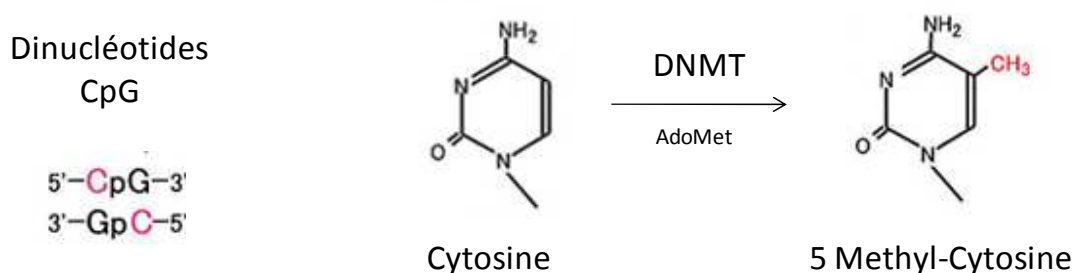
En raison de la découverte du code génétique qui suscite alors un grand intérêt, le terme épigénétique a été très peu utilisé au cours des 30 années suivant sa définition par Waddington. Cependant, de nos jours l'épigénétique est incontournable. Elle est définie comme l'ensemble des modifications de l'expression des gènes, transmises au cours des divisions cellulaires, mais n'impliquant pas de changement dans la séquence propre de l'ADN. Ainsi, l'ensemble de ces modifications épigénétiques est fréquemment regroupé sous le terme d'« épigénome » par homologie avec le terme maintenant bien connu de génome. Dans les cellules, l'ADN est enroulé autour d'un complexe protéique, le nucléosome qui est composée de 8

protéines (les histones). L'ensemble composé par les nucléosomes et l'ADN est appelé chromatine. Les mécanismes épigénétiques peuvent activer ou réprimer l'expression des gènes en agissant au niveau post-transcriptionnel, sur l'ARN, ou au niveau transcriptionnel, sur la chromatine (Figure 3). Pour cela, trois grands mécanismes sont mis en jeu : la méthylation de l'ADN, les modifications du nucléosome et les mécanismes épigénétiques liés aux ARN non codants (Figure 3). Ces trois mécanismes, sont décrits en détail dans la suite de ce chapitre. Ils agissent généralement de concert pour permettre un contrôle drastique de l'expression des gènes.

## I.2.A La méthylation de l'ADN

### 1. Définition

L'une des modifications épigénétiques les plus connues est la méthylation de l'ADN. La méthylation de l'ADN consiste en l'ajout d'un groupement méthyle ( $\text{CH}_3$ ) sur l'une des bases de l'ADN, la cytosine. Cela ne modifie en aucun cas le code génétique. Chez les mammifères, la méthylation de l'ADN a lieu au sein de dinucléotides cytosine-guanine (CpG), sur le carbone 5. La nouvelle base ainsi modifiée est appelée 5 méthyle-cytosine (5mC). La méthylation de l'ADN est orchestrée par des enzymes spécialisées, les ADN méthyltransférases souvent abrégées en DNMT (DNA methyltransferase). Elles utilisent la S-adénostyle-L-méthionine (AdoMet) comme donneur de groupement méthyle (Figure 4).



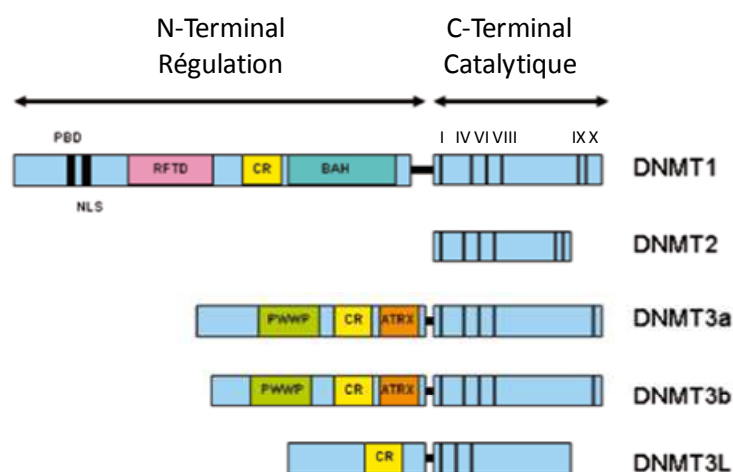
**Figure 4 : La méthylation de l'ADN.** La méthylation de l'ADN est définie par l'ajout d'un groupement méthyle ( $\text{CH}_3$ ) sur une cytosine, au sein d'un dinucléotide CpG. La réaction est catalysée par les DNMT en présence de l'AdoMet.

La méthylation de l'ADN n'est pas un mécanisme conservé parmi les eucaryotes. Ainsi, *Saccharomyces cerevisiae*, *Saccharomyces pombe* ainsi que *Caenorhabditis elegans* sont dépourvus d'une telle modification (Suzuki and Bird 2008).

La méthylation de l'ADN est un processus important, en particulier au cours du développement embryonnaire où de grandes vagues de méthylation et de déméthylation participent à la mise de place de l'empreinte parentale, thématique évoquée plus tard dans ce chapitre. De plus, chez les souris déficientes en DNMT, une létalité précoce des embryons est observée. Même si cette absence d'enzymes spécialisées est associée à une diminution de la méthylation de l'ADN, les causes de cette mortalité sont encore inconnues (Li, Bestor et al. 1992; Okano, Bell et al. 1999).

## 2. Les ADN méthyltransférases

Il existe 3 grandes familles de DNMT chez les mammifères : DNMT1, DNMT2 et DNMT3. Elles présentent toutes une structure similaire avec deux domaines distincts : une partie régulatrice variable (dans l'extrémité N-terminale) responsable de leur localisation nucléaire et contenant des sites d'interactions protéine-protéine, et une partie catalytique plus conservée (dans l'extrémité C-terminale) contenant le site catalytique de l'enzyme ainsi qu'un domaine de fixation à l'ADN (Bestor 2000) (Figure 5).



**Figure 5 : Les différentes DNMT (d'après (Siedlecki and Zielenkiewicz 2006)).** Le domaine N-terminal, qui varie selon les familles, est un domaine de régulation. Il peut ainsi contenir un site de liaison à PCNA (PBD), un signal de localisation nucléaire (NLS), un domaine riche en cystéines (CR), un domaine de ciblage vers les foyers de réplication (RFTD), un domaine BAH (bromo-adjacent homology) et PWWP impliqués dans des interactions protéiques et enfin un domaine ATRX responsable des interactions avec des HDAC (Histones Deacetylase). Le domaine C-terminal est un domaine catalytique, conservé entre les différentes familles. Les domaines les plus conservés (I, IV, VI, VIII, XI et X) sont représentés par des traits noirs : I est important pour la liaison du cofacteur, IV pour l'initiation du transfert du groupe méthyle, VI, VIII, IX et X pour la liaison au substrat.

#### a) La famille DNMT1

La DNMT1 est la première DNMT à avoir été isolée chez les mammifères (Bestor 1988). DNMT1 est décrite comme une DNMT de maintenance, c'est-à-dire qu'elle est capable de méthyle un ADN préalablement méthylé sur un brin (hémi-méthylé) (Figure 6). DNMT1 est également capable d'exercer une activité méthyltransférase de novo, sur de l'ADN non méthylé, mais sa préférence pour de l'ADN hémi-méthylé est 5 à 30 fois supérieure à celle pour de l'ADN non méthylé (Bacolla, Pradhan et al. 1999; Pradhan, Bacolla et al. 1999). En phase G1, elle est distribuée de manière uniforme dans le noyau. Au cours de la phase S, elle se relocalise dans des foyers correspondant à des sites majeurs de la réplication de l'ADN (Leonhardt, Page et al. 1992; Liu, Oakeley et al. 1998). Les souris DNMT1<sup>+/-</sup> sont viables et présentent une diminution globale, mais de faible amplitude, de la méthylation du génome. Par contre les embryons DNMT1<sup>-/-</sup>, dont le génome est fortement déméthylé, ne sont pas viables.

Il existe plusieurs isoformes de DNMT1 dont certaines sont spécifiques à certains tissus:

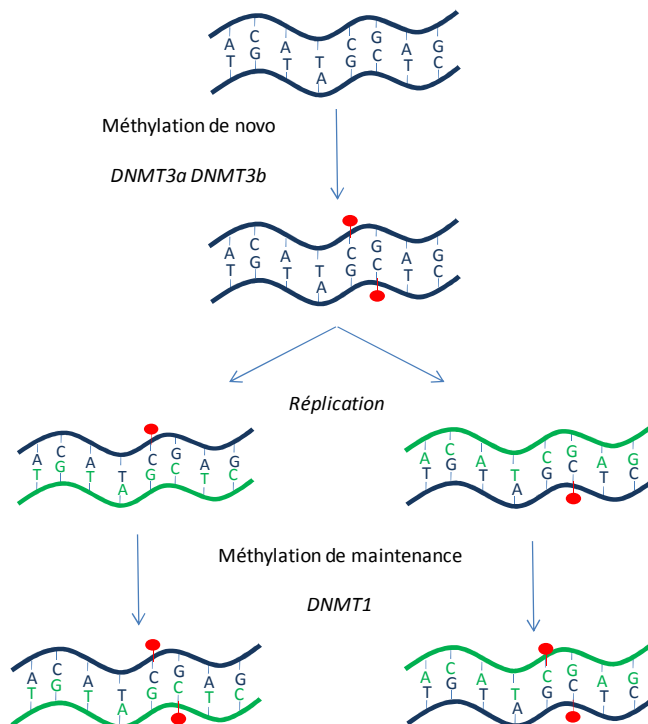
- DNMT1<sub>o</sub> est une forme tronquée de DNMT1 et est exprimée uniquement dans l'oocyte (Mertineit, Yoder et al. 1998). DNMT1<sub>o</sub> présente une localisation cytoplasmique depuis la période préimplantatoire de l'embryon jusqu'au stade 8 cellules où elle devient nucléaire afin de maintenir la méthylation des gènes particuliers, soumis à l'empreinte parentale (l'empreinte parentale est une thématique qui est abordée dans la suite de ce chapitre). En effet des délétions de DNMT1<sub>o</sub> résultent en l'abolition de la méthylation sur ces gènes (Howell, Bestor et al. 2001).

- L'ARN Dnmt1<sub>p</sub>, qui n'est pas traduit, est exprimé dans les spermatocytes au stade pachytène. Il joue un rôle important en interférant avec la machinerie de traduction et en empêchant la synthèse de DNMT1 qui est donc totalement absente à ce stade (Mertineit, Yoder et al. 1998).

#### b) La famille DNMT3

Chez les mammifères, deux DNMT3 fonctionnelles ont été identifiées par recherche d'homologie de séquences pour le domaine ADN méthyltransférase dans des banques EST (Expressed Sequence Tags), DNMT3a et DNMT3b (Okano, Xie et al. 1998). DNMT3L appartient aussi à cette famille sans toutefois posséder d'activité enzymatique.





**Figure 6 : La méthylation de l'ADN : de novo ou de maintenance.** L'ADN peut être méthylé de novo (rond rouge), c'est-à-dire en absence de méthylation de l'ADN préexistante, par DNMT3a et DNMT3b. Au cours de la réplication (vert) la méthylation de maintenance sur un ADN hémi-méthylé se fait grâce à DNMT1.

- DNMT3a et DNMT3b

DNMT3a et DNMT3b sont particulièrement importantes pour la méthylation *de novo* (Figure 6) au cours de l'embryogenèse. Leur activité est ensuite réduite au cours de la différenciation (Okano, Bell et al. 1999; Robertson, Uzvolgyi et al. 1999). Dans les tissus adultes, l'expression de DNMT3a est ubiquitaire alors que DNMT3b est exprimée à un très faible taux à l'exception du testicule, de la thyroïde et de la moelle osseuse (Xie, Wang et al. 1999).

Le gène DNMT3b humain produit cinq isoformes par épissage alternatif : DNMT3B 1 à 5 (Okano, Xie et al. 1998; Robertson, Uzvolgyi et al. 1999). DNMT3b1 est l'ARN le plus long et est considéré comme le principal produit du gène DNMT3b. DNMT3b2 possède une activité ADN méthyltransférase *in vitro* contrairement à DNMT3b3. DNMT3b4 et DNMT3b5 sont exprimées essentiellement dans le testicule. Ces deux isoformes ne possèdent pas d'activité catalytique mais les différents domaines d'interactions protéine-protéine situés en N-ter seraient fonctionnels. Leur rôle dans la spermatogenèse, s'il y en a un, est encore méconnu. Cependant, elles

pourraient être en compétition avec DNMT3b1-2 pour certains sites de liaisons de protéines cibles (Robertson 2002).

Des mutations produisant des protéines non fonctionnelles ont été réalisées chez la souris. Les souris déficientes en DNMT3a meurent 3 à 4 semaines après la naissance alors que celles déficientes en DNMT3b ne poursuivent pas leur développement embryonnaire au-delà de 9,5 jours après la fécondation. Les embryons déficients à la fois en DNMT3a et en DNMT3b sont plus sévèrement touchés puisqu'ils ne se développent pas au-delà de la gastrulation (Okano, Bell et al. 1999).

- DNMT3L

La protéine DNMT3L (DNMT3-like) présente une grande homologie de séquence avec les autres DNMT mais est elle est dépourvue d'activité catalytique. Cependant, elle n'est pas dépourvue de fonction. C'est une protéine exprimée exclusivement dans les cellules germinales au moment de l'établissement de la méthylation *de novo* de l'ADN, et en particulier sur les gènes soumis à l'empreinte parentale (Bourc'his, Xu et al. 2001). DNMT3L est capable de réguler l'activité catalytique des deux autres membres de la famille DNMT3 mais pas de DNMT1 (Suetake, Shinozaki et al. 2004). Les souris déficientes en DNMT3L sont viables mais stériles. A titre d'exemple, l'ovogenèse se déroule de manière correcte dans les souris femelles déficientes en DNMT3L mais les embryons meurent 9,5 jours après la fécondation (Bourc'his, Xu et al. 2001).

### c) La famille DNMT2

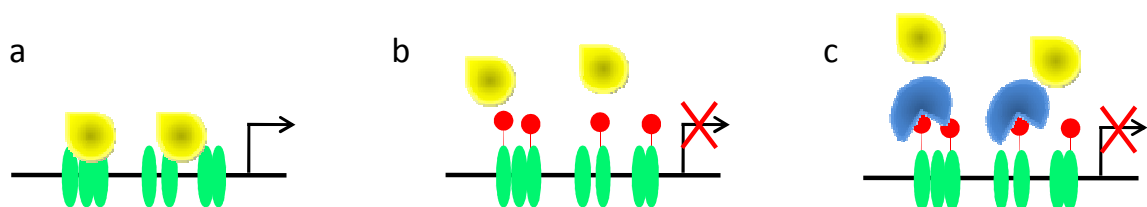
La protéine DNMT2, est la plus conservée, la plus largement représentée parmi de nombreuses espèces (dont certaines possèdent uniquement cette DNMT, comme par exemple *S.pombe* ou *D.melanogaster*) mais également la plus énigmatique des DNMT. Elle ne possède pas de domaine régulateur ni d'activité ADN méthyltransférase.

De plus, la délétion du gène codant pour DNMT2, réalisée chez la souris, n'a pas d'effet sur l'embryon. Elle n'aurait donc un rôle essentiel ni pour la méthylation de l'ADN ni au cours du développement embryonnaire (Okano, Xie et al. 1998). Il a par contre été suggéré qu'elle pourrait être impliquée dans la méthylation de l'ARN (Goll, Kirpekar et al. 2006).

### 3. Fonctions de la méthylation de l'ADN

Les fonctions biologiques de la méthylation de l'ADN ont été largement étudiées depuis plusieurs décennies. Le rôle de la méthylation de l'ADN dans la régulation des gènes a été proposé dès le milieu des années 70 et a été amplement confirmé depuis (Holliday and Pugh 1975). Ainsi, la méthylation de l'ADN est généralement associée à un état transcriptionnellement répressif (Bird and Wolffe 1999; Klose and Bird 2006). Différentes approches ont permis d'arriver à cette conclusion. En effet, l'expression de gènes dans certains tissus a été clairement corrélée à l'état de méthylation de l'ADN. C'est notamment le cas des gènes soumis à l'empreinte parentale. D'autre part, l'utilisation d'un analogue non méthylable des cytosines, tel que la 5-Aza-Cytidine, conduit à la déméthylation de l'ADN et à la transcription de gènes jusque là maintenus dans un état inactif (Attwood, Yung et al. 2002)

Il existe deux manières d'expliquer comment la méthylation de l'ADN peut exercer cet effet répressif sur la transcription (Figure 7). En effet, la présence de dinucléotides CpG méthylés peut altérer la fixation de facteurs de transcription et donc interférer directement avec les mécanismes d'activation des gènes (Watt and Molloy 1988). D'autre part, la reconnaissance des dinucléotides CpG méthylés par des protéines spécifiques conduit de manière indirecte à la répression de l'expression des gènes par recrutement de co-répresseurs. Ces protéines sont généralement connues sous le nom de MBP (Methyl-CpG binding proteins). Les MBP sont divisées en deux groupes. Les protéines du premier groupe possèdent un domaine de reconnaissance de l'ADN méthylé appelé MBD (MeCP2, MBD1, MBD2, MBD3 et MBD4). Le second groupe est constitué de protéines à doigt de zinc, ne possédant pas de domaine MDB mais pouvant également se lier à l'ADN méthylé (Kaiso, ZBTB4 et ZBTB28) (Filion, Zhenilo et al. 2006; Klose and Bird 2006).

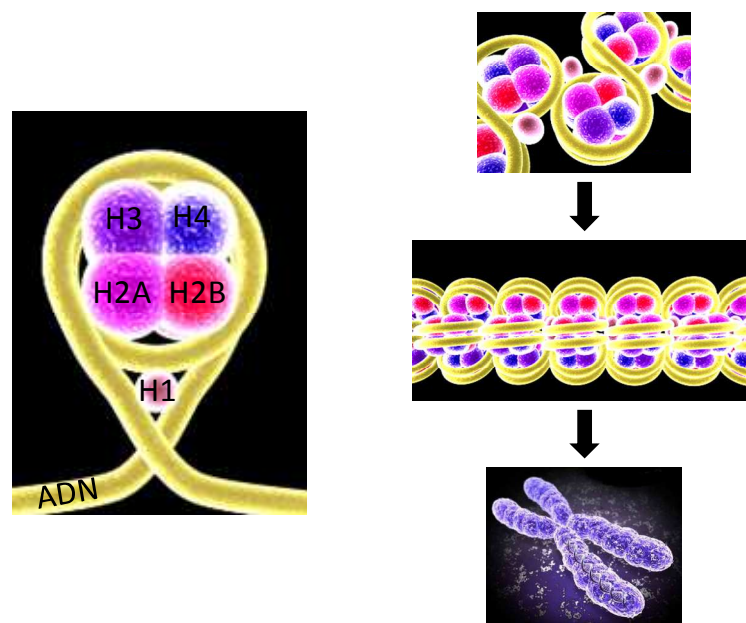


**Figure 7 : La méthylation de l'ADN réprime la transcription des gènes (d'après (Attwood, Yung et al. 2002)).** a) Les CpG (vert) ne sont pas méthylés, les facteurs de transcription (jaune) peuvent se lier à l'ADN et induire la transcription. b) Les CpG sont méthylés (rouge), les facteurs de transcription ne peuvent plus se lier à l'ADN et la transcription est réprimée (croix rouge). c) Les MBD (bleu) se lient sur les CpG méthylés, les facteurs de transcription ne peuvent plus se lier à l'ADN et la transcription est réprimée.

*En conclusion, la méthylation de l'ADN semble essentielle pour le développement et la fonctionnalité des organismes. En effet, les souris déficientes en DNMT meurent au stade embryonnaire. De plus, des problèmes dans la machinerie de méthylation de l'ADN peuvent également causer de sérieuses pathologies chez l'Homme comme c'est le cas du syndrome ICF (Immunodéficience, instabilités Centromériques et anomalies Faciales). Ce syndrome, dont nous reparlerons dans la partie résultats, est la seule pathologie humaine connue pour être en relation avec des mutations dans le gène codant pour la DNMT3b (Ehrlich, Sanchez et al. 2008). Enfin, la méthylation de l'ADN est une marque épigénétique particulièrement puissante puisqu'elle est décrite comme irréversible. En effet l'existence d'ADN déméthylase est encore à ce jour sujet à controverse (Ooi and Bestor 2008).*

## I.2.B Le nucléosome

Le génome humain est composé de quelques trois milliards de paires de bases, représentant une longueur totale de l'ordre du mètre. Or cet ADN est présent au sein du noyau dont le diamètre n'excède pas une dizaine de micromètres. Afin de contenir l'ADN dans le noyau celui-ci doit donc être compacté. Pour cela, l'ADN est associé à une structure protéique particulière, le nucléosome.



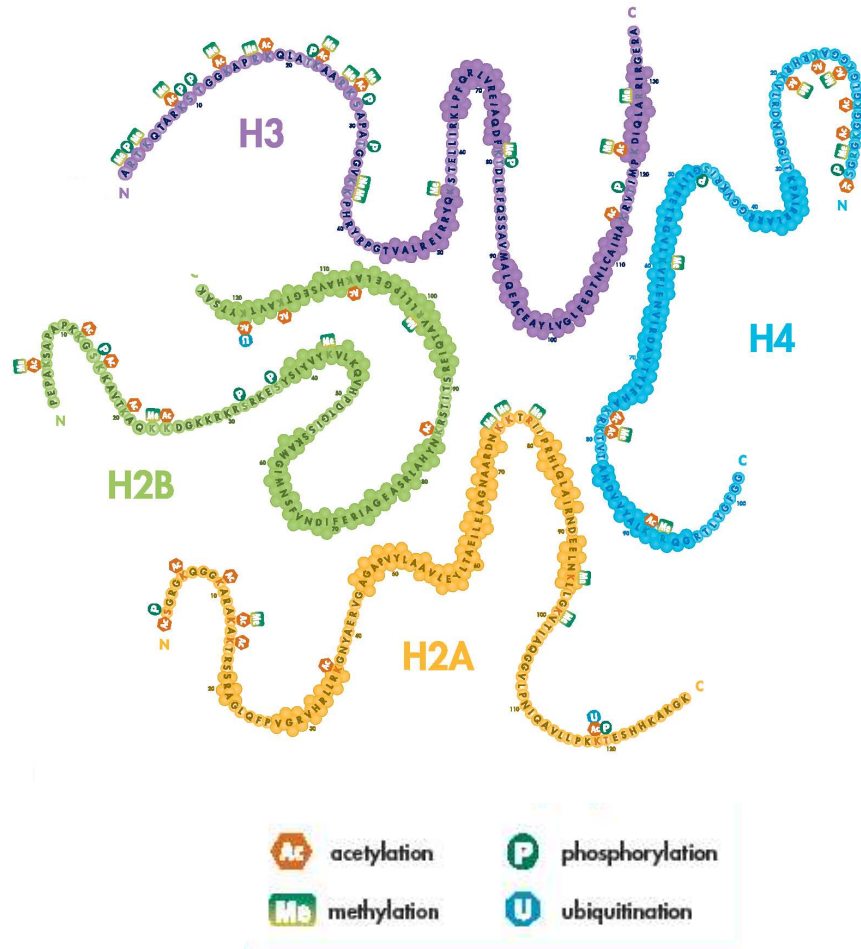
**Figure 8 : Le nucléosome et la compaction de l'ADN (avec l'autorisation de Joseph Roland).**  
 Le nucléosome est composé des histones H2A, H2B, H3 et H4 présentes chacune en deux exemplaires. Les nucléosomes, autour desquels s'enroulent l'ADN (jaune) sont reliés entre eux par une l'histone de liaison H1. Des niveaux successifs de compaction de la chromatine vont aboutir à la formation du chromosome métaphasique.

Le nucléosome est composé d'un octamère de protéines basiques, les histones H2A, H2B, H3 et H4, chacune présente en deux exemplaires. L'ADN s'enroule autour du nucléosome sur une distance d'environ 146 paires de bases (pb). La structure ainsi formée ressemble à un collier de perles régulièrement espacées le long du filament d'ADN. L'ensemble composé par l'ADN et les nucléosomes est appelé chromatine. Les nucléosomes sont reliés entre eux par un fragment d'ADN internucléosomique qui interagit avec l'histone de liaison H1. D'autres niveaux de compaction interviennent ensuite pour aboutir à la compaction ultime, la formation du chromosome métaphasique (Figure 8).

Le degré de compaction de l'ADN est associé à sa capacité transcriptionnelle. En effet, plus l'ADN est compacté et moins il est accessible à la machinerie de transcription. Ainsi le nucléosome, qui constitue la première étape de compaction de l'ADN, est la cible de modifications épigénétiques. Les mécanismes épigénétiques capables d'affecter le nucléosome sont de trois types : les modifications d'histones, les variants d'histones et les complexes de remodelage de la chromatine.

### *1. Les modifications d'histones*

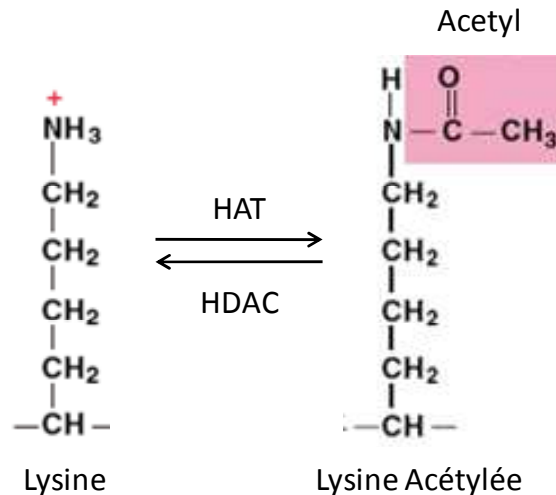
L'extrémité N-terminale des histones de cœur est riche en acides aminés basiques capables de subir des modifications post-traductionnelles. En fonction du type de modification, les interactions entre l'extrémité N-terminale des histones et l'ADN sont altérées. Ceci a pour conséquence de moduler l'accessibilité de l'ADN à des facteurs régulateurs comme ceux impliqués dans la transcription. Il s'agit alors de modifications épigénétiques. Il existe de nombreuses modifications épigénétiques que peuvent subir les histones telles que l'acétylation (sur les lysines), la méthylation (sur les lysines et les arginines), la phosphorylation (sur les sérines et les thréonines), l'ubiquitination (sur les lysines) ou la sumoylation (sur les lysines) (Figure 9). Les différentes combinaisons de ces modifications agissent de manière coordonnée pour réguler l'expression des gènes. Ainsi, à chaque état transcriptionnel pourrait être associé un « code histone » particulier (Jenuwein and Allis 2001). Dans le cadre de cette étude, nous nous intéresserons de plus près à deux de ces modifications, l'acétylation et la méthylation.



**Figure 9 : Carte des principales modifications des histones (d'après Upstate.com)**

a) L'acétylation

L'acétylation des histones joue un rôle important dans le contrôle de la transcription. En effet, celle-ci est généralement corrélée à une activité transcriptionnelle. Au contraire, la déacétylation des histones est associée à un état transcriptionnel plutôt répressif. La première évidence du lien entre l'acétylation des histones et la transcription a été faite dans les années 1960 (Allfrey, Faulkner et al. 1964). Cette étude montre en effet que les régions transcriptionnellement actives sont hyperacétylées au contraire des régions silencieuses qui sont hypoacétylées. L'acétylation des histones consiste en l'ajout d'un groupement acétyle ( $\text{COCH}_3$ ) sur un acide aminé particulier, la lysine. L'acétylation dépend de l'action de deux classes d'enzymes dont l'activité est antagoniste : les HAT (Histones Acetyl-Transferase) qui permettent l'acétylation de la lysine et les HDAC (Histones DeAcetyl-Transferase) qui produisent la réaction inverse (Figure 10).



**Figure 10 : L'acétylation des histones.** L'acétylation des histones consiste en l'ajout d'un groupement acétyle ( $\text{COCH}_3$ , en rose) sur l'extrémité N-Terminale de la lysine. La réaction d'acétylation, catalysée par des HAT, est réversible. La déacétylation des histones met en jeu des HDAC.

Les HAT ajoutent un groupement acétyle sur des lysines en utilisant l'acétyl coenzyme A comme donneur. L'acétylation des histones permet de neutraliser les charges positives des lysines et par conséquent de diminuer les liaisons électrostatiques entre les histones et l'ADN, chargé négativement. Il s'en suit alors une décompaction de la chromatine qui permet une plus grande accessibilité des gènes à la machinerie de transcription (Workman and Kingston 1998). Six familles de HAT ont été décrites à ce jour (GNAT, MYST, P300/CBP, TAFII 250, les sous unités de TFIIC et la famille des coactivateurs des récepteurs nucléaires) dont les membres agissent généralement au sein de complexes protéiques impliquant notamment la machinerie de transcription. Cependant ces enzymes ne présentent pas de spécificité quant à la lysine qu'elles acétylent. D'autre part l'activité acétyltransférase n'est pas uniquement réservée aux histones puisque les HAT peuvent également acétyler des protéines non histones telles que des facteurs de transcription (comme le suppresseur de tumeur p53) ou des protéines chaperonnes (HSP) (Yang and Seto 2007).

L'acétylation des histones est un processus parfaitement réversible. En effet, les HDAC catalysent l'hydrolyse du groupement acétyle et déacétylent ainsi les histones. Les lysines retrouvent donc leur charge positive et l'interaction entre l'ADN et les histones est augmentée. Ainsi, la déacétylation des histones est un processus associé à la répression de la transcription. La localisation des HDAC sur la chromatine dépend d'un autre signal de répression de la transcription : la méthylation de l'ADN. En effet, les HDAC peuvent être recrutées au niveau des CpG méthylés, probablement par l'intermédiaire des MBD (Mompalmer 2003).

Il existe 2 grandes familles de HDAC :

- La famille des « HDAC classiques », qui nécessitent le  $Zn^{2+}$  comme cofacteur, est composée de 3 groupes. Les HDAC de classe I comprennent les HDAC 1, 2, 3 et 8. Ces protéines sont ubiquitaires et ont une localisation nucléaire. Les HDAC de classe II sont représentées par les HDAC 4, 5, 6, 7, 9 et 10. Leur expression est plus restreinte par rapport à la classe I suggérant qu'elles pourraient avoir un rôle plus spécifique notamment au cours du développement et de la différenciation (Martin, Kettmann et al. 2007). Enfin, les HDAC de classe IV dont la seule répertoriée à ce jour est HDAC11 (Gao, Cueto et al. 2002). Elle possède des régions communes à la fois aux HDAC de classe I mais aussi de classe II. Son expression est restreinte à certains tissus (rein, cœur, cerveau, muscles squelettiques et testicule) laissant penser qu'elle pourrait avoir des fonctions particulières, distinctes de celles connues concernant les autres HDAC.

- La « famille des Sirtuines » contient les protéines apparentées à Sir-2 (Silent Information Regulateur 2). Egalement appelées HDAC de classe III, ces protéines sont au nombre de sept (SIRT1-7) dont seules SIRT1, 2, 3 et 5 semblent réellement posséder une activité déacétylase. Elles requièrent pour cette activité le cofacteur  $NAD^+$  (Saunders and Verdin 2007).

Les HDAC sont des molécules particulièrement attractives dans le cadre de certaines thérapies anticancéreuses. En effet, la tumorigénèse est associée à la fois à des altérations génétiques et épigénétiques. Or seules ces dernières sont aisément réversibles. Certains gènes suppresseurs de tumeurs sont mis sous silence pendant la tumorigénèse. L'utilisation en thérapie anticancéreuse d'inhibiteurs de HDAC, conduit à une augmentation de l'acétylation et à une réactivation de l'expression de certains gènes comme p21 qui joue un rôle dans le contrôle du cycle cellulaire (Hadnagy, Beaulieu et al. 2008). Parmi ces inhibiteurs de HDAC figurent la Trichostatine A (TSA), le Butyrate ou le nicotinamide (Yoshida, Kijima et al. 1990; Davie 2003; Porcu and Chiarugi 2005).

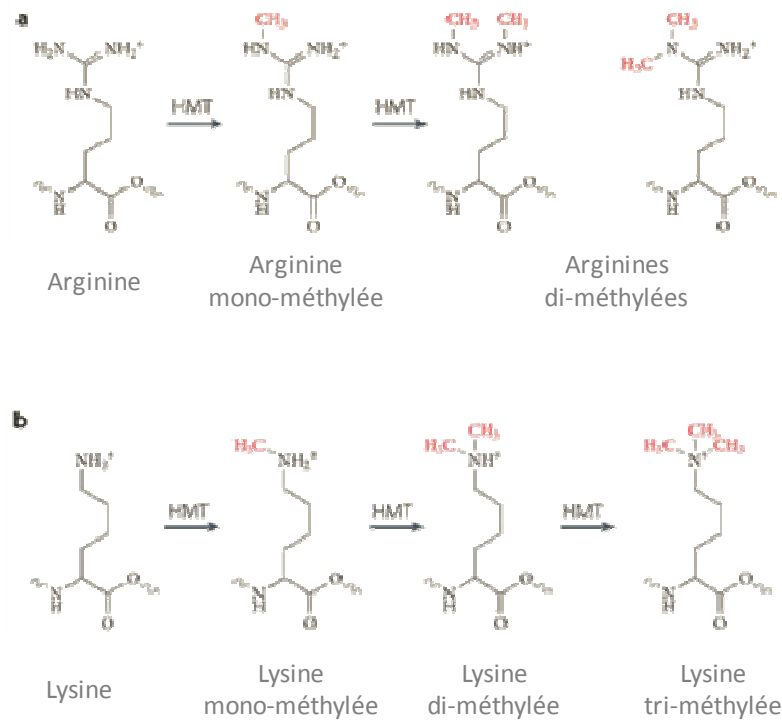
## b) La méthylation

La méthylation des histones est très probablement la modification épigénétique la plus conservée entre les espèces. Contrairement à l'acétylation, la méthylation des histones ne modifie pas la charge du nucléosome. Par contre elle joue un rôle important dans le recrutement de facteurs régulant l'accessibilité de la chromatine. Ce sont le plus souvent les extrémités N-terminales des histones H3 et H4 qui sont les



cibles de la méthylation. Les lysines (abrévées par K) peuvent être mono, di ou tri-méthylées et les arginine (abrévées par R) mono ou di méthylés (Bannister, Schneider et al. 2002). Ceci ajoute encore un degré de complexité au fameux « code histone ».

Les enzymes qui catalysent la méthylation des histones sont appelées HMT (Histones Methyl-Transferases). Les HMT ajoutent un groupement méthyle ( $\text{CH}_3$ ) sur des lysines ou des arginines, en utilisant comme substrat la S-adénosyle-méthionine (Figure 11). Le domaine responsable de l'activité enzymatique des HMT est le domaine SET (Su(var)3-0, Enhancer-of-zeste, Tritorax). Il existe deux grandes classes de HMT. Les HMT de type I, les PRMT (Protein arginine (R) Methyl-Transferase) catalysent la méthylation des arginines. Les HMT de type II, HKMT (Histone lysine (K) Methyl-Transferase) catalysent la méthylation des lysines (Bachand 2007; Pal and Sif 2007). La méthylation des histones est un processus important notamment au cours du développement embryonnaire. En effet, des souris déficientes en Suv39h1 et Suv39h2 (deux HKMT) meurent généralement au stade embryonnaire (Peters, O'Carroll et al. 2001).



**Figure 11 : Méthylation des histones (d'après (Klose and Zhang 2007)).** La méthylation des histones consiste en l'ajout d'un (mono) de deux (di) ou de trois (tri) groupements méthyles sur des arginines (a) ou sur des lysines (b).

De plus la méthylation des histones joue un rôle différent selon le résidu méthylé. En effet, contrairement aux HDAC qui sont généralement associées à un

processus de répression génique, les HMT peuvent être responsables soit d'une activation soit d'une répression des gènes, selon le résidu méthylé. Par exemple, des expériences d'immunoprécipitation de chromatine (ChIP) ont montré que les gènes actifs peuvent être méthylés sur H3K4, H3K36 et H3K79. Au contraire la méthylation sur H3K9 ou H3K27 est plutôt associée à une chromatine non permissive à la transcription (Peters and Schubeler 2005).

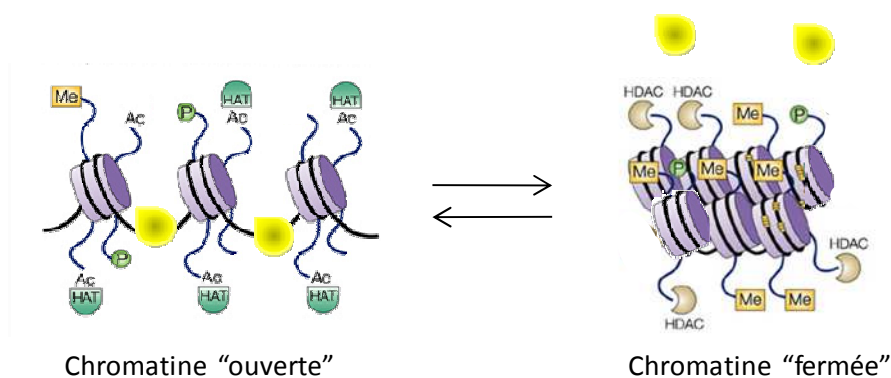
La méthylation des histones a longtemps été décrite comme une modification stable. En effet, l'absence de description d'enzymes capables d'ôter le groupement méthyle suggérait que la méthylation des histones était un processus irréversible. Cependant des histones déméthylases ont finalement été identifiées. LSD1 est capable de déméthyliser H3K4 et de participer ainsi à l'inhibition de la transcription (Shi, Lan et al. 2004). Depuis, d'autres protéines ont ainsi été identifiées comme histones déméthylases par la présence d'un domaine LSD1 (lysine specific demethylase 1) ou d'un domaine JmjC (Jumonji C) (Tian and Fang 2007). Comme les HMT, ces enzymes possèdent certaines spécificités. Par exemple, LSD1 est capable de déméthyliser H3K4m1 et H3K4m2 mais pas H3K4m3. Aucune enzyme responsable de la déméthylation des arginines n'a été identifiée. Cependant, ce processus a tout de même lieu mais d'une manière indirecte. En effet, la déimination qui conduit au remplacement de l'arginine par une citrulline non méthylable, permet d'éliminer les arginines méthylées (Cuthbert, Daujat et al. 2004; Wang, Wysocka et al. 2004).

### c) Fonctions des modifications d'histones

La réversibilité des modifications d'histones met en évidence la plasticité de l'information épigénétique, à l'inverse du code génétique qui est relativement stable. L'ensemble des modifications des histones constituerait un code, le « code histone », qui permettrait d'associer à chaque combinaison de modifications un état particulier de la chromatine. La lecture de ce code peut se faire de deux manières.

Les modifications des histones peuvent moduler la structure de la chromatine, permettant de contrôler les fonctions cellulaires telles que la transcription (Figure 12). En effet les modifications de charges peuvent modifier les contacts entre les histones et l'ADN. Ces modifications ont pour conséquence de dénouer la chromatine et de permettre ainsi un accès plus facile à la machinerie de transcription. Parmi les modifications d'histones connues, l'acétylation est particulièrement impliquée dans ce processus par la neutralisation des charges positives des lysines. En particulier l'acétylation d'H4K16 a un effet négatif sur la compaction de la chromatine (Shogren-Knaak, Ishii et al. 2006). La phosphorylation des histones, par l'ajout de charges

négatives, peut avoir un rôle similaire. Ainsi, certaines modifications épigénétiques sont associées à une activation de la transcription comme l'acétylation, la méthylation, la phosphorylation ou l'ubiquitination. Par contre, la méthylation, l'ubiquitination et la sumoylation sont généralement associées à une répression de la transcription (He and Lehming 2003).



**Figure 12 : Régulation de la transcription par les modifications d'histones (d'après (Johnstone 2002)).** L'hyperacétylation des histones (Ac) réalisée par les HAT conduit à une chromatine ouverte, accessible aux facteurs de transcription (jaune). Au contraire les HDAC déacétylent les histones et conduisent à une chromatine fermée, non accessible aux facteurs de transcription. Le rôle joué par la méthylation des histones (Me) est plus complexe. A titre d'exemple, la méthylation de H3K4 est associée à une chromatine ouverte alors que la méthylation de H3K9 est plutôt associée à une chromatine fermée.

D'autre part, la lecture du code épigénétique peut faire intervenir un intermédiaire protéique interagissant directement avec l'histone modifiée. En effet, les modifications d'histones sont reconnues par des domaines protéiques particuliers (de la Cruz, Lois et al. 2005). Ainsi, le chromodomaine et le bromodomaine reconnaissent respectivement les histones méthylées et les histones acétylées. Ces mêmes protéines peuvent également avoir une activité enzymatique capable de modifier les histones. Par exemple, le recrutement de PRC1 (Polycomb Repressive Complex 1) reconnaît, grâce à son chromodomaine, la triméthylation de H3K27 et conduit à l'ubiquitination de H2A (Kotake, Cao et al. 2007).

Il y a véritablement une interconnexion entre toutes les modifications d'histones. L'ensemble de ces modifications peut conduire à l'établissement de différents modèles épigénétiques, pouvant définir des environnements favorables ou non pour la transcription. C'est notamment le cas pour les 2 grands compartiments chromatiques que sont l'euchromatine et l'hétérochromatine dont on reparlera au chapitre suivant.

## 2. Les variants d'histones

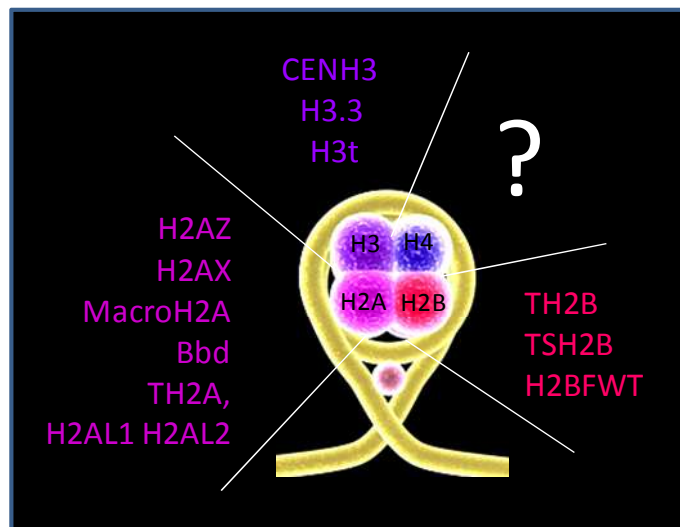
Chaque histone, à l'exception de H4, existe sous plusieurs formes protéiques dans la cellule (Figure 13). Ces différentes protéines, appelées « variants d'histones » sont codées par des gènes différents. L'incorporation de variants d'histones au sein du nucléosome introduit un degré supplémentaire de complexité. En effet, les variants d'histones peuvent induire des changements de structure du nucléosome et ainsi modifier les interactions entre l'ADN et les protéines. Ainsi, l'incorporation de variants d'histones au sein du nucléosome a pour conséquence de moduler l'accessibilité de l'ADN pour des facteurs transcription. Il s'agit donc de modifications épigénétiques. De plus, ces variants d'histones sont également la cible de modifications post-traductionnelles telles que la phosphorylation, ajoutant encore un niveau de complexité au « code histone ».

Contrairement aux histones « classiques », l'incorporation des variants d'histones au sein du nucléosome peut avoir lieu en dehors de la phase de réplication de l'ADN. De plus, la localisation et l'expression des variants d'histones sont finement régulées selon le stade du développement ou le tissu considéré, suggérant qu'ils puissent avoir des fonctions précises, particulièrement importantes. A l'appui de cette hypothèse, de nombreuses pathologies humaines, telles que l'infertilité ou le cancer, sont associées à la perte de fonction de certains variants d'histones confirmant ainsi leur importance (Boulard, Bouvet et al. 2007).

### a) H2A

Il existe 4 variants d'histones pour H2A : H2AZ, H2AX, MacroH2A, Bbd auxquels on peut ajouter les variants qui sont exprimés uniquement dans le testicule comme TH2A, H2AL1 et H2AL2. H2AZ est un variant particulièrement conservé au cours de l'évolution, même au regard de H2A. Il exerce une activité à la fois positive et négative sur l'expression des gènes. Pour exemple, chez la levure, H2AZ est associé aux sites d'initiation de la transcription des gènes inactifs. L'éviction de ce variant d'histone serait un processus rapide permettant l'activation des gènes (Guillemette, Bataille et al. 2005). H2AX est un variant d'histone qui est phosphorylé sur la sérine 139 en réponse à des cassures double brin de l'ADN (Rogakou, Pilch et al. 1998). H2AX phosphorylé sur la sérine 139 (plus communément appelé gamma-H2AX) est présent au niveau des cassures doubles brins de l'ADN et participe à la réparation de celui-ci, en collaborant avec la machinerie de réparation (Celeste, Fernandez-Capetillo et al. 2003). MacroH2A doit son nom à sa grande taille qui représente 3 fois celle de l'histone classique H2A (Pehrson and Fried 1992). Le variant MacroH2A est localisé en

particulier au niveau du chromosome X inactif (Costanzi and Pehrson 1998), également appelé corpuscule de Barr, où il est associé à la répression transcriptionnelle (Perche, Vourc'h et al. 2000). Au contraire, H2ABbd (H2A Barr Body Deficient) comme son nom l'indique est exclu du corpuscule de Barr. On le retrouve plutôt associé aux régions acétylées suggérant qu'il pourrait être associé à l'activation des gènes (Chadwick and Willard 2001; Angelov, Verdel et al. 2004). Enfin, certains variants de l'histone H2A présentent une expression restreinte au testicule TH2A (Testis specific H2A), H2AL1 et H2AL2 (H2A Like 1 et 2) (Rao, Brahmachari et al. 1983; Govin, Escoffier et al. 2007).



**Figure 13 : Les variants d'histones.**

b) H2B

Les variants d'histones H2B décrits à ce jour, sont au nombre de 3, TH2B (Testis specific H2B), TSH2B (Testis/Sperm specific H2B) et H2BFWT (H2B Family, member W, Testis specific). Ils présentent une expression restreinte aux cellules de la lignée germinale mâle. Il est important de souligner qu'au cours de la spermatogenèse la chromatine est profondément réorganisée. En effet, les histones somatiques sont en partie remplacées par des histones spécifiques du testicule, elles mêmes en partie remplacées par des protéines de transition puis par d'autres protéines appelées protamines. Au final, la chromatine des spermatozoïdes est hautement compactée (Sassone-Corsi 2002). Il n'est donc pas surprenant de découvrir des variants d'histones dont l'expression est limitée au testicule. TH2B et TSH2B sont exprimés très précocement au cours de la spermatogenèse, dès les divisions mitotiques, et cette expression est maintenue jusque dans les spermatozoïdes matures. La régulation de

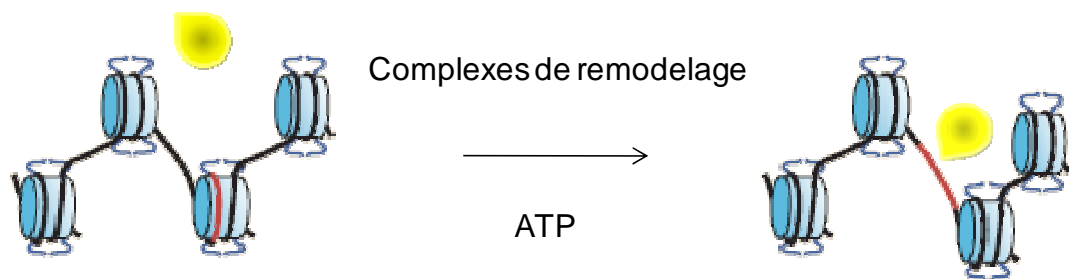
l'expression de TSH2B doit être finement contrôlée car une corrélation inverse entre la quantité de cette protéine dans le sperme et la morphologie des spermatozoïdes a été observée, suggérant que TSH2B pourrait être impliquée dans des mécanismes conduisant à l'infertilité masculine (van Roijen, Ooms et al. 1998). Enfin, le variant d'histone H2BFWT, caractérisé en 2004, n'a pas encore de fonction connue (Churikov, Siino et al. 2004; Boulard, Gautier et al. 2006)

### c) H3

Il existe 3 variants d'histones pour H3: CENH3 (Centromeric H3), H3.3 et H3t (H3 testis specific). CENH3 est localisé au niveau de régions chromosomiques particulières, les centromères, où il est indispensable à la fonction de celui-ci. Le rôle précis de CENH3 sera discuté dans un second chapitre. H3.3 est localisé dans des régions transcriptionnellement actives (Chow, Georgiou et al. 2005) Enfin, une expression restreinte aux cellules de la lignée germinale mâle a été décrite pour le variant H3t (Witt, Albig et al. 1996).

### 3. *Les complexes de remodelage*

Les modifications d'histones et l'incorporation de variants d'histones ne sont pas les seuls mécanismes capables d'agir sur la structure de la chromatine. Les complexes de remodelage, comme leur nom l'indique, assurent également un remodelage de la chromatine. Les complexes de remodelage, constitués de plusieurs protéines, permettent le mouvement des nucléosomes le long de l'ADN. Ils peuvent déplacer un nucléosome en cis par glissement (Figure 14), en trans par transfert ou même réaliser l'éviction complète du nucléosome. Ces complexes multiprotéiques comportent tous une sous-unité ATPase. C'est en effet en utilisant l'énergie fournie par l'hydrolyse de l'ATP, que les complexes de remodelage altèrent les liaisons entre les histones et l'ADN, permettant le mouvement des nucléosomes (Becker and Horz 2002). Ainsi, les complexes de remodelage sont capables de faciliter ou de restreindre l'accès de l'ADN à des facteurs de transcription et constituent le dernier mécanisme épigénétique agissant sur le nucléosome (Hogan and Varga-Weisz 2007).



**Figure 14 : Déplacement des nucléosomes par les complexes de remodelage ( d'après (Boyer, Latek et al. 2004)).** En présence d'ATP, les complexes de remodelage déplacent les nucléosomes le long de l'ADN. Ainsi, le site de fixation (ADN rouge) du facteur de transcription (jaune) devient accessible.

#### a) Les différentes familles

Les facteurs de remodelage sont classés en quatre principales familles en fonction de certains domaines additionnels présents au sein de l'unité contenant l'activité catalytique ATPase (Figure 15).

- La famille SWI/SNF :

Le domaine fonctionnel additionnel caractéristique de la famille SWI/SNF est un bromodomaine, capable de se lier aux histones acétylées. Les complexes SWI/SNF, à travers la lecture de ces marques épigénétiques, participent ainsi à l'activation de la transcription (Kasten, Szerlong et al. 2004).

- La famille ISWI

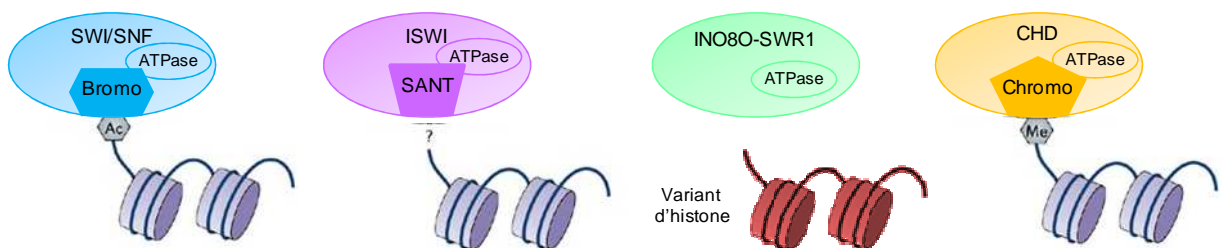
La famille ISWI présente de nombreuses similarités avec la famille SWI/SNF. Cependant, la présence d'un domaine SANT (Swi3, Ada2, N-Cor, TFIIB) dans la sous unité ATPase et l'absence d'un bromodomaine permettent de distinguer cette deuxième famille. Le domaine SANT serait un domaine de liaison aux histones non modifiées (Boyer, Latek et al. 2004). Le complexe ISWI est impliqué dans la réplication de l'ADN. Ce complexe de remodelage se lie en effet aux sites de réplication, grâce à son interaction avec la protéine de réplication PCNA, et participe à l'assemblage de la chromatine (Poot, Bozhenok et al. 2005).

- La famille INO80-SWR1:

La famille INO80-SWR1 ne possède pas de domaine particulier dans la sous unité ATPase mais une séquence espaceur au sein même de celle-ci. Le complexe INO80 est capable de se lier au variant d'histone gamma H2AX enrichi au niveau des cassures double brins de l'ADN. Ainsi, ce complexe de remodelage participerait à la réparation de l'ADN en favorisant le déplacement des nucléosomes (Cairns 2004; van Attikum, Fritsch et al. 2004). Le complexe SWR1 quant à lui facilite l'échange de l'histone H2A contre son variant H2AZ (Korber and Horz 2004).

- La famille CHD :

La famille CHD se différencie des autres par la présence d'un chromodomaine au sein de la sous unité ATPase. Ce chromodomaine permet notamment l'interaction de CHD1 avec l'histone H3 méthylée sur la lysine 4, facilitant ainsi l'ouverture de la chromatine (Sims, Chen et al. 2005). Cependant, un autre complexe appartenant à la famille CHD, le complexe NurD possède deux sous unités bien connues que sont HDAC1 et HDAC2. Le recrutement de ce complexe contribuerait donc à la répression de la transcription des gènes (Kantor, Makedonski et al. 2003).



**Figure 15 : Les différentes familles de complexes de remodelage (d'après (de la Serna, Ohkawa et al. 2006)).** Tous les facteurs de remodelage contiennent un domaine ATPase. Les membres de la famille SWI/SNF possèdent un bromodomaine capable de se lier aux histones acétylées. Les membres de la famille ISWI contiennent un domaine SANT supposé se lier aux histones non modifiées. Les membres de la famille INO80-SWR1 facilitent les échanges entre certains variants d'histones et les histones classiques. Enfin, les membres de la famille CHD possèdent un chromodomaine capable de se lier aux histones méthylées.

## b) Régulation des complexes de remodelage

Les facteurs de remodelage permettent de réguler l'accès à l'ADN à la machinerie de transcription. Mais ils peuvent aussi faciliter la réplication de l'ADN et sa réparation. Leur régulation est donc essentielle pour assurer une activité spatio-



temporelle correcte. Ainsi, au cours de la mitose, des complexes de remodelage sont inactivés et exclus de la chromatine facilitant ainsi sa compaction et inhibant par la même occasion la transcription. L'éviction du facteur de remodelage SWI/SNF est corrélée à la phosphorylation de deux de ses ATPases. Ce processus est parfaitement réversible, une fois la mitose terminée, par une simple déphosphorylation (Muchardt, Reyes et al. 1996; Sif, Stukenberg et al. 1998). Un mauvais contrôle de l'activité de ces complexes de remodelage peut se révéler désastreux. En effet, de nombreuses pathologies, telles que les cancers, sont associées à une dérégulation de l'expression des complexes de remodelage (Cho, Elizondo et al. 2004).

*En conclusion, l'unité de base de la chromatine, le nucléosome, est le siège de processus hautement dynamiques permettant notamment la régulation de l'expression génique. Il existe trois voies de modification des nucléosomes : les modifications post-traductionnelles des histones, l'incorporation de variants d'histones et enfin l'intervention de complexes de remodelage. Ces trois voies sont inter-dépendantes. Ainsi, les modifications post-traductionnelles des histones peuvent servir de cibles pour la fixation des complexes de remodelage. Les variants d'histones sont capables de réguler la fixation de ces mêmes complexes. Le variant d'histone MacroH2A, par exemple, empêche le complexe de remodelage SWI/SNF de déplacer les nucléosomes (Agelopoulos and Thanos 2006; Doyen, An et al. 2006).*

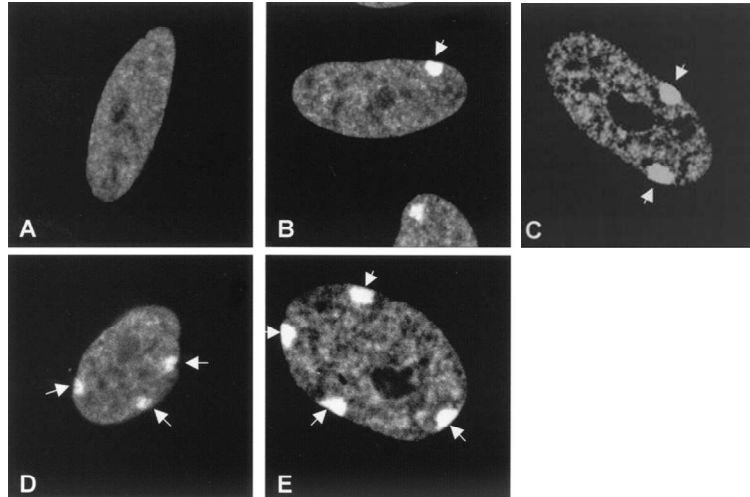
## I.2.C Les ARN non codants

La régulation de l'expression des gènes, par des mécanismes épigénétiques, ne concerne pas simplement l'ADN et sa méthylation ou encore les octamères d'histones et leurs modifications diverses. En effet, l'ARN fait partie intégrante de ce processus. Des ARN non codants (ADNnc, ADN ne codant pas pour des protéines) sont en effet impliqués dans la structure de la chromatine, la maintenant dans un état transcriptionnellement inactif. Nous abordons dans ce chapitre le rôle des ADNnc dans les mécanismes de compensation de dose, d'empreinte et d'interférence par ARN.

### 1. *Xist et l'inactivation du chromosome X*

Les mammifères femelles possèdent deux chromosomes X, contre un seul pour les mâles. Afin d'assurer la « compensation de dose » entre les sexes, l'un des deux chromosomes X de la femelle voit sa transcription réprimée. Cet état inactif sera maintenu au cours des divisions cellulaires (Lyon 1961). Par simple observation microscopique de noyaux colorés au DAPI (4',6' Di Amidino-2-Phényl Indole), une

molécule fluorescente capable de se lier à l'ADN, on peut distinguer le chromosome X inactif, également appelé Corpuscule de Barr, sous la forme d'un amas dense présent à la périphérie du noyau (Barr and Bertram 1949). L'inactivation du chromosome X se met en place au stade embryonnaire de manière aléatoire puis elle est par la suite maintenue dans les cellules somatiques.

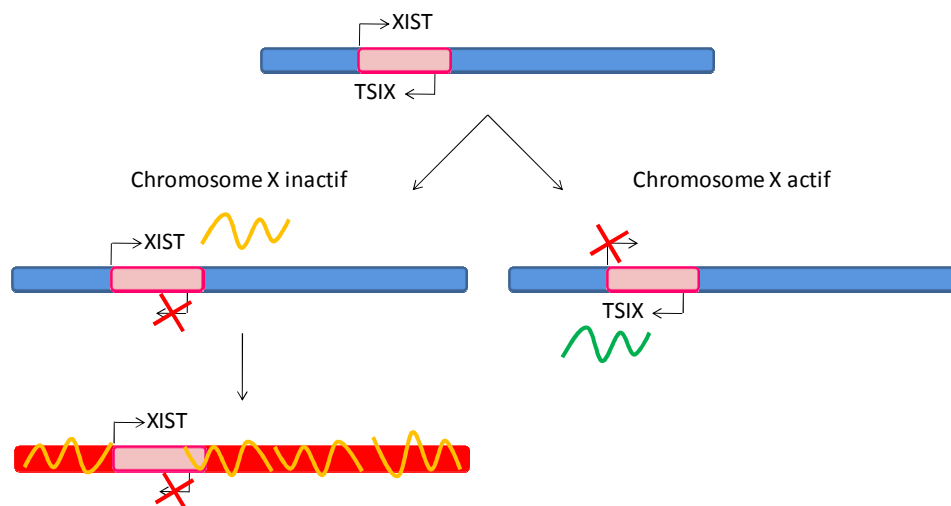


**Figure 16 : Corpuscule de Barr (Hong, Reeves et al. 2001).** Les chromosomes X inactifs, ou Corpuscules de Barr sont détectés dans des cellules possédant un (A), deux (B), trois (C), quatre (D) ou cinq (E) chromosomes X. Le nombre de Corpuscules de Barr correspond toujours au nombre de chromosome X moins un.

La première étape de l'inactivation du chromosome X débute par un comptage du nombre de chromosomes X présents dans le noyau. Ainsi dans des cellules de mammifères qui possèdent un seul chromosome X, celui-ci n'est pas inactivé. Par contre dès que le nombre de chromosomes X est supérieur à un, l'inactivation des chromosomes X est réalisée de telle sorte qu'il n'y en ait plus qu'un seul actif (Avner and Heard 2001) (Figure 16). Ce processus est régulé par un locus présent au sein du chromosome X : le centre d'inactivation Xic. Ce locus génère deux ARN non codants antiparallèles : Xist (sens) et Tsix (antisens) (Brockdorff, Ashworth et al. 1991; Brown, Ballabio et al. 1991; Lee, Davidow et al. 1999). Lors des premières divisions embryonnaires, les ARN Xist et Tsix sont faiblement exprimés. Puis, au stade blastocyste, l'expression de Tsix est réprimée sur le futur chromosome X inactif (Xi) alors que les ARN Xist sont stabilisés. Au contraire, sur le chromosome X actif (Xa), l'expression de Tsix continue et permet la dégradation du transcrit Xist (Panning, Dausman et al. 1997; Sheardown, Duthie et al. 1997).

Ainsi, Xist est activé sur le chromosome X inactif. Cet ARN va se propager en cis le long du chromosome et le recouvrir totalement, permettant son inactivation. Les mécanismes impliqués dans cette « décoration » du Xi sont encore méconnus.

L'ARN Xist est crucial pour la mise en place de l'inactivation. Cependant, il n'est pas nécessaire au maintien de cet état inactif dans les cellules somatiques, même s'il continue à y être exprimé (Wutz and Jaenisch 2000). Le maintien de l'inactivation du chromosome X implique des modifications épigénétiques mises en place très tôt au cours du développement embryonnaire. Ainsi, dans un premier temps, des modifications d'histones considérées comme activatrices, telles que la méthylation d'H3K4, sont perdues. Des modifications répressives apparaissent alors le long du Xi, comme H3K27 triméthylée (Plath, Fang et al. 2003; Kohlmaier, Savarese et al. 2004; Heard 2005). Finalement le variant d'histone MacroH2A est incorporé au sein de la chromatine du Xi, suivi par la méthylation de l'ADN sur les promoteurs des gènes (Mermoud, Costanzi et al. 1999) (Figure 17).



**Figure 17 : ARNnc et inactivation du chromosome X.** Le locus *Xic* (rose) situé sur le chromosome X génère deux ARN non codants *XIST* et *TSIX*. Sur le chromosome X inactif, l'expression de *TSIX* est réprimée (croix rouge) et les ARN *XIST* sont produits (orange). Les ARN *XIST* recouvrent ensuite le chromosome X inactif. Le chromosome X est alors la cible de modifications épigénétiques répressives (méthylation d'histones, recrutement de MacroH2A et méthylation de l'ADN). Le chromosome X est ensuite maintenu sous une forme inactive (rouge). Au contraire, sur le chromosome X actif la transcription de *TSIX* réprime celle de *XIST*. Le chromosome est alors maintenu sous une forme active (bleue).

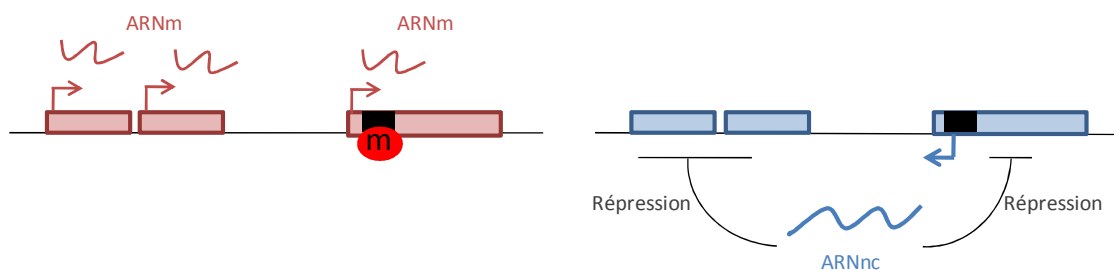
L'ARN Tsix est exprimé sur le Xa où il inhibe la répression transcriptionnelle induite par Xist. La transcription de Tsix conduit en effet à la méthylation de l'ADN du promoteur de Xist, certainement par le recrutement de DNMT3a et de MBP (Beard, Li

et al. 1995; Ohhata, Hoki et al. 2008). Ainsi, l'expression de Xist est réprimée sur le Xa.

## 2. L'empreinte parentale

L'expression d'ARN non codants (ARNnc) dans les mécanismes d'empreinte parentale est un autre exemple attestant du rôle des ARN dans la régulation de la transcription des gènes. L'empreinte parentale est un ensemble de mécanismes épigénétiques de contrôle transcriptionnel, résultant en l'expression monoallélique de certains gènes en fonction de leur origine parentale. L'empreinte parentale des gènes est mise en place très tôt au cours du développement notamment par la méthylation de l'ADN. En effet, les cellules germinales primitives sont hautement méthylées. Au cours de leur maturation, elles vont subir une phase de déméthylation globale pour effacer l'empreinte parentale héritée des parents. Puis, une phase de reméthylation globale de l'ADN a lieu au cours de la gamétogenèse pour établir de nouvelles empreintes maternelles ou paternelles (Santos and Dean 2004)

La méthylation monoallélique des gènes soumis à l'empreinte parentale a lieu au sein d'éléments de contrôle particuliers appelés DMR (Differentially Methylated Region). Certains de ces DMR sont considérés comme des centres d'empreinte (IC pour Imprinting Center) qui permettent de contrôler la transcription monoallélique en cis d'un ensemble de gènes. Sur la soixantaine de gènes connus pour être régulés par des mécanismes d'empreinte, certains expriment des ARNnc qui sont impliqués dans le contrôle de l'expression des gènes adjacents.



**Figure 18 : ARNnc et empreinte parentale.** L'allèle maternel (rose) possède un IC (boîte noire) méthylé (rouge) : les gènes sont alors exprimés. L'IC de l'allèle paternel (bleu) n'est pas méthylé : l'ARNnc est exprimé et il réprime la transcription des gènes adjacents.

L'un des exemples les plus connus est le locus IgF2R (Insulin like Growth Factor 2 Receptor). Il s'agit d'un ensemble de trois gènes (IgF2R, Slc22a2 et Slc22a3) codants pour des protéines. Ces gènes sont exprimés uniquement par l'allèle

maternel, qui possède un IC méthylé. La transcription d'un ARNnc appelé Air (Antisense of IgF2R) a lieu à partir de l'allèle paternel, dont l'IC n'est pas méthylé. Comme son nom l'indique, l'orientation de la transcription de l'ARNnc Air est opposée à celle de l'ARNm d'IgF2R. L'expression de l'ARNnc Air a un rôle direct sur l'inactivation transcriptionnelle des gènes adjacents ainsi que dans le maintien de l'empreinte (Sleutels, Zwart et al. 2002) (Figure 18). En effet, la délétion du promoteur du gène Air sur l'allèle paternel résulte en la transcription biallélique des trois gènes IgF2R, Slc22a2 et Slc22a3. Au contraire, la délétion de ce même promoteur sur l'allèle maternel n'a aucun effet sur l'empreinte.

Les mécanismes par lesquels Air régule cette inactivation transcriptionnelle sont encore très peu connus, mais plusieurs hypothèses sont envisagées. Ainsi comme Xist, l'ARNnc Air pourrait « décorer » ses gènes cibles et faciliter ainsi le recrutement de complexes répresseurs. Cependant, l'ARNnc Air pourrait ne pas avoir de rôle en lui-même. En effet, la transcription de cet ARNnc pourrait altérer la structure de la chromatine permettant ainsi le recrutement de complexes répresseurs (Pauler, Koerner et al. 2007).

### 3. *L'interférence par ARN*

Des ARNnc tels que Xist ou ceux impliqués dans l'empreinte parentale sont des ARN de grande taille qui agissent en cis, près du locus dont ils sont issus. D'autres ARN, beaucoup plus petits (20-30 nucléotides) peuvent également réguler l'expression des gènes mais en agissant cette fois à distance dans le cytoplasme ou sur des gènes cibles.

Ces petits ARN, grâce à un appariement spécifiques des bases, vont servir de guides à des protéines impliquées dans la dégradation de l'ARN messager, dans l'inhibition de la traduction ou de la transcription. Ces processus sont connus sous l'appellation d' « interférence par ARN » (ou RNAi pour RNA Interference).

L'interférence par ARN est un phénomène au cours duquel des petits ARNnc répriment une séquence homologue. Si cette séquence homologue est un ARN, la répression est alors qualifiée de post-transcriptionnelle (PTGS pour Post Transcriptional Gene Silencing). Par contre, si cette séquence est de l'ADN, dans ce cas la répression est transcriptionnelle (TGS pour Transcriptional Gene Silencing) (Figure 19). Ce processus a été mis en évidence pour la première fois chez le pétunia. En effet, des expériences ayant pour but d'intensifier la coloration fuchsia des pétales, en introduisant un système capable d'exprimer le gène produisant le pigment, ont eu

exactement l'effet inverse ! A la place de pétales fortement colorées, des pétales striées de blanc, voir intégralement blanches sont apparues (Napoli, Lemieux et al. 1990). La mise en évidence de mécanismes similaires chez d'autres espèces, et en particulier chez la levure, a eu un impact considérable sur la compréhension de la fonction des ARNnc dans la régulation de l'expression des gènes.

#### a) PTGS

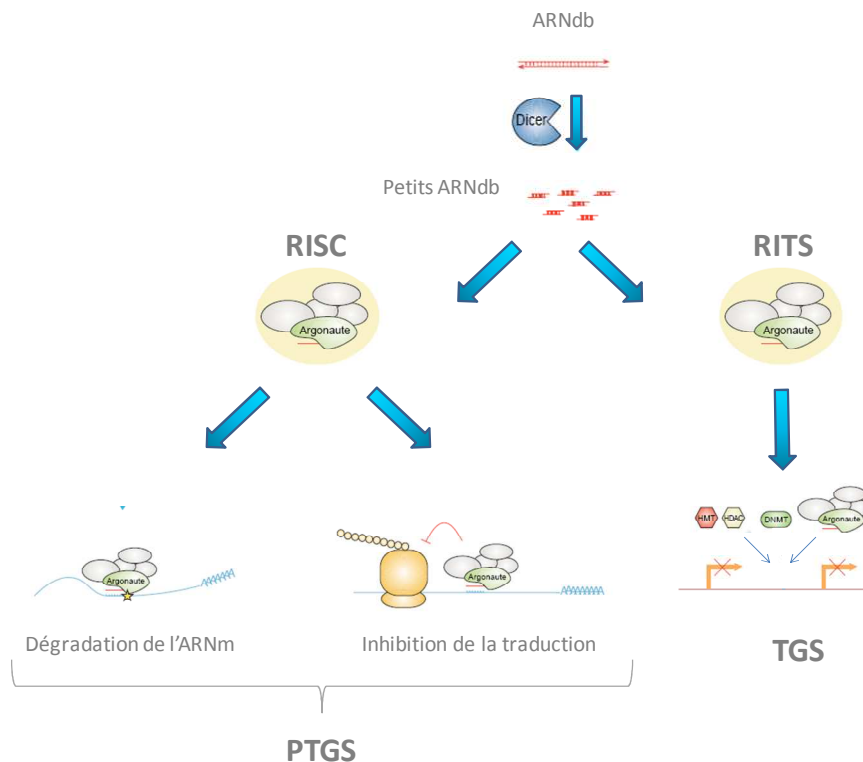
La forme la plus courante de répression par RNAi a lieu dans le cytoplasme, au niveau post-transcriptionnel. Dans ce cas, l'ARN messenger est produit mais il est soit rapidement dégradé, soit sa traduction est inhibée. Ainsi, le PTGS est un mécanisme épigénétique.

Les petits ARN impliqués dans le PTGS sont appelés siRNA (Small Interfering RNA) ou miRNA (Micro Interfering RNA) selon leur mode de production. En effet, les siRNA proviennent de longs ARN double brins (ARNdb) alors que les siRNA sont produits à partir d'un ARN précurseur d'une centaine de nucléotides de long qui forme une structure en épingle à cheveux (Cullen 2004). Cependant, les mécanismes par lesquels les siRNA et les miRNA induisent le PTGS sont les mêmes. Des ARNdb sont clivés en petits ARNdb par une RNase III, Dicer (Fire, Xu et al. 1998; Bernstein, Caudy et al. 2001). Grâce à l'énergie de l'ATP pour désappairer le duplexe, un seul brin de ces siRNA va être incorporé au complexe RISC (RNA Induced Silencing Complex) auquel appartiennent les protéines argonautes (Fagard, Boutet et al. 2000; Hammond, Bernstein et al. 2000). Le complexe RISC est alors activé et est ciblé, par un simple appariement de bases, vers un ARNm présentant une séquence homologue à celle du siRNA/miRNA. Finalement, l'ARNm est clivé par une activité RNase contenue dans RISC. Cependant, un appariement imparfait entre un petit ARN et sa séquence cible conduirait non pas à la dégradation de l'ARNm mais au blocage de sa traduction (Almeida and Allshire 2005).

#### b) TGS

Les mécanismes d'interférence par ARN peuvent également agir au niveau transcriptionnel. Il s'agit alors de TGS. Globalement, la voie qui permet de réprimer l'expression des gènes de manière post-transcriptionnelle permet également de modifier la chromatine et d'altérer sa structure et donc sa fonction. En effet, des ARNdb sont clivés par Dicer en petits ARNdb. Ces petits ARN vont cependant être intégrés au sein d'un autre complexe, RITS (RNA induced transcriptional silencing) découvert chez *S.pombe* (Verdel, Jia et al. 2004). Le complexe RITS ainsi activé est

ciblé vers l'ADN par appariement de bases où il induit des modifications de la chromatine, la rendant transcriptionnellement inactive. Ce phénomène est particulièrement bien décrit chez *S.pombe* dans la mise en place d'une structure particulière de la chromatine, l'hétérochromatine, dont nous reparlerons dans le second chapitre de cette introduction (Verdel and Moazed 2005).



**Figure 19 : L'interférence par ARN (d'après (Almeida and Allshire 2005)).** Des ARNdb sont clivés en petits ARNdb par Dicer. Ils sont ensuite pris en charge par RISC pour assurer la répression des gènes au niveau post-transcriptionnel (PTGS) en dégradant l'ARNm ou en inhibant sa traduction. Les petits ARNdb peuvent également s'associer à RITS pour induire la répression des gènes au niveau transcriptionnel (TGS), induisant des modifications de la structure de la chromatine en association avec d'autres mécanismes épigénétiques (méthylation de l'ADN, modifications d'histones).

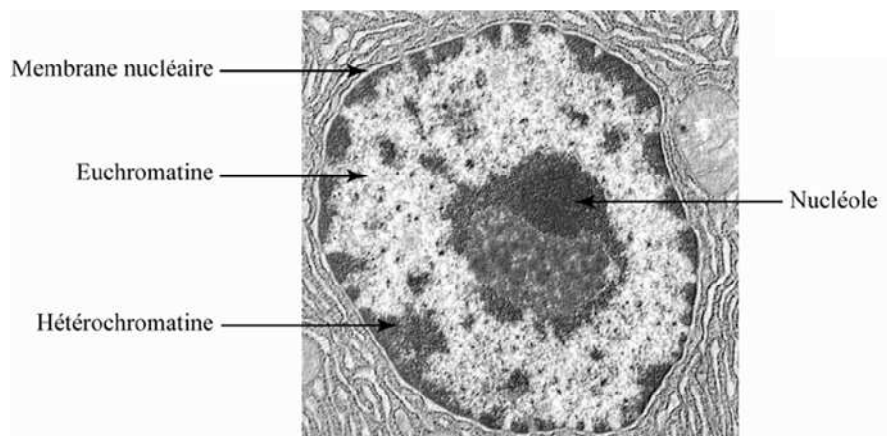
Les ARNnc font partie des mécanismes épigénétiques. Ils régulent en effet la transcription des gènes en agissant au niveau transcriptionnel ou post-transcriptionnel. L'interférence par ARN utilise des ARNdb pour permettre la régulation de la transcription. L'implication de ce processus a donc été étudiée dans les mécanismes d'empreinte parentale et d'inactivation du X. En effet, dans ces deux cas des ARN sens et antisens sont produits (*Igf2R* et *Air*, *xist* et *tsix*). Il a donc été suggéré que les ARNnc *Xist* et *Air*, impliqués dans la répression des gènes, pourraient s'apparier avec leur homologue et ainsi produire sa dégradation. Cependant, l'ARN

*interférence ne semble pas être impliqué dans ces mécanismes (Sado, Wang et al. 2001; Shibata and Lee 2003; Sleutels, Tjon et al. 2003).*



## Chapitre II Centromères & Péricentromères

L'étude de la compaction du génome a permis de différencier deux types de chromatine. En effet, le degré de compaction de l'ADN peut être mis en évidence par l'utilisation de colorants spécifiques. Ainsi, grâce à une technique d'ébullition associée à une coloration au carmin acétique, le botaniste Emile Heitz décrit, en 1928, deux types de chromatines. L'euchromatine ou « chromatine vraie » est modérément colorée en interphase au cours de laquelle elle se décondense. L'hétérochromatine ou « autre chromatine » caractérise au contraire l'ensemble des régions chromosomiques restant condensées, et donc colorées, en interphase (Figure 20).



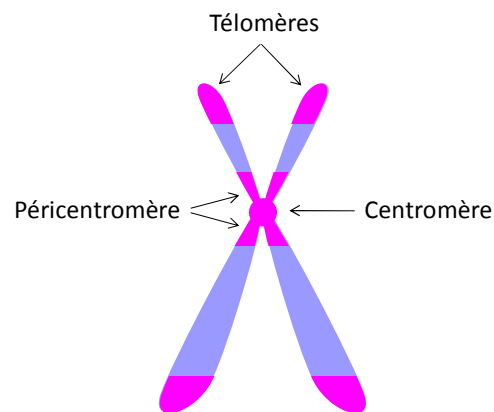
**Figure 20 : Euchromatine et Hétérochromatine (d'après <http://academics.hamilton.edu/biology/kbart/image/nucleus.jpg>).**

L'hétérochromatine est divisée en deux grandes classes :

L'hétérochromatine facultative correspond à un état transitoire et réversible. Elle contient des séquences codantes maintenues sous une forme compacte, transcriptionnellement inactive, selon le stade du développement ou le tissu considéré. Un des exemples les plus caractéristiques illustrant l'hétérochromatine facultative est le corpuscule de Barr. L'inactivation du chromosome X est en effet réversible dans les cellules germinales (Lyon 1962; Brown and Chandra 1973).

Au contraire, l'hétérochromatine constitutive est une forme stable et irréversible de la chromatine quel que soit le stade du développement et le tissu concerné. C'est

ce type d'hétérochromatine que l'on retrouve au niveau de régions particulières des chromosomes (Figure 21). La zone la plus fortement concernée est la constriction primaire des chromosomes, le centromère. Les régions entourant le centromère, appelées régions péricentromériques ou péricentromères, sont également enrichies en hétérochromatine constitutive. Enfin, les régions télomériques, c'est-à-dire les extrémités des chromosomes, et les bras courts des chromosomes acrocentriques (13, 14, 15, 21 et 22) sont également composés d'hétérochromatine constitutive. Cette hétérochromatine constitutive, principal sujet d'étude de cette thèse, n'est pas distribuée de manière aléatoire dans le noyau des cellules interphasiques. En effet, elle est localisée en périphérie du noyau et du nucléole (Rae and Franke 1972; Gagne, Laberge et al. 1973; Stahl, Hartung et al. 1976; Sparvoli, Galli et al. 1977).



**Figure 21 : L'hétérochromatine constitutive.** L'hétérochromatine constitutive (rose) est localisée au niveau des centromères, des péricentromères et des télomères des chromosomes.

Dans ce chapitre, nous présenterons les différences qui caractérisent l'euchromatine et l'hétérochromatine et en particulier la présence d'ADN non codant. Après avoir détaillé l'ensemble des séquences non codantes du génome humain, nous nous intéresserons de plus près à deux de celles-ci, les séquences centromériques et les séquences péricentromériques.

## II.1 Euchromatine et hétérochromatine

### II.1.A Les différences

Depuis la première description par Emile Heitz des deux grands domaines de la chromatine, d'autres études utilisant des colorants de l'ADN, tels que le DAPI ou l'iodure de propidium, ont confirmé la présence d'une chromatine lâche, peu colorée, et d'une chromatine condensée, fortement colorée. Ces caractéristiques histologiques qui ont permis de différencier l'euchromatine de l'hétérochromatine s'expliquent par des propriétés structurales distinctes qui ont des conséquences importantes sur la fonction de ces deux compartiments (Figure 22).

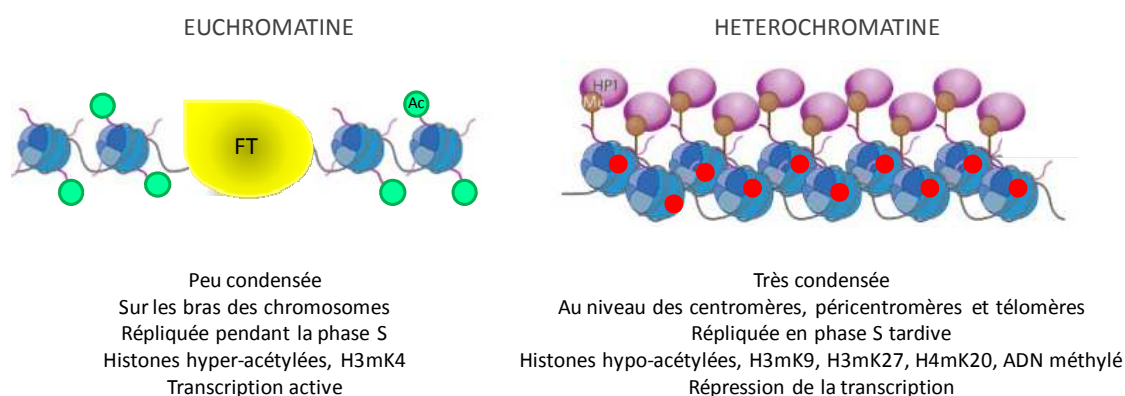
#### 1. Caractéristiques épigénétiques

La grande différence entre l'euchromatine et l'hétérochromatine réside dans le degré de compaction de l'ADN. Cette distinction a pu être faite, non seulement par l'utilisation de colorants de l'ADN, mais également par des études biochimiques. Ainsi, la résistance importante qu'offre l'hétérochromatine aux nucléases (Elgin 1996; Grewal and Elgin 2002) ainsi que la présence au sein de celle-ci d'une longueur internucléosomale réduite (Wallrath and Elgin 1995) attestent en faveur d'un haut degré de compaction de l'hétérochromatine comparativement à l'euchromatine. De plus, les modifications épigénétiques, qui sont impliquées dans le degré de compaction de la chromatine, permettent aisément de discriminer les deux domaines. En effet, l'euchromatine présente une hyperacétylation des histones (H3K9, H3K14, H4K5, H4K8, H4K12 et H4K16) et un ADN hypométhylé au niveau des îlots CpG, résultant en une chromatine peu compacte et donc faiblement marquée par les colorants nucléaires. Au contraire, le caractère chromophile de l'hétérochromatine est lié à son haut degré de compaction, impliquant une hyperméthylation de l'ADN ainsi qu'une hypoacétylation globale des histones (Miller, Schnedl et al. 1974; Jeppesen, Mitchell et al. 1992).

L'implication de la méthylation de l'ADN dans la compaction de la chromatine est mise en évidence par l'utilisation d'un analogue de la cytidine qui empêche la méthylation de l'ADN, la 5 Aza-Cytidine. Un traitement des cellules à la 5 Aza-Cytidine conduit à une hypométhylation de l'ADN particulièrement visible au sein des régions hétérochromatiques qui perdent alors leur haut degré de compaction et se décondensent (de Capoa, Menendez et al. 1996; Haaf and Schmid 2000). D'autre

part, des mutations dans les gènes codant pour les DNMT, les enzymes responsables de la méthylation de l'ADN, ont pour conséquence une hypométhylation et une décondensation de l'ADN qui affecte en premier lieu les régions hétérochromatiques. C'est en effet les caractéristiques histologiques observées dans les cellules en culture issues de patients atteints par le syndrome ICF (Ehrlich, Sanchez et al. 2008).

La relation entre l'hypoacétylation des histones et la compaction de l'ADN est également bien documentée. En effet, décrite au cours du chapitre précédent, l'acétylation des histones conduit à la neutralisation des charges positives des lysines ayant pour conséquence une configuration moins compacte de la chromatine. Ainsi, la déacétylation des histones favorise la formation d'une structure chromatiniennne plus compacte (Annunziato, Frado et al. 1988; Tse, Sera et al. 1998; Walia, Chen et al. 1998). En particulier, il a été récemment montré que l'acétylation de la lysine 16 de H4 réduit la capacité des nucléosomes à former des structures compactes *in vitro* et que certaines HDAC sont directement impliquées dans la condensation de la chromatine (Shogren-Knaak, Ishii et al. 2006; Parker, Maxson et al. 2007).



**Figure 22 : Euchromatine et Hétérochromatine (d'après (Grewal and Elgin 2007)).** Les différentes caractéristiques de l'euchromatine et de l'hétérochromatine sont indiquées. FT=Facteur de Transcription ; Ac=histone acétylée ; Me=histone méthylée ; HP1= Heterochromatin Protein 1 ; rond rouge= méthylation de l'ADN.

En plus d'être globalement hypoacétylée, l'hétérochromatine constitutive présente un schéma de méthylation des histones assez caractéristique: la triméthylation d' H3 sur la lysine 9 (H3K9), la triméthylation d' H4 sur la lysine 20 (H4K20) et la monométhylation d' H3 sur la lysine 27 (H3K27) (Peters, Kubicek et al. 2003; Kourmouli, Jeppesen et al. 2004). L'euchromatine est quant à elle méthylée sur H3K4. Certaines protéines non histones reconnaissent spécifiquement les marques hétérochromatiques et sont donc retrouvées dans ce domaine. La plus connue de ces

protéines est HP1 (Heterochromatin Protein 1) qui se lie de manière spécifique à H3K9 tri-méthylée. Cette protéine est essentielle à la formation d'une structure compacte de la chromatine (Strahl and Allis 2000; Jenuwein and Allis 2001; Lachner, O'Sullivan et al. 2003). Les détails concernant la fonction de cette protéine sont abordés dans la suite de ce chapitre.

Enfin, certains variants d'histones sont spécifiquement localisés au niveau de régions hétérochromatiques. C'est en effet le cas de CENPA, le variant de l'histone H3, qui est exclusivement présent au niveau des centromères. La fonction de ce variant d'histone est également discutée dans la suite de ce chapitre.

## 2. *Transcription*

Le cycle cellulaire se décompose en 4 phases distinctes : la phase G1, la phase S (Synthèse de l'ADN), la phase G2 et la phase M (Mitose). Comme son nom l'indique, c'est au cours de la phase S qu'a lieu la synthèse de l'ADN, c'est à dire sa réplication, avant sa répartition homogène lors de la phase M entre les deux cellules filles. Le moment de la réplication de l'ADN diffère selon la chromatine considérée. En effet, au cours de la phase S, l'euchromatine se réplique avant l'hétérochromatine (Schubeler, Scalzo et al. 2002; Woodfine, Fiegler et al. 2004). Cette différence peut s'expliquer par le fort degré de compaction de l'hétérochromatine qui serait un frein à l'accès de la machinerie de réplication. Cependant, la machinerie de réplication n'est pas la seule à avoir quelques difficultés à se frayer un chemin vers l'hétérochromatine. En effet, un des aspects les plus importants concernant l'hétérochromatine réside dans le fait qu'elle a été longtemps considérée comme une région transcriptionnellement inerte. Historiquement, cette caractérisation est liée à plusieurs observations.

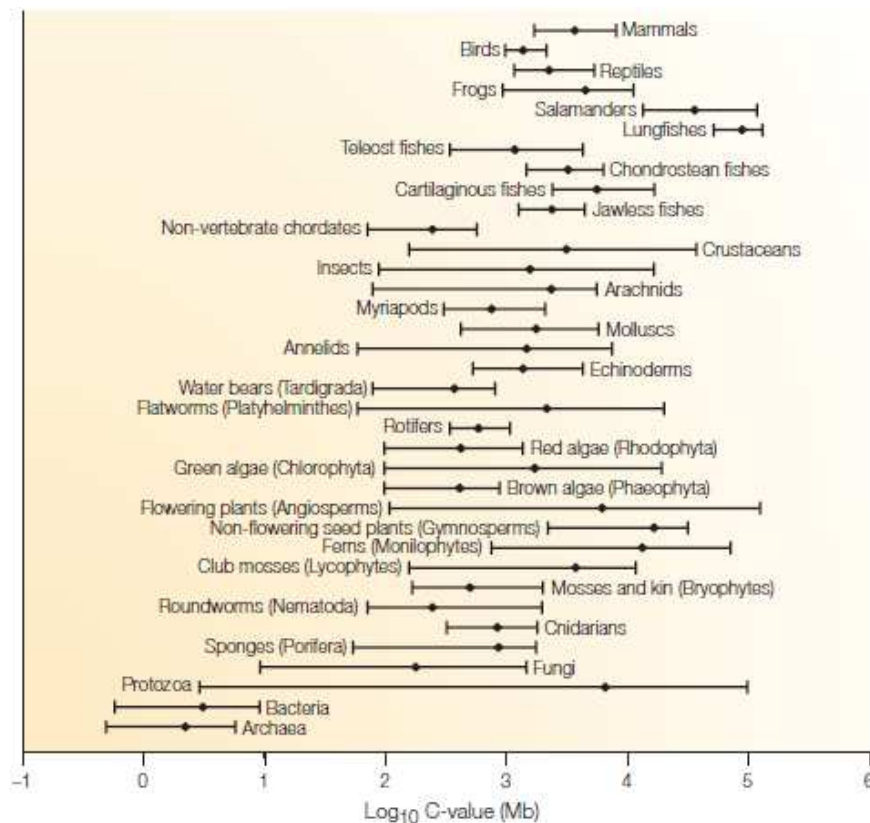
Il existe en effet une corrélation entre l'état de la chromatine et l'activité génétique. L'euchromatine, qui est constituée de chromatine peu compacte, est facilement accessible par la machinerie de transcription. Elle est donc considérée comme permissive à la transcription. Au contraire, le haut degré de condensation de l'hétérochromatine suggère qu'elle demeure relativement inaccessible aux facteurs de transcription. Plusieurs exemples attestent du fait qu'une chromatine condensée est transcriptionnellement inactive. En effet, en règle générale, l'ADN est condensé en anaphase et en métaphase, étapes du cycle cellulaire où la transcription n'a pas lieu. Par contre, la décondensation progressive de l'ADN au cours de la télophase est associée à une reprise de la transcription (Taylor 1960; Prescott and Bender 1962). D'autre part, la condensation de l'ADN peut être observée suite à un traitement des cellules à l'actinomycine D, un inhibiteur de la transcription (Izawa, Allefrey et al.

1963). D'autres observations, en particulier sur les chromosomes polyténiques de drosophile, révèlent que l'hétérochromatine n'est jamais présente dans les régions où se trouvent les bandes caractéristiques des gènes mendéliens. Par autoradiographie après marquage à l'uridine tritiée, Frenster et al. montrent que l'hétérochromatine n'est pas le site d'une transcription active dans des lymphocytes de veau (Frenster, Allfrey et al. 1963). Des observations similaires ont été par la suite réalisées chez le campagnol et la caille (Sieger, Pera et al. 1970; Mirre and Stahl 1978). D'autre part, contrairement à l'euchromatine, les délétions de l'hétérochromatine ne sont généralement pas létales et la duplication de telles régions n'engendre pas de dérégulation génétique, suggérant que l'hétérochromatine ne contient pas de gènes importants. Enfin, un dernier aspect de l'hétérochromatine atteste en faveur de son hostilité à la transcription : l'effet PEV (Position Effet Variegation). C'est un mécanisme par lequel l'hétérochromatine impose une répression transcriptionnelle aux gènes qui y sont relocalisés (Zhimulev, Belyaeva et al. 1988; Henikoff and Dreesen 1989; Talbert, LeCiel et al. 1994; Henikoff, Jackson et al. 1995; Sun, Cuaycong et al. 2001). Il est maintenant connu que cette inactivation, provoquée par l'hétérochromatine, implique des mécanismes épigénétiques (Ebert, Lein et al. 2006).

L'hétérochromatine a ainsi longtemps été considérée comme réfractaire à la transcription. Dans les années 70, le « Dogme Central » de la génétique, selon lequel un gène code pour un ARN, lui-même traduit en protéine qui elle seule assure une fonction, était bien établi (Figure 2). Ainsi, l'hétérochromatine qui n'est pas transcrite ni traduite est considérée comme inutile. L'hétérochromatine a alors été surnommée dès 1970 « garbage DNA » ou « junk DNA » qualifiant ainsi cet ADN délesté de fonction éventuelle de déchet génomique (Ohno 1972; Dover 1980; Orgel and Crick 1980).

### II.1.B Hétérochromatine et génomes

Le contenu en ADN à l'état haploïde est une valeur constante au sein des cellules d'un même organisme. Ce contenu, appelé « valeur C », confirmait parfaitement le rôle de molécule de l'hérédité qui était attribué à l'ADN. Cela a conduit les biologistes à penser que la complexité d'un organisme pouvait être liée à la taille de son génome, c'est à dire à la « valeur C ».



**Figure 23 : La « Valeur C » dans différents organismes (Gregory 2005)**

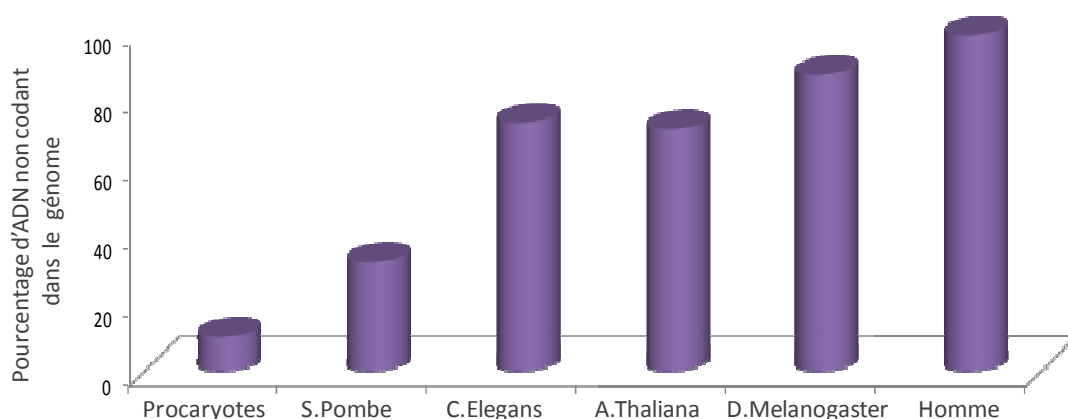
Cependant, la « valeur C » est rarement proportionnelle au nombre de gènes et donc à la complexité des organismes. En effet, si le nombre de gènes et la taille du génome sont parfaitement corrélés chez les procaryotes, il n'en est pas de même pour les eucaryotes (Figure 23). Ainsi, des organismes aussi simples que la salamandre peuvent avoir une « valeur C » 70 fois supérieure à celle d'un organisme plus complexe tel que le coq (Gregory 2005). C'est le paradoxe de la valeur C: il n'existe pas de relation entre la complexité d'un organisme et la taille de son génome.

Mais qu'en est-il des séquences non codantes ?

### 1. Comparaison des génomes

Une comparaison de la proportion globale d'ADN non codant (ADNnc) a pu être réalisée pour différentes espèces. Chez les procaryotes, tels que les bactéries, quasiment l'intégralité du génome, de l'ordre de 90%, est codant, (ce chiffre varie entre 97% pour *B.subtilis* et 49,5% pour *M.Leprae*). Ainsi, les quelques 10% de séquences non codantes fonctionnent pour la plupart comme des éléments régulateurs qui contrôlent l'expression des gènes au niveau transcriptionnel et traductionnel. Une

infime partie de ces séquences non codantes peuvent être transcrites en ARN non codants qui assurent des fonctions régulatrices (Gottesman 2005).



**Figure 24 : Evolution du pourcentage d'ADN non codant dans différents génomes.**

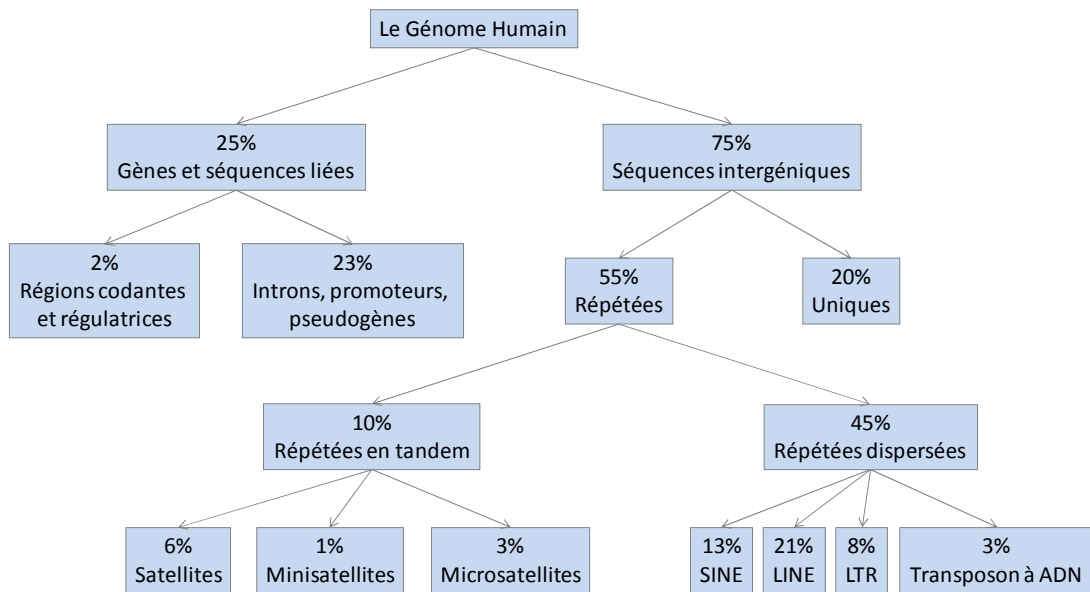
Au fil de l'évolution, la quantité d'ADN codant présent dans les différents organismes diminue (Figure 24). Bien évidemment, les organismes plus complexes possèdent une grande variété de processus de régulation, tels que les modifications épigénétiques ou l'épissage alternatif, qui ajoutent un degré de complexité. Cependant, la proportion d'ADN codant chez certaines espèces et notamment chez l'Homme reste quelque chose de tout à fait surprenant. En effet, la fraction codante de l'ADN est de 68% chez la levure *S.pombe*, 27% chez le nématode *C.elegans*, 29% pour la plante *A.thaliana*, 13% pour la mouche *D.melanogaster* et seulement 1,4% pour l'Homme! Ainsi, de manière surprenante, il existe une relation très claire entre la complexité des organismes et la quantité d'ADN non codant qu'ils possèdent. Nous allons à présent nous intéresser plus en détail à ces séquences non codantes ainsi qu'à leur classification.

## 2. Les différentes classes d'ADNnc du génome humain

Chez l'Homme près de 98 % du génome ne contient aucun gène codant pour des protéines. Parmi les séquences non codantes, 23% sont des séquences intergéniques (introns, pseudogènes). Les gènes sont en effet constitués de régions exprimées appelées « exons » intercalées avec des régions non codantes, les « introns ». Les introns, qui ne sont pas présent chez les procaryotes, sont transcrits puis éliminés par un mécanisme d'épissage lors de la maturation de l'ARN. Les pseudogènes résultent de la transcription inverse d'un ARN provenant d'un gène normal. Le l'ADN complémentaire (ADNc) ainsi obtenu est inséré en plusieurs copies dans le génome.



Ainsi les pseudogènes ressemblent à leur gène d'origine sans toutefois posséder d'éléments régulateurs ni d'introns.



**Figure 25 : Composition du génome humain (d'après (Jasinska and Krzyzosiak 2004)).**

Parmi les 98% de séquences non codantes, 75% sont des séquences intergéniques qui se distinguent en deux classes: des séquences uniques ou très peu répétées (20 % du génome) et des séquences répétées (55 % du génome). Parmi ces séquences répétées, on compte d'une part des séquences répétées en tandem qui sont de grands blocs d'ADN (environ 10 % du génome humain) et, d'autre part, des séquences répétées dispersées (45 % du génome) (Figure 25). Ce sont ces 2 grandes catégories de séquences répétées qui sont détaillées dans la suite de ce chapitre.

a) Les séquences répétées en tandem

Les séquences répétées en tandem, également appelées séquences satellites, représentent 10% du génome humain. Elles sont de taille variable et correspondent à des répétitions en cis d'un motif unitaire lui-même de longueur variable. Ainsi, selon la taille du motif et la longueur occupée par les répétitions, trois grandes catégories de séquences répétées en tandem peuvent se distinguer : les satellites, les minisatellites et les microsatellites dont le motif répété est compris entre deux et plusieurs centaines de paires de bases.

- Les satellites

Les satellites sont les séquences les plus représentées parmi les séquences répétées en tandem (Lee, Wevrick et al. 1997). Ils représentent en effet 6% du génome humain et sont divisés en six familles : les satellites 1, 2, 3, alpha, beta et gamma (Table 1).

L'ADN satellite a été mis en évidence par centrifugation sur gradient de chlorure de césium (Corneo, Ginelli et al. 1967; Corneo, Ginelli et al. 1968; Corneo, Ginelli et al. 1970; Corneo, Ginelli et al. 1971). En effet sa composition en base lui confère des propriétés de sédimentation qui divergent de celles de l'ADN génomique classique. L'ADN sédimente en effet sous la forme d'une bande unique accompagnée de bandes mineures, satellites, d'où leur nom. Ces trois bandes satellites I, II et III sont chacune composées d'une famille prédominante appelées satellites 1, 2 et 3 (Prosser, Frommer et al. 1986). Les satellites 1 sont retrouvés au niveau des chromosomes 3, 4 ainsi que sur les chromosomes acrocentriques (13, 14, 15, 21 et 22). Les satellites 2 et 3 sont majoritairement présents au niveau des régions péri-centromériques des chromosomes. Ces séquences satellites sont décrites en détail dans la suite de ce chapitre. Les alpha satellites sont présents au niveau des centromères de tous les chromosomes. Ils sont constitués de monomères de 171 paires de bases (pb), riches en GC et ils sont également présentés dans la suite du ce chapitre.

	MOTIF	LOCALISATION
Alpha satellites	170pb	Centromères
Beta satellites	68 pb	Péricentromères 1, 3, 9, 13, 14, 15, 21, 22, Y
Gamma Satellites	220pb	Centromères 8, Y
Satellites 1	42pb	Péricentromères 3, 4, 13, 14, 15, 21, 22
Satellites 2	5pb	Péricentromères 1, 2, 7, 10, 15, 16, 17, 22
Satellites 3	5pb	Péricentromères 1, 5, 10, 13, 13, 15, 17, 20, 21, 22, Y

**Table 1 : Les séquences satellites.**

Les beta satellites ont pu être isolés et caractérisés grâce à l'enzyme de restriction Sau3A (Agresti, Rainaldi et al. 1987; Waye and Willard 1989). Les beta satellites sont constitués de monomères de 68bp riches en GC (46%) mais hautement divergents. Ils sont localisés au niveau des régions péri-centromériques des chromosomes 9, 1 et 3, des bras courts des chromosomes acrocentriques ainsi qu'au

niveau du bras long du chromosome Y (Agresti, Meneveri et al. 1989; Cooper, Fisher et al. 1992; Meneveri, Agresti et al. 1993).

Les gamma satellites ont été identifiés pour la première fois sur le chromosome 8, comme une série de répétitions de monomères de 220pb, puis sur le chromosome X (Lin, Sasi et al. 1993; Lee, Li et al. 1995). Ces séquences représentent respectivement 0.013% et 0.015% du génome humain. Les séquences consensus diffèrent pour ces deux chromosomes. Cependant, elles possèdent tout de même 68% d'homologie entre elles. En particulier, un motif de 20pb est particulièrement conservé (TCAGGGGGACGTTGAGGCAG) ainsi que 8 motifs tétramériques ((G/C)AGG). Ces séquences sont localisées au niveau des centromères des chromosomes 8 et X, mais ne présentent aucune homologie avec les alpha satellites.

- Les minisatellites

Outre les grands blocs d'ADN satellite, on trouve, dispersés dans le génome, des petits blocs de motifs répétés en série : les minisatellites. Ils représentent 1% du génome humain. Ils sont constitués d'un motif compris entre 6 et 100 pb. Il existe deux grandes catégories de minisatellites.

- Les séquences minisatellites hypervariables présentent un haut degré de polymorphisme tant par la taille des unités répétées que par le nombre de répétitions. Cependant ces unités de répétitions, de 9 à 24pb, ont une séquence commune (GGGCAGGANG).

- Les séquences télomériques sont composées d'un motif hexanucléotidique répété (TTAGGG) (Meyne, Ratliff et al. 1989). Comme leur nom l'indique, elles sont localisées au niveau des extrémités des chromosomes, les télomères. Elles auraient pour rôle de protéger ces extrémités de la dégradation. En effet, l'ARN polymérase n'est pas capable de répliquer correctement les extrémités 3' des chromosomes. Ainsi, à chaque réplication, une partie des séquences télomériques est perdue. Quand la longueur des télomères devient critique, les cellules ne se répliquent plus et entrent en sénescence. Dans les cellules germinales, un mécanisme de protection est mis en place pour permettre l'élongation des télomères. La télomérase est une transcriptase inverse qui permet d'allonger les extrémités des télomères (Greider and Blackburn 1985). Elle est composée de deux éléments, TERT (telomerase reverse transcriptase) et TERC (telomerase RNA template). Alors que TERC est exprimée de manière ubiquitaire, l'expression de TERT est finement régulée et agit donc comme facteur limitant. Ainsi, dans les cellules somatiques adultes normales, l'activité de la

téломérase n'est plus détectée, à l'exception des cellules cancéreuses (Kim, Piatyszek et al. 1994).



**Figure 26 : Les télomères et la T-Loop.** La T-Loop se forme à l'extrémité des télomères par appariement de séquences entre l'extension du brin 3' (brin rouge) et sa séquence complémentaire.

Dans les cellules somatiques, des structures particulières sont donc mises en place afin de protéger les extrémités des télomères : les T-Loop (boucle télomérique) (Griffith, Comeau et al. 1999). Ces structures protectrices sont constituées à la fois de séquences télomériques et de protéines. En effet, les télomères possèdent une extension du brin 3' comprise en 150 et 200pb (Wright, Tesmer et al. 1997). La T-Loop va être créée par le repliement de cette extension télomérique qui s'apparie avec sa séquence complémentaire (Figure 26). Cette structure est par la suite renforcée par de nombreux complexes protéiques (Bianchi and Shore 2008).

- Les microsatellites

Les microsatellites représentent 3% du génome humain. Ce sont des répétitions d'un motif assez court, de 1 à 6 pb, répété un nombre limité de fois de telle sorte que la séquence totale n'excède généralement pas 150pb. La grande majorité de ces répétitions consiste en des dinucléotides (0.5% du génome) dont les plus fréquents sont AC et AT (Table 2). Ils sont classés en 3 catégories : les microsatellites parfaits qui consistent en des répétitions ininterrompues (par exemple  $(GTC)_{15}$ ), les microsatellites imparfaits qui sont composés de séquences interrompues par de courtes séquences non répétées (par exemple  $(GTC)_7$  CTCTG  $(GTC)_8$ ) et les microsatellites composés qui comme leur nom l'indique sont composés de plusieurs unités répétées différentes (par exemple  $(GTC)_3(TC)_2$ ) (Sharma, Grover et al. 2007). Contrairement aux satellites et minisatellites, les microsatellites ne sont pas localisés à des endroits précis du génome, mais ils sont dispersés. Ainsi des microsatellites

peuvent être présents au sein d'introns, d'exons ou d'autres types de séquences répétées (Li, Korol et al. 2004).

Catégorie	Motif	Quantité (% par rapport à la catégorie)
Dinucléotides	AC	50
	AT	35
	AG	15
	GC	0,1
Trinucléotides	AAT	33
	AAC	21
	ACC	4
	AGC	2
	ACT	1,5
	ACG	0,1

**Table 2 : Les microsatellites.**

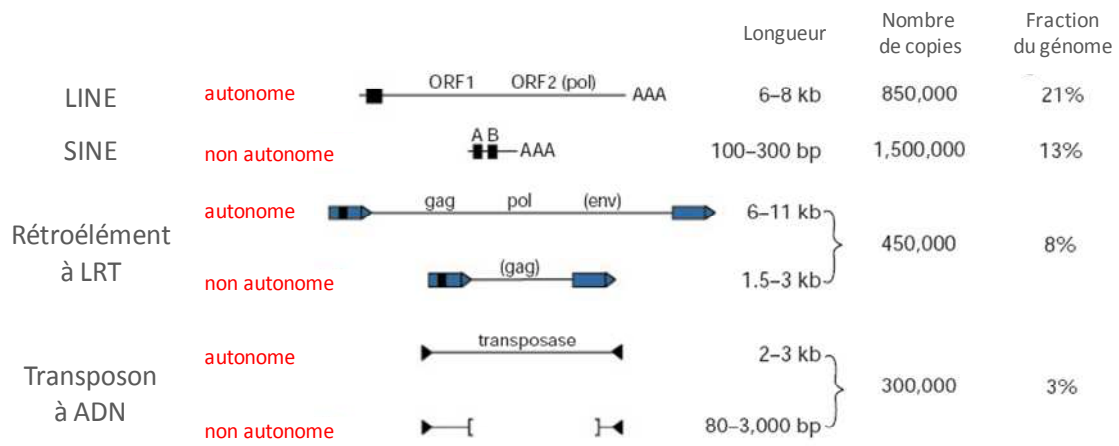
Le polymorphisme de ces microsatellites est lié au nombre de répétitions, à la taille des unités répétées mais également au motif répété. Des erreurs lors de la réplication de l'ADN (addition ou délétion d'unités de répétition) dues au glissement de l'ARN polymérase seraient à l'origine de ce polymorphisme (Freimer and Slatkin 1996). Les microsatellites constituent ainsi des marqueurs de choix pour établir des cartes génétiques. Ils sont également utilisés comme marqueurs pour l'identification d'individus, par exemple lors des tests de paternité.

Les microsatellites sont impliqués dans de nombreuses pathologies humaines. Le syndrome du X fragile est en effet associé à un grand nombre de répétitions CGG (entre 200 et 1000 fois) alors que les sujets sains en présentent un nombre réduit (entre 6 et 54 fois) (Lukusa and Fryns 2008). D'autre part, l'insertion de triplets de microsatellites dans les régions codantes est associée à des pathologies neurodégénératives (Bowater and Wells 2001).

#### b) Les éléments répétés dispersés

Les éléments répétés dispersés sont communément appelés transposons. Le terme transposon caractérise des séquences d'ADN pouvant se déplacer dans le

génomique. Ce déplacement se fait soit sous forme ADN (Transposons à ADN), soit sous forme ARN (Rétroéléments à LTR, SINE, LINE) (Figure 27). Dans ce dernier cas, le déplacement est appelé rétrotransposition. Chez les vertébrés, la plupart des éléments transposables connus sont des rétroéléments (93% des séquences répétées dispersées) qui sont dispersés dans les introns et à proximité des gènes. Les rétroéléments non LTR ont été classés en fonction de leur longueur: les éléments de quelques centaines de nucléotides ont été appelés SINE (Short INterspersed Elements), ceux de plusieurs milliers de nucléotides LINE (Long INterspersed Elements) (Lander, Linton et al. 2001).



**Figure 27 : Les différentes classes de séquences répétées dispersées dans le génome humain (d'après (Lander, Linton et al. 2001)).**

- LINE

Les rétroéléments LINE ont une longueur comprise entre 6 et 8kb de long et ils représentent 21% du génome humain, soit plus de dix fois l'ensemble des séquences codantes ! Ce sont des rétroéléments autonomes puisqu'ils codent eux-mêmes pour leur propre transposition. Ils contiennent le promoteur de l'ARN polymérase II qui code pour deux cadres de lecture (ORF) dont un correspond à la reverse transcriptase. Une fois traduit, l'ARN LINE s'assemble autour de ces deux protéines et il est importé dans le noyau où il s'insère dans l'ADN. Les nouveaux sites d'insertions sont entourés par des duplications allant de 7 à 20 pb. Il existe trois familles de LINE : LINE 1, LINE 2 et LINE 3 dont seul LINE 1 est actif et est le plus représenté (17% du génome humain). Les éléments LINE, bien que dispersés, sont préférentiellement localisés dans les séquences riches en AT (Lander, Linton et al. 2001).

- SINE

Les rétroéléments SINE sont des éléments courts, dont la taille varie entre 100 et 300 paires de bases. Ils possèdent un promoteur interne pour l'ARN polymérase III mais ne codent pour aucune protéine. Ainsi, leur transposition est un événement opportuniste puisqu'ils ne possèdent pas le gène de la transcriptase inverse et doivent par conséquent profiter de la synthèse de cette enzyme par un autre rétroélément, l'élément LINE. Ils sont donc qualifiés de non autonomes. Il existe trois familles de SINE : Alu, Mir et Ther2/Mir3. Seul l'élément Alu, qui est spécifique aux primates, est actif. Cet élément de 300 pb représente à lui seul 10% du génome humain. Les éléments SINE sont préférentiellement localisés dans les régions riches en GC (Lander, Linton et al. 2001).

- Rétroélément à LTR

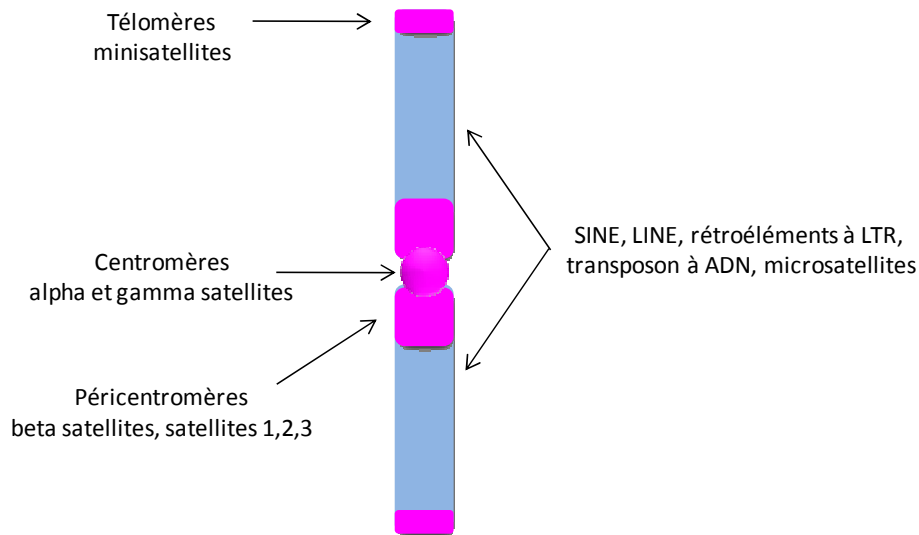
Les rétroéléments à LTR (Long Terminal Repeat), sont composés de longues répétitions terminales qui contiennent les éléments régulateurs nécessaires à la transcription. Ces éléments autonomes possèdent les gènes gag et pol qui codent pour la protéase, la reverse transcriptase, la RNaseH et l'intégrase. Ils se différencient des rétrovirus par l'absence de gène env nécessaire à la formation de l'enveloppe. Il existe une grande variété de rétroéléments à LTR mais seule la famille ERV (Endogenous RetroVirus), spécifique aux vertébrés, est active. Dans 85% des cas, les rétroéléments à LTR consistent seulement en une séquence LTR unique, la séquence interne ayant été perdue par recombinaison homologue entre les LTR. Contrairement aux séquences SINE et LINE, la distribution des rétroéléments à LTR est plutôt homogène au sein du génome (Lander, Linton et al. 2001).

- Les transposons à ADN

Les transposons à ADN ressemblent aux transposons bactériens. Ils possèdent des séquences terminales répétées inversées et codent pour la transposase. Leur mobilité est assurée par un mécanisme de « couper-coller » réalisée par la transposase qui se lie à proximité des séquences terminales. Ils sont subdivisés en plusieurs familles dont les deux groupes hAT et Te-1 qui comprennent respectivement les éléments MER1-Charlie, Zaphod et MER2-Tigger, Tc2, Mariner. Comme les rétroéléments à LTR, les transposons à ADN sont répartis de manière homogène dans le génome (Lander, Linton et al. 2001).

*Les séquences répétées représentent plus de la moitié du génome humain. Certaines d'entre elles sont dispersées le long des chromosomes mais d'autres sont*

localisées à des endroits précis (Figure 28). Ainsi, les alpha satellites forment les séquences centromériques et les satellites 1, 2 et 3 les séquences péricentromériques. Dans la suite de ce chapitre, nous allons voir que ces séquences font parties de deux domaines aux fonctions liées mais néanmoins distinctes : le centromère et le péricentromère.



**Figure 28 : Localisation des séquences répétées du génome humain.**

## II.2 Centromères et péricentromères

Les centromères (CT) et ses régions flaquantes, les péricentromères (PCT), possèdent les caractéristiques de l'hétérochromatine constitutive telles qu'elles ont été définies au chapitre précédent. Cependant, plusieurs particularités permettent de distinguer ces deux domaines. Ainsi la taille de ces régions peut être distinctive. Par exemple, les PCT murins s'étalent sur plus de 6 méga bases alors que les CT sont contenus dans 600kb. D'autre part, si d'une manière générale, l'hétérochromatine se réplique plus tardivement que l'euchromatine, il a été montré que les PCT murins se répliquent en milieu de phase S contrairement aux CT dont la réplication est plus tardive (Guenatri, Bailly et al. 2004). Un décalage temporel de la réplication a également été observé pour d'autres espèces, à l'exception de *S.cerevisiae* (Ten Hagen, Gilbert et al. 1990; O'Keefe, Henderson et al. 1992; Shelby, Monier et al. 2000; Ahmad and Henikoff 2001; Sullivan and Karpen 2001; Kim, Dubey et al. 2003). Comme nous le verrons dans la suite de ce chapitre, le type de séquences répétées,



les modifications épigénétiques et la fonction des CT et des PCT sont également différents. Pour finir, une analyse comparative des CT et des PCT de différents organismes sera présentée.

## II.2.A Les centromères

La chromatine centromérique est différente du reste des chromosomes : elle forme le centromère qui est la constriction primaire des chromosomes. Ce centromère, qui joint les deux bras des chromosomes, joue un rôle fondamental dans la ségrégation de ceux-ci au cours de la mitose et de la méiose. C'est en effet au niveau du centromère que se forme le kinétochore, la structure protéique responsable de l'attachement des microtubules qui permet le mouvement des chromosomes vers les deux cellules filles (Vagnarelli, Ribeiro et al. 2008). Des erreurs de répartition des chromosomes peuvent avoir des conséquences dramatiques pour la survie de la cellule ou de l'organisme.

### 1. Les séquences génomiques

Le kinétochore s'assemble sur une structure chromatinienne particulière localisée au niveau du centromère. Cette caractéristique, conservée au cours de l'évolution, laissait penser à l'existence au sein du centromère d'une séquence génomique spécifique. Celle-ci a longtemps été recherchée en vain. En effet, même si la séquence génomique centromérique est nécessaire et suffisante à la formation du kinétochore chez *S.cerevisiae*, ce n'est pas le cas chez les autres eucaryotes (Amor, Kalitsis et al. 2004). Cependant, à l'exception de *S.cerevisiae*, les eucaryotes contiennent au niveau de leur centromère un ADN répété (Tyler-Smith and Florida 2000). Ces séquences, au sein desquelles on retrouve également des séquences répétées dispersées, sont appelées :

- alpha satellites chez l'Homme
- satellites mineurs chez la souris
- régions AATAT et TTCTC chez la drosophile
- central core chez la levure *S.pombe*

Toutefois, la présence de séquences répétées ne suffit pas à assurer la fonction centromérique (Craig, Wong et al. 2003). En effet des études portant sur des chromosomes anormaux, caractérisés par une perte du centromère, montrent que de

nouveaux centromères peuvent se former sur des régions dépourvues d'alpha satellites mais présentant néanmoins une séquence riche en AT (du Sart, Cancilla et al. 1997). La richesse des centromères en AT est un trait commun à toutes les espèces. En effet, si le centromère de *S.cerevisiae* n'est pas composé de séquences répétées, les trois éléments Cde qui forment son centromère sont enrichis en AT (Clarke and Carbon 1985).

## 2. Les modifications épigénétiques

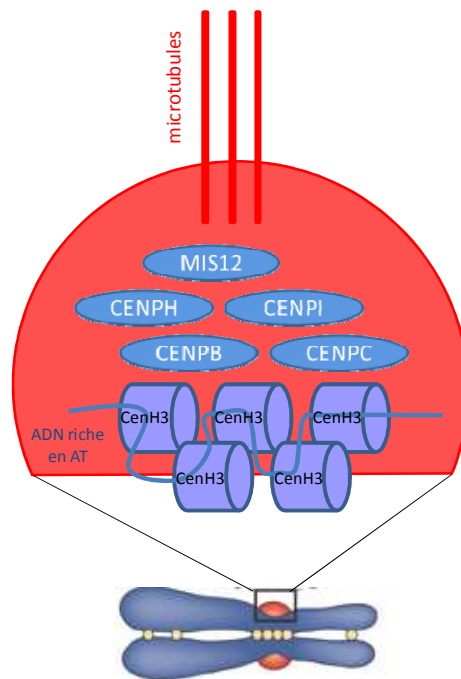
La structure des centromères ne reposant pas exclusivement sur la présence de séquences répétées, les regards se sont alors tournés vers l'implication possible de mécanismes épigénétiques. Il existe en effet un variant d'histone spécifique des centromères, CenH3. Il est nommé de manière différente selon les espèces considérées :

- hCENPA chez l'Homme
- mCENPA chez la souris
- cid chez la drosophile
- CNP1 chez *S.pombe*
- Cse4 chez *S.cerevisiae*

C'est la seule protéine centromérique qui existe chez toutes les espèces. La déplétion de CenH3 chez la levure, la drosophile, la souris ou l'Homme conduit à des défauts sévères de ségrégation des chromosomes et de localisation centromérique des autres protéines du kinétochore (Stoler, Keith et al. 1995; Valdivia, Figueroa et al. 1998; Howman, Fowler et al. 2000; Blower and Karpen 2001). De plus, des souris déficientes pour CenH3 meurent au stade embryonnaire indiquant que CenH3 aurait un rôle fondamental dans la structure du kinétochore. Bien que CenH3 soit nécessaire au recrutement d'autres protéines du kinétochore comme CENPC, elle n'est pas suffisante pour la formation d'un centromère fonctionnel. En effet, lorsque CenH3 est surexprimée, 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, Ouspenski et al. 2001).

Néanmoins, il est important de noter que l'incorporation de CENH3 dans la chromatine centromérique n'est pas exclusive. En effet, une fraction de l'histone H3

est également présente au niveau des centromères (Blower, Sullivan et al. 2002). La différence entre ces deux histones réside dans la partie N-terminale, celle-là même où ont lieu les modifications épigénétiques. La seule modification épigénétique décrite pour CENH3 est la phosphorylation de la sérine 7 par Aurora B (Zeitlin, Shelby et al. 2001). Cette modification n'est pas présente chez la souris et il est donc peu probable qu'elle joue un rôle important dans la fonction du centromère. Cependant d'autres modifications épigénétiques permettent de distinguer les CT des PCT. En effet, l'histone H3 est diméthylée sur la lysine 4, contrairement aux péricentromères qui possèdent une triméthylation de la lysine 9 (Sullivan and Karpen 2004). La présence de cette marque typique des régions euchromatiques au sein de l'hétérochromatine centromérique est surprenante. Cependant d'autres marques euchromatiques comme l'hyperacétylation d'H3 et H4 ne sont pas présentes sur le centromère (Jeppesen, Mitchell et al. 1992). Ainsi, la chromatine centromérique est différente de la chromatine péricentromérique mais également de l'euchromatine. De plus, son organisation est complexe. En effet, la diméthylation de H3K4 est présente sur la face interne du centromère, là où les chromatides sœurs sont jointes alors que CENH3 est localisée sur la face externe, là où se forme le kinétochore (Sullivan and Karpen 2004) (Figure 29).



**Figure 29 : Centromère et Kinétochore (d'après (Musacchio and Salmon 2007)).** La chromatine centromérique est enrichie en séquences « AT » ainsi qu'en CENH3. Les séquences centromériques ainsi que de nombreux complexes protéiques (dont CENPB, CENPC, CENPH, CENPI et MIS12) forment le kinétochore. Cette structure permet l'attachement des microtubules au chromosome lors de la mitose.

### 3. Les protéines centromériques

CenH3 appartient à un complexe protéique relativement bien conservé au cours de l'évolution, le kinétochore. Le kinétochore regroupe chez l'Homme, des protéines fondatrices telles que CENPB, CENPC, CENPH, CENPI et Mis12 (Figure 29). Ce complexe est localisé sur le centromère tout au long du cycle cellulaire. Au cours de la mitose, d'autres protéines (protéines de « checkpoint » ou protéines motrices) sont relocalisées au centromère pour le rendre tout à fait fonctionnel (Fukagawa 2004).

- CENPB

C'est la seule protéine centromérique qui possède une séquence de liaison spécifique (un motif de 17pb appelé « CENPB box ») présente dans les satellites mineurs de souris et les alpha satellites chez l'Homme (Masumoto, Masukata et al. 1989). Bien que la fonction de CENPB dans la formation du centromère et/ou du kinétochore soit inconnue, on sait qu'elle n'y est pas essentielle. En effet des souris déficientes en CENPB ne présentent pas de défaut mitotique et sont viables, contrairement aux souris déficientes en CenH3 (Hudson, Fowler et al. 1998). De plus les chromosomes Y murins et humains ne contiennent pas de CENPB box dans leur centromère (Craig, Earnshaw et al. 1999). Enfin, contrairement à CenH3, CENPB se lie aussi bien aux centromères actifs qu'inactifs. La formation de centromères *de novo* nécessite cependant la présence d'alpha satellites et de la CENPB box (Ohzeki, Nakano et al. 2002). Il est également à noter que chez *S.pombe*, les homologues de CENPB ne se lient pas au niveau des centromères mais plutôt dans l'hétérochromatine flanquante (Baum and Clarke 2000). Curieusement, c'est uniquement pour cette espèce qu'une corrélation entre des problèmes de ségrégation des chromosomes et une déplétion de CENPB a été décrite (Irelan, Gutkin et al. 2001). Ainsi, CENPB n'aurait peut être pas de rôle dans la formation du kinétochore mais pourrait être impliquée dans la compaction des séquences péri-centromériques (Nakagawa, Lee et al. 2002)

- CENPC

C'est une protéine très conservée au cours de l'évolution. On retrouve des homologues chez toutes les espèces à l'exception de la drosophile (Amor, Kalitsis et al. 2004). Elle est directement impliquée dans la fonction du centromère puisque sa déplétion ou sa surexpression cause des défauts de ségrégation des chromosomes (Tomkiel, Cooke et al. 1994; Kalitsis, Fowler et al. 1998). Enfin, CENPC comme CenH3

est exclusivement localisée au niveau des centromères fonctionnels (Fukagawa, Pendon et al. 1999).

- CENPH et CENPI

Ce sont des protéines relativement bien conservées au cours de l'évolution puisque des homologues ont été décrits chez la levure, la souris et l'Homme mais pas chez la drosophile. Elles sont présentes dans les centromères fonctionnels et sont essentielles à la formation du kinétochore (Fukagawa, Mikami et al. 2001; Nishihashi, Haraguchi et al. 2002)

- MIS12

C'est une protéine plutôt conservée au cours de l'évolution. La déplétion de MIS12 cause des problèmes d'alignement des chromosomes mais pas de retard mitotique contrairement à ce qui avait été observé pour les cellules déficientes en CENPA, CENPC, CENPH et CENPI (Fukagawa 2004).

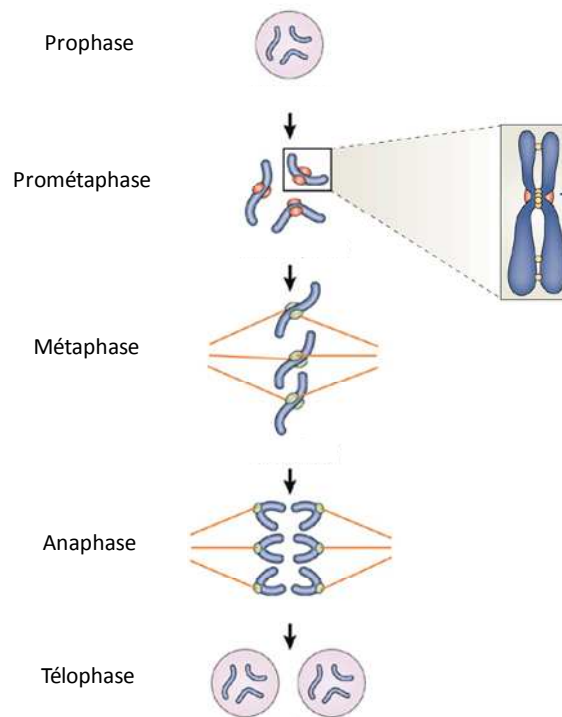
*Ainsi, même si aucune séquence génomique particulière ne permet d'expliquer la localisation du kinétochore au niveau du centromère, certaines caractéristiques de la chromatine centromérique sont conservées chez les eucaryotes : la richesse en AT et la présence du variant d'histone CenH3. D'autre part, une étude récente indique que l'hétérochromatine péri-centromérique ainsi que la machinerie d'interférence par ARN seraient nécessaires au recrutement de CENPA sur les CT (Folco, Pidoux et al. 2008). Ainsi, CT et PCT coopèrent ensemble pour assurer la bonne ségrégation des chromatides sœurs au cours de la mitose.*

## II.2.B Les péri-centromères

Les domaines chromatiques qui entourent le centromère sont connus sous différentes appellations comme hétérochromatine péri-centrique, hétérochromatine péri-centromérique ou encore péri-centromères (PCT). Ces péri-centromères sont parfois qualifiés de constriction secondaires. C'est en effet le cas pour les chromosomes 1, 9 et 16 qui possèdent une constriction secondaire juxta-centromérique en position q12, et pour le chromosome Y dont la constriction secondaire juxta-télomérique est également localisée en position q12.

Comme nous venons de le voir le CT est le siège de l'attachement des microtubules, responsables de la bonne répartition des chromatides sœurs dans les cellules filles. Cependant la cohésion des chromatides sœurs est un élément essentiel

dans l'attachement correct des chromosomes aux microtubules. Or, chez la souris, des expériences montrent que le dernier point d'attachement des chromatides sœurs entre elles à lieu au sein des PCT (Guenatri, Bailly et al. 2004). Ainsi, l'intégrité génétique indispensable au bon déroulement de la mitose est maintenue en partie grâce aux PCT qui permettent via les complexes de cohésine de conserver l'adhésion des chromatides sœurs jusqu'au début de l'anaphase (Figure 30).



**Figure 30 : Mitose et attachement des chromosomes au fuseau de microtubules (rouge)(d'après(Musacchio and Salmon 2007)).**

### 1. Les séquences génomiques

Une chromatine particulière entourant les centromères est présente chez les eucaryotes supérieurs. Celle-ci est composée de séquences répétées qui sont de nature différente de celles présentes au niveau du CT et qui portent différents noms selon l'espèce considérée :

- Satellites 1, 2 et 3 chez l'Homme
- Satellites majeurs chez la souris
- Séquences satellites chez la drosophile
- Outer repeats chez *S.pombe*

Le point commun entre ces séquences est la présence d'ADN répété, de type satellite. La comparaison de ces séquences entre différents organismes est présentée dans la suite de ce chapitre. Etant donné la variabilité des séquences génomiques présentes dans les PCT, aucune séquence particulière ne semble requise pour assurer la fonction des PCT. Une fois encore, les mécanismes épigénétiques sont privilégiés.

## 2. Epigénétique et HP1

Les PCT sont de nature hétérochromatique pour l'ensemble des espèces présentées, à l'exception de *S.cerevisiae*, et ils présentent un haut degré de compaction (Choo 2001). Cette compaction de la chromatine péri-centromérique est en accord avec la localisation préférentielle de la protéine HP1 (Heterochromatin Protein 1) à ce niveau. HP1 est une petite protéine nucléaire de 25kDa (kilodalton) qui a été initialement identifiée chez la drosophile comme impliquée dans l'effet PEV (James and Elgin 1986). Lorsque des gènes sont relocalisés à proximité de l'hétérochromatine, ils sont inactivés. Or, des mutations dans le gène *Su(var)2-5* (Suppressor of Variegation 2-5) qui code pour la protéine HP1, ont pour conséquence de supprimer l'effet PEV (Eisenberg, James et al. 1990). Ainsi, HP1 joue un rôle important dans la répression transcriptionnelle qui a lieu au sein de l'hétérochromatine. C'est une protéine essentielle puisque des mutants homozygotes pour HP1 chez la drosophile conduisent à la létalité au stade larvaire (Lu, Emtage et al. 2000).

HP1 est une protéine hautement conservée de la levure à l'Homme (Lomberk, Wallrath et al. 2006). Toutefois, aucun homologue d'HP1 n'a été décrit pour *S.cerevisiae* suggérant une fois encore l'absence de structures péri-centromériques chez cette levure. Différents isoformes d'HP1 ont été retrouvés chez différentes espèces :

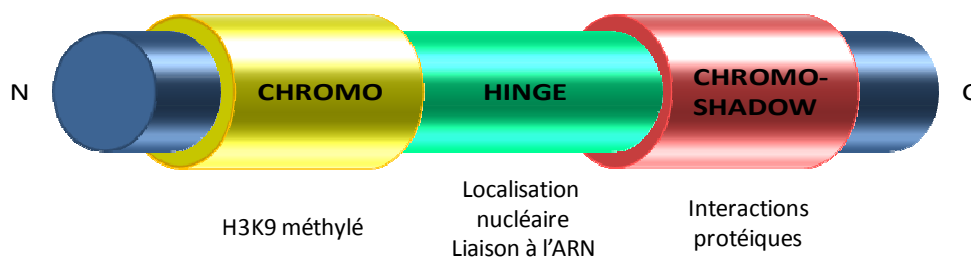
- HP1 alpha, beta et gamma chez l'Homme
- HP1 alpha, beta et gamma chez la souris
- HP1 a, b et c chez la drosophile
- Swi6 chez *S.pombe*

Les protéines HP1 présentent toutes une structure commune. Elles sont en effet composées de 3 domaines distincts (Figure 31).

- Le chromodomaine est localisé dans la moitié N-terminale de la protéine. Il est important pour la liaison de HP1 à l'histone H3 di ou triméthylée sur la lysine 9.

- Le chromoshadow domaine, localisé dans la partie C-terminale de la protéine, est impliqué dans l'homodimérisation et l'hétérodimérisation de la protéine mais également dans l'interaction avec d'autres protéines.

- Enfin, entre le chromodomaine et le chromoshadow domaine se trouve une région « hinge ». La région hinge est la partie la plus variable entre les différentes espèces mais également entre les différentes protéines d'une même espèce. Cette région contient un signal de localisation nucléaire et peut subir des modifications post-traductionnelles. Cela suggère qu'elle pourrait avoir un rôle dans la régulation de la protéine. Ce domaine permet également la liaison à l'ARN. Cette propriété est discutée dans la suite de cette introduction.



**Figure 31 : Structure d'HP1.** HP1 est composé de trois grands domaines : le chromodomaine, la région hinge et le chromoshadow domaine. Les fonctions de ces domaines sont indiquées.

HP1 est comme son nom l'indique une protéine typique de l'hétérochromatine. Cependant sa localisation n'est pas dépendante d'une séquence génomique particulière. Sa préférence pour les PCT, au détriment des CT, vient du fait qu'elle est capable de se lier via son chromodomaine à l'histone H3 méthylée sur la lysine 9 (H9k9) (Jacobs, Taverna et al. 2001). Ces modifications épigénétiques sont beaucoup



plus rares au sein du CT puisque celui-ci présente en majorité le variant de l'histone H3 CenH3. L'enzyme responsable de la méthylation d'H3K9 (Suv39H1 chez l'Homme, Suv39h1 chez la souris, SU(var)3-9 chez la drosophile et Clr4 chez *S.pombe*) est très bien conservée au cours de l'évolution (Rea, Eisenhaber et al. 2000). La perte d'activité de cette HMT conduit à une absence de méthylation d'H3K9 ainsi qu'à la perte de localisation d'HP1 au niveau des PCT (Bannister, Zegerman et al. 2001; Schotta, Ebert et al. 2002).

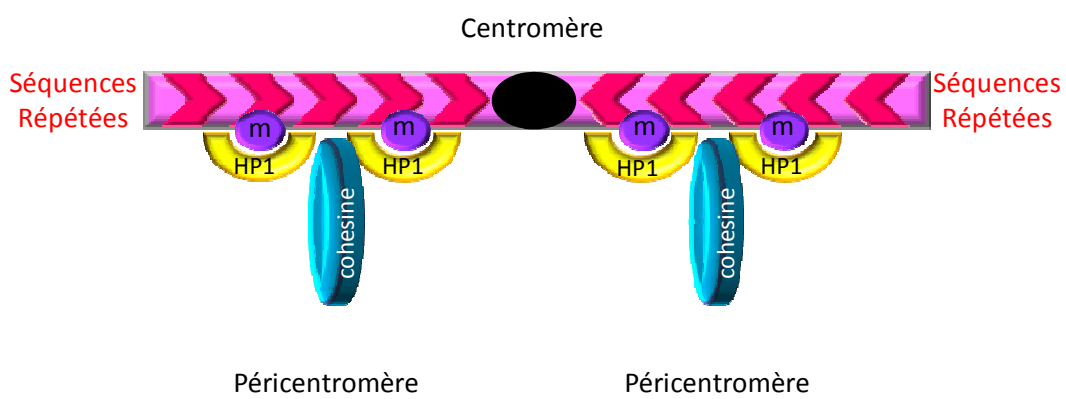
HP1 peut également être localisée dans des régions euchromatiques, selon l'isoforme considérée. Chez les mammifères, HP1 alpha et beta ont une localisation hétérochromatique alors que HP1 gamma est présente dans les deux compartiments (Minc, Courvalin et al. 2000). Cette localisation différentielle peut également être régulée par des modifications post-traductionnelles ainsi que par des interactions avec d'autres protéines (Lomberk, Bensi et al. 2006). La localisation d'HP1 dans les régions euchromatiques peut être associée à l'activation de la transcription (Hediger and Gasser 2006). Chez la souris, par exemple, la présence d'HP1 gamma et de la tri-méthylation d'H3K9 est détectée sur des gènes actifs, d'une manière dépendante de l'élongation de la transcription réalisée par l'ARN polymérase II (Vakoc, Mandat et al. 2005).

### 3. Cohésion des chromatides sœurs

La ségrégation correcte des chromosomes nécessite une série d'événements très bien contrôlés. Ainsi, le centromère, grâce au kinétochore, permet l'attachement bipolaire des chromatides sœurs aux microtubules. Cependant, l'intégrité génomique est également assurée par la régulation du dépôt d'un complexe protéique permettant le maintien de l'association des chromatides sœurs jusqu'en anaphase, la cohésine.

Le complexe multiprotéique de cohésine est composé de 4 sous unités : SMC1 (Structural Maintenance of Chromosomes), SMC3, SCC1 (également appelée rad21) et SCC3 (Sister Chromatid Cohesion) (Losada 2007). La localisation du complexe de cohésine sur le chromosome varie. En effet, la cohésine est présente dans les régions centromériques jusqu'en anaphase, contrairement au reste des chromosomes dont les complexes sont dissociés en prophase (Waizenegger, Hauf et al. 2000; Warren, Steffensen et al. 2000). A la transition métaphase-anaphase, la dégradation de la cohésine permet la dissociation des chromatides sœurs et leur migration vers des pôles opposés. Des expériences réalisées chez *S.pombe*, la souris et l'Homme indiquent qu'HP1 est impliquée dans la cohésion des chromatides sœurs (Bernard, Maure et al. 2001; Peters, O'Carroll et al. 2001; Inoue, Hyle et al. 2008) (Figure 32).

En effet, SWI6 (l'homologue d'HP1 chez *S.pombe*) qui se lie aux outer-repeats (les PCT), est indispensable à la localisation de Rad21, une protéine appartenant au complexe de la cohésine. Des mutants d'HP1 ou de Rad21 conduisent à l'apparition de chromosomes traînants le long du fuseau mitotique. Les modifications d'histones (hypoacétylation, méthylation) et la méthylation de l'ADN sont également impliqués dans la cohésion des chromatides sœurs puisque la perte de ces marques conduit à un problème de ségrégation des chromosomes (Peters, O'Carroll et al. 2001). En effet, chez *S.pombe* la perturbation de l'état hétérochromatique conduit à des défauts sévères dans la division cellulaire liés à une perte de cohésion (Bernard, Maure et al. 2001; Volpe, Kidner et al. 2002).



**Figure 32 : Péricentromère et cohésion.** La chromatine péricentromérique est enrichie en séquences répétées (rose), méthylées sur H3K9 (violet). La protéine HP1 se fixe sur H3K9 méthylée. HP1 permet alors le recrutement de la cohésine qui assure la cohésion entre les chromatides sœurs.

La liaison d'HP1 à la chromatine est dépendante de la phosphorylation de la sérine 10 d'H3 (H3phS10). Cette modification épigénétique est réalisée par la kinase Aurora B au moment où les cellules entrent en mitose et conduit à l'enlèvement d'HP1 des chromosomes mitotiques (Mateescu, England et al. 2004; Fischle, Tseng et al. 2005; Hirota, Lipp et al. 2005). Les chromatides sœurs peuvent ainsi se séparer. La phosphorylation de la sérine 10 d'H3, localisée au niveau des PCT, n'est pas présente dans les CT dont la chromatine est principalement enrichie en CenH3 (Hendzel, Wei et al. 1997). Enfin, des expériences d'hybridation *in situ* réalisées chez la souris montrent également l'implication des PCT dans la cohésion des chromatides sœurs. En effet, lors de la métaphase le CT se distingue par deux foyers individuels sur un même chromosome suggérant que les chromatides sœurs sont déjà séparées à cet endroit. Le PCT ne présente qu'un seul foyer par chromosome attestant du maintien de l'appariement des chromatides (Guenatri, Bailly et al. 2004).

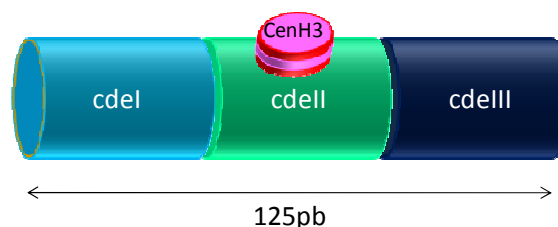
Les PCT assurent ainsi la cohésion des chromatides sœurs processus dans lequel HP1 est impliquée puisqu'elle est indispensable au recrutement des complexes de cohésine qui permettent de résister à la tension du fuseau mitotique en empêchant la séparation prématurée des chromatides sœurs.

## II.2.C Les CT et PCT dans différentes espèces.

Les CT et les PCT sont deux domaines distincts de l'hétérochromatine centromérique à la fois en terme d'épigénétique mais également de fonction. Ces caractéristiques sont assez bien conservées de la levure à l'Homme. Cependant, aucune séquence génomique ne permet de définir le centromère ou le péricentromère. Les séquences centromériques et péricentromériques ainsi que leur organisation sont décrites ci-dessous pour *S.cerevisiae*, *S.pombe*, *D.melanogaster*, *M.musculus* et finalement pour l'Homme.

### 1. *Saccharomyces cerevisiae*

La chromatine centromérique de *S.cerevisiae* est très bien caractérisée et pour cause : c'est la seule espèce pour laquelle l'élaboration du kinétochore est dépendante de la séquence primaire en ADN. Ce qui n'est pas le cas pour les autres espèces dont la fonction centromérique dépendrait plus d'une identité épigénétique.



**Figure 33 : Le centromère de *S.cerevisiae***

#### a) Le centromère

Cette séquence particulière assurant la fonction du centromère est représentée par une unité de 125pb composée de trois éléments (Figure 33):

- CdeI qui est formé de 8pb possède une séquence hautement conservée entre les 16 chromosomes. C'est un palindrome imparfait essentiel à la fonction du

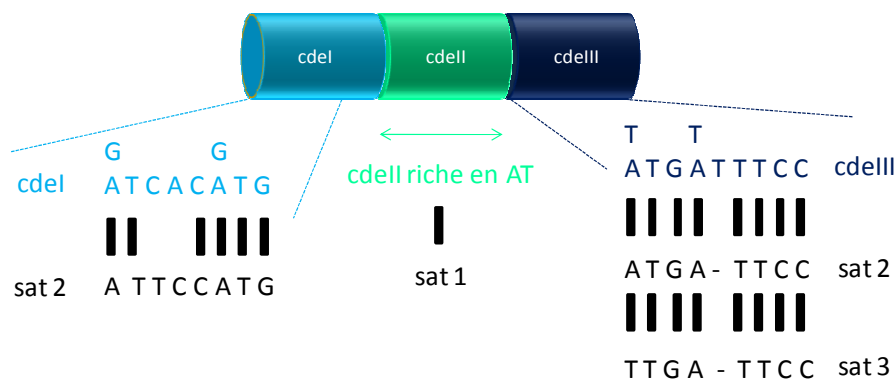
centromère puisque de simples mutations ponctuelles affectent la bonne ségrégation des chromosomes (Hegemann, Shero et al. 1988).

- CdeII est un motif qui n'est pas conservé en terme de séquence (78-86pb) mais sa longueur et son contenu, particulièrement riche en AT (90%) sont maintenus (Fitzgerald-Hayes, Clarke et al. 1982). Il permet la liaison à Cse4 (CENH3) et Mif2 (CENPC) (Masumoto, Nakano et al. 2004). Des délétions de CdeII affectent également la ségrégation des chromosomes (Sears, Hegemann et al. 1995).

- CdeIII est composée de 26pb. C'est un palindrome imparfait mais néanmoins conservé dans les différents chromosomes. Des mutations ponctuelles sont tout aussi dramatiques que celles réalisées dans CdeI, aboutissant à la destruction de la fonction centromérique (Jehn, Niedenthal et al. 1991). CdeIII permet la liaison de CBF3, une protéine essentielle pour la localisation de toutes les protéines du kinétochore (Wiens and Sorger 1998).

#### b) Le péri-centromère

La chromatine qui entoure le centromère de *S.cerevisiae* est compacte, résistante aux nucléases et ne contient pas de nucléosomes enrichis en CenH3 (Saunders, Fitzgerald-Hayes et al. 1988). Cependant, contrairement aux autres organismes décrits ci-dessous les régions qui entourent le centromère de *S.cerevisiae* ne sont pas hétérochromatiques. De plus, aucun homologue de la protéine HP1 n'a été identifié chez cette levure. Enfin la réplication de l'ADN au sein des séquences flanquantes et du centromère se déroule en même temps, en phase S précoce (Choo 2001). Ainsi, ces données suggèrent qu'il n'existe pas réellement de péri-centromères chez *S.cerevisiae*, tels qu'ils ont été définis dans la partie précédente. Cependant, des changements dans la chromatine qui entoure les centromères, qui s'étale sur 600pb, affecte la fonction du kinétochore et entraîne des problèmes de ségrégation des chromosomes lors de la mitose (Pinto and Winston 2000). Enfin, il est assez troublant de constater certaines homologies de séquences entre les éléments Cde de *S.cerevisiae* et les satellites 1, 2 et 3 humains (Grady, Ratliff et al. 1992) (Figure 34).



**Figure 34 : Homologie entre les séquences Cde de *S.cerevisiae* et les séquences satellites humaines (d'après (Grady, Ratliff et al. 1992)).** La séquence CdeI peut être comparée à celle des satellites 2. La séquence CdeII est riche en AT comme les satellites 1 humains. Et enfin, la séquence CdeIII possède des homologies de séquence avec les satellites 2 ou 3.

## 2. *Saccharomyces pombe*

L'ensemble des séquences centromériques et péricentromériques sont de taille variable pour les trois chromosomes de *S.pombe*: 35kb pour le chromosome 1, 65kb pour le chromosome 2 et 110 kb pour le chromosome 3 (Wood, Gwilliam et al. 2002).

### a) Le centromère

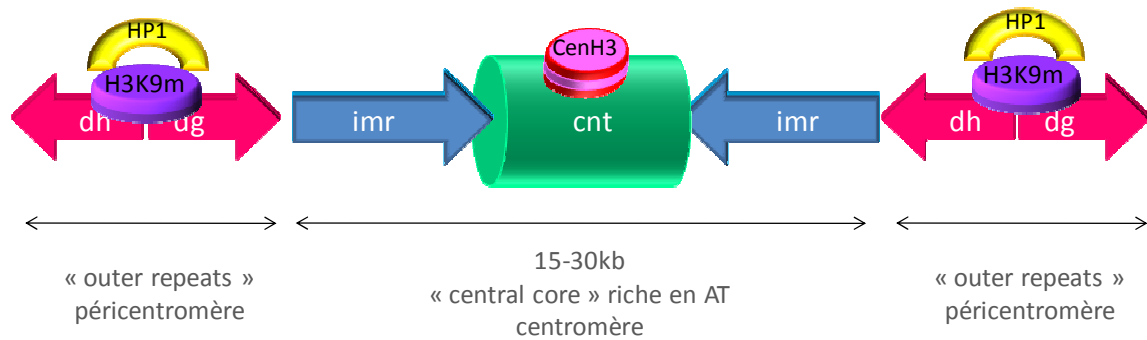
Les trois centromères de *S.pombe* sont assez similaires, bien que leur organisation soit différente. Ces centromères s'étendent sur une zone allant de 15 à 30 kb soit une centaine de fois plus longue que la séquence centromérique de *S.cerevisiae*. Cette région, appelée « central core region », est composée de deux éléments distincts (Takahashi, Murakami et al. 1992) (Figure 35):

- le centre à proprement parler (cnt) qui est une séquence de 4kb d'ADN non répété. La séquence génomique n'est pas conservée entre les trois chromosomes. Ce domaine cnt est riche en AT.

- Les « innermost » repeats (imr), séquences d'environ 6kb entourent le domaine central. Ce sont des répétitions inversées uniques pour chaque centromère.

Les imr et le cnt sont nécessaires à la fonction du centromère. Ces séquences permettent la liaison au fuseau mitotique (Hahnenberger, Carbon et al. 1991; Baum,

Ngan et al. 1994; Nakaseko, Goshima et al. 2001). CENH3 s'associe au cnt ainsi qu'aux imr et cette localisation est dépendante de CENPI. Cependant, pour les autres organismes décrits ci-dessous, la localisation de CENH3 est indépendante des autres CENPs (Masumoto, Nakano et al. 2004).



**Figure 35 : Le centromère et les péricentromères de *S.pombe***

#### b) Le péricentromère

La « central core region » est entourée d'hétérochromatine péricentromérique représentée par les « outer repeats » (Figure 35). Ces régions péricentromériques sont composées de séquences répétées inversées dg/dh, séquences de 4,4 et 4,8 Kb (Wood, Gwilliam et al. 2002). Les séquences dg et dh sont de longues répétitions dont le nombre de copies varie selon le chromosome considéré. Ainsi, le chromosome 1 est celui qui en possède le moins et le chromosome 3 celui qui en possède le plus. Cette répartition correspond parfaitement avec la longueur totale des séquences centromériques et péricentromériques (Nakaseko, Adachi et al. 1986; Nakaseko, Kinoshita et al. 1987). Une forte homologie de séquence a pu être définie entre les répétitions dg d'une part (99%) et entre les séquences dh d'autre part (80%). Les séquences dg et dh ne présentent aucune homologie entre elles. Cependant, elles contiennent toutes les deux de courtes unités de répétitions (Table 3). Ces séquences ne sont pas organisées aléatoirement dans les péricentromères mais sont disposées tête bêche l'une par rapport à l'autre et sont symétriques par rapport au centromère (Chikashige, Kinoshita et al. 1989). Ces séquences sont importantes pour le maintien de la cohésion des chromatides sœurs (Clarke and Baum 1990).

Finalement, *S.pombe* a été particulièrement étudiée comme modèle de mise en place de l'hétérochromatine péricentromérique, sujet qui est développé dans un autre chapitre.

Type de séquences présentes dans dg	Nombre de répétitions	Type de séquences présentes dans dh	Nombre de répétition
CAAA	8	TTA	8
GATGAAAT	2	TA	25
TGTGGT	2	GTTT	4
CTAGTGCA	5	TATAATGNA	3
GTACTAGTGTCATCTAGTGCA	2	TTATCA	0
CATCA	2	ATTCAA	2
CCA(A/C)(A/C)	13	AAAGTTT	2
TGCGTCT(A/C)GGTATCC	2	TTTC	2
AAGTGG	2	TTTTG/C	3
TAGTAT	2	(T/C)CATAT(T/C)(A/C)A(G/T)(G/T)	4
CATCATT(A/C)AAA	2	TCTCTATA	0
		TT(A/G)(T/C)A(G/A)	0
		G(C/G)ATGG(A/G)TATGGACACA	2
		CA(A/T)G	4
		G(A/T)ATTA	3

**Table 3 : Types de séquences répétées dg et dh présentes dans les péricentromères du chromosome 1 (d'après (Nakaseko, Adachi et al. 1986; Nakaseko, Kinoshita et al. 1987))**

### 3. *Drosophila melanogaster*

Chez la drosophile, les régions centromériques et péricentromériques sont composées de séquences satellites qui varient selon les chromosomes. Par ailleurs, les séquences précises ainsi que leur agencement le long du chromosome ne sont pas totalement définies (Adams, Celniker et al. 2000). Ainsi, il est difficile de déterminer des séquences consensus composant les CT et les PCT comme cela peut être fait pour la souris ou l'Homme.

On peut cependant classer les satellites de drosophile en deux groupes (Palomeque and Lorite 2008): les séquences répétées en tandem simples et les satellites plus complexes.

- Les séquences simples sont subdivisées en trois catégories correspondant aux trois bandes, parmi les quatre bandes isolées sur gradients de chlorure de césium : 1.672g/ml, 1,686 g/ml et 1.705g/ml (Gall, Cohen et al. 1971). Elles sont caractérisées par un motif répété de 5, 7 ou 10 paires de bases dont la formule générale est la suivante :  $(RRN)_m(RN)_n$  dans laquelle R peut représenter A ou G et N n'importe quel nucléotide (Table 4). Il existe toutefois une exception à la règle. En effet, les dodéca

satellites sont des palindromes imparfaits dont la séquence ne respecte pas la formule donnée ci-dessus (Abad, Carmena et al. 1992).

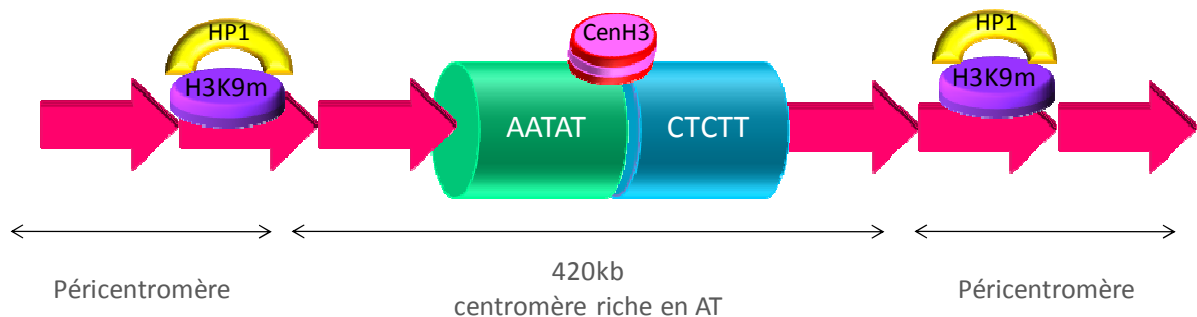
- Enfin, les satellites correspondant à la bande 1.688 g/ml sont majoritairement composés de séquences de 359pb et sont localisés sur le chromosome X (Bonaccorsi and Lohe 1991).

Types de satellites	Séquences
1,672 g/mL	AATAT, AATAG, AATAC, AATAAC, AATAGAC
1, 686 g/mL	AAGAC, AATAACATAG Dodeca satellites: CCCGTACTGGT
1, 705 g/mL	AAGAGAG, AAGAGAGAG
1, 688 g/mL	359 pb

**Table 4. Les séquences satellites de *D.melanogaster* (d'après (Lohe, Hilliker et al. 1993)**

a) Le centromère

Le CT de drosophile a pu être identifié grâce à l'analyse d'un minichromosome (Murphy and Karpen 1995; Sun, Wahlstrom et al. 1997). Le centromère est composé de 2 blocs adjacents de microsatellites hautement conservés AATAT et TTCTC, répétés en tandem, au sein desquels sont insérés des transposons (Sun, Le et al. 2003). C'est donc également une structure riche en AT qui définit le centromère de la drosophile et qui permet la liaison de CENH3, seule protéine centromérique identifiée dans cette espèce.



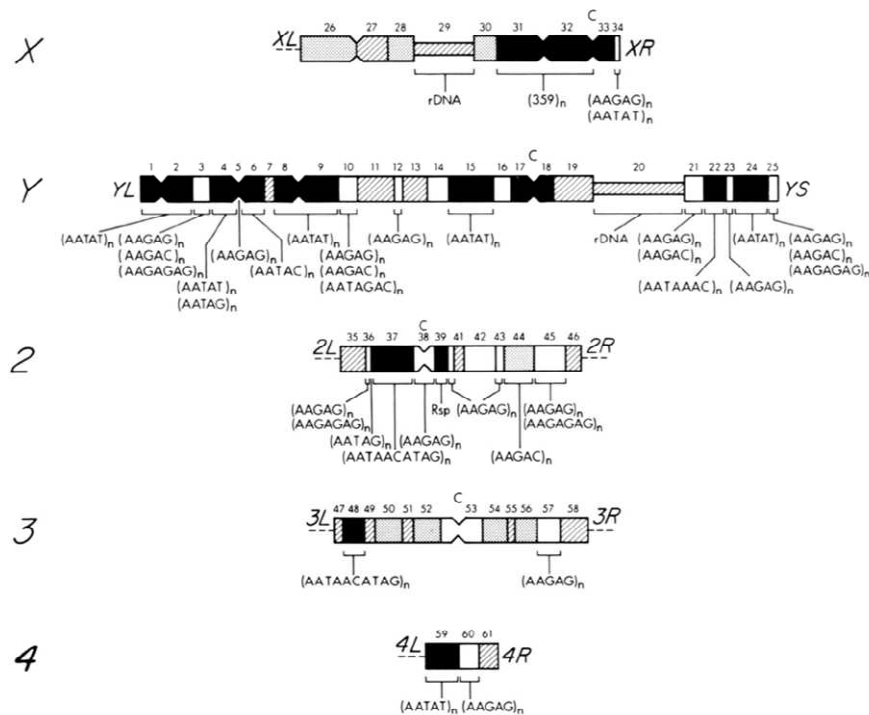
**Figure 36 : Le centromère et les péricentromères de *D.melanogaster***



b) Le péricentromère

Tout comme les CT, les séquences satellites présentes au niveau des PCT varient d'un chromosome à l'autre. Les dodeca satellites sont riches en GC et leur localisation pourrait correspondre aux régions péricentromériques. Cette localisation n'est pas encore totalement définie. Cependant, une homologie de séquence a été décrite entre les dodeca satellites et les satellites 3 humains localisés dans les régions péricentromériques. En effet, les signaux obtenus en hybridation *in situ*, sur des chromosomes métaphasiques humains en utilisant comme sonde les dodeca satellites, sont majoritairement localisés sur les chromosomes 9, 15 et Y dans des régions connues pour être enrichies en séquences satellites 3. D'autre part, il existe des variants des satellites correspondant à la bande 1.688 g/ml qui sont présents dans tous les PCT des chromosomes (Abad, Agudo et al. 2000; Abad and Villasante 2000).

Même si aucune séquence consensus n'est définie pour les PCT de drosophile (Figure 37), la fonction de ceux-ci dans le maintien des chromatides sœurs ne fait aucun doute (Kellum and Alberts 1995).



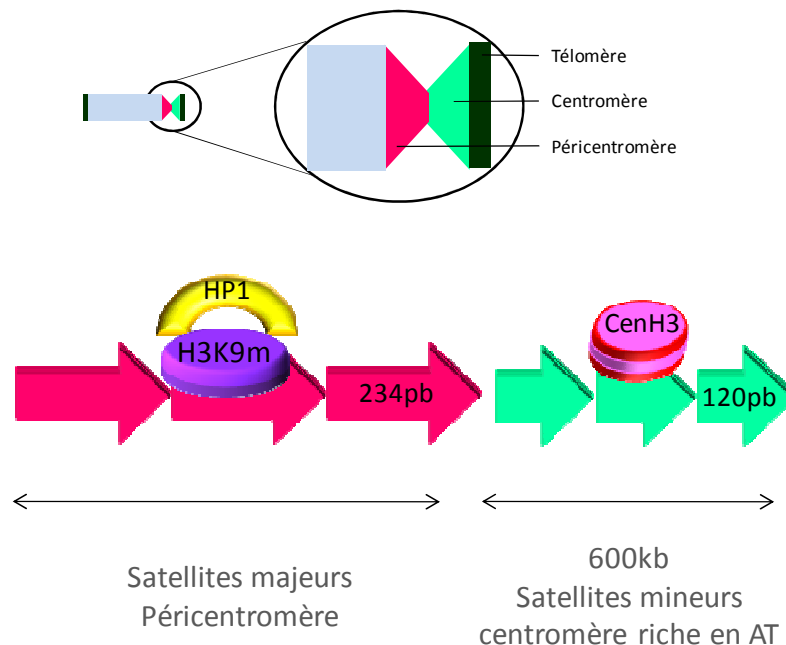
**Figure 37 : Localisation des différents satellites le long des péricentromères de *D.melanogaster* (Lohe, Hilliker et al. 1993).**

#### 4. *Mus musculus*

Les différences entre les régions centromériques et péricentromériques de *M.musculus* ont été particulièrement bien étudiées (Guenatri, Bailly et al. 2004). Leur composition en séquences répétées est très simple : les satellites mineurs sont présents au niveau des CT et les satellites majeurs sont localisés au niveau des PCT.

##### a) Le centromère

Le CT des chromosomes acrocentriques murins est composé de répétitions riches en AT appelées satellites mineurs et localisés à proximité des télomères (Figure 38). Ces satellites sont composés de multiples répétitions d'une unité de 120pb sur près de 600kb et sont présents sur tous les chromosomes (Choo 1997) (Table 5). Ils représentent 0,45 % du génome murin. Cette séquence de 120pb possède une grande homologie de séquence (79%, 31/39 nucléotides) entre ses deux moitiés de 60 pb (Wong and Rattner 1988). Certains monomères contiennent une séquence spécifique de 17pb : la CENPB box, à l'exception du chromosome Y qui en est dépourvu. C'est l'unique séquence génomique centromérique conservée entre les rongeurs et les primates.



**Figure 38 : Le centromère et le péricentromère des chromosomes acrocentriques de *M.musculus***

## b) Le péricentromère

Les péricentromères murins, situés dans la région adjacente des centromères sont composés de séquences appelées satellites majeurs (Figure 38). Ces répétitions sont riches en AT et le motif de base, long de 234pb, s'étale sur plus de 6Mb (Wong and Rattner 1988) (Table 5). Ces séquences représentent 3% du génome murin (Martens, O'Sullivan et al. 2005).

Types de satellites	Séquence
Satellites mineurs 120 pb	GGAAATGATAAAACCACACTGTAGAACATATTAGATGAGTGAGTTACACT GAAAAACACATTTCGTTGGAACGGGATTTGTAGAACAGTGTATATCAATGAG TTACAATGAGAAACAT
Satellites majeurs 234pb	GGACCTGGAATATGGCGAGAAACTGAAAATCACGGAAAATGAGAAATACA CACTTTAGGACGTGAAATATGGCGAGGAAAAGTAAAAAGTTGAAAAATTT AGAAATGTCCACTGTAGGACGTGGAATATGGCAAGAAAAGTAAAAATCATG GAAAATGAGAATCCACTTGACGACTTGAAAATGACGAAATCACTAAAA AACGTGAAAATGAGAAATGCACACTGAA

**Table 5: Séquences centromériques et péricentromériques de *M.musculus* (d'après (Wong and Rattner 1988)).**

Les péricentromères murins ont été utilisés comme modèle d'étude de la mise en place de l'hétérochromatine chez les mammifères. Les avancées dans ce domaine sont présentées au chapitre suivant.

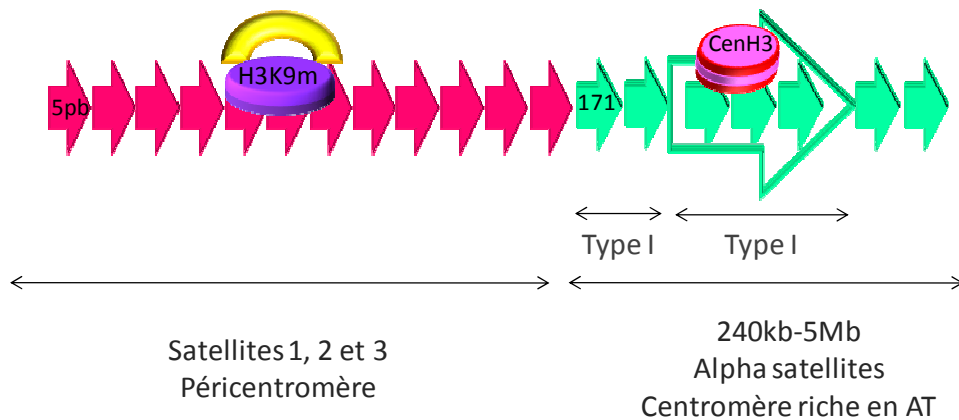
## 5. L'Homme

Enfin, les CT et les PCT humains se distinguent par la présence de différents types de séquences répétées. Alors que les alpha satellites sont présents au niveau des CT, les satellites de type 1, 2 et 3 sont localisés au niveau des PCT.

### a) Le centromère

Tous les CT humains sont composés de répétitions d'alpha satellites dont le monomère est constitué de 171 pb (Manuelidis and Wu 1978; Vissel and Choo 1987) (Table 7). Les monomères peuvent être organisés en deux types d'alpha satellites : soit en structure hautement organisée (alpha satellites de type I) soit en simples monomères (alpha satellites de type II). Les alpha satellites de type I sont eux aussi composés de monomères arrangés dans une configuration particulière, elle-même répétée. Et cette organisation varie d'un centromère à l'autre allant de 240kb à 5Mb (Wevrick and Willard 1989) (Figure 39). Des expériences dans le but de former des chromosomes artificiels montrent que les alpha satellites de type I, mais pas de type

II, sont étroitement associés à la formation de centromères fonctionnels (Masumoto, Nakano et al. 2004).



**Figure 39 : Le centromère et le péri-centromère humain**

La protéine CENH3 est particulièrement enrichie au niveau des séquences alpha satellites. CENPB se lie uniquement aux alpha satellites de type II sur lesquels on retrouve la CENPB box, à l'exception du chromosome Y qui est en dépourvu (Masumoto, Masukata et al. 1989).

#### b) Le péri-centromère

Les PCT humains sont composés de trois types majeurs de séquences répétées : les satellites 1, 2 et 3.

Les satellites 1 sont riches en AT (environ 70% d'enrichissement) et sont composés d'une alternance de monomères de 17 et 25 pb (Prosser, Frommer et al. 1986; Tagarro, Fernandez-Peralta et al. 1993) (Table 7). Ces séquences se distinguent des satellites 2 et 3 par leur résistance à différentes enzymes de restriction (Frommer, Prosser et al. 1982) (Table 6). La localisation des séquences satellites 1 est restreinte à de petits domaines péri-centromériques essentiellement localisés au niveau des chromosomes 3 et 4 et des chromosomes acrocentriques (Meyne, Goodwin et al. 1994; Tagarro, Wiegant et al. 1994).

	Satellites 1	Satellites 2	Satellites 3
EcoR1	Résistant	Sensible	Sensible
HaeIII	Résistant	Sensible	Sensible
Taq1	Résistant	Sensible	Résistant
HlnF1	Résistant	Sensible Fragments courts <80pb	Sensible Fragments longs (15-250pb)

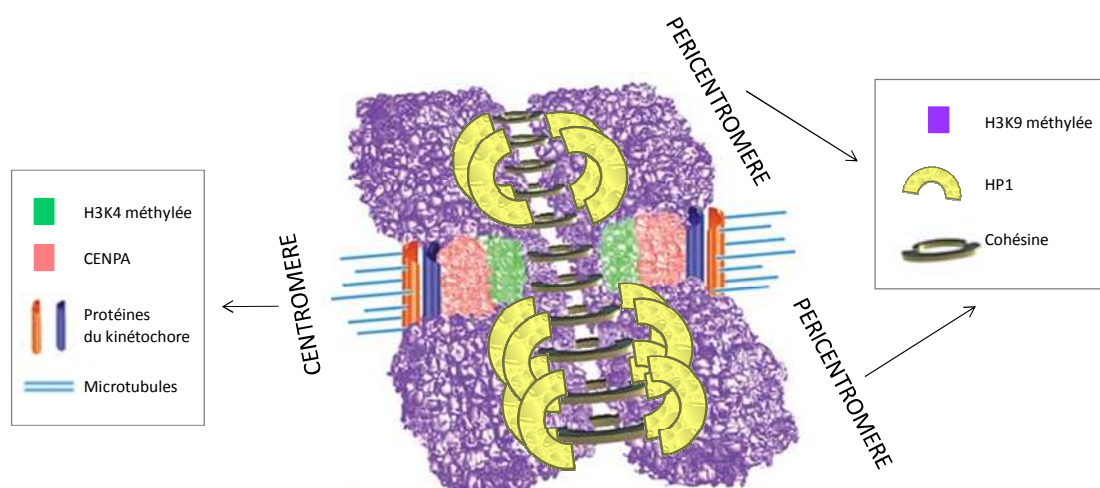
**Table 6 : Les profils de digestions enzymatiques permettent de différencier les satellites 1, 2 et 3 (d'après (Frommer, Prosser et al. 1982))**

Les satellites 2 et 3 (2% et 1.5% du génome) sont beaucoup plus abondants que les satellites 1 (0.5%) (Jones, Purdom et al. 1974; Gosden, Mitchell et al. 1975; Mitchell, Beauchamp et al. 1979). Les satellites 2 et 3 sont composés d'un motif pentamérique GGAAT, dont la séquence est quelque peu dégénérée pour les satellites 2 (Mitchell, Beauchamp et al. 1979; Frommer, Prosser et al. 1982; Prosser, Frommer et al. 1986; Hollis and Hindley 1988) Les répétitions de ce pentamère sont espacées par des séquences spécifiques CATCATCGA(A/G)T pour les satellites 2 et CAACCCGA(A/G)T pour les satellites 3 (Table 7). Des techniques de digestion enzymatique permettent également de différencier ces deux types de satellites (Table 6). En effet, alors que les satellites 2 sont sensibles à l'enzyme de restriction Taq1, les satellites 3 y sont insensibles. De même, si les deux séquences sont sensibles à l'enzyme HlnF1, leur profil de digestion est loin d'être équivalent (Frommer, Prosser et al. 1982). Enfin, par hybridation *in situ*, la localisation de ces séquences a pu être établie. Les satellites 2 sont majoritairement présents au niveau des régions péricentromériques des chromosomes 1 et 16. Des sites mineurs sont également détectés sur les péricentromères des chromosomes 2, 7, 10, 15, 17 et 22. Les satellites 3 sont majoritairement présents au niveau des régions péricentromériques du chromosome 9 mais également détectés sur les chromosomes 1, 5, 10, 13, 14, 15, 17, 20, 21, 22 et Y (Bobrow, Madan et al. 1972; Frommer, Paul et al. 1988; Tagarro, Fernandez-Peralta et al. 1994). Ces localisations ayant été établies par hybridation *in situ*, une technique relativement peu sensible, il n'est pas exclu que d'autres chromosomes puisse également être enrichis, dans une moindre mesure, en séquences satellites 2 et 3.

	Types de satellites	Séquences
Centromères	Alpha satellites (171pb)	CTTCTGTCTAGTTTTATATGAAGATATCCCGTTTCCAACCAAG GCCTCAAAGCGGTCCAAATATCCACAAGCTGATTCTACAAAA GAGTGTTTCAAACCTGCTCTATGAAAAGGAAGTTCAACTCTG TGAGTTGAATGTATACATCACAAGAAGTTTCTGAGAATG
Péricentromères	Satellites 1 (42pb)	(ACTTTC/GATATTTATGT) (ACAAGTATATAATA(C/T)A(C/T)ATTTGGGT)
	Satellites 2 (5pb)	(CGAATGGAAT) <sub>2</sub> + (CAT) <sub>1 ou 2</sub>
	Satellites 3 (5pb)	(GGAAT) <sub>n</sub> +/- (CAACCCGA(C/A)T)

**Table 7. Les séquences centromériques et péricentromériques humaines.**

Les centromères et les péricentromères, structures assez bien conservées au cours de l'évolution, présentent une nature hétérochromatique différente ainsi que des fonctions distinctes. Le centromère est généralement composé de séquences répétées, riches en AT, et il contient le variant d'histone CenH3. Les centromères assurent l'attachement des chromosomes sur le fuseau mitotique, via le kinétochore. Les péricentromères, sont composés de séquences répétées et sont enrichis en histone H3 méthylée sur la lysine 9. Cette modification épigénétique est reconnue par HP1 qui permet ensuite le recrutement du complexe de cohésine au niveau des péricentromères. Ces derniers assurent le dernier point de liaison des chromatides sœurs entre elles avant leur séparation en anaphase. Ainsi, centromère et péricentromère contribuent ensemble à la bonne ségrégation des chromosomes (Figure 40). Nous allons voir dans le chapitre suivant qu'un autre élément est nécessaire à cette fonction



**Figure 40: Fonction du centromère et du péricentromère (d'après (Sullivan and Karpén 2004)).** La principale fonction du centromère est d'assurer la formation du kinétochore. Le péricentromère permet la cohésion des chromatides sœurs.

## Chapitre III Hétérochromatine & Transcription

Afin d'assurer une fonction particulière, les séquences génomiques n'ont pas forcément besoin d'être transcrites puis traduites. En effet, dans un contexte particulier, elles peuvent avoir des rôles fondamentaux tels que ceux présentés pour les centromères et les péri-centromères. Ainsi, ces séquences sont loin d'être inutiles.

Cependant, l'utilisation de l'expression « ADN poubelle » reflète généralement l'absence de transcription et de traduction, deux mécanismes qui font toute la noblesse des gènes. Il va pourtant falloir se résoudre à se débarrasser des qualificatifs péjoratifs attribués à l'ADN non codant. En effet, cet ADN non codant, et en particulier celui présent au niveau des régions d'hétérochromatine constitutive, peut être transcrit. Dans ce chapitre, nous présenterons le rôle de ces ARN non codant (ARNnc) au regard de l'hétérochromatine péri-centromérique avant de résumer les conditions dans lesquelles les séquences centromériques et péri-centromériques peuvent être exprimées chez les mammifères.

### III.1 Hétérochromatine et ARN non codant

L'hétérochromatine péri-centromérique est particulièrement importante pour la cohésion des chromatides sœurs. Cette fonction ne semble pas être liée à une séquence génomique particulière mais plutôt à des caractéristiques épigénétiques. Ainsi, la méthylation de l'ADN et l'hypoacétylation des histones font de l'hétérochromatine péri-centromérique un domaine compact susceptible de résister aux tensions exercées par le fuseau mitotique. De plus, la méthylation de la lysine 9 de l'histone H3 permet le recrutement d'HP1 qui est impliquée dans le maintien de la cohésion des chromatides sœurs. Mais l'étude de la mise en place et du maintien de l'hétérochromatine, réalisée chez la levure *S.pombe*, a révélé l'existence d'un facteur inattendu : un ARN non codant.

### III.1.A L'hétérochromatine péricentromérique de *S.pombe*

L'étude des mécanismes impliqués dans la formation et le maintien de l'hétérochromatine a eu lieu en grande partie chez *S.pombe*. Ce domaine de recherche a été favorisé par la découverte des mécanismes d'interférence par ARN. D'abord décrit comme un mécanisme de répression post transcriptionnelle des gènes, impliquant l'intervention d'ARNdb dans la destruction d'ARNm cibles, d'autres fonctions ont depuis été assignées au mécanisme d'interférence par ARN.

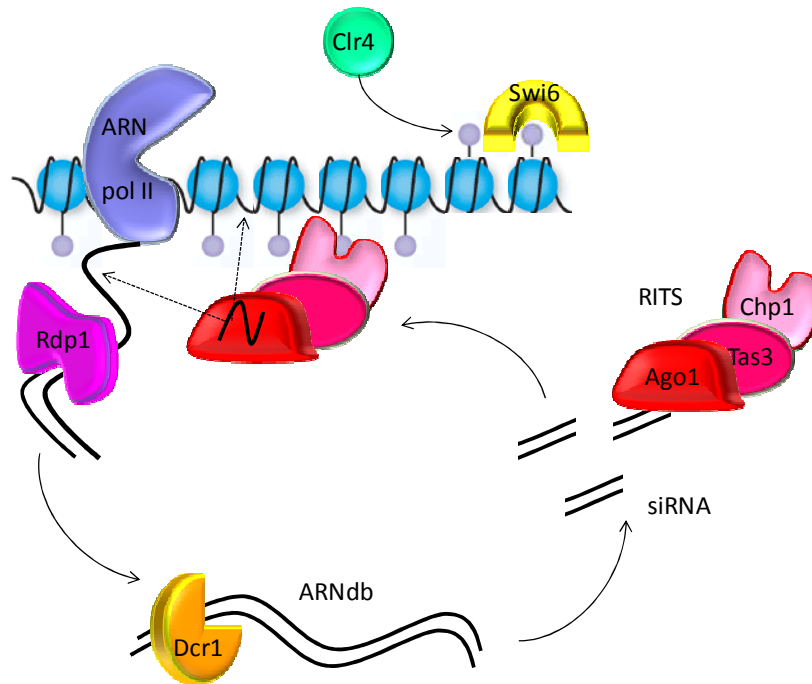
En effet, chez *S.pombe*, les mutants déficients en la protéine Dicer, présentent un ralentissement de la croissance cellulaire, associé à l'apparition de chromosomes traînants (Provost, Silverstein et al. 2002) Ce défaut est généralement le signe d'une altération fonctionnelle de l'hétérochromatine péricentromérique. Les mutants pour la protéine Swi6 (l'homologue d'HP1 chez la levure) présentent ainsi des caractéristiques similaires, liées à une perte de la cohésion des centromères (Ekwall, Javerzat et al. 1995; Bernard, Maure et al. 2001). De plus, des mutations dans les gènes codant pour des protéines impliquées dans les mécanismes d'interférence par ARN telles que Dicer (Dcr1), mais également argonaute (Ago1) ou Rdp1 (RNA dependent RNA polymerase 1) conduisent à une désorganisation de l'hétérochromatine. Ce bouleversement est traduit par une diminution de la localisation hétérochromatique de la protéine Rad 21, une sous unité du complexe de la cohésine (Hall, Noma et al. 2003). Un lien existe donc entre la machinerie d'interférence par ARN et la structure de l'hétérochromatine péricentromérique.

Les mutants de la machinerie d'interférence par ARN sont incapables de générer des petits ARN. Ainsi, une accumulation de longs ARN, sous forme sens et antisens, peut être observée. Ces ARN sont issus de la transcription de l'hétérochromatine et en particulier des séquences répétées dh et dg (Hall, Shankaranarayana et al. 2002; Provost, Silverstein et al. 2002; Volpe, Kidner et al. 2002) Ces résultats suggèrent donc un rôle des mécanismes d'interférence par ARN dans la stabilisation et la répression transcriptionnelle de l'hétérochromatine.

La mise en évidence du complexe RITS a permis d'identifier le chaînon manquant entre l'hétérochromatine et les mécanismes d'interférence par ARN (Verdel, Jia et al. 2004). RITS est composé de trois protéines : Tas 3, Ago 1, Chp1 (Chromodomain Protein 1) et un siRNA. Ago1 permet la liaison au siRNA, Chp1 via son chromodomaine se lie à H3mK9 méthylée et Tas3 agit comme protéine de liaison entre les deux. Les siARN associés à RITS proviennent du clivage par Dicer de complexes d'ARN doubles brins issus des séquences répétées péricentromériques dh et dg (Cam, Sugiyama et



al. 2005). Ces petits ARN permettraient donc le ciblage du complexe RITS au niveau des séquences répétées dh et dg par complémentarité de bases, soit par interaction directe siRNA-ADN soit via l'ARN naissant. En parallèle, Chp1 renforce l'interaction entre RITS et l'hétérochromatine péricentromérique en se liant à H3K9 méthylée (Partridge, Scott et al. 2002; Noma, Sugiyama et al. 2004). Enfin, un mécanisme de rétro-contrôle positif est également impliqué dans le maintien de l'inactivation transcriptionnelle des séquences répétées. En effet la protéine Rdp1 qui est une ARN polymérase dépendante de l'ARN, permet d'amplifier les siRNA sans avoir recours à la transcription. (Motamedi, Verdel et al. 2004). RDRC et RITS interagissent ensemble et colocalisent au niveau des outer repeats, les péricentromères de *S.pombe*. Cette interaction est dépendante de Clr4 (l'HMT responsable de la méthylation d'H3K9) et la méthylation des histones est elle-même dépendante de la machinerie d'interférence par ARN (Volpe, Kidner et al. 2002).



**Figure 41 : Modèle de mise en place de l'hétérochromatine chez *S.pombe* (d'après (Ekwall 2007))** L'ARN polymérase II initie la transcription des PCT de *S.pombe* pour générer un ARN précurseur. Cet ARN servirait de matrice à Rdp1 pour la synthèse du brin complémentaire. Ce long ARN double brin serait ensuite clivé par Dcr1. en siRNA. Ces siRNA permettraient le ciblage du complexe RITS vers l'hétérochromatine péricentromérique par interaction directe soit avec l'ARN naissant soit avec l'ADN. La stabilisation du complexe RITS sur l'hétérochromatine serait permise pour la liaison de Chp1 à l'histone H3 méthylée sur la lysine 9 par Clr4. Cette modification d'histone est également reconnue par Swi6.

Ainsi, le modèle de mise en place de l'hétérochromatine n'est pas totalement défini. En effet, de part l'interdépendance de certains facteurs, il se confronte à la

problématique de la poule et de l'œuf et le mystère n'est pas résolu.(Grewal and Elgin 2007) (Figure 41).

### III.1.B L'hétérochromatine péricentromérique murine

Même si l'implication de la machinerie d'interférence par ARN dans la structure de l'hétérochromatine est clairement démontrée chez *S.pombe*, ce n'est pas le cas chez les mammifères. Cependant un ARN a pu être identifié comme participant à la structure de l'hétérochromatine chez la souris.

La localisation d'HP1 au sein de l'hétérochromatine péricentromérique est très importante. En effet, elle représente l'unique lien entre la chromatine et le complexe contenant la cohésine, permettant le maintien de l'attachement des chromatides sœurs entre elles. L'implication de la protéine HP1 et de l'histone méthyltransférase d'H3K9 (Suv39H) dans la bonne ségrégation des chromosomes est aujourd'hui bien décrite. Cependant la simple reconnaissance de l'histone H3K9 méthylée ne suffit pas à cibler HP1 vers les PCT. En effet, H3K9 méthylée et HP1 sont présents non seulement dans les domaines hétérochromatiques péricentromériques mais peuvent également être associés à la régulation de l'expression de certains gènes (Vakoc, Mandat et al. 2005). Ceci suggère donc l'existence d'autres composants nécessaires à la mise en place du statut hétérochromatique des PCT.

En utilisant un anticorps reconnaissant une structure particulière de l'histone H3K9 méthylée, Maison et al montrent chez la souris que l'hétérochromatine péricentromérique est hautement organisée (Maison, Bailly et al. 2002). L'altération de cette structure par des inhibiteurs de HDAC, tels que la TSA, conduit à la perte de localisation d'HP1 sur les PCT sans affecter le taux d'H3K9 méthylée. La reconnaissance d'H3K9 méthylée par HP1 serait donc dépendante d'une structure particulière de la chromatine. Des traitements à la RNase A conduisent également à la perte de la localisation d'HP1 au niveau des PCT, sans affecter le taux de H3K9 méthylée, suggérant une déstabilisation de la structure de l'hétérochromatine péricentromérique. Ainsi, un ARN serait impliqué dans la structure de l'hétérochromatine. Or, la localisation d'HP1 sur les PCT est dépendante non seulement du chromodomaine mais également du domaine Hinge qui est capable de se lier à un ARN (Muchardt, Guilleme et al. 2002). La séquence de cet ARN n'est pas identifiée. On sait cependant qu'il s'agit d'un ARN simple brin d'origine nucléaire (Maison, Bailly et al. 2002).

Le lien entre les mécanismes d'interférence par ARN et la formation de l'hétérochromatine n'a pas pu être identifié à ce jour chez les mammifères. Il est toutefois possible que ceux-ci soient seulement impliqués dans l'initiation de la mise en place de l'hétérochromatine. En effet, Dicer est présente en quantité plus importante dans les cellules souches embryonnaires murines que dans les cellules différenciées (Billy, Brondani et al. 2001). D'autre part, à la différence de *S.pombe*, les mammifères ont un ADN méthylé. On peut ainsi imaginer que la méthylation de l'ADN, pourrait être impliquée dans le maintien du statut hétérochromatique des séquences péri-centromériques.

## **III.2 Transcription des séquences centromériques et péri-centromériques chez les mammifères**

La conservation de l'organisation des CT et des PCT au cours de l'évolution ainsi que la mise en évidence, chez *S.pombe*, de l'implication d'ARNnc péri-centromériques dans la structure de l'hétérochromatine permet d'envisager l'existence de tels ARNnc chez les mammifères. La présence d'un ARN au sein de l'hétérochromatine péri-centromérique murine témoigne en faveur de cette hypothèse. De plus, de nombreux exemples attestent de la transcription des séquences répétées centromériques et péri-centromériques chez l'Homme et la souris. Nous présenterons ainsi dans un premier temps les conditions sous lesquelles la transcription des CT et des PCT peut avoir lieu dans ces deux espèces avant de discuter des mécanismes potentiellement impliqués dans leur expression.

### **III.2.A Conditions d'expression des séquences répétées**

La difficulté à mettre en évidence la transcription des séquences centromériques et péri-centromériques vient du fait que celle-ci est soumise à certaines conditions telles que la réponse cellulaire au stress, le développement, la différenciation, le cycle cellulaire ou encore la sénescence. Dans certaines de ces conditions, l'orientation de la transcription est finement régulée. Ainsi, quand cela sera possible nous indiquerons l'orientation des ARN détectés, en accord avec les indications données dans la Table 8

ARN-PCT		Souris	Homme
SENS	➔	Riche en T	Riche en GGAAT
ANTISENS	➔	Riche en A	Riche en ATTCC

ARN-CT		Souris
SENS	➔	Riche en A
ANTISENS	➔	Riche en T

**Table 8. Nomenclature concernant l'orientation des ARN issus des séquences péricentromériques (ARN-PCT) et centromériques (ARN-CT), chez la souris et l'Homme.**

### 1. La réponse cellulaire au stress

L'exemple le plus caractéristique de l'expression des séquences répétées est celle qui a lieu en réponse à un stress cellulaire. Dans de nombreuses cellules normales et pathologiques, le stress thermique conduit à la transcription de séquences répétées de type satellites 3 localisées principalement au niveau du PCT du chromosome 9. (Denegri, Moralli et al. 2002; Jolly, Metz et al. 2004; Rizzi, Denegri et al. 2004) Ces ARNc issus des séquences satellites 3 (ARN-sat3) sont polyadénylés mais des controverses subsistent encore concernant leur taille. En effet, ils sont soit détectés sous forme très longue soit comme un ensemble de transcrits dont la taille varie entre 2 et 5 kb (Rizzi, Denegri et al. 2004). Dans les cellules soumises à un choc thermique, les séquences satellites 3 sont principalement transcrites dans une orientation sens, bien qu'une faible transcription dans une orientation antisens ait été décrite. Des faibles taux d'ARN-sat3 sens et antisens ont également été détectés dans des cellules en absence de stress (Valgardsdottir, Chiodi et al. 2008). La transcription des ARN-sat3 peut également être induite par d'autres types de stress tels que l'utilisation d'agents endommageant l'ADN (Valgardsdottir, Chiodi et al. 2008).

Le rôle fonctionnel des ARN-sat3 est inconnu. Cependant plusieurs hypothèses sont envisagées.

- Par analogie aux études réalisées chez *S. pombe*, ces ARNc pourraient être clivés en petits ARN double brin et être impliqués dans la structure de l'hétérochromatine dans des conditions normales de température. Dans ce scénario les ARN-sat3 s'accumuleraient au cours de la réponse au stress sous une forme longue en raison de l'inhibition ou de la saturation de la machinerie d'ARN interférence (Kloc, Zaratiegui et al. 2008).

-A l'instar du transcrit Xist, les ARN-sat3 sous une forme longue pourraient être impliqués dans la maintenance ou la mise en place d'une structure chromatinienne particulière au cours du stress. En effet, dans les cellules soumises à un choc thermique, les ARN-sat3, une fois synthétisés, restent dans le noyau à proximité du locus 9q12. Ainsi, ils pourraient être impliqués soit dans la protection du locus 9q12, soit dans la régénération de la structure de la chromatine de ce locus après le stress (Jolly, Metz et al. 2004).

- Les ARN-sat3 pourraient également avoir un rôle dans le contrôle de la transcription et de l'épissage au cours de stress en permettant la séquestration transitoire des facteurs clés impliqués dans ces mécanismes (Biamonti 2004; Jolly and Lakhotia 2006). En effet, plusieurs facteurs de transcription et d'épissage co-localisent et/ou interagissent avec les ARN-sat3 au cours du choc thermique (Chiodi, Biggiogera et al. 2000; Denegri, Chiodi et al. 2001; Jolly, Metz et al. 2004; Metz, Soret et al. 2004).

- Enfin, sachant que l'hétérochromatine peut avoir une influence répressive sur les gènes situés à proximité par effet PEV, l'activation transcriptionnelle de ces énormes blocs d'ADN méthylé pourrait contrer leur influence répressive (Fisher and Merkschlager 2002)

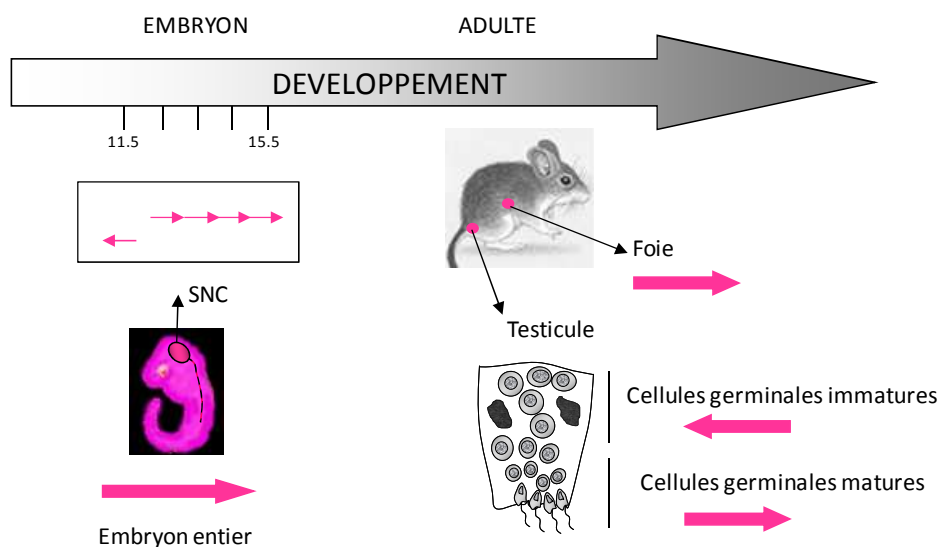
Ainsi, l'activation transcriptionnelle des régions péricentromériques pourrait représenter un moyen de modifier de façon transitoire l'organisation structurale et fonctionnelle du noyau.

## *2. Développement et différenciation cellulaire*

Des analyses d'expression réalisées dans les embryons et les tissus adultes de souris montrent que la transcription des PCT est finement contrôlée (Rudert, Bronner et al. 1995).

Ainsi, dans des embryons de 11.5-15.5 jpc (jour post coïtum), des ARN sens issus des régions péricentromériques (ARN-PCT sens) sont détectés dans différents tissus. En particulier, un haut niveau d'expression des séquences péricentromériques est observé dans le système nerveux central (SNC) à partir de 12,5 jpc. De manière intéressante, des ARN péricentromériques antisens (ARN-PCT antisens) sont uniquement observés dans des embryons de 11,5 jpc, dans des cellules du SNC correspondant certainement à celles dans lesquelles les ARN-PCT sens ont été détectés (Figure 42). Ainsi, dans les cellules du SNC, la transcription des PCT sous une

orientation antisens pourrait précéder la transcription de ces mêmes séquences sous une orientation sens.



**Figure 42 : Transcription des séquences périceptomériques de souris au cours du développement.** Dans des embryons de 11,5 et 15,5 jpc, des ARN-PCT sont détectés dans tous les tissus embryonnaires. Un fort taux d'ARN-PCT est détecté dans les cellules du système nerveux central (CNS). L'orientation des ARN-PCT varie en fonction du stade du développement. Chez l'adulte, les ARN-PCT sont détectés dans le testicule et le foie. Dans le testicule, l'orientation des ARN-PCT varie également en fonction de l'état de différenciation des cellules germinales.

Dans les tissus adultes, des ARN-PCT sens sont détectée dans le foie et les testicules, mais pas dans d'autres tissus tels que le cerveau, le côlon, la rate, le cœur ou les poumons. De plus, des ARN-PCT antisens sont également détectés dans les testicules, mais seulement dans les tubes séminifères dépourvus de spermatozoïdes matures. Ainsi, cette étude révèle que l'expression des PCT murins est régulée en fonction du stade embryonnaire et du type cellulaire considéré. Plus récemment, des ARN-PCT antisens ont été détectés dans les testicules de l'Homme, suggérant qu'ils pourraient jouer un rôle essentiel dans la différenciation des cellules germinales, en régulant l'expression de certains gènes (Jehan, Vallinayagam et al. 2007).

Des ARN issus des CT (ARN-CT) de souris ont été observés au cours de la différenciation des cellules musculaires (Terranova, Sauer et al. 2005; Bouzinba-Segard, Guais et al. 2006). Des ARN-CT, dont la taille est comprise entre 1 et plusieurs kb, sont présents dans des cellules non différenciées. Au cours de la différenciation, des ARN-CT de 120 nucléotides de long (120nt) sont alors détectés. Ces petits ARN pourraient correspondre à l'unité de base des satellites mineurs de souris (Table 5). Le rôle de ces ARNnc au cours de la différenciation cellulaire est

inconnu. Néanmoins, il est intéressant de noter que dans les cellules musculaires, la différenciation est associée au regroupement progressif des CT ainsi qu'à l'augmentation de deux modifications épigénétiques répressives sur les PCT: la méthylation d'H3K9 et d'H4K20. On peut donc supposer que l'accumulation des ARN-CT participe à la formation de l'hétérochromatine au cours de la différenciation. (Terranova, Sauer et al. 2005; Bouzinba-Segard, Guais et al. 2006). D'autres arguments sont en faveur d'un rôle des ARN-CT et ARN-PCT dans la différenciation cellulaire. En effet, une sur-expression d'ARN-CT sens de 120nt affecte la ségrégation des chromosomes et la cohésion des chromatides sœurs. Les ARN-CT pourraient ainsi interagir avec des facteurs essentiels impliqués dans la structure et la fonction des CT. La sur-expression des ARN-CT pourrait donc provoquer un titrage de ces facteurs conduisant à des altérations des régions centromériques (Bouzinba-Segard, Guais et al. 2006).

L'augmentation de la transcription des PCT est observée dans des cellules souches embryonnaires (ES) amenées à se différencier par un traitement à l'acide rétinoïque. Cependant, un résultat inverse a été obtenu en traitant des cellules pluripotentes P19, issues de carcinomes embryonnaires, avec de l'acide rétinoïque (Rudert, Bronner et al. 1995; Martens, O'Sullivan et al. 2005). Ces résultats contradictoires sont susceptibles de révéler des différences dans le statut épigénétique des cellules P19 et des cellules ES et/ou une expression différentielle de facteurs requis pour la transcription des séquences répétées péri-centromériques.

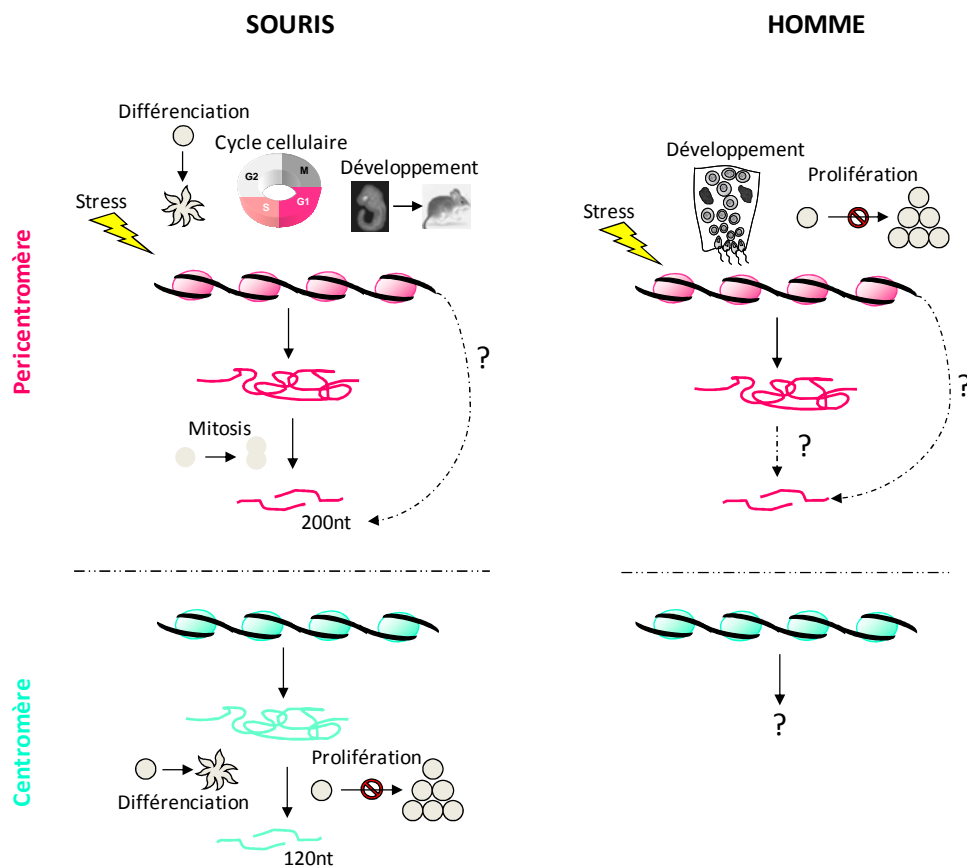
### 3. Cycle cellulaire

L'analyse de l'expression des PCT au cours du cycle cellulaire apporte de nouveaux éléments en faveur d'un rôle des ARN-PCT dans la structure de l'hétérochromatine. Dans les cellules prolifératives de souris, le taux d'ARN-PCT augmente au cours de la phase G1, atteint un pic d'expression à la transition G1/S avant de décroître. La diminution de la quantité des ARN-PCT coïncide avec la réplication de l'hétérochromatine péri-centromérique (Lu and Gilbert 2007). Ces ARN-PCT dont la taille varie entre 1 kb et plus de 8 kb, s'accumulent sur la surface externe des chromocentres (région d'association des régions péri-centromériques), là où la réplication des séquences péri-centromériques a lieu (Quivy, Roche et al. 2004). Ainsi, de manière similaire à ce qui est décrit chez *S. pombe*, les ARN-PCT pourraient jouer un rôle dans la reformation de la chromatine péri-centromérique (Lu and Gilbert 2007; Chen, Zhang et al. 2008).

L'accumulation d'ARN-PCT de plus petites tailles (200 nucléotides) est détectée au début de la mitose, phase au cours de laquelle les facteurs de transcription sont décrochés de la chromatine (Prasanth, Sacco-Bubulya et al. 2003). Il est probable que la transcription des séquences péricentromériques puisse également renforcer la structure de l'hétérochromatine au cours des dernières étapes de la mitose, lorsque la plupart des molécules de cohésine ont été éliminées. Par ailleurs, cette transcription pourrait participer au dépôt d'HP1, processus qui se produit au cours de l'anaphase (Wu, Singh et al. 2006).

#### 4. Sénescence

L'expression d'ARN polyadénylés issus des PCT, sous forme sens, est également observée dans les passages tardifs des fibroblastes primaires sénescents. L'augmentation de l'expression des ARN de la région péricentromérique 1q12 est corrélée avec une déméthylation et une décondensation du locus (Enukashvily, Donev et al. 2007).



**Figure 43 : Résumé des contextes physiologiques pour lesquels des ARN-CT (vert) et des ARN-PCT (rose) ont été détectés, chez la souris et l'Homme.**



### III.2.B Mécanismes impliqués dans la transcription des CT et des PCT

Comme nous venons de le voir, le schéma d'expression des séquences centromériques et péri-centromériques humaines et murines est complexe (Figure 43). Cette information suggère l'existence d'une régulation transcriptionnelle ou post-transcriptionnelle au moins aussi complexe. Le fait que des ARN sens et antisens ne soient pas nécessairement présents en quantité égale dans la cellule laisse penser, de manière similaire à ce qui se produit chez *S. pombe*, que la transcription pourrait être gouvernée par des complexes de régulation transcriptionnelle spécifiques (Volpe, Kidner et al. 2002; Nicolas, Yamada et al. 2007). Nous allons à présent voir quels sont les mécanismes connus pour être impliqués dans la transcription des séquences répétées humaines et murines (Figure 44)

#### 1. Facteurs de transcription spécifiques

Les seuls facteurs de transcription qui ont été formellement identifiés dans le contrôle de l'expression des PCT sont HSF1 (Heat-Shock Factor 1) et TonEBP (tonicity Enhancer-Binding Protein) (Jolly, Metz et al. 2004; Valgardsdottir, Chiodi et al. 2008). En réponse au stress thermique, HSF1 se lie directement aux séquences satellites 3 du locus 9q12. Son absence empêche l'accumulation des ARN-sat3 dans les cellules soumises à un choc thermique (Jolly, Metz et al. 2004). HSF2, qui colocalise avec HSF1, pourrait également jouer un rôle dans l'activation transcriptionnelle des PCT (Alastalo, Hellesuo et al. 2003). TonEBP s'accumule également au locus 9q12, en réponse à un stress hyper-osmotique. La présence de ce facteur de transcription est indispensable à l'activation transcriptionnelle des séquences satellites 3 dans les cellules soumises à un stress hyper-osmotique (Valgardsdottir, Chiodi et al. 2008). Dans les deux cas, les séquences d'ADN correspondant à des sites potentiels de liaison pour HSF1 et TonEBP ont été identifiées dans les séquences péri-centromériques (Valgardsdottir, Chiodi et al. 2008). De plus, le nombre élevé de sites de liaison potentiels d'HSF1 dans les régions régissant l'expression des PCT sous une forme sens expliquerait l'accumulation préférentielle des ARN-sat3 sens par rapport aux ARN-sat3 antisens.

De même, l'expression ou la répression des ARN-PCT en réponse à l'acide rétinoïque chez la souris pourrait refléter la présence de sites de liaison pour les récepteurs de l'acide rétinoïque, au sein des PCT (Rudert, Bronner et al. 1995; Martens, O'Sullivan et al. 2005). Cette hypothèse est renforcée par le fait que les PCT possèdent en effet la capacité de lier les récepteurs de l'acide rétinoïque (Rudert and

Gronemeyer 1993). La liaison de l'acide rétinoïque sur de tels sites pourrait dépendre du contexte cellulaire ainsi que de l'action de cofacteurs spécifiques.

## 2. Remodelage de la chromatine

Les différentes situations dans lesquelles une variation de l'expression des séquences centromériques et péri-centromériques a été observée, correspondent souvent à des états particuliers de la chromatine. C'est en effet le cas pour le développement et la différenciation cellulaire, le cancer, le vieillissement, la réplication et la réponse au stress, des mécanismes connus pour entraîner des changements de méthylation de l'ADN ainsi que des modifications d'histones (Gerbi and Bielinsky 2002; Rasmussen 2003; Jolly, Metz et al. 2004; Brock, Herman et al. 2007; Fraga and Esteller 2007; Probst and Almouzni 2008). L'hétérochromatine constitutive est caractérisée par un niveau élevé de méthylation de l'ADN, une méthylation des histones, en particulier sur H3K9, H3K27 et H4K20, ainsi que par un faible niveau d'acétylation des histones, en particulier sur H3K9. La relation entre la nature de ces marques épigénétiques et le statut transcriptionnel des CT et des PCT a donc été examinée.

### a) Méthylation de l'ADN

L'effet de la méthylation de l'ADN sur l'expression des séquences satellites a été analysé dans des cellules ES murines déficientes en DNMT1 ou DNMT3B. Aucune variation de l'expression des CT et des PCT n'a été observée dans ces cellules par rapport aux cellules ES contrôles (Lehnertz, Ueda et al. 2003; Martens, O'Sullivan et al. 2005). L'effet de la méthylation de l'ADN a également été abordé dans des cultures de cellules eurythroleucémiques murines traitées avec la 5 Aza-Cytidine, un puissant inhibiteur de la méthylation de l'ADN. Les cellules ainsi traitées affichent une augmentation de l'expression des séquences centromériques caractérisée par une accumulation d'ARN-CT de 120nt (Bouzinba-Segard, Guais et al. 2006). Cette observation est en accord avec des données suggérant que l'expression des séquences satellites, observée dans des cellules sénescents et cancéreuses, pourrait être facilitée par une déméthylation des PCT. En effet, une expression constitutive des séquences péri-centromériques du locus 1q12 est observée dans des cellules A431 issues d'un carcinome épithélial, ainsi que dans des cellules MRC5 (cellules embryonnaires sénescents de poumon) dont le génome est globalement hypométhylé (Erukashvily, Donev et al. 2007). Les cellules issues de patients atteints par le syndrome ICF constituent un puissant modèle pour étudier la relation entre la

méthylation de l'ADN et l'expression de séquences péri-centromériques. Le syndrome ICF est caractérisé par une hypométhylation sévère des PCT. Les chromosomes issus de lymphocytes de patients ICF présentent une décondensation locale des PCT associée à des réarrangements chromosomiques de ces régions (Ehrlich, 2003). En revanche, l'absence d'une forte expression constitutive des séquences satellites 3 suggère que la déméthylation de l'ADN n'est pas suffisante pour induire leur expression (Alexiadis, Ballestas et al. 2007).

Bien que la méthylation de l'ADN soit un processus associé à la répression transcriptionnelle, le lien celle-ci et l'expression des séquences répétées n'est pas encore très clair. Une étude récente montre que l'inhibition de SIRT1 (une HDAC de classe III) provoque une réactivation des gènes normalement réprimés. Cette activation transcriptionnelle n'est pas associée à une perte de méthylation de l'ADN au niveau de leur promoteur (Pruitt, Zinn et al. 2006). Ces résultats permettent d'envisager la contribution d'autres mécanismes dans l'activation transcriptionnelle des CT et des PCT.

#### b) Modifications d'histones

Dans les cellules ES de souris, la transcription des PCT se produit en dépit d'un niveau élevé de H3K9m3, H4K20m3 et H3K27m3, marques répressives des histones (Martens, O'Sullivan et al. 2005). De même, au cours de la différenciation musculaire, l'activation transcriptionnelle des CT et des PCT se produit malgré une augmentation de H3K9me3 et H4K20me3 dans ces régions (Terranova, Sauer et al. 2005). Ces observations laissent penser que les ARN-CT et les ARN-PCT pourraient être impliqués dans l'établissement et la maintenance de marques hétérochromatiques spécifiques au cours de la différenciation des cellules musculaires. Bien que la perte de H3K9me3 ne soit pas associée à l'activation transcriptionnelle des CT et des PCT, la perte de Suv39H, l'histone méthyltransférase impliquée dans la méthylation de H3K9, facilite néanmoins l'expression et/ou la stabilisation des transcrits issus des CT et des PCT qui s'accumulent alors sous forme double brin (Lehnertz, Ueda et al. 2003; Martens, O'Sullivan et al. 2005).

L'idée qu'une perte des marques épigénétiques répressives pourrait faciliter la transcription des séquences répétées hétérochromatiques a également été confortée par des observations faites dans les fibroblastes de patients touchés par le syndrome de la Progéria (HGPS pour Hutchinson Gilford Progeria Syndrome). Ce syndrome, causé par la présence d'une forme mutante de la Lamine A, est caractérisé par une apparition rapide du vieillissement au cours de l'enfance. Dans les cellules issues de

ces patients, une perte complète des marques hétérochromatiques H3K9me3, H3K27me3, H4K20me3 et HP1 est corrélée à l'expression constitutive des PCT du chromosome 9. L'absence d'expression des CT dans ces cellules suggère que la transcription des CT et des PCT met en jeu des mécanismes épigénétiques différents (Shumaker, Dechat et al. 2006).

Etant donné que l'information épigénétique doit être maintenue lors de la mitose, des altérations du mécanisme impliqué dans la transmission des marques hétérochromatiques lors de celle-ci sont susceptibles d'avoir une incidence sur l'expression des séquences répétées. C'est en effet le cas pour Np95, une protéine qui lie les histones et qui est régulée au cours du cycle cellulaire. Np95, qui permet la localisation d'HDAC1 sur des promoteurs des gènes, est impliquée dans la déacétylation de H4K5 et H4K12 après leur intégration dans les régions hétérochromatiques lors de la réplication. L'expression de Np95 débute à la transition G1/S et persiste jusqu'à la fin de la mitose. L'élimination de Np95 se traduit par l'augmentation de l'expression des séquences péri-centromériques. Ce résultat suggère ainsi que l'altération d'une première étape post-répllicative, impliquée dans la formation de l'hétérochromatine, conduit à une dérégulation des PCT (Papait, Pistore et al. 2007). Cette transcription est toutefois limitée aux PCT et ne porte pas atteinte aux CT, illustrant encore une fois une régulation transcriptionnelle différente entre les PCT et les CT (Papait, Pistore et al. 2007). Ainsi, Np95 et Suv39H apparaissent comme des acteurs essentiels de la formation et du maintien de l'hétérochromatine. Il est probable que des modifications de la structure de l'hétérochromatine, en réponse à l'absence de l'un de ces acteurs, soient perçues comme un signal positif pour l'activation transcriptionnelle des CT et/ou des PCT en vue de faciliter la reformation de l'hétérochromatine.

Finalement, le remodelage de la chromatine est aussi impliqué dans la transcription des séquences satellites 3 induite par le choc thermique. En effet, dans les cellules soumises à un choc thermique, la liaison d'HSF1 au PCT du chromosome 9 initie une série d'événements conduisant à l'activation transcriptionnelle de ces séquences. La manifestation la plus frappante de ce remodelage est l'apparition de foyers d'histones hyperacétylées au locus 9q12 dans les cellules soumises à un choc thermique (Jolly, Metz et al. 2004; Rizzi, Denegri et al. 2004). Cependant, on ne sait pas si la formation de ces foyers acétylés résulte d'une acétylation *de novo* ou d'une déacétylation globale du génome, épargnant le locus 9q12. L'histone acétyltransférase CBP, lorsqu'elle est surexprimée, s'accumule au locus 9q12 dans des cellules soumises à un choc thermique, favorisant ainsi l'hypothèse d'une acétylation *de novo* (Jolly, Metz et al. 2004; Rizzi, Denegri et al. 2004).

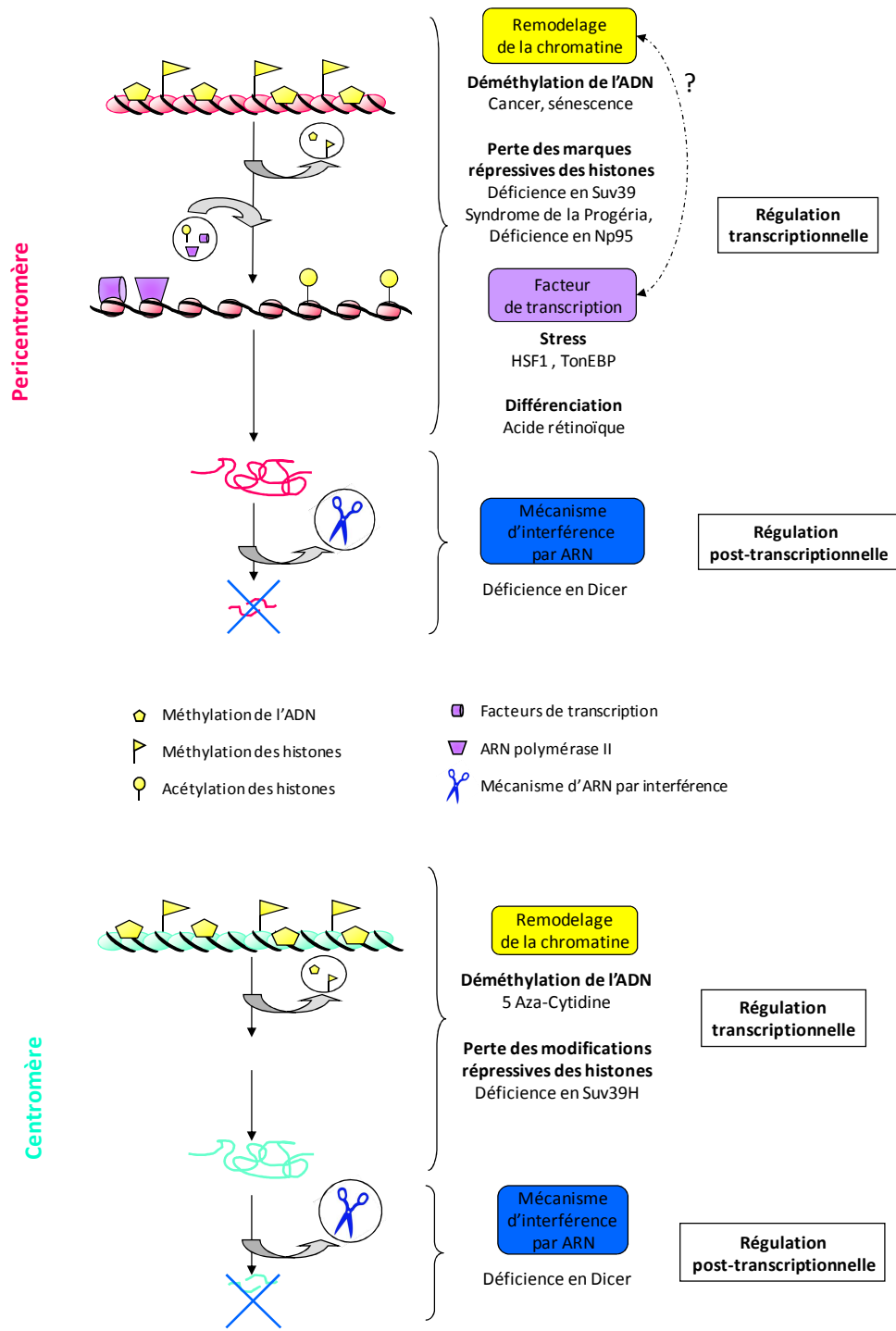
### 3. Régulation post-transcriptionnelle

Chez *S. pombe*, des siRNA issus des PCT sont générés par Dicer. Ils jouent un rôle important dans la structure et la fonction des régions hétérochromatiques, grâce au ciblage réalisé par RITS (Verdel, Jia et al. 2004; Buhler and Moazed 2007). Ainsi, par analogie à ce qui se produit chez cette levure modèle, les longs transcrits détectés dans des cellules humaines et murines pourraient représenter les précurseurs de petits ARN double brin. Il est toutefois important de noter que la validation de cette hypothèse dans les cellules de mammifères exige de tenir compte du fait que la levure ne possède pas de méthylation de l'ADN. Ainsi, par analogie avec l'ARN Xist, il est également possible que les longs ARN-CT et ARN-PCT puissent jouer un rôle dans le maintien de l'hétérochromatine.

Un rôle pour les petits ARN double brins de 20 à 30 nucléotides de long dans la structure de l'hétérochromatine humaine et murine, a toutefois été suggéré par des données obtenues dans deux modèles cellulaires déficients pour Dicer. Dans des hybrides Homme-poulet contenant le chromosome 21 humain, la perte d'expression de Dicer conduit à l'accumulation d'ARN-CT et d'ARN-PCT dont la taille varie entre 20-30 nucléotides et plusieurs kb (Fukagawa, Nogami et al. 2004). Cette accumulation d'ARNnc conduit à la séparation prématurée des chromatides sœurs et finalement à la mort cellulaire. De même, une accumulation de longs ARN-CT et ARN-PCT, associée à des défauts sévères de différenciation, est également détectée dans des cellules ES de souris déficientes pour Dicer. Des petits ARN double-brins dont la taille varie entre 20-30 et 150 nucléotides sont uniquement présents dans les cellules exprimant Dicer (Kanellopoulou, Muljo et al. 2005). La présence de ces transcrits de 20-30 nucléotides de long suggère que ces ARN pourraient être générés par Dicer et qu'ils représenteraient ainsi l'équivalent fonctionnel des siRNA issus des PCT de *S.pombe*. L'absence de détection de ces petits ARN dans d'autres systèmes cellulaires pourrait tout simplement être le reflet de leur faible niveau d'expression. En effet, les cellules de mammifères ne possèdent pas de système d'amplification tels que l'ARN polymérase dépendante de l'ARN (RDRP) présente chez la levure (Bouzinba-Segard, Guais et al. 2006). Enfin, de manière intéressante, une corrélation inverse entre le taux d'ARN-PCT et l'expression de Dicer a été décrite chez la souris (Terranova, Sauer et al. 2005).

*Les ARN-CT et les ARN-PCT de mammifères sont pour la plupart détectés sous la forme de longs ARN (>100 pb). Leur accumulation peut dépendre de mécanismes à la fois transcriptionnels et post-transcriptionnels. L'implication de facteurs de transcription spécifiques suggère qu'ils pourraient assumer des fonctions particulières,*

telles que la séquestration de facteurs de transcription et d'épissage au cours de la réponse au stress. Cependant, l'implication d'ARN non codants issus des séquences centromériques et péricentromériques dans la structuration de l'hétérochromatine chez les mammifères n'est toujours pas prouvée.



**Figure 44 : Résumé des mécanismes moléculaires conduisant à l'expression des CT (vert) et des PCT (rose) chez la souris et l'Homme.**

# Résultats

## OBJECTIFS

Chapitre I Identification de nouvelles cibles chromosomiques d'HSF1

Chapitre II Transcription des séquences satellites du génome humain

Chapitre III Le syndrome ICF

# OBJECTIFS

Le génome humain ne contient que 2% de séquences codantes. Une grande majorité de celui-ci est en effet constitué de séquences répétées, parmi lesquelles on retrouve celles localisées au niveau de l'hétérochromatine centromérique et péricentromérique. L'hétérochromatine constitutive a longtemps été considérée comme transcriptionnellement inerte. Or, il a été récemment montré, chez la levure, que la mise en place et le maintien de l'hétérochromatine péricentromérique ainsi que son rôle dans la cohésion des chromatides sœurs au cours de la mitose, nécessitait la participation d'ARN non codants produits à partir de ces mêmes séquences. Cependant, peu de choses sont connues concernant la capacité transcriptionnelle des séquences centromériques et péricentromériques chez l'Homme.

L'exemple de transcription des séquences péricentromériques le plus documenté à ce jour, dans les cellules humaines, concerne les séquences répétées de type satellites 3 du locus 9q12. En effet, au cours de la réponse cellulaire au choc thermique, le facteur de transcription HSF1, élément clé de la réponse au stress, se relocalise principalement au niveau de la région péricentromérique du chromosome 9, pour former les « granules de stress ». A mon arrivée dans le laboratoire du Dr Claire Vourc'h, l'équipe venait de montrer que ces granules de stress correspondaient à des sites de transcription actifs des séquences répétées de type satellites 3.

## **Mon travail de thèse a consisté à caractériser le potentiel transcriptionnel des séquences centromériques et péricentromériques du génome humain.**

Dans un premier temps, je me suis attachée à identifier de nouvelles cibles chromosomiques du facteur HSF1. En effet, les séquences satellites 3 sont présentes dans les régions péricentromériques de nombreux autres chromosomes. Il semblait ainsi assez étonnant que seules celles localisées au locus 9q12 soient transcrites. D'autre part, afin d'étudier le potentiel transcriptionnel de l'ensemble des séquences répétées du génome humain, nous avons cherché à développer un outil sensible, dédié à l'expression des séquences centromériques et péricentromériques : la RepChip. Grâce à cet outil, j'ai évalué l'expression de ces séquences dans des tissus sains et pathologiques. De plus, je me suis particulièrement intéressée à l'impact de modifications épigénétiques sur la capacité transcriptionnelle des séquences



centromériques et péricentromériques. C'est également cette même question qui m'a conduit à étudier le syndrome ICF, caractérisé par la déméthylation de l'ADN au niveau des régions péricentromériques des chromosomes 1, 9 et 16.

L'ensemble de ces résultats seront présentés en trois chapitres incluant chacun un article en cours de préparation.

# Chapitre I Identification de nouvelles cibles chromosomiques d'HSF1

Les résultats présentés dans ce chapitre, font l'objet d'un article en cours de préparation :

« *Identification of novel heterochromatic targets for the Heat Shock Factor 1* »

*Angéline Eymery, Catherine Souchier, Claire Vourc'h and Caroline Jolly*

Après avoir brièvement introduit la réponse au stress et le contexte dans lequel j'ai commencé mes travaux de thèse, les résultats seront présentés puis discutés.

## **I.1 Introduction : la réponse cellulaire au stress**

Certaines situations environnementales (élévation de température, métaux lourds...) ou physiopathologiques (inflammation, fièvre, cancer...) sont capables d'induire une réponse cellulaire ubiquitaire appelée réponse au stress (Lindquist 1986). L'exposition des cellules à un stress conduit à l'activation de gènes spécifiques dont ceux codant pour des protéines hautement conservées, les HSP (Heat Shock Proteins) (Trinklein, Murray et al. 2004). Les gènes hsp ont été initialement identifiés sur les chromosomes de glandes salivaires de drosophile oubliés sur un radiateur. Les gènes hsp sont nombreux et généralement nommés selon la taille moléculaire des protéines pour lesquelles ils codent, comme par exemple les gènes hsp70, hsp90 ou encore hsp40.

Les protéines HSP sont des protéines chaperonnes qui protègent les cellules des effets néfastes du stress permettant ainsi la survie cellulaire. Certains stress, tel que l'augmentation de température, conduit à une altération de la structure des protéines et donc à la perte de leur fonction biologique. Les protéines chaperonnes ont donc pour rôle d'empêcher l'agrégation des protéines dénaturées et de permettre leur repliement correct (Nguyen, Morange et al. 1989; Burel, Mezger et al. 1992; Lee, Roseman et al. 1997; Oh, Chen et al. 1997). D'autre part, afin d'éviter l'accumulation de protéines dénaturées un mécanisme d'inhibition globale de la transcription et de la

traduction est également mis en place dans les cellules stressées (Spradling, Penman et al. 1975; Jamrich, Greenleaf et al. 1977; Panniers 1994).

### I.1.A Le facteur de transcription HSF1

Chez les eucaryotes, l'expression inductible des gènes hsp est sous le contrôle de facteurs de transcription spécifiques appelés HSF pour « Heat Shock Factor » (Pirkkala, Nykanen et al. 2001). Le gène hsf a été initialement isolé chez la levure *S. cerevisiae* comme un gène unique, tout comme chez la Drosophile (Sorger and Pelham 1988; Clos, Westwood et al. 1990). Par la suite, trois gènes hsf (hsf1, hsf2, et hsf4) ont été identifiés chez les mammifères (Rabindran, Giorgi et al. 1991; Sarge, Zimarino et al. 1991; Schuetz, Gallo et al. 1991; Nakai, Tanabe et al. 1997). Chez *S. cerevisiae*, l'unique protéine HSF est essentielle à la survie cellulaire (Sorger and Pelham 1988; Wiederrecht, Seto et al. 1988). Chez la drosophile, HSF est nécessaire au bon déroulement de l'ovogénèse et du développement précoce (Jedlicka, Mortin et al. 1997). Les souris déficientes en HSF1 présentent un développement placentaire anormal, un retard de croissance et une infertilité des femelles (Xiao, Zuo et al. 1999; Christians, Davis et al. 2000).

#### 1. Structure de la protéine

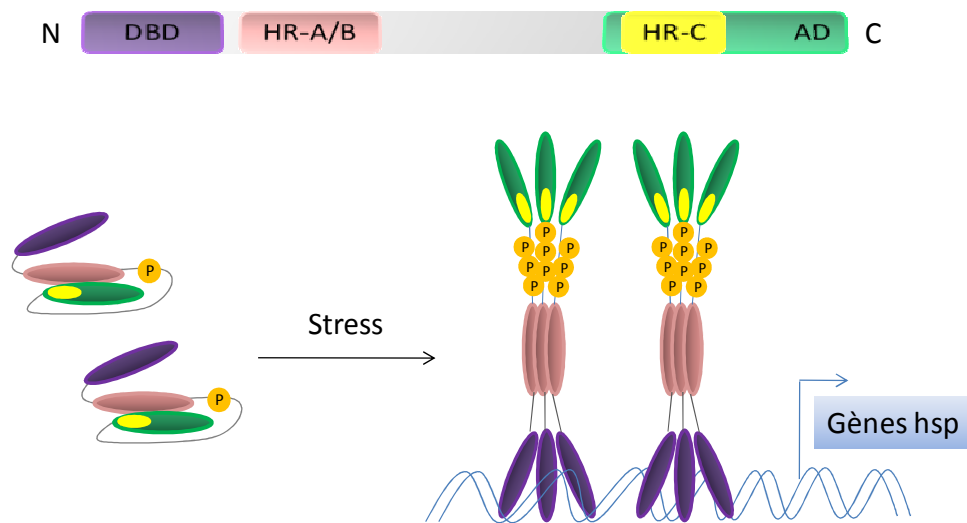
L'organisation structurale des protéines HSF est particulièrement bien conservée au cours de l'évolution (Voellmy 2004). Chez les vertébrés, HSF1 est le facteur clé de la régulation des gènes HSP. Il est divisé en trois domaines fonctionnels (Figure 45) :

- Le domaine DBD (DNA binding Domain) situé dans la partie N-terminale de la protéine permet la liaison d'HSF1 à l'ADN. HSF1 se lie sur des séquences consensus présentes sur les promoteurs des gènes hsp dont il active la transcription (Wu 1995). Ces séquences, appelées HSE (Heat Shock Element), sont composées de 3 pentamères nGAAn disposés tête-bêche.

- Les régions HR-A/B (Hydrophob Region) permettent la trimérisation d'HSF1. En effet, en présence de stress, les trois régions hydrophobes HR-A/B présentes sur les trois monomères s'associent entre elles (Rabindran, Haroun et al. 1993).

- Le domaine AD (Activation Domain) localisé dans la partie C-terminale de la protéine permet la trans-activation du facteur. Ce domaine possède également une région hydrophobe HR-C très similaire aux régions HR-A/B. En absence de stress,

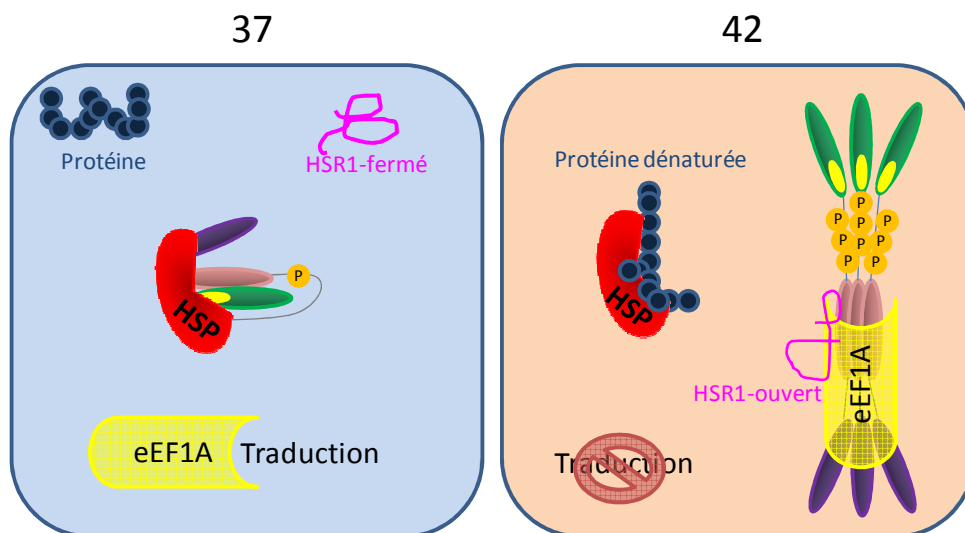
HSF1 est maintenu dans une structure monomérique inactive grâce à des interactions entre HR-A/B et HR-C.



**Figure 45 : Structure de la protéine HSF1.** HSF1 est composé de trois domaines, le domaine de liaison à l'ADN (DBD), le domaine de trimérisation (HR-A/B) et le domaine de transactivation (AD). En absence de stress, HSF1 est monomérique, non lié à l'ADN. Au cours du stress thermique, HSF1 est hyperphosphorylé, il se trimérise, se lie à l'ADN et induit la transcription des gènes hsp.

## 2. Régulation de l'activation d'HSF1

HSF1 est exprimé de manière constitutive et son activation, au cours du stress, passe uniquement par des mécanismes post-traductionnels. HSF1 est présent dans les cellules non stressées à l'état inactif, c'est à dire monomérique, non lié à l'ADN et distribué de manière diffuse dans le noyau. Lors d'un stress, il se trimérise et est relocalisé dans le noyau où il se lie aux promoteurs des gènes hsp (Holmberg, Tran et al. 2002). D'autre part, l'activation d'HSF1 dépend également de modifications post-traductionnelles. En effet, HSF1, qui est constitutivement phosphorylé au niveau des sérines 303, 307 et 363, subit des phosphorylations activatrices au cours du stress, dont les cibles ne sont pas toutes identifiées (Pirkkala, Nykanen et al. 2001; Guettouche, Boellmann et al. 2005). En particulier, la sérine 230, phosphorylée spécifiquement au cours du stress, semble jouer un rôle essentiel dans l'activité transcriptionnelle du facteur mais pas dans sa liaison à l'ADN (Holmberg, Hietakangas et al. 2001).

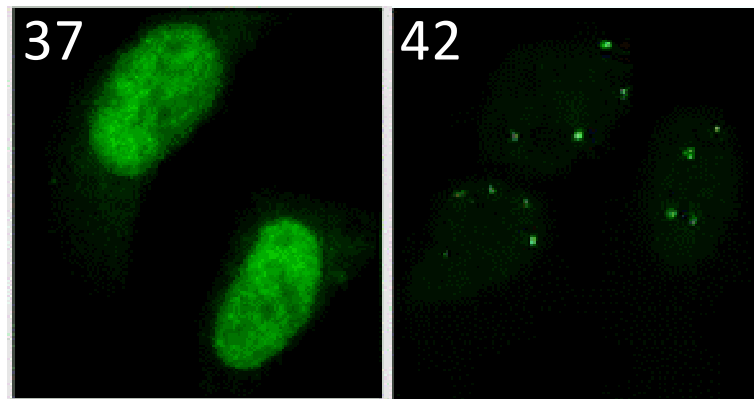


**Figure 46 : Modèle d'activation du facteur HSF1 au cours de la réponse au stress (d'après (Shamovsky and Nudler 2008)).** Dans des conditions physiologiques (37°C, panneau bleu), eEF1A est engagé dans l'élongation de la traduction et l'ARN non codant HSR1 est en position fermée, inactive. Le monomère d'HSF1 est maintenu dans un état inactif par des protéines chaperonnes, les HSP. Au cours de la réponse au stress thermique (42°C panneau rouge), les protéines sont dénaturées et prises en charge par les HSP qui libèrent ainsi le monomère d'HSF1. Les mécanismes de traduction sont inhibés, permettant la libération d'eEF1A. Le choc thermique induit également un changement conformationnel de HSR1 qui, dans une position ouverte active, participe avec eEF1A à l'activation du facteur HSF1.

Cependant une activation excessive et prolongée d'HSF1 pourrait être délétère. En effet, des souris mâles surexprimant un HSF1 actif sont infertiles (Nakai et al. 2000). Un contrôle précis de l'activation d'HSF1 est donc indispensable (Figure 46). Ainsi après un stress, HSF1 est inactivé par un mécanisme de rétrocontrôle négatif assuré par des protéines HSP (Sangster, Lindquist et al. 2004). Lorsque ces protéines sont transitoirement mobilisées, HSF1 est alors libéré. L'activation d'HSF1 nécessite deux acteurs: eEF1a, un facteur d'élongation de la traduction et HSR1 un long ARN non codant (Shamovsky, Ivannikov et al. 2006). eEF1a est une protéine particulièrement conservée au cours de l'évolution. Elle est impliquée dans la traduction ainsi que dans l'architecture du cytosquelette d'actine (Negrutskii and El'skaya 1998; Gross and Kinzy 2005). Sous l'effet d'un stress, la synthèse des protéines diminue et l'organisation du cytosquelette est perturbée, permettant ainsi de libérer eEF1a et de favoriser sa liaison à HSF1 (Welch and Suhan 1985; Panniers 1994). Dans ce modèle, EF1a et HSR1, qui pourrait de changer de conformation sous l'effet du stress, apparaissent comme d'excellents candidats pour le rôle de senseur du stress (Shamovsky and Nudler 2008).

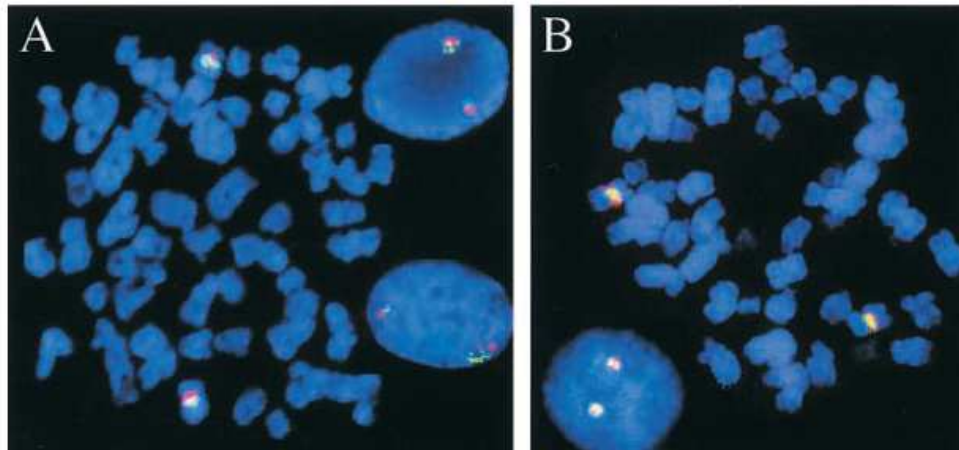
### I.1.B Les granules de stress

Parmi les nombreux effets engendrés par le stress, l'un des plus spectaculaire est la concentration massive du facteur HSF1, chez l'Homme mais pas chez la souris, au sein de foyers nucléaires visibles en immunofluorescence et appelés granules de stress ou nSBs (nuclear Stress Bodies) (Sarge, Murphy et al. 1993; Jolly, Usson et al. 1999; Biamonti 2004) (Figure 47).



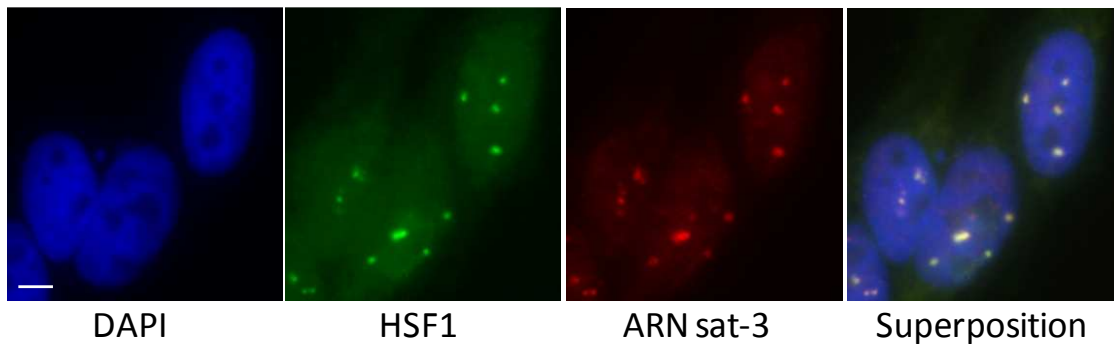
**Figure 47 : Localisation d'HSF1.** A 37°C, HSF1 est présent de manière diffuse dans le noyau. Après un choc thermique (42°C), il se relocalise dans des foyers nucléaires appelés granules de stress.

Il est important de noter que ces granules ne correspondent en aucun cas aux gènes hsp, gènes cibles du facteur HSF1 (Jolly, Konecny et al. 2002). En effet, la cible chromosomique de ces foyers a été caractérisée. Il s'agit de la région péricentromérique du chromosome 9 humain et plus particulièrement la région 9q12, connue pour être une zone d'hétérochromatine constitutive contenant des séquences répétées de type satellites 3 (Jolly, Konecny et al. 2002) (Figure 48). HSF1 interagit de manière directe avec ces séquences, qui contiennent de nombreux sites de liaison pour HSF1. En effet, les séquences satellites 3 sont composées de répétitions du motif GGAAT commun à la séquence consensus HSE (nGAAn).



**Figure 48 : Identification de la cible chromosomique d'HSF1 (Jolly, Konecny et al. 2002).** La protéine HSF1 recombinante (vert) est détectée par immunofluorescence sur des étalements de chromosomes métaphasiques avec soit une sonde spécifique du centromère du chromosome 9 (A-rouge) soit avec une sonde spécifique des séquences satellites 3 du locus 9q12 (B-rouge). Alors que les signaux sont juxtaposés en A, ils colocalisent parfaitement en B.

HSF1 n'est pas le seul composant de ces granules de stress. En effet, après la fixation d'HSF1, des foyers d'histones acétylées apparaissent au locus 9q12. L'ARN polymérase II est ensuite recrutée sur ce même locus (Jolly, Metz et al. 2004). Ainsi, la présence d'un facteur de transcription, de marques épigénétiques associées à une chromatine permissive à la transcription et le recrutement de l'ARN polymérase II sur les séquences répétées du locus 9q12 suggèrent très fortement que ces régions, bien qu'hétérochromatiques, puisse être transcrites. La présence de facteurs d'épissage dans ces granules de stress constitue un autre argument en faveur de cette hypothèse. Nous avons montré que ces séquences satellites 3 sont en effet transcrites au cours du choc thermique (Figure 49). Les ARN ainsi générés sont polyadénylés, de grande taille et non codants. Ces ARN restent en effet localisés à proximité du locus 9q12 (Jolly, Metz et al. 2004; Rizzi, Denegri et al. 2004). La fonction de ces ARNnc n'est pas définie. Mais plusieurs hypothèses ont été proposées. Les granules de stress et les ARNnc pourraient séquestrer les facteurs de transcription et d'épissage, participant ainsi à l'inhibition de la transcription au cours de la réponse au stress (Jolly and Lakhota 2006). D'autre part, ces ARNnc pourraient également être impliqués dans le maintien de la structure chromatinienne en période de stress (Biamonti 2004).



**Figure 49 : Colocalisation d'HSF1 et des ARN issus des séquences satellites 3 (ARN-sat3) détectés par immunofluorescence et FISH ARN dans des cellules HeLa soumises à un choc thermique.**

Cependant alors que les granules de stress ont été principalement détectées au niveau du locus 9q12, deux autres sites de liaison d'HSF1 ont été identifiés dans des cellules hybrides Homme-hamster (Denegri, Moralli et al. 2002). Sur la base de ces informations, nous avons cherché à identifier l'ensemble des cibles chromosomiques d'HSF1 et à déterminer leur capacité transcriptionnelle.

## I.2 Matériel et Méthodes

### 1. Culture cellulaire et choc thermique

Les cellules HeLa sont cultivées dans du DMEM additionné de 10% de SVF (sérum de veau fœtal) ; les cellules HCT116 normales et mutantes (obtenues par Bert Vogelstein, USA) sont cultivées dans du McCoy's 5A additionné de 10% SVF ; les IMR90 (Institut Coriel, USA) et les fibroblastes de peau (SF) sont cultivés dans du EMEM 15% de SVF. Tous les milieux sont complétés avec de la glutamine 4mM.

Pour les expériences de transfections transitoires, les cellules IMR90 sont cultivées dans des chambres de culture en verre (Labteck Nunc, Danemark) puis transfectées avec 2 µg de plasmide mHSF1-eGFP (obtenu par le Dr R. Morimoto, USA) en présence de Lipofectamine utilisée selon les instructions du fabricant (Cotto, Fox et al. 1997). Pour les expériences d'analyses microscopiques sur cellules vivantes, les cellules HeLa sont cultivées dans des chambres de culture en verre (Labteck Nunc, Danemark) et le choc thermique est réalisé directement sur la platine.



Pour les expériences d'immunofluorescence et de FISH, les cellules sont cultivées sur des lames de verre. Le choc thermique est réalisé par immersion des lames (ou des flasques pour les extractions de protéines et d'ARN) dans un bain marie pendant une heure à 43°C pour les cellules tumorales ou une heure à 45°C pour les cellules normales. Concernant les expériences de récupération, les cellules ont été soumises à un choc thermique d'une heure puis placées en récupération à 37°C pendant 3 heures.

## *2. Immunofluorescence*

Les cellules sont fixées 10 minutes dans une solution de 4% formaldéhyde/ PBS. Les membranes cellulaires sont perméabilisées par trois bains de 5 minutes dans 0,5% saponine/ 0,5% Triton / PBS. Après un bref rinçage au PBS, les sites aspécifiques sont bloqués par incubation des lames dans une solution de blocage (10% SVF/ 0,3% Triton / PBS) pendant une heure à 37°C. Les cellules sont alors incubées pendant 1h30 à 37°C avec l'anticorps primaire dilué dans une solution de détection (2% SVF/ 0,3% Triton/ PBS). Les anticorps non fixés sont éliminés par trois lavages dans une solution de 2%SVF/ 0,3%Triton/ PBS à 45°C. Les cellules sont incubées pendant 1 heure à 37°C avec un anticorps secondaire couplé à un fluorochrome (FITC ou TRITC) et dilué dans la solution de détection. Trois lavages sont ensuite réalisés afin d'éliminer les anticorps non liés. Finalement les lames sont montées dans une solution de DAPI (250 ng/ml 4',6-diamidino-phenylindole:2HCl) – DABCO (90% glycerol and 2.3% diaza-bicyclo-octane) qui permet de colorer l'ADN tout en diminuant les phénomènes de blanchiment du fluorochrome. Les anticorps qui ont été utilisés dans cette étude ainsi que leur dilution sont les suivants : lapin anti-HSF1 1/300 (Stressgen), lapin anti-lysines acétylées 1/200 (Upstate), souris anti-ARN polymérase II 1/400 (obtenu par M.Vigneron, France) et souris anti-SF2/ASF 1/50 (Zymed).

Concernant les expériences de transfections, les cellules IMR90 transfectées avec le plasmide HSF1-eEGFP, sont fixées dans du formaldéhyde et les lames sont directement montées avec du DAPI-DABCO.

### 3. *Hybridation in situ fluorescente pour la détection d'ARN et d'ADN (FISH ARN et FISH ADN)*

#### a) FISH ADN

Marquage des sondes : le marquage des sondes par incorporation de biotine se fait par amorçage aléatoire à l'aide du kit Bioprime (InVitrogen). Le principe consiste à couper l'ADN par l'enzyme de klenow et à incorporer dans la séquence un nucléotide couplé à la biotine. Le mélange comprenant 300ng d'ADN, 20µL d'eau et 20µL d'amorces aléatoires est dénaturé 5 minutes dans de l'eau bouillante avant d'être refroidi immédiatement sur glace. 5µL de dNTP, 4µL d'eau et 1µL de fragment de klenow sont ajoutés et la réaction est incubée pendant 1 heure à 37°C. La réaction est ensuite arrêtée par ajout de 5µL de tampon Stop.

Préparation des sondes : 100 ng de sonde marquée sont précipités à -80°C en présence de 50 µg d'ADN de sperme de saumon, d'AcNa 0,1M et d'EtOH 100%. Après une centrifugation (15min à 12500 rpm et 4°C) le culot est lavé par de l'EtOH 70%, séché et resuspendu dans 50% formamide/ 10% dextran sulfate/ 2SSC pour un volume final de 10 µl.

Préparation des lames : Les lames sont fixées 10 minutes dans une solution de 4% formaldéhyde/ PBS puis incubées 20 minutes dans du glycérol 20%/ PBS, un cryoprotectant. La perméabilisation des membranes cellulaires se fait par quatre cycles successifs de congélation/décongélation par immersion dans de l'azote liquide. Les cellules sont ensuite déshydratées par 3 bains successifs d'EtOH 70%, 90% et 100%. Après séchage, la sonde est déposée sur la lame et la dénaturation se fait sur une plaque chauffante à 86°C pendant 6 minutes. L'hybridation se fait sur la nuit à 37°C, en chambre humide. Le lendemain, les lames sont rincées trois fois 5 minutes dans du 50% formamide/ 2SSC à 45°C puis trois fois 5 minutes dans du 0,1% SSC à 65°C. Elles sont ensuite incubées avec la solution de blocage (3% BSA / 0,1% tween 20/ 4XSSC) pendant 30 minutes à 37°C. La détection de la sonde est réalisée par incubation pendant 30 minutes à 37°C dans une solution de détection (1% BSA/ 0,1% tween 20/ 4X SSC) en présence d'avidine-TRITC (Sigma). Après trois lavages dans du 4XSSC / 0,1% Tween 20 à 45°C, permettant d'éliminer les fluorochromes non liés, les lames sont montées dans la solution de DAPI-DABCO.

b) FISH ARN avec oligonucléotide.

La préparation de la sonde oligonucléotidique biotinylée spécifique des séquences satellites 3 (Tagarro, Fernandez-Peralta et al. 1994) est réalisée d'une manière similaire à celle présentée pour la FISH ADN. Cependant le culot est resuspendu dans 20% formamide/ 10% dextran sulfate/ 2SSC pour un volume final de 10 µl. L'oligonucléotide est directement déposé sur les lames déshydratées et l'hybridation se fait sur la nuit à 37°C en chambre humide. Le lendemain les lames sont rincées trois fois 5 minutes dans du 2SSC à température ambiante. Elles sont ensuite incubées avec la solution de blocage (5% lait/ 4SSC) pendant 30 minutes à 37°C. La détection de la sonde est réalisée par incubation pendant 30 minutes à 37°C dans une solution de détection (5% lait/ 4SSC) en présence d'avidine-FITC (Sigma). Après trois lavages dans du 4XSSC / 0,05% Tween 20 à température ambiante, les lames sont montées dans la solution de DAPI-DABCO.

Pour les expériences combinées de FISH ADN et d'immunofluorescence, la détection d'HSF1 est réalisée en premier lieu. Les cellules sont fixées une nouvelle fois dans du formaldéhyde et la FISH ADN est ensuite réalisée. Pour les expériences combinées de FISH ARN et d'immunofluorescence, la FISH ARN est d'abord réalisée, suivie de l'immunofluorescence. Un anticorps anti-lapin-FITC ou TRITC est ensuite ajouté pour augmenter le signal d'immunofluorescence.

4. *Reconstitution de granules sur chromosomes métaphasiques*

a) Préparation de chromosomes métaphasiques

Dans un T25, 5 ml de sang périphérique sont cultivés dans du milieu RPMI 1640 supplémenté par 15% SVF, 100 µg/mL de pénicilline, 100 U/ml de streptomycine ainsi que 2mM de glutamine. De la phytohémaglutinine (9 mg/ml), permettant d'accélérer les divisions cellulaires, est ajoutée dans la flasque qui est alors incubée dans une étuve à 37°C, en atmosphère 5% CO<sub>2</sub>, pendant 72 heures. Les cellules sont alors incubées pendant 1h30 avec une solution de colchicine à 10 µg/ml afin de bloquer les cellules en métaphase et en prométaphase. Les cellules sont ensuite brièvement centrifugées (5 minutes à 1200 rpm et à 4°C) et le culot est resuspendu dans un milieu hypotonique préchauffé à 37°C (KCl à 75 mM) déposé goutte à goutte et sous agitation. Les cultures sont maintenues ainsi à 37°C pendant 20 minutes. Quelques gouttes de fixateur glacé (méthanol - acide acétique, 3 :1) sont ajoutées aux cellules et une incubation de 5 minutes est réalisée, avant une nouvelle centrifugation dans les

même conditions que précédemment. Le culot cellulaire subit trois fixations successives dans le fixateur glacé. Les étalements sont réalisés en lâchant une goutte de la suspension cellulaire sur des lames de verre.

#### b) Reconstitution de granules

Les lames de chromosomes métaphasiques sont déshydratées par des bains d'éthanol successifs (70%, 90% et 100%). Après un bref rinçage au PBS, les sites aspécifiques sont bloqués par incubation dans une solution de blocage (10% SVF/ 0,3% Triton/ PBS) pendant une heure à 37°C. Après un rapide rinçage de la lame dans du PBS, différentes quantités de protéine HSF1 recombinante (Stressgen), diluées dans un mélange Hepes 50 mM/ EDTA 0,1mM/ NaCl 200mM/ Eau ; pH 7,4, sont déposées sur la lame et l'incubation se fait à 37°C pendant 3 heures. Les étapes suivantes de détection et de lavages sont identiques à celles décrites dans le protocole d'immunofluorescence.

### 5. *Western blot*

#### a) Extraction de protéines

Les cellules sont lavées et grattées dans du PBS avant d'être culottées par une centrifugation rapide. Le culot cellulaire est repris dans de l'urée 8M et soniqué sur glace pendant 5 minutes à 200J (30 secondes de sonication en alternance avec 30 secondes de pause). Après centrifugation, le surnageant qui constitue l'extrait protéique est dosé par la méthode de Bradford.

#### b) Western Blot

10 µg d'extraits protéiques sont dénaturés 5 minutes à 100°C et déposés sur un gel d'acrylamide / SDS. La migration se fait à 160V pendant 1h30. Les protéines sont alors transférées sur une membrane de nitrocellulose par transfert électrique en milieu liquide (100V pendant 1 heure). Afin de saturer les sites de liaison non spécifiques, la membrane est incubée dans un bain de 3% lait/ PBS pendant 1 heure, avant d'être mise en présence de l'anticorps primaire dilué dans 3% lait/ PBS, pendant 1 heure également. Après trois lavages dans du PBS/ 0,1% tween 20, la membrane est incubée avec l'anticorps secondaire couplé à la peroxydase (HRP) et dilué dans 3% lait/ PBS. Après des étapes de lavages, la révélation est réalisée par chemiluminescence avec le kit ECL+ (Amersham). Les anticorps utilisés dans cette

étude ainsi que leur dilution sont les suivant : lapin anti-HSF1 1/5000 (Stressgen), lapin anti-H2B 1/5000 (Upstate) et anti lapin-HRP 1/5000 (Roche).

## 6. *Gel retard*

1 $\mu$ M d'oligonucléotide (sens) sont marqués par phosphorylation de l'extrémité 3' en présence de gamma <sup>32</sup>PATP et de kinase T4 pendant 15 minutes à 37°C puis purifiés sur colonne Sephadex G25 pour éliminer la radioactivité non incorporée. L'oligonucléotide ainsi marqué est incubé avec le brin antisens pendant 5 minutes à 95°C puis la réaction est refroidie lentement. 100 000 cpm de sonde marquée sont incubés avec 0,5 $\mu$ g de poly dI/dC, 0,04% BSA, du tampon d'hybridation (10mM Tris pH 7,8, 50mM NaCl, 0,5mM EDTA, 5% glycérol) et 100 ou 500 ng de protéine HSF1 recombinante dans un volume final de 25 $\mu$ L pendant 20 minutes à température ambiante. La réaction est arrêtée par ajout de 2,5 $\mu$ L de sample dye (50% glycérol, 0,2% xylène cyanol, 0,2% bleu de bromophénol). Les échantillons sont déposés sur un gel d'acrylamide 4%. La migration se fait dans du TBE à 160V et 250mA jusqu'à ce que le bleu ait parcouru les  $\frac{3}{4}$  du gel. Le gel est ensuite transféré sur un papier Whatman, séché pendant 2 heures sous vide à chaud et exposé dans une cassette de PhosphorImager. Les séquences des sondes utilisées sont les suivantes :

HSE : TCGGCTGGAATATTCCCGACCTGGCAGCCGA

sat1 : ATATCAAAGTACACCAAATACATATTATATACTGTACATAAAATATCAAA

sat2 cons 1 : GACTCGAATGGAATCGAATAGAATCATCGAATGAAATCGAATGGAATCAT

sat2 cons 2 : ATCATCGAATGGAATCGAATGGAATCATCATCGAATGGAATCGAATGGAA

sat2 1q12 : GACTCGAATGGAATAATCATTGAACGGAATCGAATGGAATCATCATCGGA

sat2 16q12 : ATCACCAAATGGAATCGAATGGAATCATCTAATGAACTTGAA

Sat3 cons : TCCAYTCGGTTGATTTCCAYTCGGTTGATT

sat3 9q12 : ATTGGAATGGAATGGAATGGAATCAACCCGAGTACAGGAATGGAATGGAA

## 7. *Microscopie et analyse d'image*

Les images ont été acquises sur le système de microscopie fluorescent Axiovision (Zeiss, Allemagne) en utilisant l'objectif 63X à immersion dans l'huile (NA=1.25). La quantification a été réalisée en utilisant le logiciel d'analyse Metamorph (Molecular Devices, Sunnyvale, USA). Premièrement, la densité de volume (surface totale des granules / surface nucléaire) a été évaluée sur la section Z centrale. Deuxièmement, une projection des différentes coupes a été réalisée de façon à ce que toutes les granules soient présentes sur une même image. Pour chaque noyau, la surface, le

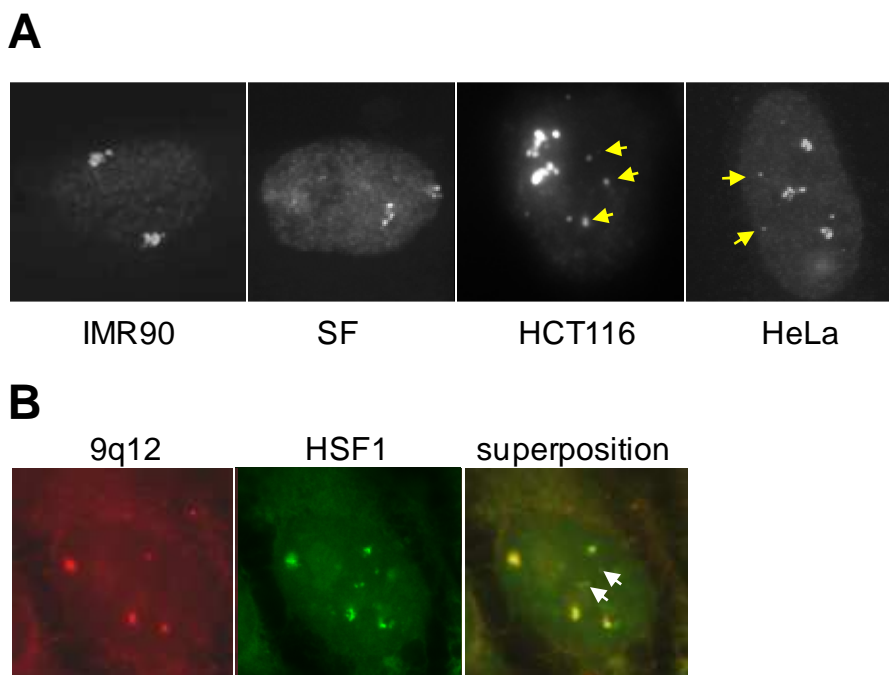
nombre ainsi que la taille des granules ont été mesurés et les plus gros foyers considérés comme des granules de stress primaires. 70 noyaux, au minimum, ont été analysés pour chaque paramètre. Les séries d'images de cellules vivantes ont été obtenues grâce à un microscope confocal (Confocor 2, Zeiss, Allemagne) en utilisant un objectif 40X à immersion dans l'eau (C-Apochromat ; NA=1.2). La température et le CO<sub>2</sub> ont été contrôlés grâce au système d'incubation et à la platine chauffante (Zeiss, Allemagne).

## **I.3 Résultats**

### **I.3.A Identification de granules de stress secondaires.**

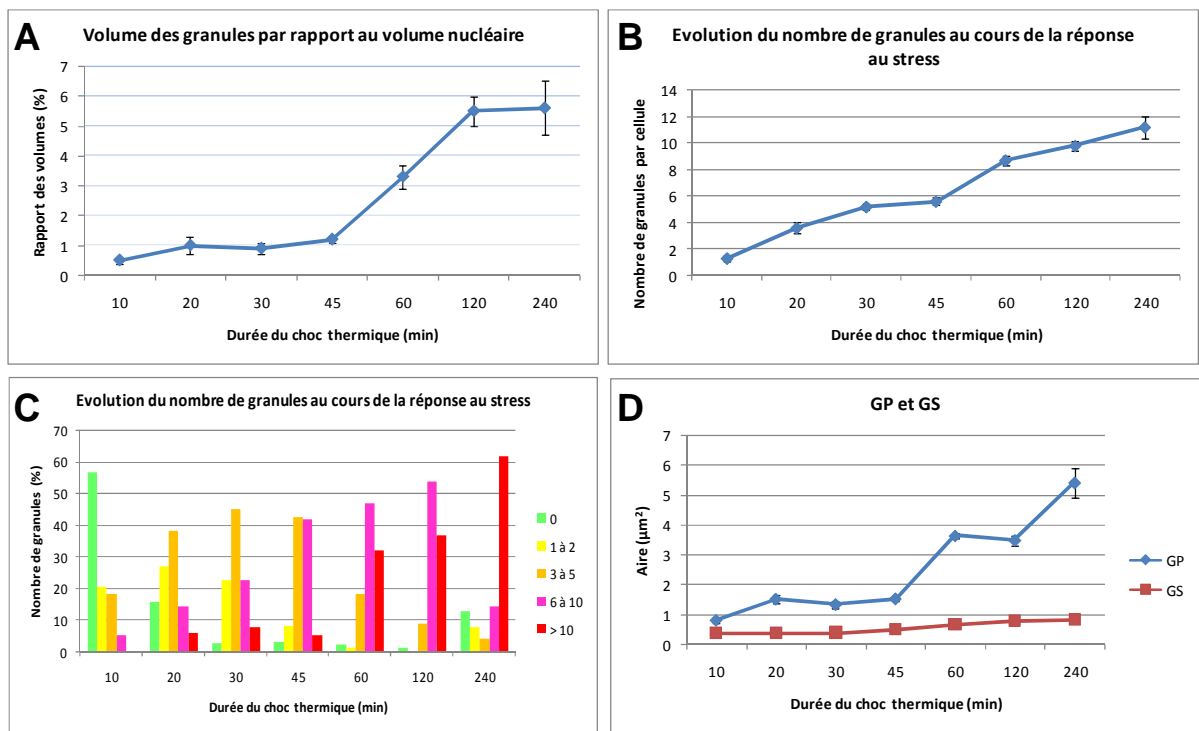
Afin de confirmer la présence d'autres sites de liaison d'HSF1 dans des cellules soumises à un choc thermique, nous avons analysé, par immunofluorescence, la distribution d'HSF1 dans différentes lignées cellulaires normales et tumorales. Nous nous avons utilisé des cellules tumorales HeLa (cancer du col de l'utérus) et HCT116 (cellules diploïdes humaines de carcinome colorectal) ainsi que deux cultures primaires de cellules normales, des fibroblastes embryonnaires de poumon appelées IMR90 et des fibroblastes de peau (SF) d'un individu adulte en bonne santé.

À 37°C, un marquage nucléaire diffus est observé aussi bien dans des cellules normales que dans les cellules tumorales (Figure 50-A). Après un choc thermique, des foyers nucléaires d'HSF1 sont visibles dans toutes les lignées cellulaires. Comme décrit précédemment, deux granules de stress correspondant aux deux loci 9q12 sont détectés dans les cellules normales (Cotto, Fox et al. 1997; Jolly, Morimoto et al. 1997). Dans environ 25% de ces cellules, un petit foyer surnuméraire peut être décelé. Les cellules HeLa présentent de 2 à 4 grosses granules de stress, désignées sous le terme « granules de stress primaires » ou GP, ainsi que de nombreux foyers de taille très inférieure que nous avons appelés « granules de stress secondaires » ou GS (flèches). La codétection d'HSF1 avec le centromère du chromosome 9, réalisée par FISH ADN dans des cellules HeLa soumises à un choc thermique, confirme que seuls les GP correspondent aux locus 9q12 (Figure 50-B). Dans les cellules diploïdes HCT116, deux GP sont présentes ainsi que de nombreuses GS, observées dans plus de 90% des cellules (flèches). Ces observations montrent donc qu'HSF1 est capable de se lier à des sites secondaires distincts des loci 9q12.



**Figure 50 :** **(A)** Analyse de la distribution d'HSF1 dans différentes lignées cellulaires humaines exposées à un choc thermique. La protéine HSF1 a été détectée par immunofluorescence dans deux cultures primaires normales, les IMR90 et des fibroblastes de peau (SF), ainsi que dans deux lignées cellulaires tumorales, HeLa et HCT116, exposées à une heure de choc thermique à 43°C (cellules tumorales) ou à 45°C (les cellules normales). HSF1 forme deux foyers dans le noyau des cellules normales, alors que dans les cellules tumorales trois à quatre gros foyers sont présents, avec en plus de nombreux autres petits foyers (flèches). **(B)** Codétection d'HSF1 et du locus 9q12 par immunofluorescence (vert) et FISH ADN (rouge) dans des cellules HeLa exposées à une heure de choc thermique. Les quatre gros foyers HSF1 colocalisent avec les quatre copies du locus 9q12 présents dans ces cellules pluriploïdes, mais ce n'est pas le cas des petits foyers (flèches).

Afin de mieux décrire les GP et les GS, une analyse quantitative des granules de stress a été réalisée dans des cellules HeLa. HSF1 a été détecté par immunofluorescence à différents moments au cours du choc thermique. Le nombre et la taille moyenne des GP et des GS ont été évalués. Les résultats sont présentés dans la Figure 51.



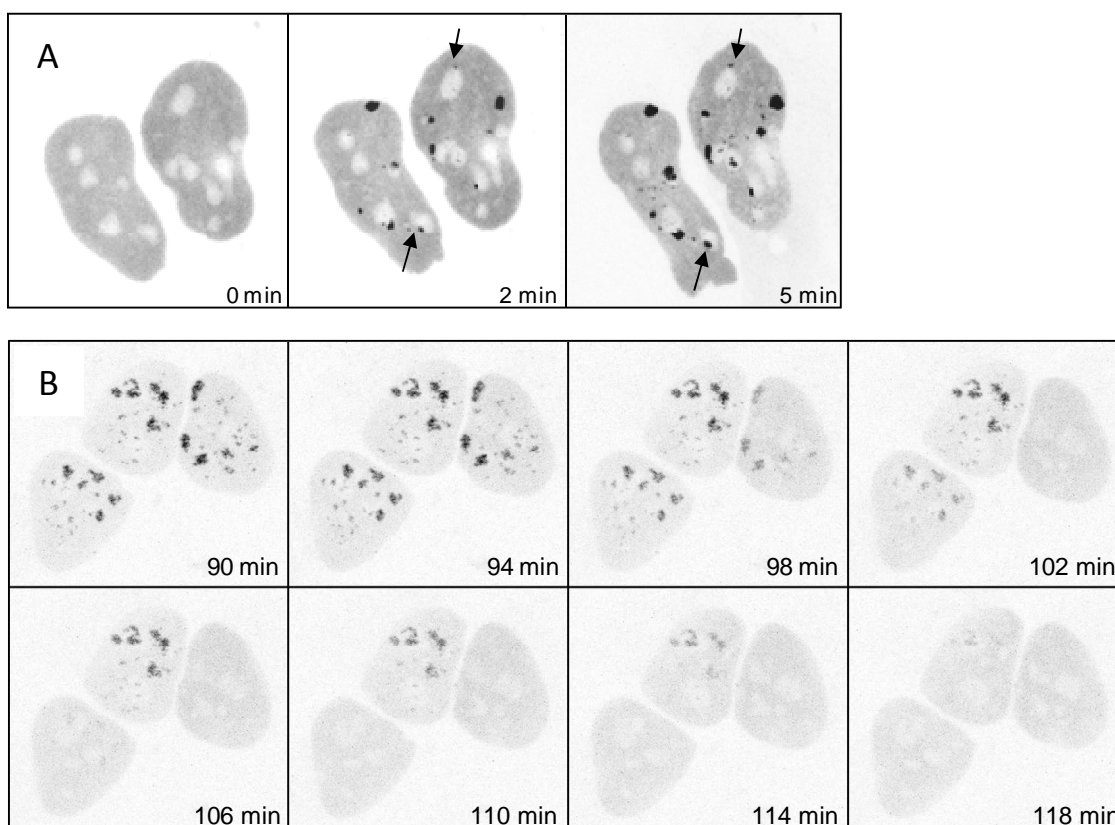
**Figure 51 :** Dynamique de localisation d'HSF1 dans les granules de stress primaires (GP) et secondaires (GS). Le nombre et la taille des GP et des GS ont été évalués après différentes périodes d'exposition au choc thermique (de 10 min à 4h) (A) Densité de volume des granules (volume total des granules / volume nucléaire). (B) Evolution du nombre moyen de granules. (C) Distribution du nombre de granules (D) Surface moyenne des GP et des GS.

Comme le montre la Figure 51-A, le volume des granules de stress augmente de manière significative au cours du choc thermique. En effet, les granules de stress occupent moins de 1% du volume nucléaire total après un choc thermique court (20 min), et plus de 5% après deux heures d'exposition au stress. Cela est dû à la fois à une augmentation du nombre et de la taille des GP et des GS. En effet, le nombre moyen de granules de stress par noyau est de 1,3 après 10 min de choc thermique, et de plus 11 après quatre heures de choc thermique (Figure 51-B). De plus, la superficie moyenne des granules de stress passe d'environ  $0,9 \mu\text{m}^2$  après 10 minutes de choc thermique à  $5,8 \mu\text{m}^2$  après 4 heures de choc thermique (Figure 51-D). Des cellules dépourvues de granules de stress ont été principalement observées après de courtes périodes de choc thermique tandis que les cellules avec plus de 10 granules de stress sont détectées après des expositions prolongées au stress (Figure 51-C). Des GP et des GS sont toutes les deux détectées, au plus tôt, après 10 minutes de choc thermique, suggérant une apparition concomitante de ces foyers (Figure 51-C). Enfin, la taille maximale des GS, observée après une longue exposition au stress (2-4h), est égale à la taille moyenne de toutes les granules observées pendant les dix premières



minutes de choc thermique (Figure 51-D). Ces résultats montrent donc que les GP et les GS apparaissent simultanément et augmentent en taille au cours de la réponse stress pour atteindre un maximum après 2 à 4 heures de choc thermique.

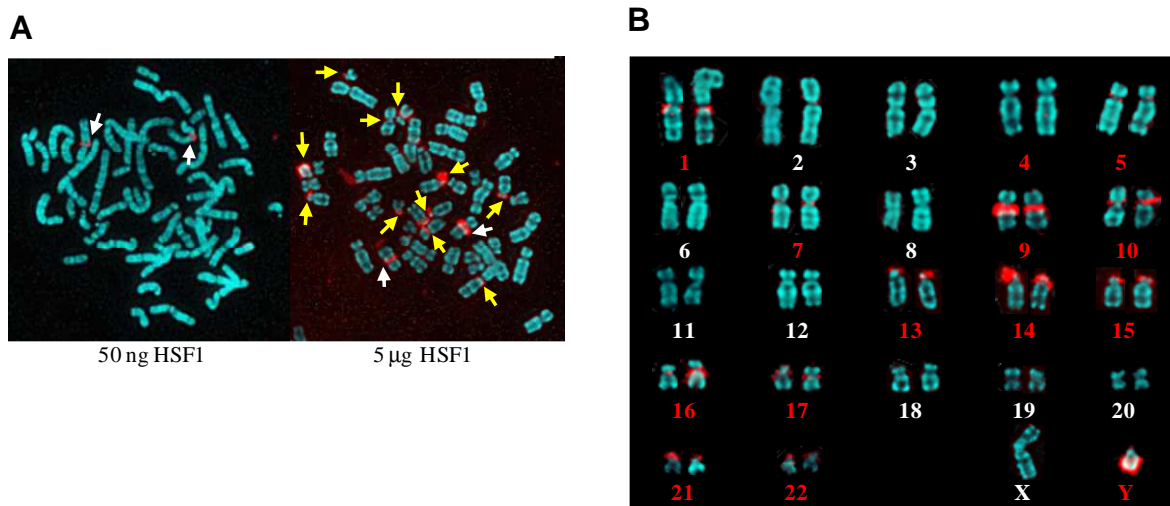
Nous avons également étudié la dynamique d'apparition et de disparition des GP et des GS dans des cellules vivantes. Des cellules HeLa exprimant de manière transitoire une protéine HSF1 couplée à la GFP (HSF1-GFP) ont été soumises à un choc thermique de 30 minutes à 42°C sur la platine du microscope puis placées en récupération à 37°C. Les images ont été acquises à différents moments au cours de l'exposition au stress et lors de la récupération. Comme le montre la Figure 52-A, les GP et les GS (flèches) apparaissent simultanément pendant les premières minutes du choc thermique. De même, les GP et les GS disparaissent de manière concomitante lors de la récupération après une heure de choc thermique (Figure 52-B).



**Figure 52** : Cinétique d'apparition et de disparition des granules de stress sur des cellules vivantes transfectées avec HSF1-GFP. **(A)** Chronologie de formation au cours de l'exposition au choc thermique. Le choc thermique de 43°C a été induit sur la platine du microscope. **(B)** Chronologie de disparition des granules de stress au cours de la période de récupération après 30 minutes de choc thermique à 43°C.

### I.3.B HSF1 se lie aux satellites 2 et satellites 3

Afin d'identifier les chromosomes cibles de ces GS, une expérience de reconstitution de granules *in vitro* a été effectuée. Le principe de cette approche est la suivante: une protéine HSF1 recombinante est incubée sur des chromosomes métaphasiques avant d'être détectée par immunofluorescence (Jolly, Konecny et al. 2002). De part l'absence de régulateurs négatifs, la protéine HSF1 recombinante possède les mêmes propriétés de liaison à l'ADN que les trimères actifs d'HSF1 induits par le stress (Kroeger, Sarge et al. 1993). De manière très intéressante, les résultats varient en fonction de la quantité de protéine HSF1 utilisée. Ainsi, seulement deux signaux sont observés sur les deux chromosomes 9 en présence de faibles quantités de protéines (50-100 ng) (Jolly, Konecny et al. 2002) (Figure 53-A). En revanche, lorsque des quantités plus importantes de protéines HSF1 ont été utilisées (2-5 µg), plusieurs signaux fluorescents sont détectés sur d'autres chromosomes (Figure 53-A).

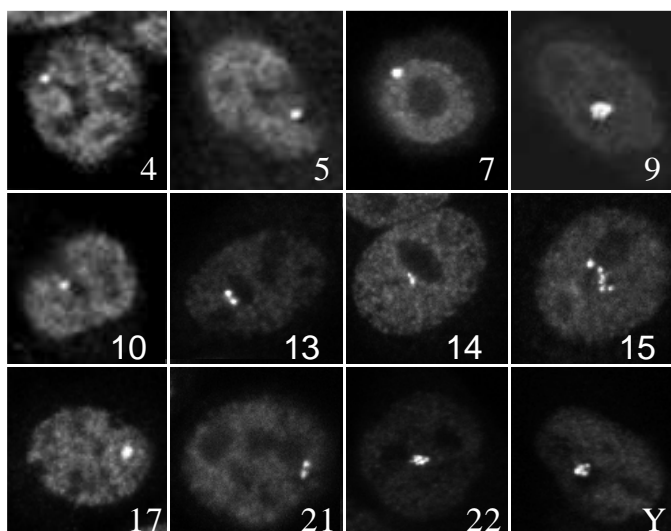


**Figure 53 :** (A) Reconstitution *in vitro* des granules de stress. Les chromosomes métaphasiques humains ont été incubés avec 500 ng (à gauche) ou 5 µg (à droite) de protéine HSF1 recombinante. Celle-ci a ensuite été détectée par immunofluorescence (rouge). Avec 500 ng de protéines, deux signaux fluorescents sont détectés sur les deux chromosomes 9 (flèches blanches). Avec 5 µg de protéines, de nombreux signaux sont observés sur les régions péricentromériques de plusieurs autres chromosomes (flèches jaunes). (B) Caryotype de la métaphase présentée figure A (à droite). Les chromosomes présentant un signal fluorescent sont numérotés en rouge.

Grâce aux caryotypes réalisés à partir de trois expériences indépendantes, nous avons identifié les différents chromosomes capables de lier HSF1. Il s'agit des chromosomes 1, 5, 7, 9, 10, 13, 14, 15, 16, 17, 21, 22 et Y (Figure 53-B). Ces

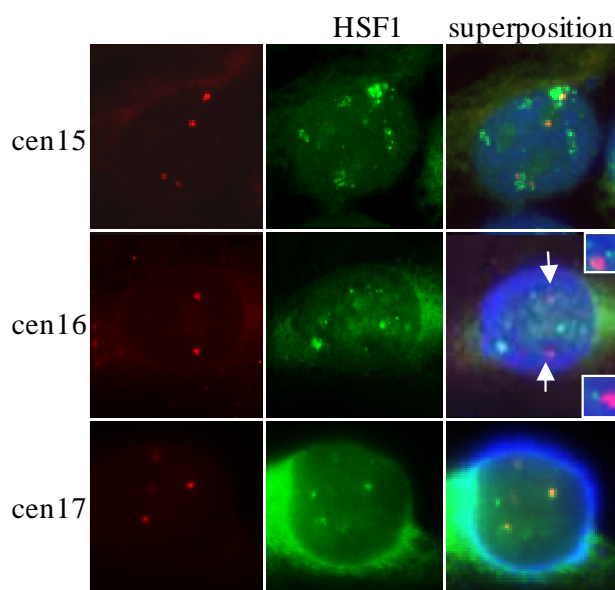
chromosomes peuvent être classés en fonction de la taille des signaux obtenus pour HSF1 : 9, Y, 1, chromosomes acrocentriques, 16, 17, 10, 7 et 5.

Une approche *in vivo* a également été utilisée pour confirmer ces observations *in vitro*. Profitant du fait que les granules de stress ne se forment pas dans les cellules de rongeurs, nous avons analysé par immunofluorescence la distribution du facteur HSF1 après un choc thermique dans des lignées hybrides Homme-hamster qui ne contiennent qu'un seul chromosome humain. Comme le montre la Figure 54, les chromosomes capables de lier HSF1 dans l'expérience de reconstitution de granules *in vitro* (à l'exception des chromosomes 1 et 16) sont également en mesure de diriger la formation de foyers HSF1 dans ces lignées cellulaires hybrides. Contrairement à l'expérience de reconstitution *in vitro*, le chromosome 4 semblerait être également une cible chromosomique d'HSF1. Cette exception peut s'expliquer par le fait que cette lignée hybride en particulier aurait pu conserver une partie d'un autre chromosome humain, comme cela a été décrit pour d'autres lignées humaines hybrides (Antonacci, Marzella et al. 1995). De même, l'incapacité des chromosomes 1 et 16 à diriger la formation des foyers HSF1, pourrait être liée à des réarrangements génomiques conduisant à la perte des régions péri-centromériques. De plus, certaines lignées cellulaires hybrides sont instables et peuvent perdre le chromosome humain assez facilement lorsqu'elles sont mise en culture, et ce même sous des conditions de pression de sélection.



**Figure 54** : Analyse par immunofluorescence de la distribution d'HSF1 dans des lignées cellulaires hybrides Homme-hamster exposées à une heure de choc thermique à 43 °C. Une granule de stress est observée dans le noyau des cellules de rongeurs qui contiennent l'un des chromosomes suivants : 4, 5, 7, 9, 10, 13, 14, 15, 17, 21, 22 ou Y.

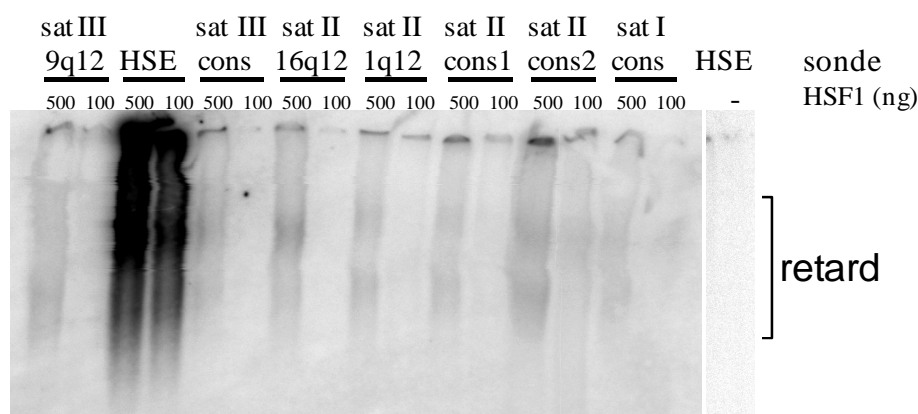
Pour confirmer que les chromosomes identifiés comme des cibles d'HSF1 sont en effet capables de lier HSF1 *in vivo* dans des cellules humaines, nous avons combiné des expériences d'immunofluorescence et de FISH ADN pour la détection simultanée d'HSF1 et des centromères de certains chromosomes cibles. Comme le montre la Figure 55, HSF1 est détecté à proximité des centromères des chromosomes 15, 16 et 17, confirmant ainsi que ces chromosomes sont capables de lier HSF1 *in vivo*.



**Figure 55 :** Codétection d'HSF1 et du centromère du chromosome 15 (panneau supérieur), 16 (panneau médian) ou 17 (panneau inférieur) par immunofluorescence (vert) et par FISH ADN (rouge) dans des cellules HeLa exposées à une heure de choc thermique. Les granules de stress sont toujours détectées à proximité de ces centromères.

Il est important de noter que, dans les expériences de reconstitution de granules de stress *in vitro*, toutes les cibles d'HSF1 sont situées au niveau des régions péri-centromériques des chromosomes, à l'exception du chromosome Y pour lequel HSF1 se lie sur le bras long (Figure 53-A). Toutes ces régions sont connues pour être enrichies en séquences répétées de type satellites 2 et/ou 3 (Frommer, Paul et al. 1988; Tagarro, Fernandez-Peralta et al. 1994). Pour confirmer le fait qu'HSF1 puisse en effet lier les satellites 2 et 3, nous avons réalisé une expérience de gel retard en utilisant la protéine HSF1 recombinante ainsi que plusieurs sondes. Des sondes représentant les satellites 1, 2 et 3 ainsi que des sondes reconnaissant de manière spécifique les satellites 2 des chromosomes 1 et 16. D'autre part, la sonde HSE qui contiennent le motif consensus de liaisons à l'ADN d'HSF1 ainsi qu'une sonde spécifique des satellites 3 du locus 9q12 ont été utilisées comme contrôles positifs

(Jolly, Konecny et al. 2002). Comme le montre la Figure 56, en présence d'une grande quantité d'HSF1, un retard de migration très net est observé avec les différentes sondes satellites 2 et 3 alors qu'un retard modéré est détecté pour les séquences satellites 1. Ainsi ces observations confirment l'affinité d'HSF1 pour les séquences satellites 2 et 3.

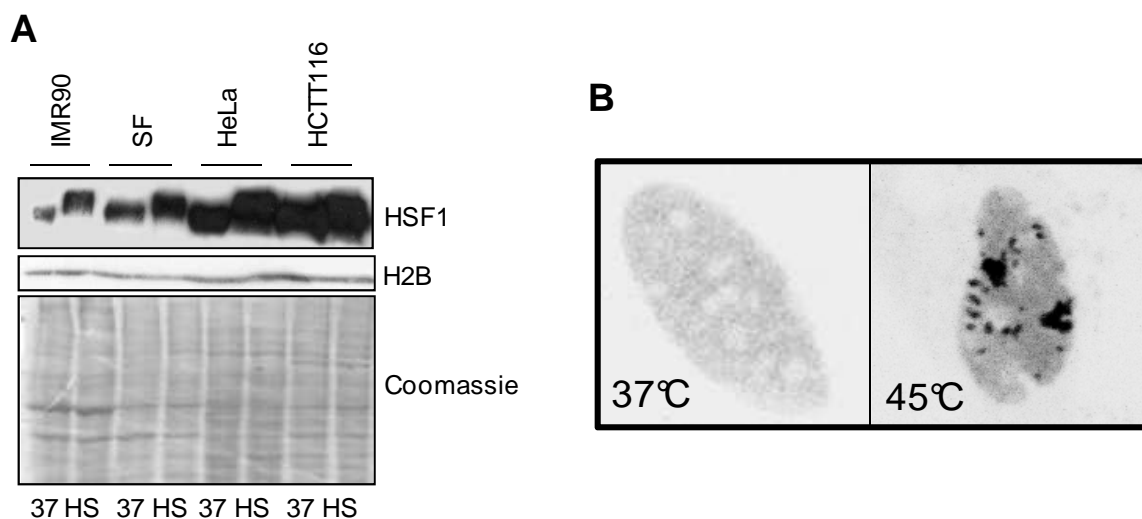


**Figure 56 :** Gel retard réalisé avec différentes sondes représentant les satellites 1, 2 ou 3. La sonde HSE est utilisée en contrôle positif. Chaque sonde a été incubée avec 100 ou 500 ng de protéine recombinante HSF1. La sonde HSE libre est présentée dans la dernière piste. Un retard de migration est détecté pour toutes les sondes satellites 2 et 3 en présence de 500 ng de protéine. Un retard mineur est observé avec la sonde spécifique des satellites 1

### I.3.C La formation des granules de stress secondaires est dépendante de la quantité d'HSF1.

Les résultats obtenus lors des expériences de reconstitution de granules de stress *in vitro* suggèrent que l'apparition des GS pourrait dépendre de la quantité d'HSF1. On pourrait donc imaginer que l'augmentation du nombre de GS dans des cellules tumorales (Figure 50) pourrait être due à une quantité plus importante d'HSF1 dans ces cellules par rapport aux cellules normales. En accord avec cette proposition, une surexpression d'HSF1 a été décrite dans certains types de cancers tels que le cancer colorectal ou encore dans des adénocarcinomes de prostate (Hoang, Huang et al. 2000; Cen, Zheng et al. 2004). Pour tester cette hypothèse, la quantité d'HSF1 présente dans les lignées normales et tumorales a été analysée par western blot. Comme le montre la Figure 57-A, HSF1 est détectée dans toutes les lignées cellulaires. Le retard de migration d'HSF1 observé dans les échantillons soumis à un choc thermique est du à des modifications post-traductionnelles de la protéine (Holmberg, Tran et al. 2002). Ce retard de migration est le même pour toutes les

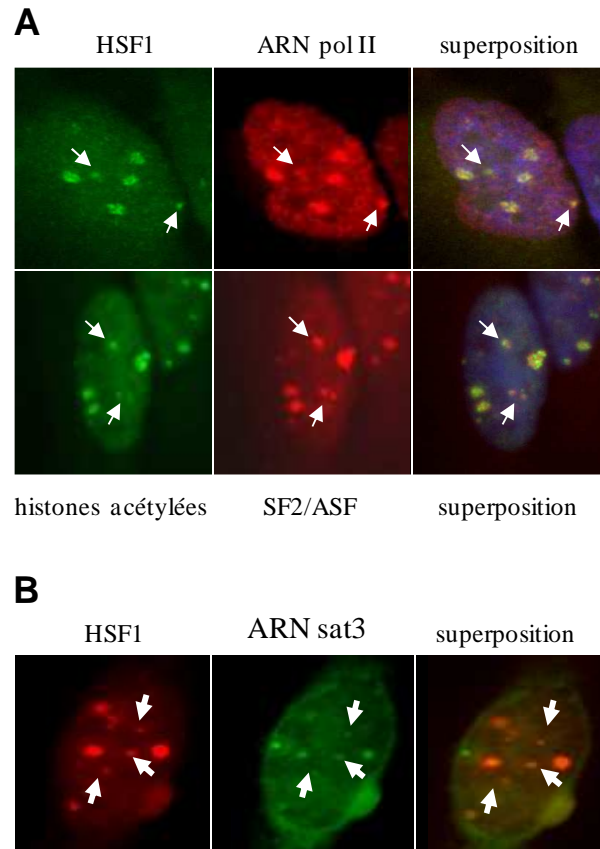
cellules étudiées, confirmant ainsi que suite au choc thermique, HSF1 est activé de façon similaire dans tous les types de cellules. Toutefois, pour une quantité égale de protéines totales, HSF1 est présente en quantité plus élevée dans les cellules tumorales par rapport aux cellules normales (Figure 57-A).



**Figure 57 :** (A) Analyse par western blot de la quantité d'HSF1 dans différentes lignées cellulaires. 10 mg d'extraits protéiques totaux issus des cellules IMR90, SF, HeLa et HCT116 soumises ou non à un choc thermique d'une heure ont été déposés sur un gel d'acrylamide 8%, transférés sur nitrocellulose et hybridés avec un anticorps anti-HSF1 (panneau supérieur). Le contrôle de charge a été effectué par la détection de l'histone H2B (panneau médian) et par la coloration au bleu de Coomassie (panneau inférieur). Les lignées cellulaires tumorales HeLa et HCT116 présentent une quantité d'HSF1 plus importante que celle des cellules normales SF et IMR90. (B) Analyse de la distribution d'HSF1 dans des cellules IMR90 exprimant de manière transitoire une protéine HSF1 fusionnée à la GFP. La protéine marquée présente une localisation diffuse à 37° (à gauche). Après une heure de choc thermique, elle se relocalise dans plusieurs foyers de différentes tailles (à droite).

Ainsi, ce résultat suggère que la formation des GS est dépendante de la quantité d'HSF1. Afin de confirmer cette hypothèse, nous avons surexprimé une protéine HSF1 couplée à la GFP (HSF1-GFP) dans des fibroblastes normaux (IMR90) et analysé la distribution de la protéine après un choc thermique. À 37°C, un marquage nucléaire diffus est observé (Figure 57-B). Cependant, deux granules majoritaires, accompagnées par de plus petites qui ressemblent aux GS présentes dans les cellules tumorales, sont détectées. Ce résultat confirme donc que la formation des GS est bien dépendante de la quantité d'HSF1 présente dans la cellule.

### I.3.D Les granules de stress secondaires sont des sites de transcription



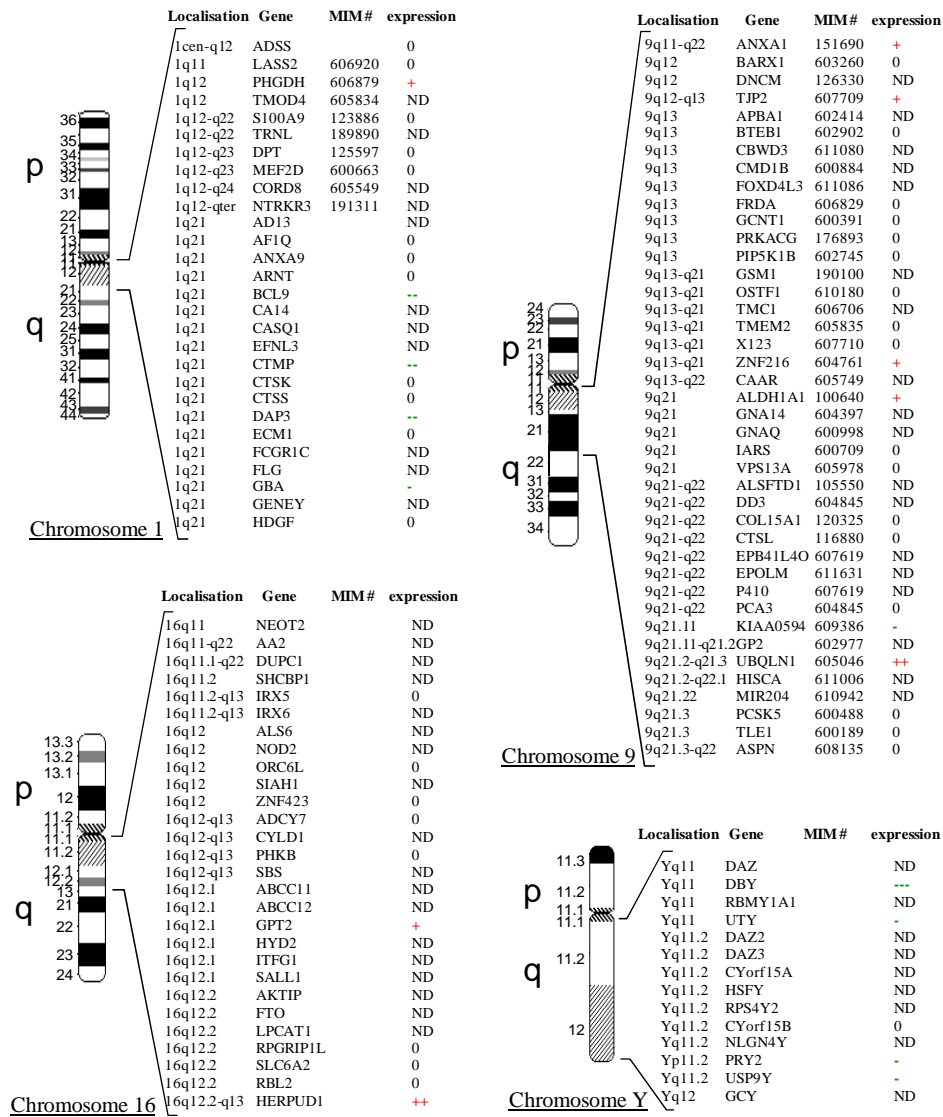
**Figure 58 :** (A) Analyse de la localisation d'HSF1 (vert, panneau supérieur), de l'ARN polymérase II (rouge, panneau supérieur), des histones acétylées (vert, panneau inférieur) et du facteur d'épissage SF2/ASF (rouge, panneau inférieur) dans des cellules HeLa exposées à une heure de choc thermique à 43°C. Toutes les protéines sont présentes dans les GS (flèches). (B) Analyse par FISH ARN des ARN satellites 3 (rouge) par rapport aux foyers HSF1 détectés par immunofluorescence (vert) dans des cellules HeLa exposées à une heure de choc thermique à 43°C. Les ARN satellites 3, détectés avec une sonde consensus représentant l'ensemble des séquences satellites 3, sont présents dans les GS (flèches).

Nous avons précédemment montré qu'au cours du choc thermique, les séquences satellites 3 du locus 9q12 sont activement transcrites par l'ARN polymérase II et d'une manière dépendante d'HSF1 (Jolly, Metz et al. 2004). Nous avons donc cherché à déterminer si les cibles secondaires d'HSF1 pouvaient également être transcrites après un stress. Pour répondre à cette question, nous avons d'abord examiné par immunofluorescence, dans les GS, la présence d'autres marqueurs de la transcription. L'ARN polymérase II, les histones acétylées, et le facteur d'épissage SF2/ASF ont ainsi été co-détectés avec HSF1 dans des cellules HeLa soumises à un

choc thermique. Comme le montre la Figure 58-A, nous avons observé la présence de ces protéines dans toutes les GS (flèches). Ainsi, les GS pourraient également correspondre des sites de transcription. Afin de tester cette hypothèse, une FISH ARN a été réalisée en utilisant une sonde spécifique des séquences satellites 3 (Tagarro, Fernandez-Peralta et al. 1994). Une immunofluorescence dirigée contre HSF1 a été réalisée en parallèle. Comme le montre la Figure 58-B des ARN satellites 3 sont effectivement détectés dans les GS. De manière intéressante, les intensités relatives des marquages fluorescents indiquent que les GP contiennent une plus grande quantité d'HSF1, d'ARN polymérase II, d'ARN satellites 3 et de facteur d'épissage SF2/ASF que les GS. Ce résultat suggère donc que la capacité transcriptionnelle des GP est plus importante que celle des GS.

Sachant que le choc thermique induit l'activation de plusieurs régions péricentromériques, nous avons émis l'hypothèse selon laquelle cette activité transcriptionnelle pourrait influencer localement la transcription des gènes adjacents. Pour tester cette proposition, nous avons exploré la base de données de Murray et al ([http://microarray-pubs.stanford.edu/human\\_stress/Home.shtml](http://microarray-pubs.stanford.edu/human_stress/Home.shtml)) (Murray, Whitfield et al. 2004). Dans ce document, les auteurs ont analysé par puce à ADN, l'expression de nombreux gènes en réponse à différents stress, y compris de choc thermique et ce, dans différentes lignées cellulaires humaines. L'ensemble de leurs résultats est disponible sous la forme d'une base de données facilement exploitable pour tout gène d'intérêt. Ainsi, nous avons d'abord identifié dans la base de données OMIM (<http://www.ncbi.nlm.nih.gov/sites/entrez?db=omim>) les gènes adjacents à la région péricentromérique des chromosomes 1, 9, 13, 14, 15, 16, 21, 22 et Y. Nous avons ensuite étudié dans la base de données de Murray et al, l'expression de chacun de ces gènes en fonction de la réponse au stress. Les résultats obtenus pour les chromosomes 1, 9, 16 et Y, qui contiennent les plus grands blocs hétérochromatiques enrichis en satellites 2 et/ou 3, sont présentés dans la Figure 59. Aucun changement significatif d'expression des gènes n'a été observé pour l'ensemble des chromosomes analysés suggérant ainsi que la transcription des régions péricentromériques sous l'effet du choc thermique n'affecte pas l'expression des gènes adjacents.





**Figure 59** : Analyse du profil d'expression, au cours de la réponse au stress thermique, de gènes adjacents aux péricentromères des chromosomes 1, 9, 16 et Y. Cette analyse a été réalisée grâce à l'exploitation de la banque de données de Murray et al (Murray, Whitfield et al. 2004). Les modifications d'expression des gènes au cours du stress sont indiquées : + ou ++ gènes dont l'expression est augmentée ; - ou -- gènes dont l'expression est diminuée ; 0 pas de changement d'expression des gènes ; ND non déterminé.

## I.4 Conclusion

Dans ce travail, nous avons cherché à identifier des cibles chromosomiques potentielles d'HSF1, distinctes des séquences satellites 3 du locus 9q12. Nous montrons que le choc thermique induit en effet la liaison d'HSF1 à la région

péricentromérique, enrichies en satellites 2 et 3, de plusieurs chromosomes. Nous avons nommé ces nouvelles cibles « granules de stress secondaires » (GS).

De manière très intéressante, les chromosomes cibles d'HSF1 que nous avons caractérisés coïncident parfaitement avec les chromosomes décrits pour être enrichis en séquences satellites 2 et/ou 3. Par exemple, Frommer et al ont analysé la distribution génomique des séquences satellites 3 par hybridation *in situ*, en utilisant une sonde consensus (Frommer, Paul et al. 1988). Ces auteurs ont montré que lorsque la sonde est utilisée à faible concentration, l'hybridation a eu lieu exclusivement sur le locus 9q12. Cependant, avec des quantités croissantes de sonde, l'intensité du marquage s'intensifie sur le chromosome 9 et l'hybridation est également détectée sur les chromosomes Y, 1, et sur les chromosomes acrocentriques. À plus forte concentration encore, des signaux ont également été observés sur les chromosomes 16, 17, 20 et occasionnellement sur les chromosomes 5 et 10. Tagarro et al ont également montré par FISH ADN, en utilisant une sonde satellite 3, qu'avec une forte concentration de sonde, des signaux d'hybridation sont détectés sur les chromosomes Y, 1, 5, 10, 17, 20 et sur les chromosomes acrocentriques, en plus du locus 9q12 (Tagarro, Fernandez-Peralta et al. 1994). Enfin, la technique du Giemsa-11, qui permet de marquer certaines régions hétérochromatiques, révèle un marquage particulier des chromosomes 1, 5, 7, 10, 13, 14, 15, 17, 20, 21, 22 et Y en plus du chromosome 9 (Bobrow, Madan et al. 1972). Par FISH ADN, les satellites 2 ont été principalement localisés sur les chromosomes 1 et 16, et à plus haute concentration de sonde sur les chromosomes 2, 10, 7, 15, 17 et 22 (Tagarro, Fernandez-Peralta et al. 1994). Ces observations suggèrent donc que le nombre de copies des séquences satellites 2 et 3 est différent sur chacun de ces chromosomes.

Dans ce contexte, l'augmentation du nombre et la variabilité de la taille des cibles chromosomiques d'HSF1 dans l'expérience de reconstitution de granules *in vitro* soutient l'idée que la taille des GS pourrait corrélérer avec le nombre de copies de séquences satellites 2 et/ou 3. En effet, le classement des chromosomes selon la taille des signaux d'HSF1 est très similaire à celui obtenu sur la base des signaux de FISH ADN pour les séquences satellites 3 (Frommer, Paul et al. 1988). De plus, l'apparition et la disparition concomitantes des GP et des GS montrent sans ambiguïté que la formation des GS n'est pas due à une saturation des principales cibles d'HSF1, le locus 9q12. Cette hypothèse est renforcée par le gel retard sur lequel HSF1 affiche une affinité similaire pour les séquences satellites 2 et 3, quelle que soit la sonde utilisée. Les séquences satellites 2 et 3 sont caractérisées par un élément consensus commun, GGAAT, qui contient l'élément HSE (nGAAn) (Tagarro, Fernandez-Peralta et al. 1994).

L'affinité d'HSF1 pour les séquences satellites 2 et 3 est très probablement due à cette similitude. De plus, l'expérience de surexpression d'HSF1 dans les cellules normales indique clairement que la taille et le nombre des GS dépend directement de la quantité d'HSF1. Ainsi, nos données suggèrent fortement que l'apparition des GS dépend directement du nombre de séquences satellites présent dans le locus cible, ainsi que de la quantité d'HSF1 présente dans la cellule. Dans ce contexte, le fait que des GS ne soient pas observées sur le chromosome 20, ni dans la reconstitution de granules *in vitro*, ni dans les hybrides Homme-hamster, pourrait s'expliquer par la présence d'un nombre de copies des satellites 3 moins importantes sur ce chromosome par rapport aux autres.

D'autre part, les régions péricentromériques des cellules tumorales pourraient également présenter une structure chromatinienne différente de celle des cellules normales. Ceci pourrait entraîner une plus grande accessibilité des séquences répétées à HSF1 dans les cellules tumorales. L'hypométhylation de l'ADN est en effet une caractéristique commune des cellules cancéreuses, en particulier dans les régions hétérochromatiques (Narayan, Ji et al. 1998; Ehrlich 2002; Feinberg and Tycko 2004; Wilson, Power et al. 2007). On peut ainsi imaginer que les régions péricentromériques hypométhylées, et donc formées de chromatine moins compacte, puissent être plus accessibles dans les cellules tumorales que dans les cellules normales.

Une conclusion importante de ce travail est le fait que les GS sont des sites de transcription actifs. Ces constatations sont importantes en ce qui concerne le nombre croissant de preuves montrant que des séquences répétées jouent un rôle clé dans la régulation des gènes (Huttenhofer, Schattner et al. 2005; Yazgan and Krebs 2007; Amaral and Mattick 2008; Tomilin 2008). Dans ce contexte, le fait que le choc thermique induit une forte activité transcriptionnelle des séquences péricentromériques suggère que cette transcription pourrait affecter des fonctions cellulaires et en particulier la transcription. Notre analyse des données transcriptomiques de Murray et al montrent déjà que cette transcription massive n'affecte pas la transcription des gènes adjacents (Murray, Whitfield et al. 2004). Cependant, cette transcription pourrait influencer l'expression de gènes plus distants, qui pourraient être localisés sur d'autres chromosomes, en créant des zones de transcription actives au sein du noyau. Les futures études visant à comprendre la fonction de ces ARN issus des séquences répétées nous permettront de mieux appréhender leur rôle dans la régulation de la réponse au stress.

# Chapitre II Transcription des séquences satellites du génome humain

Les résultats présentés dans ce chapitre, font l'objet d'un deuxième article en cours de préparation :

« *The RepChip: a sensitive tool dedicated to satellite sequences expression* »

*Angéline Eymery, Béatrice Horard, Michèle El Atifi, Geneviève Fourel, Nikita Vassetsky, Sophie Rousseaux, Saadi Khochbin, Eric Gilson and Claire Vourc'h*

Ce travail a été réalisé en étroite collaboration avec les groupes des Dr Saadi Khochbin, Eric Gilson et François Berger. Après avoir introduit cette étude et le contexte qui la rendait nécessaire, les résultats seront présentés et discutés. Enfin, des résultats complémentaires seront ajoutés parmi lesquels certaines données intéressantes qui seront publiées dans un troisième article pour lequel je figure en second auteur :

« *Global analysis of expression and epigenetic status of human repetitive DNA* »

*Béatrice Horard, Angéline Eymery, Geneviève Fourel, Nikita Vassetzky, Jacques Puechberty, Gérard Roizes, Kevin Lebrigand, Pascal Barbry, Aurélie Laugraud, Christian Gauthier, Elsa Bensimon, Claire Vourc'h and Eric Gilson*

## II.1 Introduction

Nous avons montré que le locus 9q12 n'est pas l'unique cible hétérochromatique d'HSF1 sous l'effet du choc thermique. Par des études réalisées *in situ*, nous avons en effet caractérisé d'autres régions chromosomiques capables de lier HSF1. Ces séquences correspondent à des régions péri-centromériques enrichies en séquences répétées de type satellites 2 et 3. Nous nous sommes donc intéressés au potentiel transcriptionnel des séquences répétées du génome humain présentes non seulement au niveau des péri-centromères (PCT) mais également au niveau des centromères (CT). En effet, depuis la mise en évidence, chez *S.pombe*, de l'implication d'ARN non codants péri-centromériques dans la structure et le maintien de l'hétérochromatine, la

recherche, chez les mammifères, de ce type de transcrits et des mécanismes impliqués dans cette transcription constituait un sujet d'une ampleur importante.

Malgré l'importance fonctionnelle des CT et des PCT au cours de la mitose, très peu de choses sont connues quant à leur expression chez les mammifères et en particulier chez l'Homme. En effet, le haut degré de polymorphisme et l'absence de séquençage de ces séquences répétées sont autant d'obstacles à une analyse transcriptionnelle correcte. Quelques exemples attestent cependant d'une transcription de ces séquences (voir chapitre 3 de l'introduction), mais ces études sont généralement focalisées sur un type de séquence particulier et dans une condition physio(patho)logique donnée. Les techniques d'analyses qui sont couramment utilisées, telles que les northern blot ou les FISH ARN, ne permettent pas une étude à grande échelle qui serait pourtant nécessaire aux vues du polymorphisme de ces séquences. De plus, des techniques sensibles, comme la Q-PCR, ne sont pas adaptées à l'analyse transcriptomique des séquences répétées.

Dans ce contexte général, notre objectif a été de fournir un outil permettant une analyse quantitative de l'expression des séquences répétées centromériques (alpha satellites) et péricentromériques (satellites 2 et 3), en tenant compte de leurs spécificités de séquences. De plus cet outil devait être suffisamment sensible pour détecter de faibles taux d'expression. Nous avons ainsi développé une approche transcriptomique dédiée à l'analyse de l'expression des séquences centromériques et péricentromériques. Nous avons baptisé cette macro-array « RepChip ». Cette approche présente plusieurs avantages par rapport à d'autres méthodes d'analyse d'expression des séquences centromériques et péricentromériques. Elle permet en effet une vision globale de la capacité transcriptionnelle de ces séquences tout en évitant la sélection d'amorces (susceptibles de restreindre le domaine de l'analyse), et l'utilisation de procédures de pré-amplification par PCR, méthodes difficiles à gérer lorsqu'il s'agit des séquences répétées hautement polymorphes. De plus, nous avons choisi de développer cet outil en utilisant comme support des membranes de nylon et un marquage radioactif. Cette combinaison a été préférée à celle du marquage fluorescent sur lame de verre car elle permet d'obtenir une sensibilité de détection accrue due à la fois au marquage radioactif mais également aux capacités d'absorption du nylon. Ce travail pour but de présenter à la fois la validation de la RepChip mais également de dresser un portrait de conditions capables d'induire la transcription des séquences répétées alpha satellites (CT), satellites 2 et satellites 3 (PCT).

## II.2 Matériel et Méthodes

### 1. Culture cellulaire et traitements

Lignées cellulaires : Les cellules HeLa, provenant d'un adénocarcinome de l'utérus, et les cellules A431, provenant d'un carcinome de la vulve, ont été fournies par l'ATCC (American Type Culture Collection, USA). Les fibroblastes primaires IMR90, provenant de poumon fœtal, ont été fournis par l'Institut Coriell (USA). Les cellules HCT116, provenant d'un carcinome de colon, et les cellules U2OS, issues d'un ostéosarcome, ainsi que les cellules déficientes en HSF1 (KD HSF1) nous ont été gracieusement fournies par le Dr Bert Vogelstein (USA), Karine Monier (France) et Lea Sistonen (Finlande) respectivement.

Conditions de stress : Les expériences de choc thermique ont été réalisées par immersion des flasques de culture dans un bain-marie, pendant une heure à 43°C pour les cellules tumorales et une heure à 45°C pour les IMR90.

Utilisation des drogues : La 5 Aza-Cytidine (Aza-C) (Sigma) est utilisée pendant une période de 72h à une concentration de 5µM. Le traitement est renouvelé toutes les 24 heures. La Trichostatine A (TSA) (Sigma) et le butyrate sont utilisés à des concentrations de 330nM et 10mM respectivement et les durées d'incubation sont de 2h (TSA) et 7h (butyrate).

siRNA : Les séquences des siRNA dirigés contre Dicer sont les suivantes : DicerA 5'UUUGUUGCGAGGCUGAUUC3' et DicerB 5'UCUAUUAGCACCUUGAUGU3' (Chendrimada, Gregory et al. 2005). Les doubles transfections (DicerA + DicerB) ont été réalisées dans des cellules HeLa avec de l'oligofectamine (Invitrogen). 80µL de siRNA à 20µM sont incubés 5 minutes à température ambiante avec 1400µL d'optimem (Invitrogen). En parallèle une deuxième réaction est préparée avec 32µL d'oligofectamine et 120µL d'optimem dans laquelle sera ajoutée la préparation contenant les siRNA. L'incubation se fait à température ambiante pendant 20 minutes. La réaction est alors déposée sur une culture de 500000 cellules, dont le milieu (6,4 mL) est composé de DMEM et de 2% de glutamine. L'incubation se fait pendant 4h à 37°C. 4mL de milieu composé de DMEM, 3% pénicilline/streptomycine, 30% SVF et 2% glutamine sont ensuite ajoutés sur les cellules et l'incubation se fait sur la nuit à 37°C. Le lendemain, après un rapide lavage des cellules, une seconde transfection est réalisée selon le même protocole. Enfin, le troisième jour les cellules sont récupérées et les ARN et les protéines peuvent être extraits.

## 2. Extractions d'ARN

Lignées cellulaires : Les cellules sont lavées dans du PBS, puis à l'aide d'un grattoir stérile, décollées dans 1 ml de PBS. La suspension cellulaire ainsi obtenue subit une centrifugation rapide, le culot est repris dans du PBS puis les cellules sont perméabilisées grâce à une incubation de 5 minutes dans du Tri-Reagent (Sigma). Du chloroforme est ensuite ajouté au mélange précédent et après une incubation de 15 minutes à température ambiante et une centrifugation de 15 minutes à 12000 rpm et à 4°C, trois phases apparaissent nettement. La phase aqueuse supérieure, qui contient les ARN, est récupérée, et un volume d'isopropanol est ajouté. Après une incubation à température ambiante de 10 minutes et une centrifugation de 10 minutes dans les mêmes conditions que précédemment, le culot est brièvement rincé à l'EtOH 70%, séché, puis resuspendu dans de l'eau. La qualité des ARN est contrôlée par électrophorèse sur gel d'agarose où l'intégrité des ARNr est vérifiée. La quantification des ARN totaux est réalisée par spectrophotométrie.

Tissus normaux : Les ARN totaux provenant des glandes surrénales, de moelle osseuse, de cerveau, de cerveau fœtal, de foie, de foie fœtal, de cœur, de rein, de poumon, de placenta, de prostate, des glandes salivaires, de muscle squelettique, de testicule, de thymus, de thyroïde, de trachée, d'utérus, de colon et de moelle épinière ont été obtenus par Clontech (Human Total RNA Master Panel II, 636643).

ARN normal et tumoral : Les couples d'ARN provenant de tissu tumoral et de tissu normal adjacent d'un même patient ont été obtenus par Ambion. Ovaire: AM7256 (femme de 32 ans); Foie: AM7244 (Homme de 3 mois); Poumon: AM7224 (Homme de 60 ans); Testicule AM7260 (Homme de 43 ans).

## 3. RepChip

Macro-array : La combinaison de membrane de nylon et du marquage radioactif permet de détecter de faibles quantités d'ARN grâce à la forte capacité d'absorption du nylon et à la sensibilité du marquage radioactif (Bertucci, Bernard et al. 1999). 100nM d'oligonucléotides d'une longueur moyenne de 50 bases sont déposés dans des plaques 96 puits en présence de bleu de bromophénol afin de détecter un éventuel problème de spottage. Les oligonucléotides sont spottés en duplicat sur des membranes de nylon (Hybond N+, Amersham) par le système automatisé GMS 417. Après le spottage, les oligonucléotides sont liés de manière covalente sur la membrane de nylon grâce à une irradiation aux UV à 260 nm.

Marquage et hybridation : 3 µg d'ARN totaux sont transcrits de manière inverse en utilisant comme amorces à la fois des hexanucléotides aléatoires (Roche) et du dt25 (Invitrogen Life Technologies) ainsi qu'un marquage à l' $\alpha$ 33P-dCTP (Amersham ou Perkin-Elmer). Les ARN et les amorces sont incubés pendant 8 minutes à 70°C avant de subir un refroidissement lent à température ambiante. La transcription inverse est réalisée avec un mélange de 40U de RNasin (40 U, RNasin Ribonuclease Inhibitor, Promega), de 30 µCi d'  $\alpha$ 33P-dCTP, de 0.4 mM de dATP, dTTP et dGTP, de 2.4 µM de dCTP (Promega), de tampon de l'enzyme (Promega) et de 400U de Reverse Transcriptase (Promega) et la réaction est incubée 2h à 42°C. Les cDNA marqués de manière radioactive sont purifiés en utilisant une colonne d'exclusion (YM-30, Millipore). L'efficacité de la transcription inverse est évaluée par un comptage (Spectromètre Beckman) d'1µL de la réaction, déposé dans des cupules (Ready Cap™). L'ARN résiduel est éliminé par un traitement à 68°C pendant 30 minutes en présence de 10% SDS, 0.5M EDTA et 3M NaOH. La neutralisation de la réaction est ensuite effectuée par ajout de 1M Tris-HCl et 3N HCl. Les sondes sont ensuite dénaturées à 100°C pendant 5 minutes. L'hybridation des sondes marquées sur la membrane se fait dans un tampon d'hybridation (5X SSC, 5X Denhardt's, 12% BSA) pendant 5 jours à 50°C. Les membranes sont alors lavées pendant 2 heures dans du tampon de lavage (5X SSC, 0.5% SDS) dans un bain marie à 45°C sous agitation. Les membranes sont ensuite séchées et exposées dans une cassette (Fujifilm BAS cassette) pendant 24h puis celle-ci est scannée grâce à un Radio-Imager (BAS 5000, Fuji).

Analyse : L'intensité des différents signaux d'hybridation est quantifiée grâce au Radio-Imager (BAS 5000, Fuji) et aux logiciels Image et Array Gauge. Les résultats sont ensuite corrigés de manière à ce que toutes les données dont l'intensité est inférieure à celle du bruit de fond plus 3 écarts types soient exclues. Les normalisations effectuées entre les différentes séries sont indiquées dans le texte. Les résultats sont présentés sous la forme de box-plots : la boîte rectangulaire représente 75% des valeurs ; la médiane correspond au trait horizontal long au sein de la boîte rectangulaire ; la moyenne est représentée par la marque plus (+) au sein de la boîte ; les extrémités inférieure et supérieure des moustaches sont représentées par des marques en forme de tiret (-) et correspondant respectivement à la plus petite donnée supérieure à  $q1 - 1.5 * (q3 - q1)$ , et à la plus grande donnée inférieure à  $q3 + 1.5 * (q3 - q1)$  ; les minimum et maximum sont représentées par des marques en forme de cercle (o). Si le minimum ou le maximum n'est pas confondu avec le tiret d'extrémité de moustache, c'est qu'il s'agit d'une valeur atypique. Les valeurs P sont calculées avec un test de Student d'échantillons appariés.



#### 4. Immunoprécipitation d'ADN méthylé (MeDIP)

L'ADN génomique est préparé en utilisant le kit DNeasy Tissue (Qiagen) selon les instructions du fabricant. La qualité et la quantité de l'ADN sont ensuite contrôlées par électrophorèse sur gel d'agarose et spectrophotométrie. L'ADN est soniqué puis dénaturé 10 minutes à 94°C et immunoprécipité avec un anticorps anti-5-Méthyl Cytosine (Megabase Research Products, USA) sur la nuit à 4°C. L'immunoprécipitation contrôle est réalisée en utilisant un anticorps anti-IgG (Sigma, France). La fraction liée est isolée par une chromatographie protéine A et l'ADN méthylé est ensuite récupéré (Weber, Davies et al. 2005). La PCR en temps réel est réalisée en utilisant le kit QuantiTect SYBR Green PCR kit (Quiagen, France). Le taux d'enrichissement relatif en 5-Méthyl Cytosine pour chacune des séquences cibles est ensuite calculé en comparant le nombre de cycles d'amplification entre les échantillons immunoprécipités avec l'anticorps anti-5-Méthyl Cytosine et l'ADN non immunoprécipité.

#### 5. Immunofluorescence

Les expériences d'immunofluorescence anti-HSF1 ont été réalisées dans des cellules HeLa traitées ou non à la 5 Aza-Cytidine et dans les cellules U2OS selon la technique présentée dans l'article n°1.

#### 6. Western blot

Les expériences de western blot ont été réalisées avec un anticorps de lapin anti-HSF1 (1/5000, Stressgen) et de souris anti-Dicer (1/100, Abcam) selon le protocole décrit dans l'article n°1.

#### 7. RT PCR

1µg d'ARN total est digéré par la DNase RQ1 (RNase free), pendant 30 minutes à 37°C. L'action de la DNase est inactivée en présence d'EDTA par incubation à 65°C pendant 10 minutes. La transcription inverse est réalisée grâce au kit Super Script III first strand synthesis system for PCR (Invitrogen). Brièvement, les ARN sont incubés à 65°C pendant 5 minutes en présence d'oligodT et de dNTP. Puis, la réaction de transcription est permise grâce à l'ajout de la reverse transcriptase (SSIII), en présence du tampon de l'enzyme, de MgCl<sub>2</sub>, de DTT et de RNase OUT. L'incubation se fait pendant 50 minutes à 50°C. La réaction est arrêtée par un traitement de 5 minutes à 85°C. L'ARN est ensuite éliminé par traitement à la RNase H, pendant 20

minutes à 37°C. Les cDNA obtenus sont ensuite amplifiés par PCR comme suit : les cDNA sont mélangés à des dNTP, du MgCL<sub>2</sub>, le tampon de l'enzyme, de l'ADN polymérase (Taq) et des amorces. Les conditions de la PCR sont les suivantes : 2 minutes à 94°C, puis 30 cycles des trois étapes suivantes 30 secondes à 94°C, 30 secondes à 55°C, 30 secondes à 72°C, et 10 minutes à 72°C d'élongation finale. Les amorces permettant l'amplification de Dicer sont les suivantes : 5'CATGGATAGTGGGATGTAC3' et 5'CTACTTCCAGAGTGACTCTG3' (Chendrimada, Gregory et al. 2005). Les produits de PCR sont ensuite déposés sur un gel d'agarose.

## II.3 Résultats

### 1. Conception des oligonucléotides

Les séquences centromériques et péricentromériques appartiennent à des régions chromosomiques dont la structure n'a pas été entièrement caractérisée. Ces régions ont échappé aux grands programmes de séquençage du génome humain, et seuls quelques clones génomiques ont été séquencés et cartographiés par des approches *in situ*. Jusqu'à très récemment, les EST (Expressed Sequence Tag) correspondant aux séquences centromériques et péricentromériques étaient considérés comme le résultat d'une contamination par de l'ADN génomique et étaient donc généralement retirés des banques de données. Cependant, certaines de ces séquences sont malgré tout présentes dans ces bases de données ce qui pourrait révéler une transcription des séquences centromériques et péricentromériques, au moins dans certaines cellules et certains tissus.

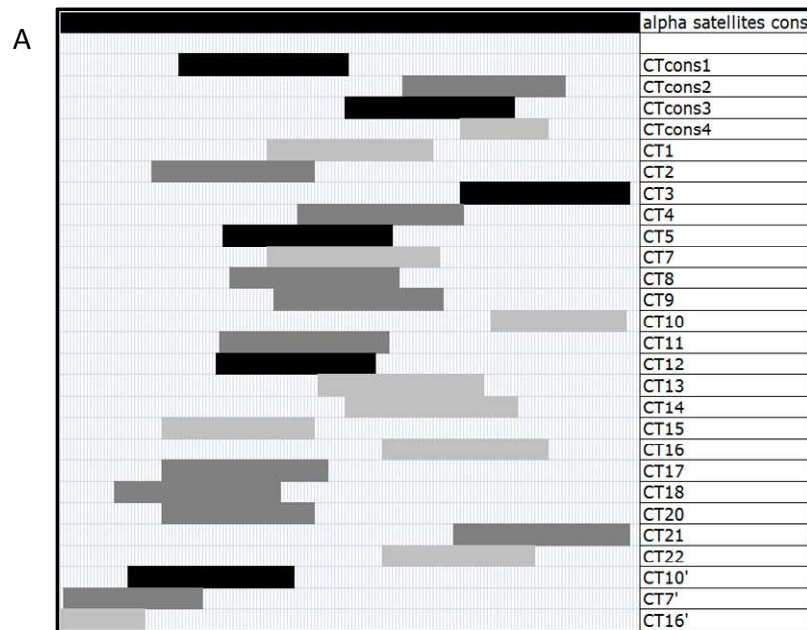
Dans ce contexte général, la conception d'oligonucléotides représentant les séquences centromériques (alpha satellites) et péricentromériques (satellites 2 et 3) présentes dans le génome humain a été réalisée à l'aide de séquences consensus connues (KCS) ou nouvellement conçues (ND) ainsi que de séquences spécifiques de certains chromosomes obtenues par alignements de plusieurs séquences présentes dans les bases de données. D'autre part, certains de ces oligonucléotides ont été obtenus par des analyses de blasts dans des banques de données EST avec des séquences qui avaient déjà été caractérisées *in situ*, par FISH ADN (F/EST). Enfin, certaines séquences spécifiques du CT correspondant à des sites préférentiels d'intégration génomique du VIH latent ont également été ajoutées, afin de déterminer si l'intégration virale au sein de ces séquences et/ou la réactivation du virus,



Bien que certains oligonucléotides présents sur la RepChip aient été conçus à partir de séquences dont la spécificité chromosomique a été validée par FISH ADN, le recours à des séquences plus petites ainsi que les conditions d'hybridation utilisées ne garantissent pas le maintien de cette spécificité. Cependant, l'objectif général était de fournir un outil sensible, permettant une analyse globale de la capacité transcriptionnelle des CT et des PCT et ce sous différentes conditions. À cet égard, bien que l'utilisation de multiples séquences appartenant à une même catégorie de séquences répétées augmente la probabilité d'obtenir des données redondantes, elle permet de réduire le risque de passer à côté de certains ARN issus des CT et des PCT.

Les séquences centromériques (alpha satellites) et péricentromériques (satellites 2 et 3) sont présentées dans la Table 9, avec celles des gènes contrôles (actine B, histone H4, ARNr 18S et 28S) et des gènes hsp (hsp70-1, hsp70-2 et hsp40). La position et le score de chaque alignement de séquence centromérique avec la séquence alpha satellite consensus de 171 pb (Vissel and Choo 1987) est présenté dans le Table 10-A. Enfin, bien que les séquences péricentromériques satellites 2 et 3 contiennent toutes les deux le motif GGAAT, celui-ci est moins conservé pour les satellites 2. Le taux d'homologie des différents oligonucléotides spécifiques des PCT avec à la fois les clones génomiques pHuR98 et pHuR195, identifiés dans la littérature comme contenant exclusivement des satellites 3 et des satellites 2 respectivement est indiqué dans la Table 10-B (Grady, Ratliff et al. 1992).

Compte tenu de notre connaissance limitée de l'activité transcriptionnelle des CT et des PCT *in vivo*, les oligonucléotides représentant les CT et les PCT ont été conçus dans les deux orientations et les termes « sens » et « antisens » sont bien sûr arbitraires. Pour les ARN centromériques, le terme «sens» fait référence à la transcription dans la même orientation que la séquence consensus alpha satellites de 171pb (Vissel and Choo 1987). Pour les ARN péricentromériques, le terme «sens» fait référence aux ARN enrichis en motifs GGAAT.



**B**

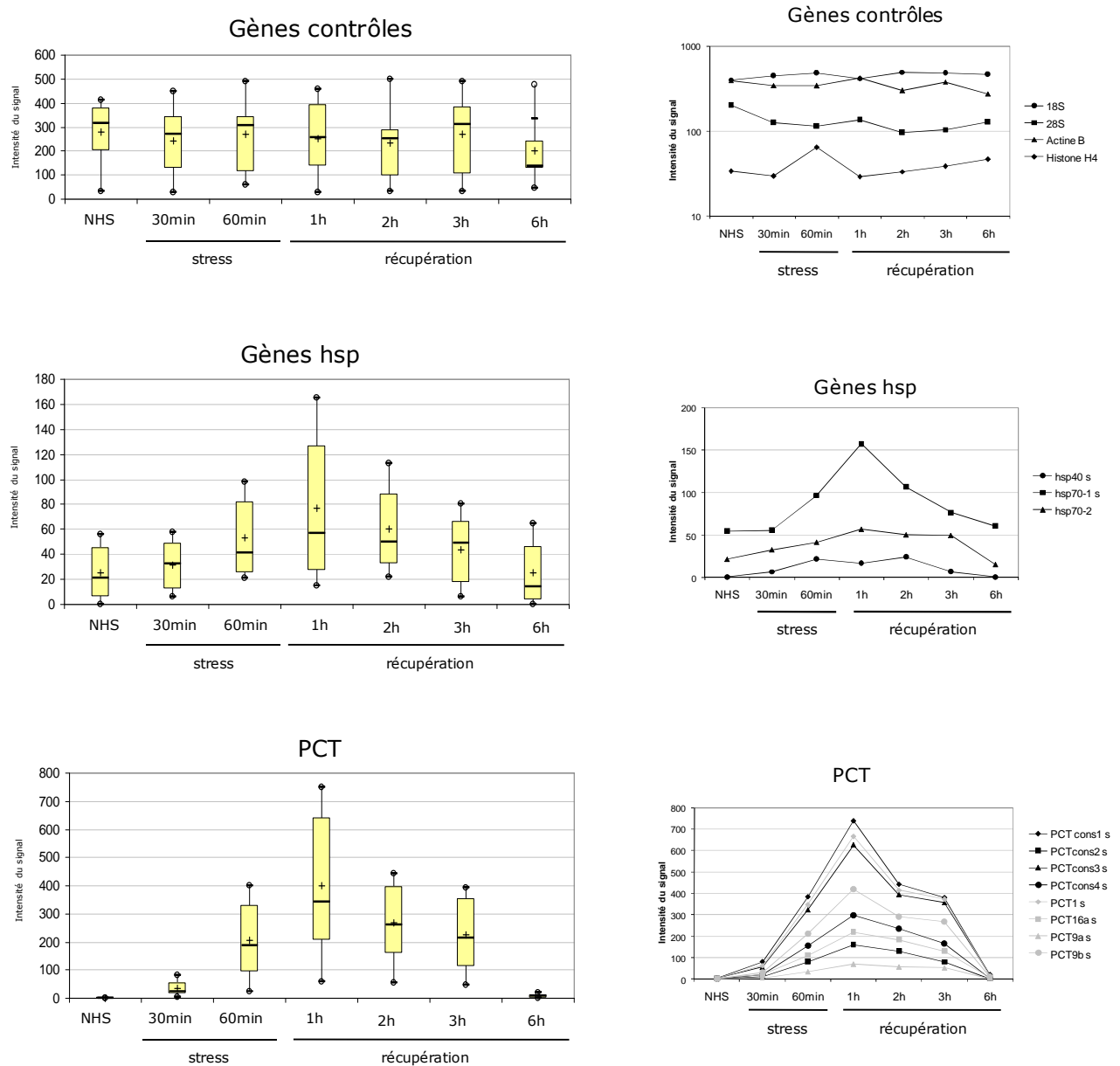
	pHuR98	GGAAT	pHuR195	
PCTcons1		3		Sat2/Sat3
PCTcons3		10		
PCTcons4		2		
PCT1		3		Sat2
PCT16a		2		
PCT16b		1, 2		Sat3
PCT9a		2		
PCT9b		6		
PCT14		2		
PCTncsa		3		
PCTncsb		4		

**Table 10 : A** Localisation et homologie des différents oligonucléotides spécifiques des centromères, par rapport à la sonde consensus de référence des alpha satellites (Vissel and Choo 1987). Les différentes homologies ont été calculées grâce au programme Lalign. Gris clair= faible homologie ( $>10^{-3}$ ); gris foncé = homologie moyenne ( $10^{-7} < x < 10^{-4}$ ); noir= forte homologie ( $< 10^{-10}$ ) **B** Homologies de séquences entre les différents oligonucléotides spécifiques des péri-centromères et la sonde pHuR98 (enrichie en satellites 3) et la sonde pHuR95 (enrichie en satellites 2). Le nombre de motifs GGAAT est aussi indiqué en blanc. Gris clair= faible homologie ( $>10^{-3}$ ); gris foncé = homologie moyenne ( $10^{-7} < x < 10^{-4}$ ); noir= forte homologie ( $< 10^{-8}$ ).

## 2. La réponse au stress thermique: un modèle de validation pour la RepChip

Le choc thermique étant connu pour induire l'expression des séquences péri-centromériques, il représente donc un bon modèle pour valider notre RepChip (Jolly, Metz et al. 2004; Rizzi, Denegri et al. 2004; Valgardsdottir, Chiodi et al. 2008). Notre objectif est d'effectuer une analyse quantitative de l'expression des PCT, dans les orientations sens et antisens, et de tester la capacité transcriptionnelle éventuelle des séquences centromériques. En effet, alors que l'activation transcriptionnelle des

séquences péri-centromériques spécifiques du chromosome 9 au cours de la réponse au choc thermique a été clairement documentée, la capacité transcriptionnelle des séquences centromériques chez l'Homme est totalement inconnue.



**Figure 60 :** Cinétique d'expression des ARN-PCT au cours de la réponse au choc thermique. Les cellules HeLa ont été soumises à 30 ou 60 minutes de choc thermique ; ou à une heure de choc thermique suivie d'une période de récupération à 37°C allant de 1 à 6h (REC1-6). Des cellules non soumises au stress sont utilisées comme contrôle (NHS). Pour chaque classe de séquences (gènes contrôles, gènes hsp et séquences péri-centromériques), les résultats obtenus avec l'ensemble des oligonucléotides représentant cette classe sont présentés sous forme de box-plots (panneau de gauche). Un pic d'expression des gènes hsp et des séquences péri-centromériques est observé après une heure de choc thermique suivi d'une heure de récupération à 37°C. Les résultats présentés ici concernent les ARN-PCT sous une forme sens. Aucune variation d'expression des gènes contrôles en relation avec la réponse au stress n'est observée. A chaque box-plot correspond également une représentation graphique linéaire montrant l'évolution individuelle de l'expression de chaque oligonucléotide (panneau de droite).

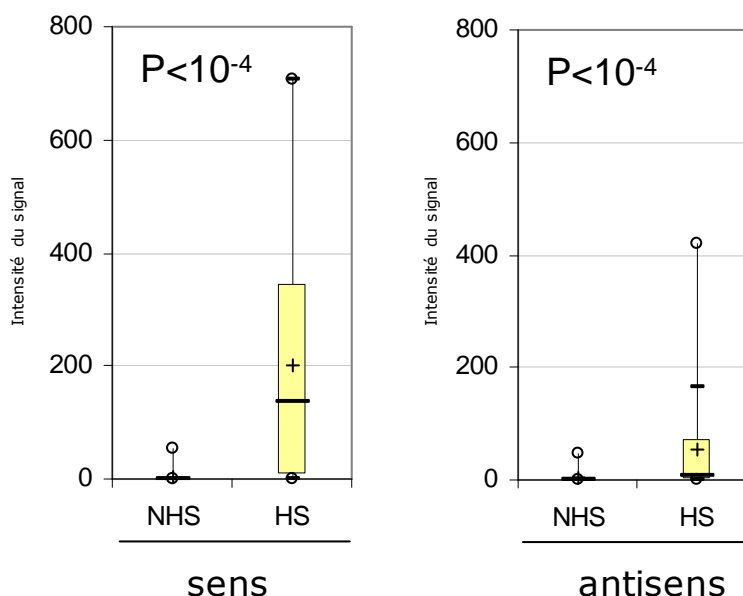
L'expression des séquences du CT et du PCT a été évaluée dans des cellules HeLa soumises à une cinétique de choc thermique. Les cellules ont été stressées de 30 minutes à une heure à 42°C, ou alors soumises à une heure de choc thermique suivie par une période de récupération allant de une à six heures à 37°C. Les oligonucléotides ont été regroupés en quatre catégories : gènes contrôles, gènes de choc thermique (hsp), séquences péricentromériques (PCT), séquences centromériques (CT). Les intensités de signal obtenues pour les oligonucléotides d'une même catégorie sont présentées sous la forme de box-plots.

Comme le montre la Figure 60, l'expression des gènes contrôles varie légèrement mais d'une manière indépendante de la réponse au stress. En revanche, une accumulation d'ARN issus des séquences péricentromériques (ARN-PCT), dans l'orientation sens, a été détectée avec l'ensemble des oligonucléotides spécifiques de ces séquences. Le profil d'expression des séquences péricentromériques est similaire à celui des gènes hsp, attestant clairement de la corrélation entre la réponse au stress et la transcription des séquences péricentromériques. Ainsi, alors qu'une très faible quantité d'ARN-PCT est détecté en absence de stress, ceux-ci sont nettement observés après 30 minutes de choc thermique. Le taux d'ARN-PCT atteint un pic après une heure de récupération à 37°C puis diminue jusqu'à 6h de récupération où ne l'on détecte quasiment plus d'ARN-PCT. A 6h de récupération, le taux d'ARN hsp est comparable à celui obtenu pour les cellules non stressées (comparer NHS et REC6) suggérant que les cellules ont parfaitement récupéré du stress. Aucune corrélation n'a été mise en évidence entre l'intensité du signal obtenu avec les différents oligonucléotides représentant les PCT et le pourcentage d'homologie avec une sonde (PCTcons3) qui consiste en une succession du motif consensus GGAAT présents dans les séquences satellites 2 et 3.

L'impact du choc thermique sur l'expression des séquences péricentromériques a été étendu à l'analyse des séquences centromériques. Chez la souris, l'expression des séquences du PCT et du CT a été rapportée dans différents modèles de différenciation et de prolifération cellulaire, alors qu'aucune évidence d'expression des CT n'a été signalée jusqu'ici dans les cellules humaines. Cependant, nous ne détectons aucun ARN issus des séquences centromériques (ARN-CT) ni dans les cellules soumises au stress ni de manière constitutive. Ce résultat permet d'envisager l'existence de mécanismes indépendants de contrôle de l'expression des CT et des PCT au cours du choc thermique.

Afin de comparer les niveaux d'expression des PCT, en sens et en antisens, nous avons analysé les taux d'ARN-PCT dans des cellules soumises ou non à un choc

thermique. Les résultats obtenus sont présentés dans la Figure 61. Une légère transcription constitutive de ces séquences, dans les deux orientations, peut être détectée. Au cours de la réponse au stress, l'induction de la transcription des PCT sous une orientation sens et antisens est de 44.9 et 11.6 fois respectivement. Ainsi, au cours du choc thermique les séquences du PCT sont majoritairement induites sous une forme sens.



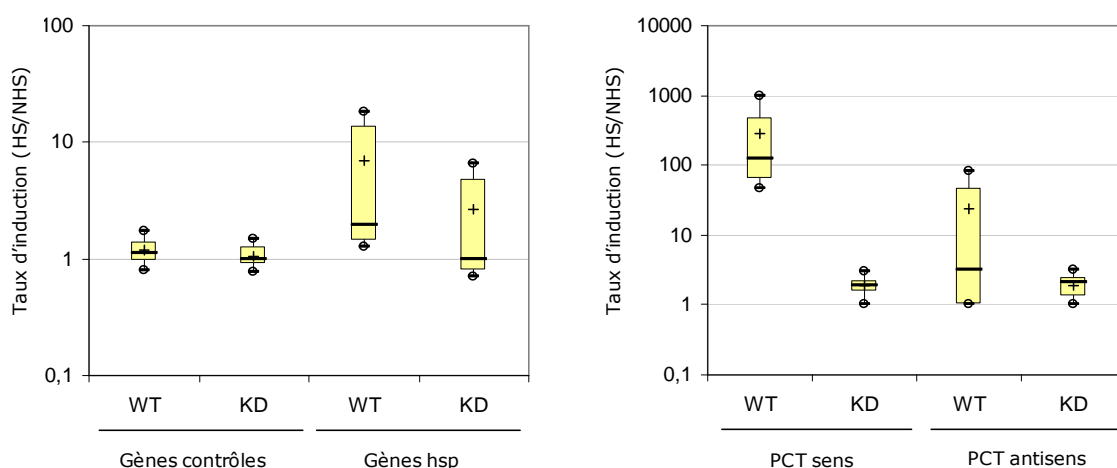
**Figure 61** : Orientation de la transcription des séquences péri-centromériques. Des cellules HeLa ont été soumises (HS) ou non (NHS) à une heure de choc thermique. Les intensités de signal obtenues pour l'ensemble des oligonucléotides sens ou antisens de cinq expériences indépendantes, sont représentées sous forme de box-plot. Les valeurs p ont été calculées grâce à un test de Student d'échantillons appariés. Les taux d'induction des séquences satellites en sens et en antisens au cours du choc thermique sont respectivement de 44.9 et 11.6 fois.

L'expression constitutive des séquences péri-centromériques n'a jamais été détectée par des expériences de Northern blot ou de FISH ARN, laissant penser que ces séquences pourraient être exprimées à un taux très bas et/ou sous la forme de petits ARN difficiles à visualiser avec de telles approches. Nos données confirment cependant les résultats obtenus par RT-PCR et QRT-PCR avec les oligonucléotides spécifiques des séquences satellites 3 du chromosome 9, dans des cellules HeLa (Jolly, Metz et al. 2004; Rizzi, Denegri et al. 2004; Valgardsdottir, Chiodi et al. 2008). De plus, les taux d'induction moyens que nous avons obtenus pour l'ensemble des séquences péri-centromériques au cours du stress sont en accord ceux obtenus par QRT-PCR pour l'analyse des séquences satellites 3 du chromosome 9. En effet, après une heure de choc thermique, les taux d'induction de ces séquences sous une



orientation sens et antisens étaient estimés respectivement à 100 et 10 fois (Jolly, Metz et al. 2004; Rizzi, Denegri et al. 2004; Valgardsdottir, Chiodi et al. 2008).

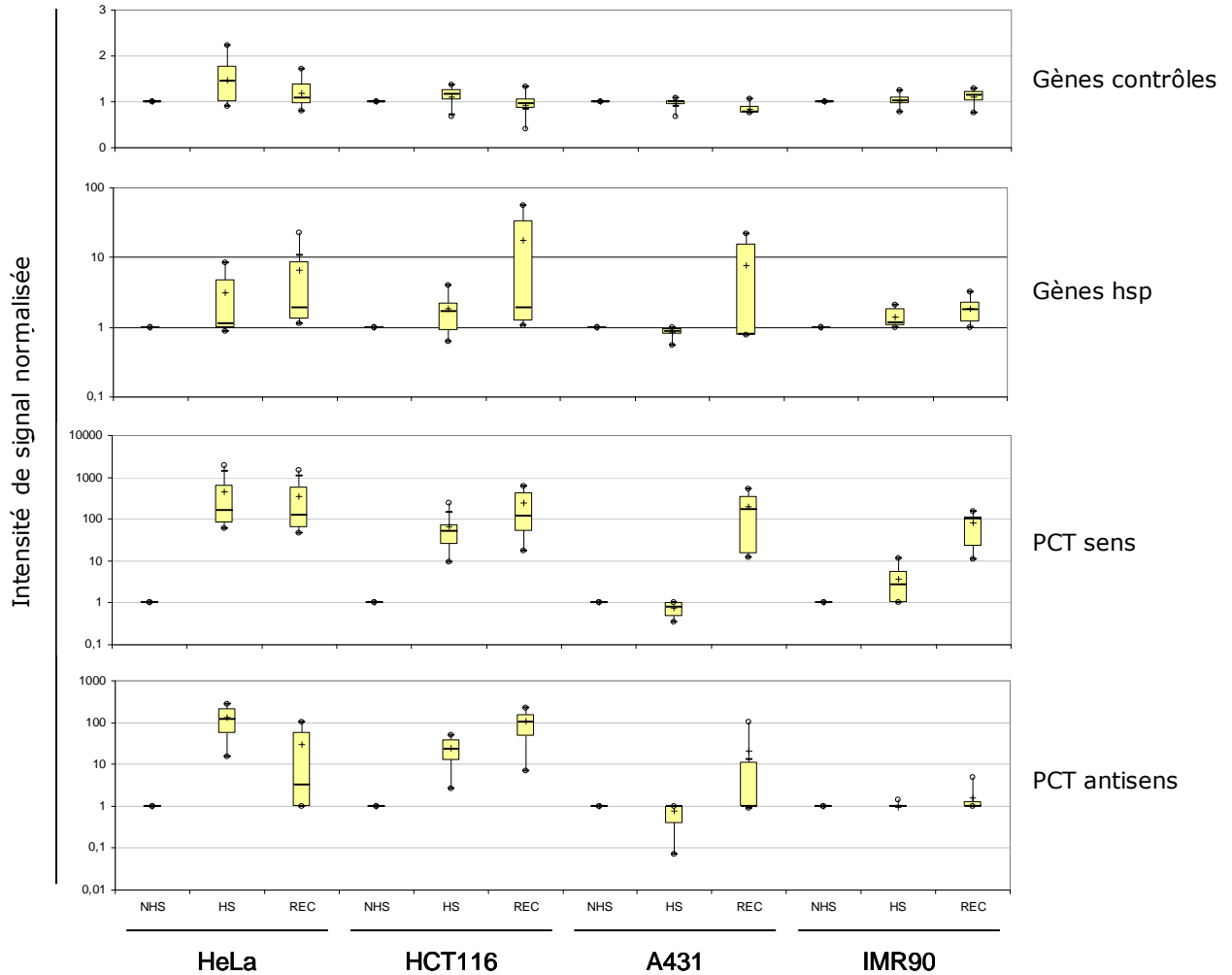
La possibilité que ces ARN-PCT antisens, détectés en faible abondance, puissent représenter des artefacts de la transcription inverse a été examinée grâce aux transcrits satellites 3 du chromosome 9 (Perocchi, Xu et al. 2007). Nous avons réalisé des expériences de transcription *in vitro*, en sens et en antisens, de la sonde pHuR98. Les ARN ainsi produits ont été transcrits de manière inverse puis hybridés avec les oligonucléotides sens et antisens correspondant aux séquences satellites 3 du chromosome 9, dans les conditions classiques utilisées pour la RepChIP. Lorsque la transcription a été réalisée sous forme sens, les produits issus de la transcription inverse ne se sont hybridés qu'avec les oligonucléotides sens. Et inversement, lorsque la transcription a été réalisée sous forme antisens, les produits issus de la transcription inverse ne se sont hybridés qu'avec les oligonucléotides antisens. Ainsi, la présence d'ARN-PCT sous forme antisens constituerait une véritable entité biologique plutôt qu'un défaut de la transcription inverse.



**Figure 62 :** Impact d'HSF1 sur la transcription des séquences péri-centromériques. L'expression des séquences péri-centromériques a été évaluée dans des cellules HeLa exprimant (WT) ou non (KD) le facteur de transcription HSF1. Les résultats obtenus sont présentés comme le ratio des valeurs obtenues après le choc thermique sur celles obtenues sans choc thermique (HS/NHS). La transcription des séquences péri-centromériques induite sous l'effet du choc thermique est sévèrement inhibée en absence d'HSF1.

HSF1 est connu pour être un facteur déterminant dans l'induction de la transcription des séquences péri-centromériques du chromosome 9 au cours de la réponse au stress (Jolly et al. 2004). Ainsi, une analyse comparative de l'expression des PCT dans des cellules HeLa soumises ou non à un choc thermique et exprimant (WT) ou non (KD) HSF1 a été réalisée (Figure 62). Comme cela était prévisible,

l'accumulation des ARN-PCT au cours de la réponse au stress est sévèrement compromise dans les cellules déficientes en HSF1, avec une diminution de l'induction de près de 99% et 92% en sens et en antisens respectivement.



**Figure 63 :** Expression des séquences péri-centromériques dans des cellules normales et tumorales au cours d'une cinétique de choc thermique. La quantité d'ARN-PCT a été analysée dans des cellules tumorales HeLa, HCT116 et A431 ainsi que dans des cellules normales IMR90. Ces cellules ont subi un choc thermique d'une heure (HS) ou d'une heure suivie de 3 heures de récupération à 37°C (REC). Des cellules non stressées sont utilisées comme contrôle (NHS). Pour chaque type cellulaire, les résultats sont normalisés par rapport à ceux obtenus dans les cellules contrôles non stressées. Le stress induit l'accumulation d'ARN-PCT, en sens et en antisens, dans toutes les lignées cellulaires étudiées avec cependant des différences dans la cinétique d'expression.

Le seul modèle cellulaire pour lequel la transcription des PCT a été formellement démontrée au cours de la réponse au stress est celui des cellules HeLa. La capacité d'autres types de cellules à exprimer les séquences péri-centromériques et éventuellement les séquences centromériques a donc été testée (Figure 63). L'expression des séquences centromériques et péri-centromériques a été évaluée dans

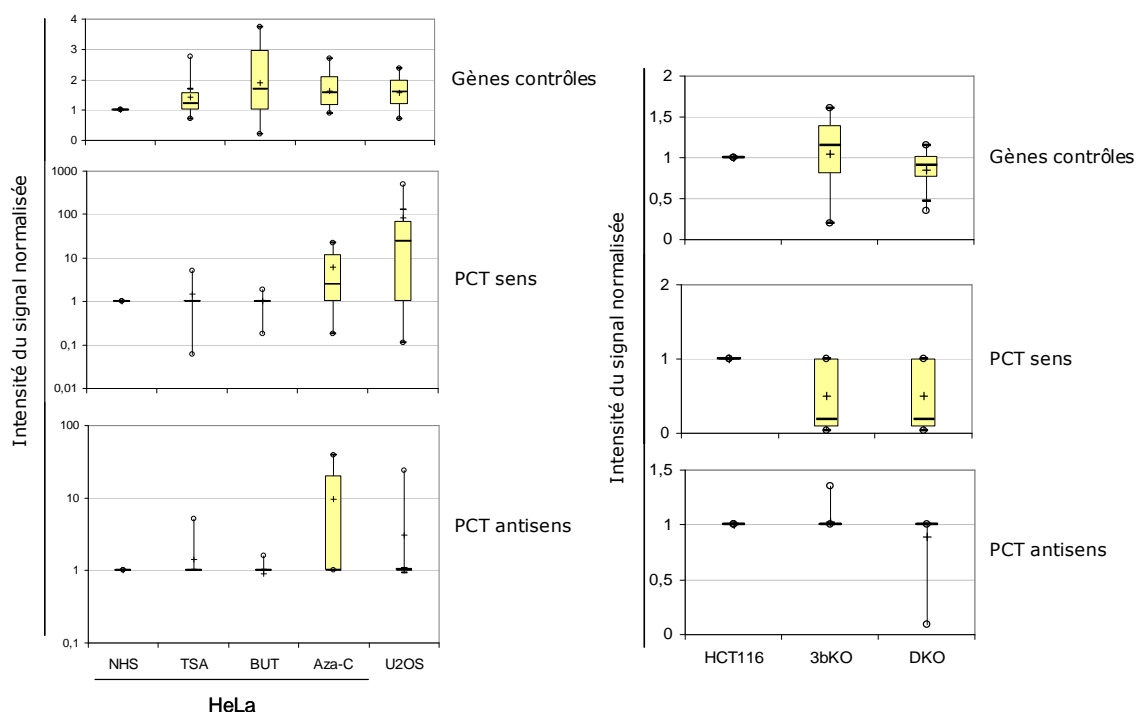
des lignées cellulaires tumorales telles que les cellules de carcinome du col de l'utérus (cellules HeLa, utilisées comme contrôle), de cancer du côlon (HCT116) et de carcinome de la vulve (A431) ainsi que dans des lignées primaires de fibroblastes (IMR90). Toutes ces cellules ont été soumises à une cinétique de choc thermique. La transcription des séquences péri-centromériques est une caractéristique commune de toutes les cellules soumises à un choc thermique. Ainsi, la capacité des cellules à exprimer les séquences péri-centromériques en condition de stress ne se limite pas aux cellules cancéreuses. Cependant, nos résultats indiquent que la cinétique d'accumulation des ARN-PCT varie en fonction du type cellulaire. En effet, une forte accumulation des ARN-PCT est observée après trois heures de récupération dans les cellules A431 et IMR90 alors qu'elle est détectée seulement après une heure de choc thermique dans les cellules HeLa et HCT116. La différence dans la cinétique d'accumulation des ARN-PCT dans les cellules A431 et IMR90 d'une part, et HeLa et HCT116 d'autre part, pourrait être liée à des différences de sensibilité de ces cellules au choc thermique. A l'appui de cette hypothèse, la cinétique de l'accumulation des ARN-PCT est similaire à celle des ARN hsp.

La présence d'ARN-CT n'a pas été observée ni dans les IMR90 ni dans les HeLa ni dans les HCT116. Cependant, une expression sporadique de 3 séquences centromériques, parmi les 29 oligonucléotides présents sur la RepChip, a été mise en évidence dans les A431. Ces ARN-CT (CTcons3, CT10 et CT16) sont présents dans une orientation sens dont le taux ne varie pas de manière significative au cours de la réponse au stress.

Finalement, ces premiers résultats permettent de valider la RepChip ainsi que d'apporter des idées nouvelles sur la relation entre le stress et la transcription des CT et des PCT. Ainsi, par cette approche, nous confirmons que l'expression des séquences péri-centromériques est corrélée à la réponse au stress, qu'elle est dépendante du facteur de transcription HSF1 et qu'elle a lieu principalement sous une orientation sens. De plus, nous montrons que le choc thermique n'induit pas une dérépression de toutes les séquences satellites puisque l'induction concomitante, au cours du stress, des séquences centromériques et péri-centromériques n'est pas détectée. Enfin, nous montrons que la transcription des séquences péri-centromériques au cours du choc thermique est un processus qui a lieu à la fois dans les cellules normales et tumorales.

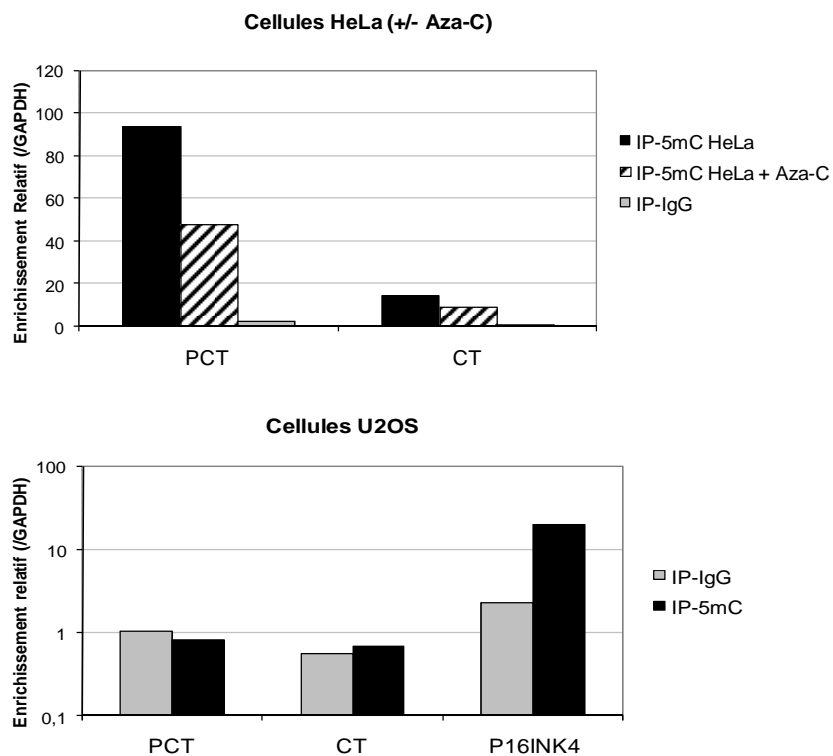
### 3. Impact du remodelage de la chromatine sur la transcription des séquences satellites

Le choc thermique est associé à des événements de remodelage de la chromatine. Un des meilleurs exemples est l'apparition de foyers d'histones hyperacétylées qui se forment au niveau des régions péri-centromériques transcrites. Sachant que l'hyper-acétylation des histones et la déméthylation de l'ADN sont caractéristiques d'une chromatine permissive pour la transcription, nous avons cherché à déterminer quel pouvait être l'impact de ces modifications sur la transcription des séquences répétées.



**Figure 64 :** La déméthylation de l'ADN peut conduire à l'expression des séquences péri-centromériques. L'analyse de l'expression des séquences péri-centromériques a été réalisée dans des cellules HeLa traitées à la Trichostatine A (TSA), au butyrate (BUT), à la 5 Aza-Cytidine (Aza-C) ainsi que dans des cellules U20S (panneau de gauche) et des cellules déficientes en DNMT3b seule (3bKO) ou en combinaison avec la DNMT1 (DKO) (panneau de droite). Les résultats ont été normalisés par rapport à des cellules HeLa non traitées (panneau de gauche) ou par rapport aux cellules HCT116 (cellules dans lesquelles les délétions des dnmt ont été réalisées) (panneau de droite). La transcription des séquences péri-centromériques est uniquement détectée sous l'effet de la déméthylation de l'ADN, dans les cellules traitées à la 5 Aza-Cytidine et dans les U20S mais pas dans les HCT116 déficientes en DNMT. Le traitement des cellules avec des agents inhibiteurs des HDAC n'a aucun effet sur l'expression des séquences péri-centromériques

L'impact de la trichostatine A (TSA), un puissant inhibiteur des HDAC de classes I et II, ainsi que celui du butyrate (BUT), un inhibiteur des HDAC de classe I, sur l'expression des CT et des PCT ont été analysés dans des cellules HeLa. Comme le montre la Figure 64, aucune augmentation significative de l'expression constitutive des séquences péri-centromériques n'a été observée dans les cellules traitées avec de la TSA ou du butyrate, par rapport aux cellules HeLa contrôles, non traitées. Ces données suggèrent donc que la TSA et le butyrate ne sont pas suffisants pour induire l'expression des séquences péri-centromériques. D'autre part, ces drogues ne sont pas non plus capables d'induire la transcription des séquences centromériques puisqu'aucun ARN-CT n'a été détecté dans les cellules traitées à la TSA ou au butyrate.



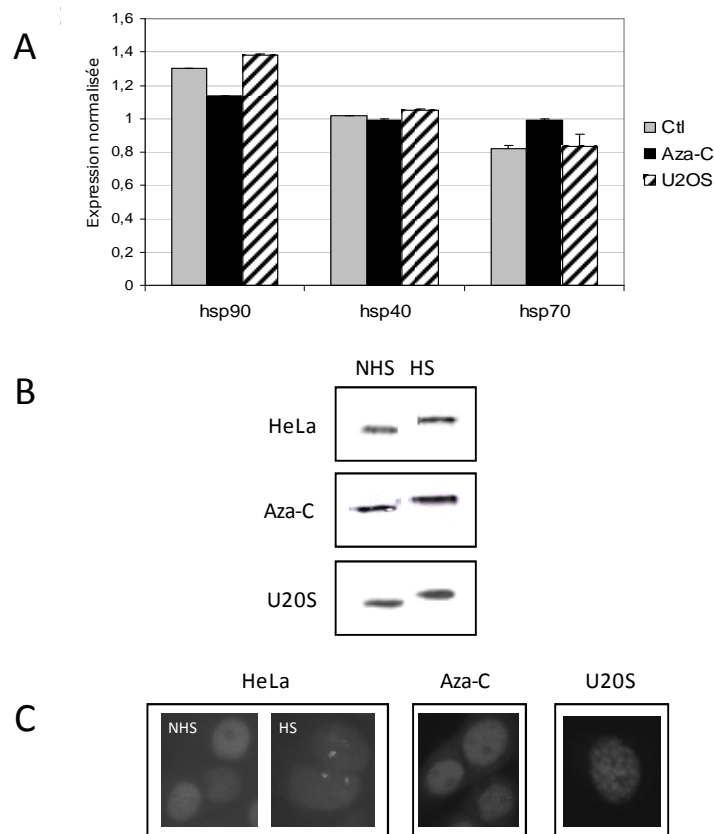
**Figure 65 :** Les séquences péri-centromériques et centromériques sont hypométhylées dans les cellules U2OS et dans les cellules traitées à la 5 Aza-Cytidine. Des expériences de MeDIP ont été réalisées dans des cellules HeLa traitées ou non à la 5 Aza-Cytidine (panneau du haut) ainsi que dans les cellules U2OS (panneau du bas). Les histogrammes indiquent le taux relatif d'enrichissement en 5-méthyl cytosine des séquences répétées par rapport au gène contrôle (GAPDH). Pour les cellules U2OS, le gène P16INK4 est utilisé comme témoin, ce gène étant normalement méthylé dans ces cellules. IP-IgG: immunoprécipitation réalisée avec des IgG non spécifiques, IP5mC: immunoprécipitation réalisée avec un anticorps spécifique de la 5-méthyl cytosine.

L'effet de l'hypométhylation de l'ADN sur l'expression des séquences centromériques et péri-centromériques a été analysé d'une part dans des cellules HeLa traitées avec la 5 Aza-Cytidine (5 Aza-C), un puissant agent déméthylant et d'autre part dans des cellules U2OS, décrites comme étant hypométhylées sur les PCT (communication personnelle de Karine Monier) ainsi que dans des cellules déficientes en DNMT3b seule (3bKO) ou en combinaison avec la DNMT1 (DKO). Le statut hypométhylé des PCT dans les cellules U2OS ainsi que dans les cellules HeLa traitées à la 5 Aza-Cytidine a été confirmé par une expérience de MeDIP (immunoprécipitation d'ADN méthylé suivie d'une PCR avec des sondes spécifiques des séquences à amplifier ; ici les séquences satellites centromériques et péri-centromériques) (Figure 65). La déméthylation des PCT dans les cellules 3bKO et DKO a été validée par l'approche Rep'Array (article n°3 : Horard, Eymery et al, en préparation). Comme le montre la Figure 64, l'accumulation d'ARN-PCT est observée dans les cellules traitées à la 5 Aza-C, dans les cellules U2OS, mais pas dans les cellules 3bKO ni DKO. Dans les cellules traitées à la 5 Aza-C, la présence d'ARN-PCT en sens et en antisens suggère qu'ils pourraient résulter d'une transcription dite « read through », non spécifique. Au contraire, dans les cellules U2OS, l'expression des séquences péri-centromériques uniquement sous une forme sens laisse penser que cette transcription pourrait être finement contrôlée. Cependant, l'absence d'ARN-PCT dans les cellules 3bKO et DKO indique clairement que la déméthylation de l'ADN n'est pas suffisante pour induire l'expression des séquences répétées. Ainsi, la dérégulation d'autres mécanismes épigénétiques est peut être impliquée dans la transcription des séquences péri-centromériques observée dans les cellules traitées à la 5 Aza-C et dans les cellules U2OS.

L'absence d'ARN-CT dont les séquences sont pourtant hypométhylées dans les cellules U2OS et les cellules traitées à la 5 Aza-C suggère une nouvelle fois que l'expression des séquences centromériques et péri-centromériques est régulée par des mécanismes différents (Figure 65).

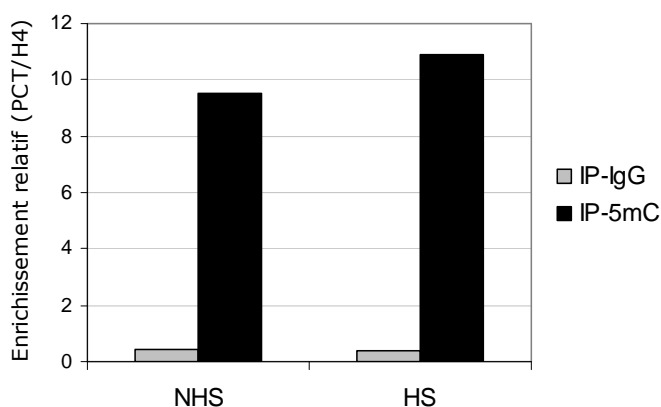
L'activation transcriptionnelle des séquences péri-centromériques dans les cellules traitées à la 5 Aza-C et dans les cellules U2OS pourrait être liée à une activation de la réponse au stress plutôt qu'à la déméthylation de ces séquences. Cette hypothèse a donc été étudiée. L'expression des gènes hsp, induits au cours de la réponse au stress, a été évaluée par QRT-PCR. Comme le montre la Figure 66-A l'expression des gènes hsp n'est pas induite ni dans les cellules traitées à la 5 Aza-C ni dans les cellules U2OS. D'autre part, nous avons étudié l'état d'activation d'HSF1 dans ces cellules, par western blot et immunofluorescence. En effet, lorsqu'il est activé, le facteur HSF1 subit des modifications post-traductionnelles, résultant en un retard de migration sur

gel, et forme des granules de stress aisément détectables par immunofluorescence. Comme on le voit sur la Figure 66-B, la mobilité réduite d'HSF1 témoignant de son activité n'est présente que dans les cellules soumises à un choc thermique. En absence de stress, la forme activée d'HSF1 n'est pas détectée ni dans les cellules U2OS ni dans les cellules traitées à la 5 Aza-C. Enfin, les granules de stress ne sont pas détectés dans ces cellules (Figure 66-C). Ainsi, l'expression des séquences péricentromériques dans les cellules traitées à la 5 Aza-Cytidine et dans les cellules U2OS n'implique pas les mécanismes de la réponse au stress.



**Figure 66 :** L'expression des ARN-PCT dans les cellules U2OS et les cellules traitées à la 5 Aza-C n'est pas liée à la réponse au stress. **A** L'expression des gènes hsp40, hsp70 et hsp90 a été évaluée par QRT-PCR dans des cellules HeLa traitées à la 5 Aza-C et dans les cellules U2OS. Les résultats ont été normalisés par rapport à trois gènes contrôles (GAPDH, U6 et H4) et comparés à ceux obtenus pour des cellules HeLa non traitées, utilisées comme contrôles. Aucune activation des gènes hsp n'est détectée ni dans les cellules U2OS ni dans les cellules traitées à la 5 Aza-C. L'état d'activation d'HSF1 dans des cellules HeLa traitées à la 5 Aza-C et dans les cellules U2OS a été évalué par western blot (**B**) et par immunofluorescence (**C**). Des cellules HeLa soumises (HS) ou non (NHS) à un choc thermique sont utilisées comme contrôles. Lorsqu'il est activé, HSF1 subit des modifications post-traductionnelles qui retarde sa migration sur gel. Ce retard est observé pour toutes les cellules lorsqu'elles sont soumises à un choc thermique, mais pas dans les cellules non stressées. D'autre part, HSF1 activé forme des granules de stress nucléaires dont on ne détecte pas la présence ni dans les cellules traitées à la 5 Aza-C ni dans les cellules U2OS.

Etant donné que l'hypométhylation des séquences péricentromériques permet l'accumulation d'ARN-PCT, nous avons cherché à savoir si la transcription de ces mêmes séquences, au cours du choc thermique, pouvait impliquer la déméthylation de ces régions. Cette hypothèse a été testée par des approches MeDIP avec de l'ADN issus de cellules HeLa soumises (HS) ou non (NHS) à un choc thermique. Comme le montre la Figure 67, aucune variation drastique de la méthylation de l'ADN des séquences péricentromériques n'a été observée entre les cellules non stressées et celles soumises au choc thermique. Ceci suggère donc que la déméthylation des séquences péricentromériques n'est pas une condition préalable à leur expression. Ainsi, la transcription des séquences péricentromériques en réponse à une hypométhylation de l'ADN d'une part et à un stress thermique d'autre part met en jeu des mécanismes indépendants.



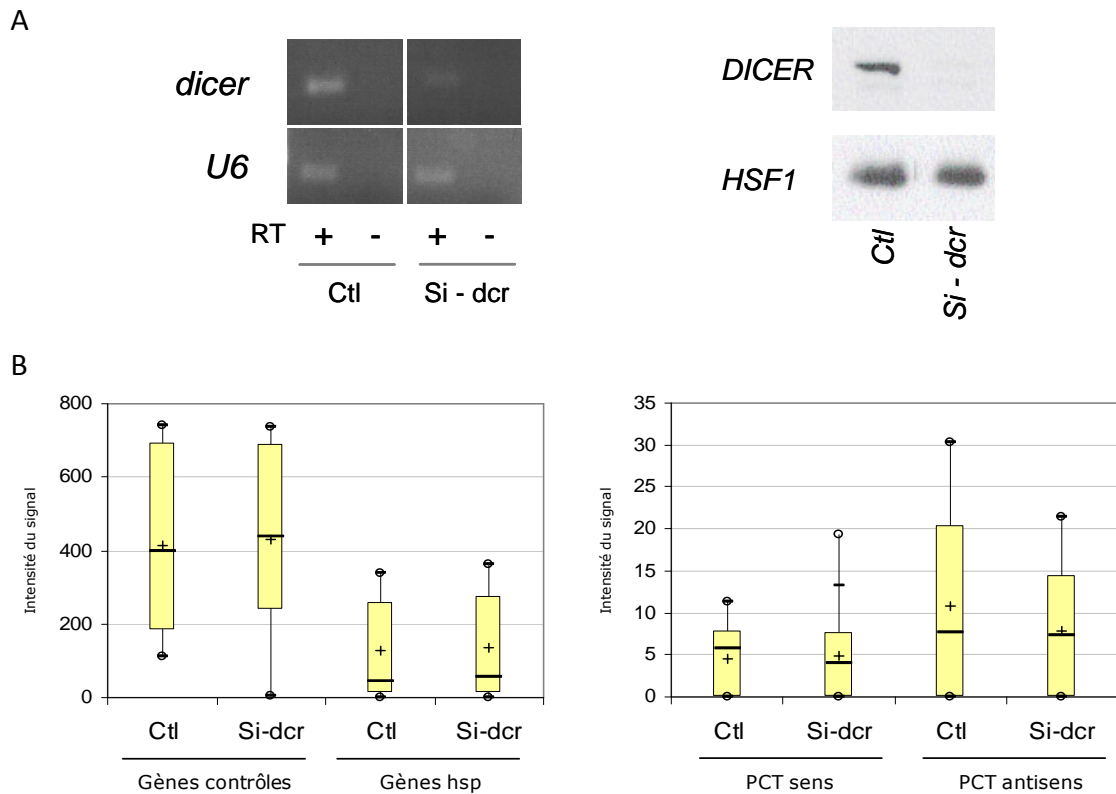
**Figure 67 :** L'expression des séquences péricentromériques au cours de la réponse au stress n'est pas corrélée à leur hypométhylation. Une analyse par MeDIP des séquences péricentromériques dans des cellules HeLa soumises (HS) ou non (NHS) à un choc thermique a été réalisée. Les résultats ont été normalisés par rapport à un gène contrôle (GAPDH).

#### 4. Impact de Dicer sur l'expression des séquences satellites.

Chez *S.pombe*, des petits ARN double brins issus des régions péricentromériques et générés par Dicer, permettent, grâce au complexe RITS, de jouer un rôle dans la structure des régions hétérochromatiques (Verdel, Jia et al. 2004; Buhler and Moazed 2007). Ainsi, par analogie à ce qui se produit chez cette levure modèle, les longs transcrits détectés chez l'Homme, pourraient représenter les précurseurs de ces petits ARN double brin. Cependant, il n'existe aucune preuve à ce jour de l'implication de Dicer dans la structure de l'hétérochromatine dans les cellules humaines. La seule évidence de la présence de petits ARN impliqués dans la structure de l'hétérochromatine chez l'Homme provient d'études réalisées dans des cellules



hybrides Homme-poulet contenant le chromosome 21 humain. Dans ces cellules, la perte d'expression de Dicer conduit à l'accumulation d'ARN-PCT et d'ARN-CT dont la taille varie entre 20-30 nucléotides de longs et plusieurs kb (Fukagawa, Nogami et al. 2004). De même, dans des cellules ES de souris déficientes en Dicer, des ARN-CT et des ARN-PCT sont détectés sous la forme de petits ARN double brin dont la taille varie entre 25-30 et 50 nucléotides (Kanellopoulou, Muljo et al. 2005).



**Figure 68 :** Impact de l'inhibition de DICER (*dcr*) sur l'expression des séquences répétées. **A** L'efficacité des siRNA dirigés contre *dicer* a été validée par RT-PCR (image de gauche) et par western blot (image de droite) sur des cellules traitées avec le siRNA (*si-dcr*) et des cellules contrôles (*ctl*). L'expression de U6 est utilisée comme contrôle dans l'analyse par RT-PCR et l'anticorps anti-HSF1 sert de témoin de charge dans le western blot **B** L'inhibition de Dicer n'est pas associée à la transcription des séquences répétées. L'expression des séquences péricentromériques a été analysée dans des cellules traitées avec un siRNA dirigé contre Dicer (*si-dcr*) et dans des cellules contrôles (*ctl*). Aucune variation significative de l'expression des séquences répétées n'est observée entre les cellules *ctl* et *si-dcr*.

Ainsi, pour étudier l'hypothèse selon laquelle l'accumulation de longs ARN, issus des séquences centromériques et péricentromériques, pourrait être favorisée par l'absence de Dicer, des cellules HeLa ont été traitées avec des siRNA ciblant de manière spécifique Dicer. L'efficacité des siRNA a été déterminée au niveau transcriptionnel et post-transcriptionnel (Figure 68-A). L'absence de détection d'ARN-CT et d'ARN-PCT en

absence de Dicer (Figure 68-B) suggère que l'accumulation des ARN-PCT sous forme longue ne résulte pas d'une stabilité de ces ARN liée à l'altération de Dicer.

### 5. Expression des séquences satellites *in vivo*, dans des tissus normaux et tumoraux.

Afin d'obtenir un aperçu de la capacité transcriptionnelle des CT et des PCT *in vivo*, l'expression de ces séquences a été analysée dans des fractions d'ARN provenant de différents tissus normaux. Ces échantillons proviennent de tissus adultes de différentes origines (système immunitaire, système nerveux central, muscles, système endocrinien, système digestif, système cardio-vasculaire, système reproducteur, système respiratoire) ainsi que de tissus fœtaux (foie, cerveau) et d'un tissu d'origine extraembryonnaire (placenta) (Table 11).

	Système immunitaire		Système nerveux central et muscles effecteurs			Système endocrinien		Système digestif		
	thymus	moelle osseuse	cerveau	moelle épinière	muscles squelettiques	tyroïde	glandes surrénales	glandes salivaires	foie	colon
PCT	non	non	non	non	non	non	non	non	non	non
CT	non	non	non	non	non	non	non	non	non	non

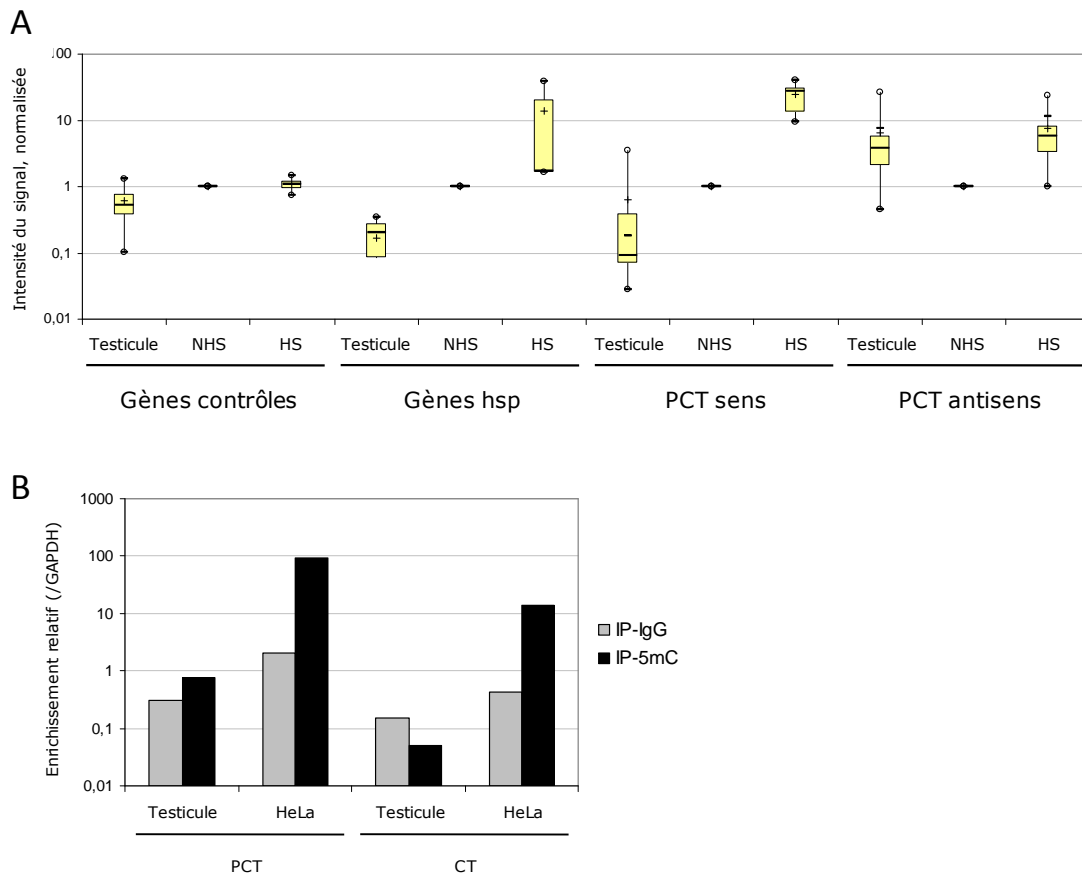
	Système cardiovasculaire	Système reproducteur			Système respiratoire		Embryon et placenta			
	cœur	testicule	prostate	utérus	ovaire	trachée	poumon	foie	cerveau	placenta
PCT	non	OUI	non	non	non	non	non	non	non	non
CT	non	non	non	non	OUI*	non	non	OUI*	non	OUI*

**Table 11 :** Expression des séquences centromériques et péricentromériques dans des tissus normaux. L'expression des séquences centromériques et péricentromériques a été analysée par RepChip dans 21 tissus normaux. Des ARN-PCT sont uniquement détectés dans le testicule (sous forme antisens). Des ARN-CT sont détectés dans le foie fœtal, le placenta et les ovaires : \* CT7' sous forme sens et CT10' sous forme antisens.

Aucune expression des séquences centromériques n'a été détectée dans ces tissus normaux à l'exception du placenta, du foie fœtal et des ovaires. Dans ces trois tissus, la transcription correspond à deux séquences différentes de CT décrites comme des sites d'intégration du VIH : CT7' dans une orientation sens et CT10' dans une orientation antisens. Ainsi, contrairement aux séquences péricentromériques qui sont toujours toutes exprimées ensemble, l'expression des séquences centromériques est sporadique. La raison de cette expression reste encore à déterminer. Il est toutefois possible que l'expression des séquences centromériques dans le placenta puisse être liée à leur statut hypométhylé (Zagradisnik and Kokalj-Vokac 2000).

Concernant les séquences péricentromériques, aucun ARN-PCT n'a été détecté dans les tissus mentionnés ci-dessus, à l'exception du testicule. En effet, de manière particulièrement intéressante, la transcription des séquences péricentromériques dans le testicule a lieu mais uniquement sous une orientation antisens (Figure 69-A). Cette

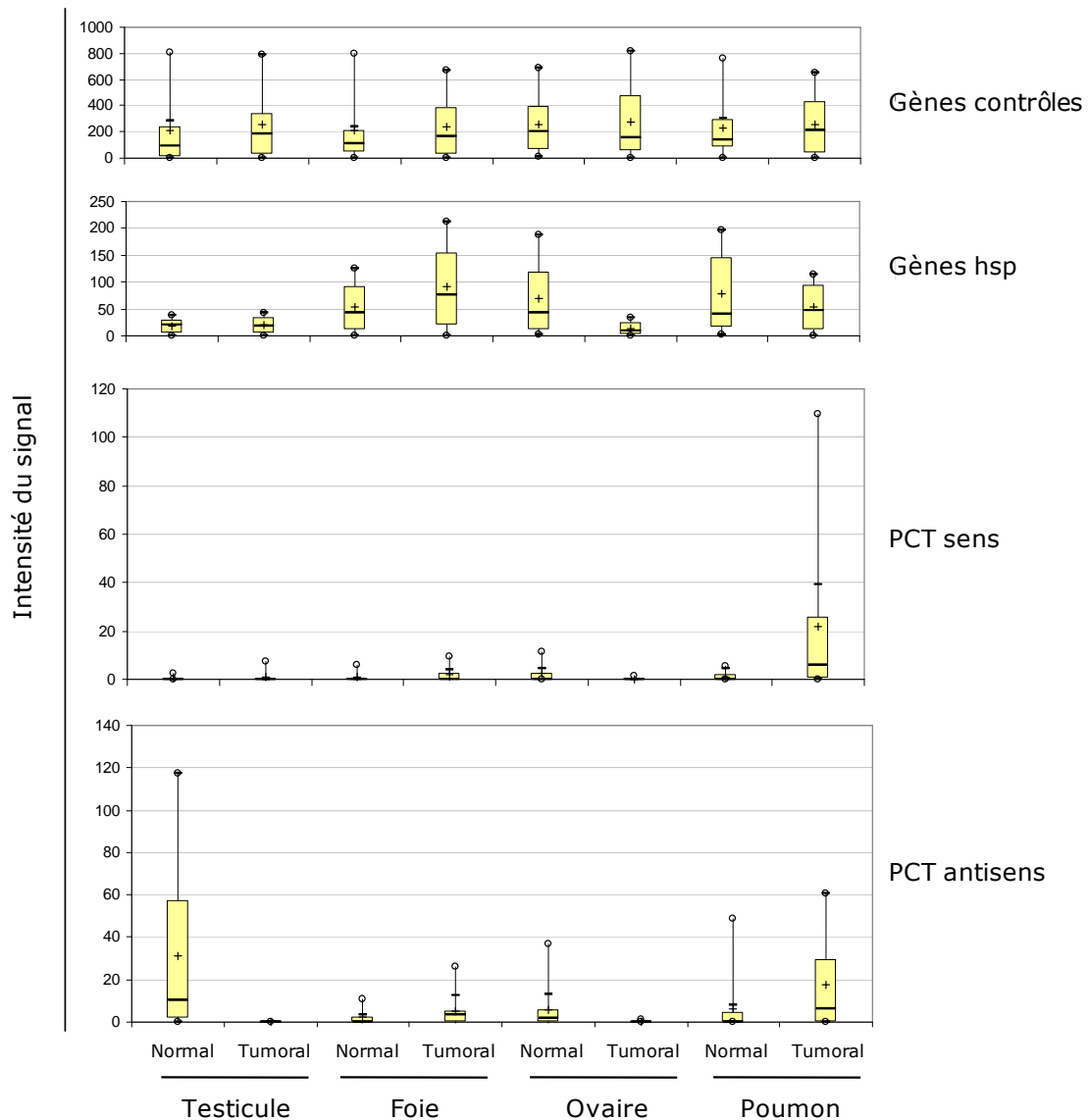
observation a été confirmée dans deux autres échantillons indépendants d'ARN testiculaire. Malgré l'orientation de la transcription opposée à celle qui a lieu au cours de la réponse au stress, nous avons tout de même vérifié l'implication possible de la réponse au stress dans cette transcription. L'absence d'induction des gènes hsp dans les testicules, suggère que l'accumulation des ARN-PCT antisens n'implique pas les mécanismes de la réponse au stress (Figure 69-A).



**Figure 69 :** Expression des séquences péri-centromériques dans le testicule. **A** L'expression des séquences péri-centromériques dans le testicule ainsi que dans des cellules HeLa soumises (HS) ou non (NHS) à un choc thermique a été réalisée. Tous les résultats ont été normalisés par rapport aux cellules HeLa non stressées. Contrairement aux cellules HeLa soumises à un choc thermique, dans lesquelles les ARN-PCT sont principalement détectés sous une forme sens, les ARN-PCT détectés dans le testicule sont sous forme antisens. L'accumulation des ARN-PCT dans le testicule n'est pas corrélée avec une induction des gènes hsp. **B** Les séquences péri-centromériques et centromériques sont hypométhylées dans le testicule. Des expériences de MeDIP ont été réalisées pour le testicule et les cellules HeLa soumises ou non à un choc thermique. L'histogramme indique le taux relatif d'enrichissement en 5-méthyl cytosine des séquences répétées par rapport au gène contrôle (GAPDH)

L'expression des séquences péri-centromériques dans les testicules, pourrait être facilitée par leur statut hypométhylé révélé par MeDIP (Figure 69-B). Cependant, le fait que la transcription se fasse uniquement dans l'orientation antisens suggère que

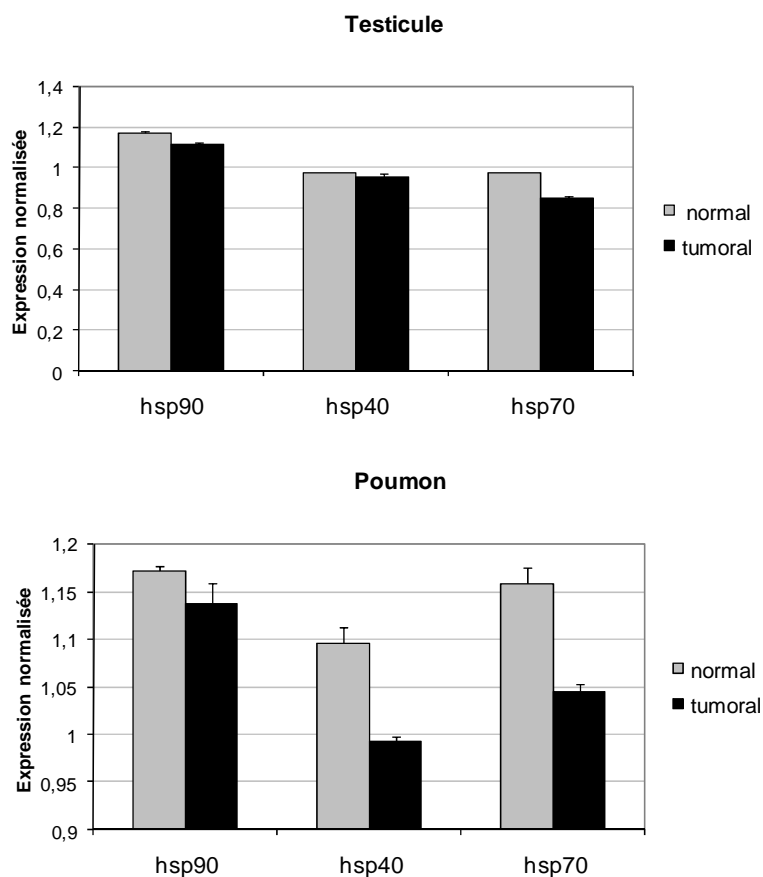
celle-ci n'est pas la simple conséquence d'une hypométhylation des ces séquences, mais qu'elle serait plutôt contrôlée par un mécanisme spécifique.



**Figure 70 :** Variation d'expression des séquences péri-centromériques dans des tissus sains et tumoraux de testicule, foie, ovaire et poumon. Chaque couple de tissu (normal-tumoral) provient d'un même patient. L'accumulation des ARN-PCT sous la forme antisens est détectée dans le tissu testiculaire normal mais pas dans le tissu tumoral. D'autre part, des ARN-PCT, dans les deux orientations, sont abondants dans le tissu pulmonaire tumoral mais pas dans le tissu normal.

Compte tenu de l'impact de l'hypométhylation de l'ADN sur l'expression des séquences péri-centromériques et du fait que ces modifications épigénétiques sont fréquemment rencontrées dans les cellules cancéreuses (Brock, Herman et al. 2007), l'expression des séquences centromériques et péri-centromériques a été testée dans des fractions d'ARN provenant de tissus tumoraux et de tissus sains. Afin de permettre

l'analyse comparative des tissus sains et tumoraux, un même fond génétique était nécessaire. Nous avons donc choisi d'utiliser des tissus sains et tumoraux provenant de mêmes patients. Ces couples de tissus sont les suivants : testicule, foie, ovaire et poumon. Les échantillons de tissus sains ont été prélevés à la périphérie de la tumeur.



**Figure 71** :L'accumulation des ARN-PCT dans les tissus testiculaires et pulmonaires n'est pas corrélée avec une induction des gènes hsp. L'expression des gènes hsp40, hsp70 et hsp90 a été évaluée par QRT-PCR dans les différents tissus. Les résultats ont été normalisés par rapport à trois gènes contrôles (GAPDH, U6 et H4).

Concernant les tissus ovariens et hépatiques, normaux et tumoraux, aucune expression significative des séquences péricentromériques n'a été détectée (Figure 70). Des différences d'expression de séquences satellites entre tissu normal et tumoral a été observée pour les échantillons de testicules et de poumons. En effet, une perte d'expression des séquences péricentromériques sous la forme antisens est observée dans le tissu testiculaire tumoral, comparativement au tissu normal. D'autre part, une accumulation d'ARN-PCT en sens et en antisens, est détectée dans le tissu tumoral pulmonaire mais pas dans le tissu normal. Ce résultat indique que, contrairement aux tissus testiculaires, le phénotype tumoral est dans ce cas associé à une induction de

l'expression des séquences péricentromériques. Une fois de plus, l'absence d'induction des gènes hsp dans les tissus pulmonaires suggère que l'accumulation des ARN-PCT, dans le tissu tumoral, n'implique pas les mécanismes de la réponse au stress (Figure 71).

De manière intéressante, la forte expression des séquences péricentromériques dans le tissu pulmonaire tumoral, est accompagnée par une accumulation d'ARN-CT dans les deux orientations (Table 12). Ainsi, nous montrons pour la première fois une induction concomitante de l'expression des séquences centromériques et péricentromériques. La transcription des séquences centromériques a également été mise en évidence dans le tissu ovarien normal ainsi que dans les tissus hépatiques, normal et tumoral. La présence d'ARN-CT dans le foie normal pourrait constituer une caractéristique spécifique des cellules immatures du foie. En effet, les tissus normaux et tumoraux hépatiques proviennent d'un bébé de trois mois. Or, l'expression des séquences centromériques a également été observée dans le foie fœtal normal (Table 11).

	Testicule		Foie		Ovaire		Poumon	
	normal	tumoral	normal	tumoral	normal	tumoral	normal	tumoral
CTcons1 (s)								
CTcons2 (s)								
CTcons3 (s)								
CTcons3 (as)								
CT7' (s)								
CT10' (as)								
CT8 (s)								
CT16 (s)								
CT21 (s)								

**Table 12 :** Comparaison d'expression des séquences centromériques dans des tissus sains et tumoraux. Blanc=non exprimé, Noir=exprimé. s= sens ; as= antisens.

Étant donné le faible nombre de cas analysés, il est impossible de tirer des conclusions générales à partir des analyses comparatives des tissus normaux et tumoraux. Il est toutefois possible que l'expression (ou la répression) des séquences centromériques et péricentromériques puisse révéler des modifications de l'épigénome dans les cellules cancéreuses. D'autre part, la présence d'ARN-PCT uniquement détectée dans le testicule mais pas dans les autres tissus normaux testés, suggère un rôle possible des modifications épigénétiques dans l'expression de ces séquences. En effet, les cellules de ce tissu subissent un remodelage épigénétique drastique au cours de la spermatogenèse.

## II.4 Conclusion :

Très peu de données sont disponibles aujourd'hui concernant les conditions d'expression des séquences centromériques et péricentromériques chez les mammifères. Dans les cellules humaines, le haut niveau de polymorphisme au sein des séquences péricentromériques et centromériques, combiné à une mauvaise connaissance de la structure générale de ces séquences, rendent les analyses d'expression particulièrement difficiles (Valgardsdottir, Chiodi et al. 2008). Notre objectif était de développer une macro-array dédiée à l'analyse de l'expression des séquences satellites centromériques et péricentromériques du génome humain. Pour cela nous avons utilisé des oligonucléotides susceptibles d'être représentatifs de l'ensemble de ces séquences. Nous montrons que cette approche permet la détection d'ARN sans avoir recours à leur amplification préalable par des techniques de PCR, particulièrement difficiles à réaliser lorsqu'elles sont appliquées aux séquences répétées humaines. Nos résultats concernant l'expression des séquences centromériques et péricentromériques au cours de la réponse au stress, et dans des tissus sains et tumoraux, nous ont permis non seulement de valider la RepChip comme un outil dédié à la transcription des séquences satellites humaines mais également de mettre en évidence des mécanismes indépendants de régulation de cette transcription.

Le fait que toutes les séquences péricentromériques présentes sur la RepChip arborent toutes soit un signal d'hybridation positif (témoignant de la transcription de ces séquences) soit un signal d'hybridation négatif (témoignant d'une absence de transcription de ces séquences) suggère qu'elles se comportent comme une population homogène en ce qui concerne leur mode d'expression. Bien que certains oligonucléotides présents sur la RepChip aient été conçus à partir de clones génomiques spécifiques de séquences péricentromériques d'un chromosome donné, cela n'exclue pas une possible hybridation croisée entre ces séquences. L'expression sporadique des séquences centromériques suggère au contraire un certain degré de spécificité. Cependant, l'existence de sites secondaires de fixation d'HSF1, autres que le locus 9q12, dans les régions péricentromériques de chromosomes enrichis en séquences répétées de type satellites 2 ou 3 est en faveur de l'hypothèse d'une transcription homogène des séquences péricentromériques (Denegri, Moralli et al. 2002). De plus ces nouvelles cibles chromosomiques d'HSF1 correspondent à des sites de transcription. (Eymery et al, article n°1 en cours de préparation). Enfin, nous montrons que la transcription des séquences centromériques n'a pas lieu au cours de

la réponse au stress. Ce résultat permet d'envisager l'existence de mécanismes distincts contrôlant l'expression des séquences centromériques et péricentromériques.

Les différentes situations dans lesquelles l'accumulation des ARN-CT et des ARN-PCT a été observée chez les mammifères, sont souvent associées à des modifications épigénétiques de ces régions. C'est en effet le cas pour le développement et la différenciation cellulaire, le cancer, le vieillissement et la réplication, autant de mécanismes connus pour entraîner des changements dans la méthylation de l'ADN et/ou des modifications d'histones (Rasmussen 2003; Brock, Herman et al. 2007; Fraga and Esteller 2007; Probst and Almouzni 2008). A travers ce travail, nous montrons que l'hypométhylation de l'ADN peut favoriser l'expression des séquences péricentromériques alors que l'utilisation d'inhibiteurs de HDAC n'a aucun effet.

Ainsi, il existe une corrélation entre l'hypométhylation de l'ADN et la transcription des séquences péricentromériques dans les cellules traitées à la 5 Aza-Cytidine et dans les cellules U2OS. Ce résultat est en accord avec des données publiées récemment suggérant que, dans les cellules humaines, l'expression des séquences péricentromériques est associée leur état déméthylé, en particulier lors de cancer (Erukashvily, Donev et al. 2007). Cependant, l'impact de la déméthylation de l'ADN sur la transcription des séquences centromériques et péricentromériques n'est pas complètement défini. En effet, dans les cellules déficientes en DNMT qui présentent une hypométhylation des régions centromériques et péricentromériques, aucun ARN issus de ces régions n'a pu être mis en évidence. Ce résultat concorde avec ceux obtenus dans des cellules ES murines déficientes en DNMT1 ou DNMT3B (Lehnertz, Ueda et al. 2003; Martens, O'Sullivan et al. 2005).

Ainsi, la déméthylation de l'ADN ne suffit pas en elle-même à induire la transcription des séquences péricentromériques. Il est possible que l'expression de ces séquences implique des interventions croisées de différents mécanismes épigénétiques. A ce titre il est intéressant de souligner que l'expression des séquences péricentromériques n'est pas régulée de la même manière dans les U2OS et dans les cellules traitées à la 5 Aza-Cytidine. En effet, des ARN-PCT sens sont détectés dans les cellules U2OS alors qu'ils sont présents en sens et en antisens dans les cellules traitées avec la 5 Aza-Cytidine. Cette drogue pourrait avoir des conséquences drastiques sur la structure de l'hétérochromatisme, en modifiant le schéma épigénétique global de ces régions. Il a en effet été montré que la 5 Aza Cytidine pouvait avoir un impact sur la méthylation ou l'acétylation des histones (Takebayashi, Nakao et al. 2001; Nguyen, Weisenberger et al. 2002; Zhao, Soejima et al. 2005). La raison de l'hypométhylation des PCT dans les cellules U2OS n'est pas expliquée.



Cependant une différence de statut épigénétique des PCT entre les cellules U2OS et les cellules traitées à la 5 Aza-Cytidine est parfaitement possible. Ceci expliquerait la différence d'expression des séquences péricentromériques observés pour ces cellules. De plus, dans l'hypothèse selon laquelle les ARN-PCT seraient impliqués dans la structure de l'hétérochromatine, la transcription de ces régions dans les cellules U2OS pourrait représenter une réponse physiologique à l'hypométhylation de ces séquences. Concernant l'absence de transcription des séquences péricentromériques dans les cellules déficientes en DNMT, on ne peut pas exclure l'implication éventuelle de mécanismes permettant de compenser la perte de ces enzymes. De manière intéressante, les séquences péricentromériques sont déméthylées dans les testicules. Le testicule est connu pour être le site d'événements post-méiotiques impliquant des mécanismes drastiques de remodelage de la chromatine. En particulier, l'éviction des histones est précédée par leur hyperacétylation dans les régions correspondant aux péricentromères. Dans ce cas, le remodelage de la chromatine en combinaison avec un ADN hypométhylé pourrait expliquer la présence d'ARN-PCT dans ce tissu. Enfin, il est possible que la transcription des séquences centromériques et péricentromériques observée dans le tissu pulmonaire tumoral implique également des modifications épigénétiques telles que la déméthylation de l'ADN.

Nous montrons que l'induction de l'expression des séquences péricentromériques sous l'effet du stress n'est pas associée à la déméthylation de ces régions. Par conséquent, l'hypométhylation de l'ADN n'est pas un pré-requis pour la transcription des PCT au cours du stress. Cette observation n'est pas surprenante puisqu'à ce jour aucun mécanisme permettant la déméthylation de l'ADN d'une manière indépendante de la réplication de l'ADN n'a été décrit en dehors des cellules embryonnaires. Ce résultat illustre encore une fois la complexité relationnelle qui existe entre la méthylation de l'ADN et l'expression des séquences satellites. L'existence de deux mécanismes indépendants permettant l'expression des séquences péricentromériques, l'un impliquant HSF1 et l'autre une hypométhylation de l'ADN, n'est probablement qu'une part infime de la complexité des mécanismes impliqués dans le contrôle de l'expression de ces séquences. A titre d'exemple, une publication récente montre qu'un autre facteur de transcription tonEBP est impliqué dans le contrôle de l'expression des séquences péricentromériques (Valgardsdottir, Chiodi et al. 2008).

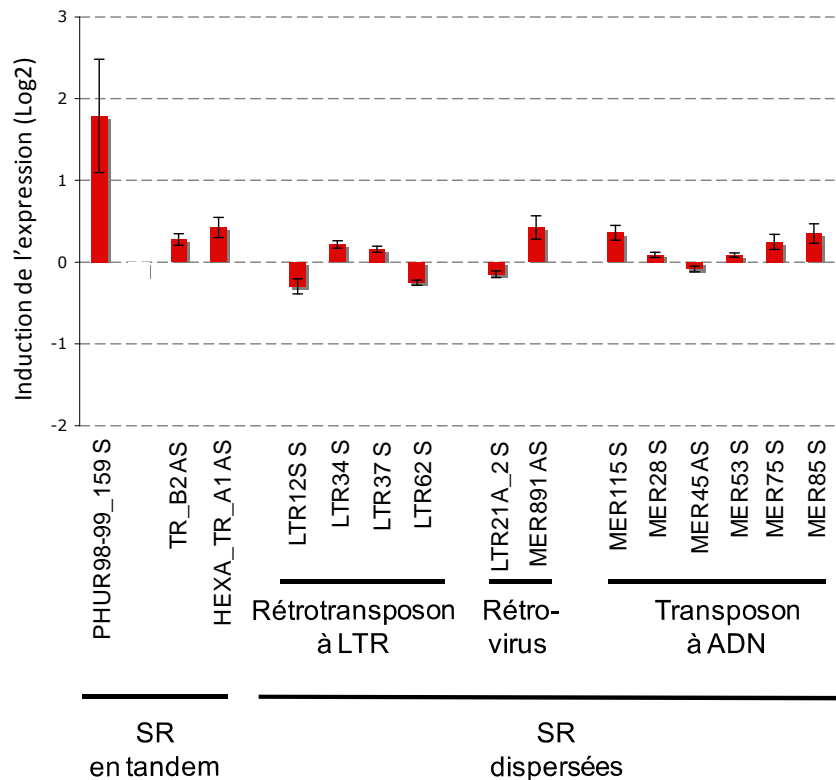
En conclusion, notre étude représente la première description de l'expression globale des séquences centromériques et péricentromériques, à la fois dans des tissus sains et pathologiques. Ce travail a permis de mettre en évidence la complexité des mécanismes impliqués dans la régulation de l'expression de ces séquences. Cependant aucune conclusion ne peut être donnée concernant la relation entre la dérégulation

des séquences centromériques et péricentromériques et le développement de cancers. L'analyse de l'expression de ces séquences dans des échantillons de plus grande ampleur est bien évidemment nécessaire. Ce travail suggère néanmoins que les ARN-CT et les ARN-PCT pourraient être de bons senseurs des modifications de l'épigénome qui se produisent en réponse aux conditions environnementales ou dans différents contextes de différenciation ou de transformation cellulaire.

## **II.5 Résultats complémentaires**

### *1. Rep'Array*

Parallèlement à la RepChip, l'équipe du Dr Eric Gilson a développé un outil, appelé Rep'array. Cette micro-array possède plus de 200 oligonucléotides spécifiques de différents types de séquences répétées en tandem et de séquences répétées dispersées. Ces deux outils présentent des différences majeures, non seulement dans leur composition en oligonucléotides, mais également dans la technique de fabrication et d'analyse. En effet, la Rep'Array a été développée en combinant des techniques modernes de mise en place de micro-array, c'est-à-dire l'utilisation de lame de verre entant que support ainsi qu'un marquage fluorescent. Cet outil a été validé pour l'analyse globale du statut épigénétique des séquences répétées du génome humain (Article n°3 : Horard B, Eymery A et al, en préparation). D'autre part, nous avons validé l'utilisation de cette micro-array pour l'analyse transcriptionnelle des séquences répétées, en utilisant comme modèle la réponse cellulaire au stress thermique. Comme cela était attendu, nous avons pu mettre en évidence une induction de la transcription des séquences répétées péricentromériques au cours de la réponse au stress (PHUR98-99\_159). De plus, en accord avec des données récemment publiées utilisant des expériences de FISH ARN, nous détectons une légère induction de la transcription des séquences télomériques (HEXA\_TR\_A1) au cours de la réponse au stress (Schoeftner and Blasco 2008) Enfin, cette expérience révèle également des modifications d'expression de microsatellites (TR\_B2) et de séquences répétées dispersées (Figure 72).



**Figure 72 :** Expression des séquences répétées (SR) en tandem et dispersées au cours de la réponse cellulaire au stress. En plus de la séquence représentant les satellites 3 de la région péri-centromérique 9q12 (PHUR-98\_99), qui présente la plus forte augmentation d'expression, nous avons détecté une légère induction des séquences télomériques (HEXA\_TR\_A1) de microsatellites (trinuécléotides TR\_B2) mais également de rétrotransposons à LTR, de rétrovirus et de transposons à ADN. Les taux d'induction sont présentés comme la moyenne des rapports HS/NHS de trois expériences indépendantes (+/- écart type).

## 2. Irradiation et différenciation cellulaire

Jusqu'à présent, nous avons présenté un certain nombre de conditions capables d'induire la transcription des séquences répétées. Cependant, grâce à la RepChIP, nous avons également mis en évidence des conditions particulières ne conduisant pas à l'expression de ces séquences.

### a) Irradiation

Nous avons montré que le choc thermique, à travers le facteur de transcription HSF1 est capable d'induire la transcription des séquences répétées. De même, le stress hyper-osmotique conduit également à la transcription des séquences satellites 3 du locus 9q12, via le facteur de transcription TonEBP (Valgardsdottir, Chiodi et al. 2008). Nous avons cherché à caractériser l'impact d'un autre type de stress, un stress

génomique, sur l'expression des séquences répétées. Pour cela, nous avons étudié, en collaboration avec le Dr Jérôme Lamartine, l'effet de différentes doses d'irradiation aux rayons gamma sur des kératinocytes (Franco, Lamartine et al. 2005).

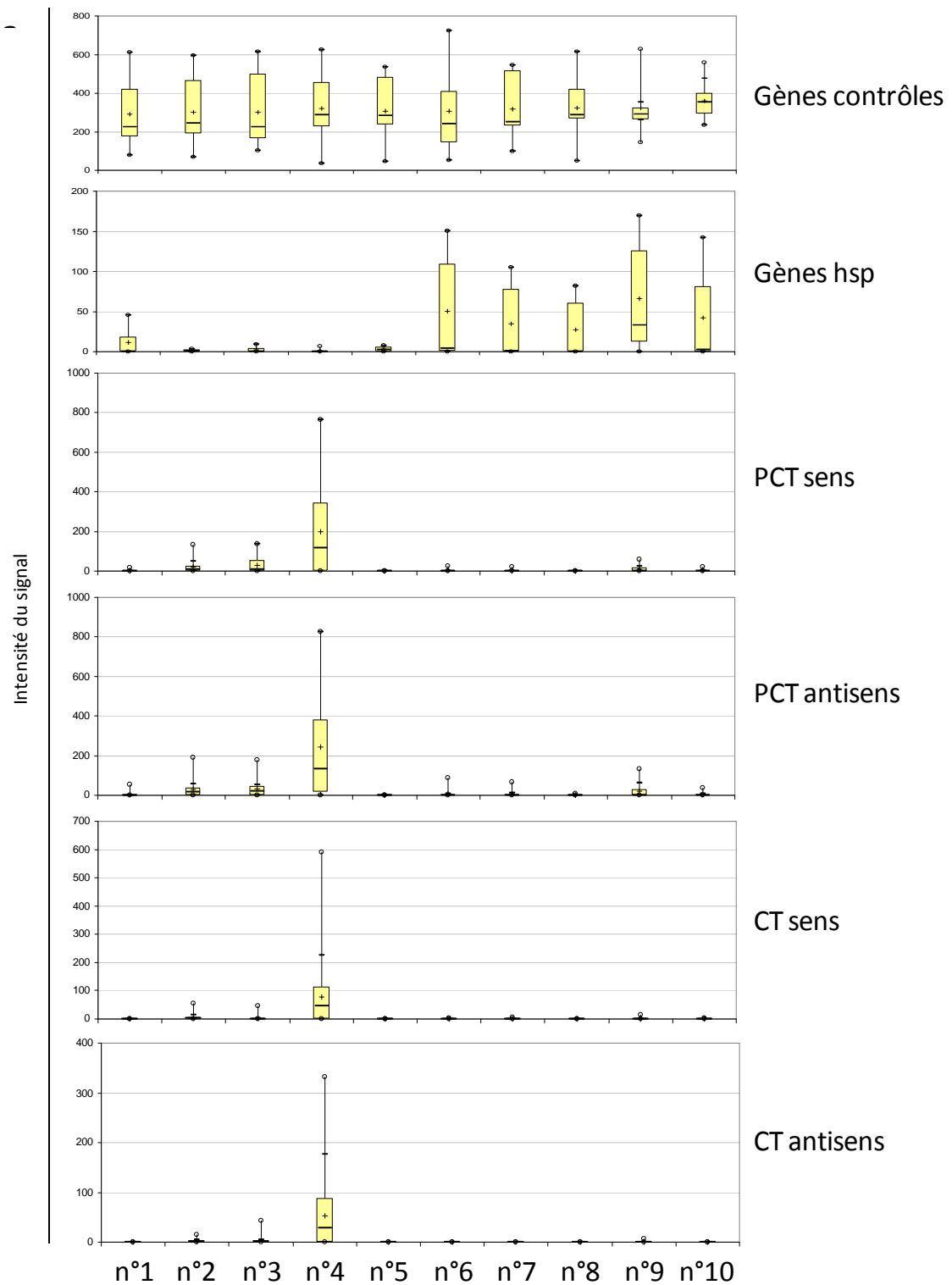
Les cellules ont été irradiées à des doses allant de 0.5Gy à 2Gy Les ARN ont ensuite été extraits 3h ou 24h après l'irradiation. Il a en effet été montré que si les fortes doses ont un effet immédiat, dans les heures qui suivent le stress, de plus faibles doses entraînent une réponse tardive, de 24 à 72h après l'irradiation. L'expression des séquences répétées a ensuite été analysée par RepChip. Cependant, aucune induction de l'expression de ces séquences n'a pu être mise en évidence, quelque soit la condition testée.

#### b) Différenciation cellulaire

Nous nous sommes intéressés à l'implication de la différenciation cellulaire sur l'expression des séquences répétées. Pour cela, nous avons utilisé comme modèle d'étude des lymphocytes B spléniques stimulés ou non par le CD40 ligand (CD40L). Le CD40 est exprimé de manière constitutive par les LB. Au contraire, l'expression du CD40L est induite lors de l'activation des lymphocytes T. L'interaction entre le CD40 et le CD40L joue un rôle crucial dans la mise en place de la réponse immunitaire. En effet, cette interaction stimule la prolifération et la différenciation des LB (D'Orlando 2007). Ainsi, dans ce travail, réalisé en collaboration avec le groupe des Dr Mary Callanan et Dominique Leroux, nous avons reproduit ces conditions d'activation *in vitro*. L'expression des séquences répétées dans les LB stimulés ou non par le CD40L a ensuite été analysée par RepChip. Cependant, aucune induction de l'expression de ces séquences n'a pu être mise en évidence.

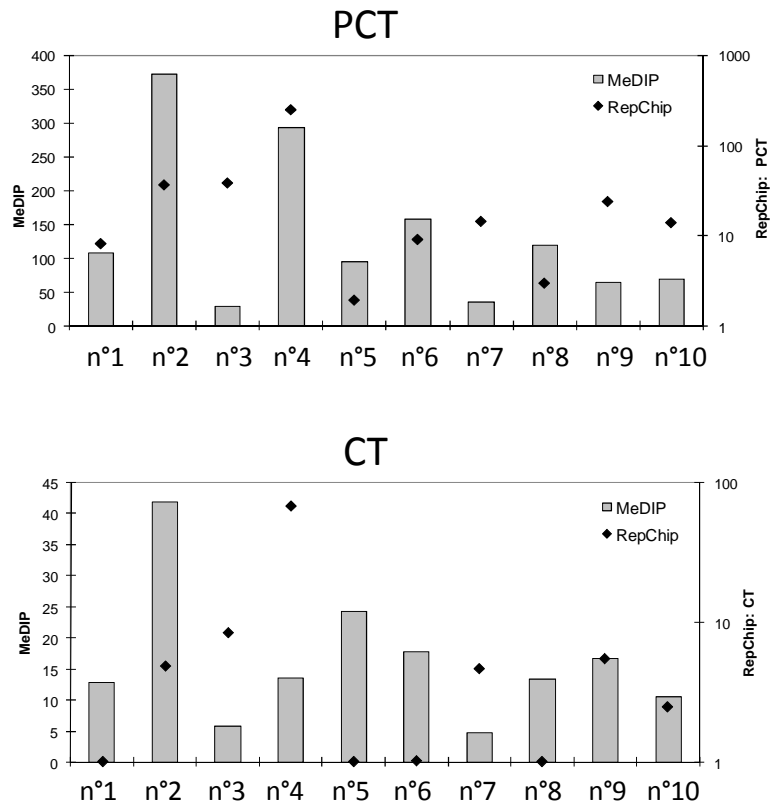
### *3. Etude de l'expression des séquences répétées au cours d'un stress physiopathologique*

Nous avons abordé le potentiel transcriptionnel des séquences répétées au cours d'un stress physiopathologique, un lymphome B. Ce travail a été réalisé en collaboration avec le groupe des Dr Mary Callanan et Dominique Leroux. Nous avons à notre disposition un échantillon de 10 patients présentant un lymphome B. Les résultats que nous avons obtenus avec la RepChip sont présentés dans la Figure 73.



**Figure 73 :** Etude par approche RepChip du potentiel transcriptionnel des séquences répétées centromériques (CT) et péricentromériques (PCT) dans dix patients (N°1 à n°10) présentant un lymphome B. Parmi les dix patients testés, l'expression des séquences centromériques et péricentromériques à été détectée dans le patient n°4 et dans une moindre mesure dans les patients n°2 et n°3.

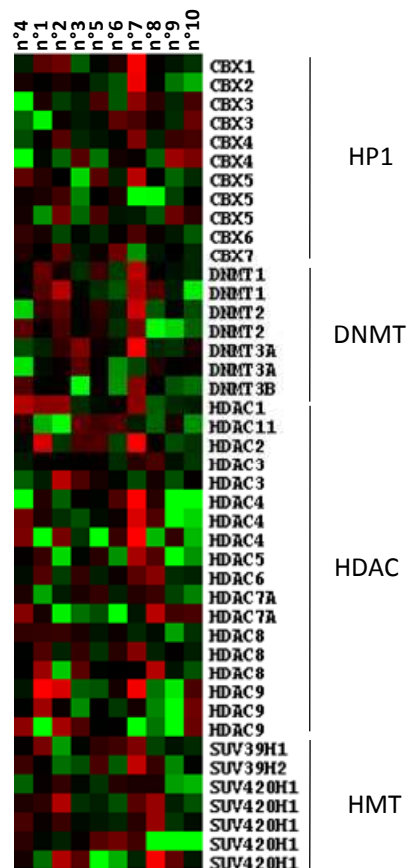
Parmi les 10 patients testés, le patient n°4 est particulièrement intéressant. En effet, pour ce patient on observe une forte expression des séquences répétées péri-centromériques, dans les deux orientations, mais également de TOUTES les séquences centromériques, également dans les deux orientations. On assiste ainsi à une dérégulation totale de l'expression des séquences centromériques et péri-centromériques chez ce patient. Dans une moindre mesure, les patients n°2 et 3 expriment également à la fois les séquences centromériques et péri-centromériques.



**Figure 74 :** Relation entre la méthylation des séquences centromériques et péri-centromériques et leur taux de transcription chez dix patients présentant un lymphome B. Des expériences de MeDIP ont été réalisées pour les dix patients. L'histogramme gris indique le taux relatif d'enrichissement en 5-méthyl cytosine des séquences répétées par rapport au gène contrôle (GAPDH). Les taux d'expression des séquences répétées sont également indiqués (losanges noirs). Aucune corrélation n'a pu être mise en évidence entre le statut méthylé de ces séquences et le taux d'expression des séquences centromériques ou péri-centromériques.

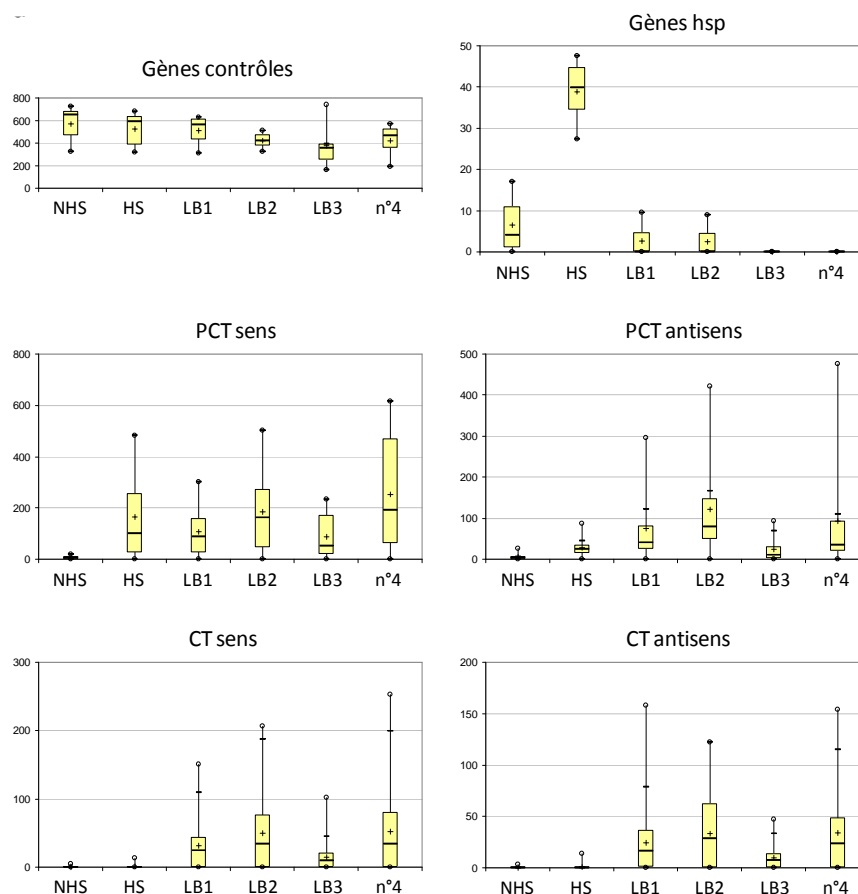
Afin de déterminer si cette transcription mettait en jeu la réponse au stress, nous avons analysé l'expression des gènes hsp pour l'ensemble de ces patients. Cependant aucune corrélation entre le niveau d'activation des gènes hsp et l'état transcriptionnel des séquences répétées n'a pu être établi (Figure 73). De plus, de manière assez spectaculaire, les gènes hsp sont faiblement exprimés chez les patients N°2, 3, 4 et 5 par rapport aux autres patients.

Sachant que nous avons mis en évidence deux voies indépendantes, conduisant à l'expression des séquences répétées péri-centromériques, nous avons cherché à déterminer quel était l'état de méthylation des séquences centromériques et péri-centromériques chez ces patients. Nous avons donc réalisé des expériences de MeDIP pour chacun d'entre eux, à la fois sur les séquences péri-centromériques et centromériques (Figure 74). Cependant aucune corrélation n'a pu être établie entre l'expression de ces séquences et leur état de méthylation. Ainsi, les mécanismes mis en jeu dans cette transcription sont certainement plus complexes. Profitant du fait que des analyses Affymetrix (puce U133 2.0 Plus) avaient été réalisées pour l'ensemble de ces patients, nous avons exploité ces données dans le but de mettre en évidence une corrélation entre l'expression de certains remodeleurs de la chromatine, tels que les HAT, HDAC ou encore les HMT, et l'expression des séquences répétées. Cependant, encore une fois, aucune corrélation n'a pu être établie (Figure 75).



**Figure 75 :** Résultat d'analyse de puce Affymetrix pour l'expression de différents acteurs épigénétiques dans dix patients présentant un lymphome B. Les résultats ont été normalisés grâce aux résultats obtenus pour le gène de l'actine. (Vert= gènes faiblement exprimés, rouge= gènes fortement exprimés).

Dans cette étude, nous ne disposons pas de témoins adéquats permettant de savoir si l'expression des séquences répétées est globalement induite ou réprimée au cours de la tumorigenèse. Nous avons cependant comparé la capacité transcriptionnelle des séquences répétées du patient n°4 à celles de trois échantillons différents de lymphocytes B circulants provenant du sang périphérique d'individus sains (LB1, LB2, LB3) (Figure 76). Des cellules HeLa soumises (HS) ou non (NHS) à un stress thermique sont présentées comme cellules contrôles. Ce travail, réalisé en collaboration avec le groupe du Dr Joël Plumas, indique que dans les LB périphériques du sang circulant, les séquences répétées centromériques et péri-centromériques sont exprimées, d'une manière similaire à ce qui a été observé pour le patient n°4. Cependant, ces LB normaux circulants sont des cellules quiescentes et en raison de leurs caractéristiques particulières, ils ne constituent pas forcément de bons témoins dans cette expérience pour les patients atteints de lymphome B (Baxter, Sauer et al. 2004).



**Figure 76 :** Expression des séquences centromériques et péri-centromériques dans un patient présentant un lymphome B (n°4) et dans trois échantillons de LB circulants provenant de sujets sains (LB1, LB2 et LB3). Les cellules HeLa soumises (HS) ou non (NSH) à un choc thermique sont utilisées comme témoins. L'expression des séquences centromériques et péri-centromériques est observée pour les quatre échantillons de LB, normaux ou tumoraux.



Bien qu'aucune conclusion ne peut être apportée concernant la relation entre l'expression de ces séquences répétées et le cancer, cette étude nous a permis de mettre en évidence une totale dérégulation de l'ensemble des séquences centromériques et péri-centromériques. Des analyses complémentaires sont bien évidemment à envisager afin de comprendre les mécanismes impliqués dans cette transcription et d'établir le rôle de celle-ci dans la pathologie

## **II.6 Perspectives**

Les perspectives de ce travail sont nombreuses. En effet, après avoir validé la RepChip, nous avons à peine effleuré certains des mécanismes qui pourraient potentiellement être impliqués dans l'expression des séquences répétées. Ainsi, concernant les mécanismes conduisant à l'expression des séquences répétées, bien d'autres modifications épigénétiques restent à tester avec en tête de liste l'implication de Suv39H dans ce processus. En effet, dans des cellules ES murines déficientes en Suv39H une forte induction de l'expression des satellites majeurs et mineurs a été observée (Lehnertz, Ueda et al. 2003). Nous avons utilisé des approches d'ARN interférence dirigées contre Suv39H pour déterminer son impact sur la transcription des séquences répétées mais la diminution de l'expression de cette protéine n'est pas suffisante pour abolir la méthylation d'H3K9. Il est impossible de lister l'ensemble des mécanismes épigénétiques qu'il serait important d'étudier. Cependant afin de compléter la liste des inhibiteurs de HDAC, il serait intéressant de tester l'effet d'inhibiteurs des sirtuines tels que le nicotinamide, sur la transcription des séquences répétées (Avalos, Bever et al. 2005). Il a en effet été montré que l'inhibition de ces HDAC de classe III pouvait conduire à la réactivation de certains gènes, sans affecter la méthylation de l'ADN (Pruitt, Zinn et al. 2006). Dresser la liste des modifications épigénétiques capables d'induire ou non la transcription des séquences répétées permettrait de mieux comprendre les mécanismes impliqués dans ce processus. Pour cela, il serait également très avantageux de combiner de manière systématique l'approche transcriptomique de la RepChip avec l'analyse globale du statut épigénétique qui pourrait être obtenue avec la Rep'Array.

D'autre part, étant donné que la transcription des séquences répétées chez *S.pombe* et chez la souris est régulée au cours du cycle cellulaire, il serait particulièrement intéressant d'analyser le taux d'ARN-CT et d'ARN-PCT dans des cellules issues de chaque phase du cycle cellulaire (Lu and Gilbert 2007; Kloc, Zaratiegui et al. 2008; Lu and Gilbert 2008). En faveur d'une régulation conservée de

la transcription des séquences répétées au cours du cycle cellulaire, nous avons pu mettre en évidence la transcription des séquences centromériques et péricentromériques dans certains tissus tumoraux. Ainsi l'étude de la capacité transcriptionnelle des séquences répétées en fonction de l'état prolifératif des cellules est à envisager. Concernant l'analyse de patients atteints de cancer, la problématique repose à la fois sur le choix de tissus sains de référence et sur le nombre de cas étudiés. Dans le cadre du réseau Epimed associant plusieurs laboratoires de l'Institut Albert Bonniot, sous la direction du Pr Sophie Rousseaux, il serait intéressant de valider les données que nous avons obtenues pour le cancer du poumon sur une grande série de patients. Ce travail pourrait se faire en collaboration avec le groupe du Dr Sylvie Gazzeri.

Enfin, le dernier aspect, mais non le moindre, concerne la fonction de ces ARN issus de séquences répétées. L'étude précise de la corrélation entre différents stades de la progression tumorale et l'expression de ces séquences pourrait permettre d'établir de nouvelles hypothèses. A l'issue de ce travail, nous pourrions espérer pouvoir conclure quant à l'utilisation de ces ARN issus des séquences répétées comme marqueurs de diagnostic, voir de pronostic. A l'appui de cette hypothèse, la ressemblance entre le schéma d'expression des séquences répétées et celle des « Cancer Testis Antigens » est frappante. Ces protéines ne sont en effet pas exprimées dans les cellules somatiques mais dans le testicule ainsi que dans des tumeurs. Ils sont à l'heure actuelle à l'étude quant à leur utilisation comme marqueurs de diagnostic, pronostic et même comme de nouvelles cibles potentielles en immunothérapie (Scanlan, Gure et al. 2002; Almeida, Sakabe et al. 2008; Gjerstorff, Harkness et al. 2008; Grau, Oltra et al. 2008). Il serait ainsi intéressant de pouvoir comparer les conditions sous lesquelles ces Cancer Testis Antigens s'expriment avec celles induisant la transcription des séquences centromériques et péricentromériques. Cette étude nous permettrait de déterminer si leur expression est régulée par des mécanismes communs. Enfin, un rôle pour les ARN issus des séquences satellites 3 du locus Yq12 a été récemment suggéré. Ces ARN-PCT particuliers pourraient en effet être impliqués dans des mécanismes de trans-épissage impliquant l'ARN CDC2L2 (Cell Division Cycle 2-Like 2) (Jehan, Vallinayagam et al. 2007). Etant donné que certaines isoformes de cette kinase CDC2L2 sont associés aux cyclines avant de devenir des substrats pour les caspases, les auteurs de cet article suggèrent que les événements de trans-épissage, et donc les ARN-PCT, pourraient être importants pour la régulation du cycle cellulaire ainsi que de l'apoptose, autant de mécanismes dérégulés dans les cancers.

## Chapitre III Le syndrome ICF

Les résultats présentés dans ce chapitre, font l'objet d'un quatrième article en cours de préparation :

« *ICF syndrome, DNA methylation and the stress response* »

*Angéline Eymery, Caroline Jolly, Evani Viegas-Pequignot and Claire Vourc'h*

Ce travail a été réalisé en collaboration avec le groupe du Dr Evani Viegas-Pequignot pour l'obtention de cellules provenant de patients atteints par le syndrome ICF. Après avoir introduit le contexte de cette étude, les résultats seront présentés et discutés.

### III.1 Introduction

Nous venons de voir que le choc thermique induit la transcription des séquences satellites 2 et 3. Dans certains cas, la déméthylation de l'ADN semble être un processus capable de favoriser l'expression de ces mêmes séquences. Afin de poursuivre notre étude sur l'existence d'une relation potentielle entre la méthylation de l'ADN et la transcription des séquences répétées, nous avons élargi notre champ d'investigation au syndrome ICF.

Le syndrome ICF (Immunodéficience, instabilité Centromérique et anomalies Faciales) (OMIM 24242860) est une pathologie particulièrement intéressante pour examiner l'impact de la méthylation de l'ADN sur la capacité transcriptionnelle des séquences répétées de type satellites. En effet, le syndrome ICF a été originellement caractérisé comme la seule pathologie humaine associée à des mutations dans le gène codant pour une ADN méthyltransférase, la DNMT3b (Hansen, Wijmenga et al. 1999; Xu, Bestor et al. 1999). Cette maladie récessive est particulièrement rare. En effet, depuis sa découverte dans les années 1970, moins de 50 cas ont été répertoriés (Tiepolo, Maraschio et al. 1979). Le diagnostic de cette pathologie associe à l'immunodéficience, des remaniements chromosomiques impliquant les régions péri-centromériques, ainsi que des anomalies faciales. L'immunodéficience est

généralement la cause du décès prématuré, avant l'adolescence, des enfants atteints par ce syndrome.

Les chromosomes issus de cellules de patients ICF présentent une hypométhylation des séquences satellites 2 et 3 principalement au niveau des régions péricentromériques des chromosomes 1, 9 et 16 (Maraschio, Zuffardi et al. 1988). De plus, dans les chromosomes de lymphocytes issus de patients ICF, de nombreuses anomalies cytogénétiques peuvent être mises en évidence telles que des délétions, des cassures ou encore l'apparition de formes particulières appelées chromosomes multibranchés impliquant l'association des péricentromères des chromosomes 1 et 16. Le chromosome 9 est rarement impliqué dans ces remaniements.

Très peu de choses sont connues quant aux conséquences fonctionnelles de la déméthylation de l'ADN chez ces patients atteints du syndrome ICF, en particulier sur l'expression des gènes (Ehrlich, Sanchez et al. 2008). Cependant, la déméthylation dans les cellules ICF des séquences non codantes NBL2, présentes en plusieurs exemplaires dans la région péricentromérique des chromosomes 9, 13, 14 et 21, est associée à une augmentation de leur expression (Kondo, Bobek et al. 2000). Dans ce contexte général, nous avons cherché à évaluer l'impact de la déméthylation de l'ADN sur l'expression des séquences satellites 3 du locus 9q12, dans ces cellules issues de patients ICF, soumises ou non à un choc thermique.

## **III.2 Matériel et méthodes**

### *1. Culture cellulaire et choc thermique*

Les fibroblastes primaires de deux patients ICF (référéncés comme les patients R et G d'après (Xu, Bestor et al. 1999)) ainsi que les fibroblastes primaires de patients ICF obtenus par l'Institut Coriell (USA) (référéncés sous GM08747A) ont été utilisés. Cependant, étant donné que des résultats équivalents ont été obtenus pour ces trois lignées cellulaires, seulement les résultats obtenus avec le patient G seront présentés. Des fibroblastes primaires provenant d'un individu sain ont été utilisés comme contrôles. Le choc thermique a été réalisé par immersion des boîtes de Pétri contenant les lames dans un bain-marie. Différentes conditions de choc thermique ont été testées : la température (de 42°C à 45°C) et la durée du choc thermique (de 1h à 3h) ainsi que celle de la période de récupération à 37°C (de 1h à 6h) suivant le choc thermique. L'existence de cellules révertantes dans les ICF ainsi que l'utilisation de

cellules témoins nous ont permis de déterminer les conditions optimales de choc thermique pour lesquelles les granules de stress et la transcription des séquences répétées sont détectées : 1h à 45°C suivie de 3h de récupération à 37°C.

Les cellules HCT116 parentales ou mutantes ont été soumises à un choc thermique de 43°C pendant une heure, suivi ou non d'une période de 3h récupération à 37°C.

## *2. Immunofluorescence, et hybridation in situ fluorescente*

L'Immunofluorescence et la FISH ADN ont été réalisées selon le protocole décrit dans l'article n°1. Pour la FISH ARN, la préparation des sondes pHuR98 (Grady, Ratliff et al. 1992) et PH2.3 (obtenue par R.I Morimoto, USA) correspondant respectivement aux séquences satellites 3 du locus 9q12 et à la séquence codante du gène hsp70, a été réalisée comme décrit dans l'article n°1. Les sondes sont ensuite dénaturées 5 minutes à 75°C avant d'être déposées sur les lames déshydratées. L'incubation se fait à 37°C pendant une nuit. Le lendemain, les lames sont rincées 3 fois 5 minutes dans 60% formamide/2SSC à 45°C puis trois fois 5 minutes dans du 2SSC à température ambiante. Le blocage, la détection et le montage des lames se font ensuite de manière similaire à la technique décrite pour la FISH ADN.

## *3. Western Blot*

Les extractions de protéines et le western blot ont été réalisés selon le protocole décrit dans l'article n°1. Les anticorps utilisés dans cette étude, ainsi que leur dilution, sont les suivants : lapin anti-HSF1 1/5000, lapin anti-HSP70 1/2500, lapin anti-actine 1/500 et lapin anti-tubuline 1/2000.

## *4. Reconstitution de granules sur chromosomes métaphasiques*

Les étalements de chromosomes métaphasiques issus de sang périphérique de patients ICF ont été réalisés comme décrit dans l'article n°1 (Ronne, Andersen et al. 1979). La quantité de protéine HSF1 recombinante utilisée dans cette expérience est de 200 ng.

## 5. RepChip

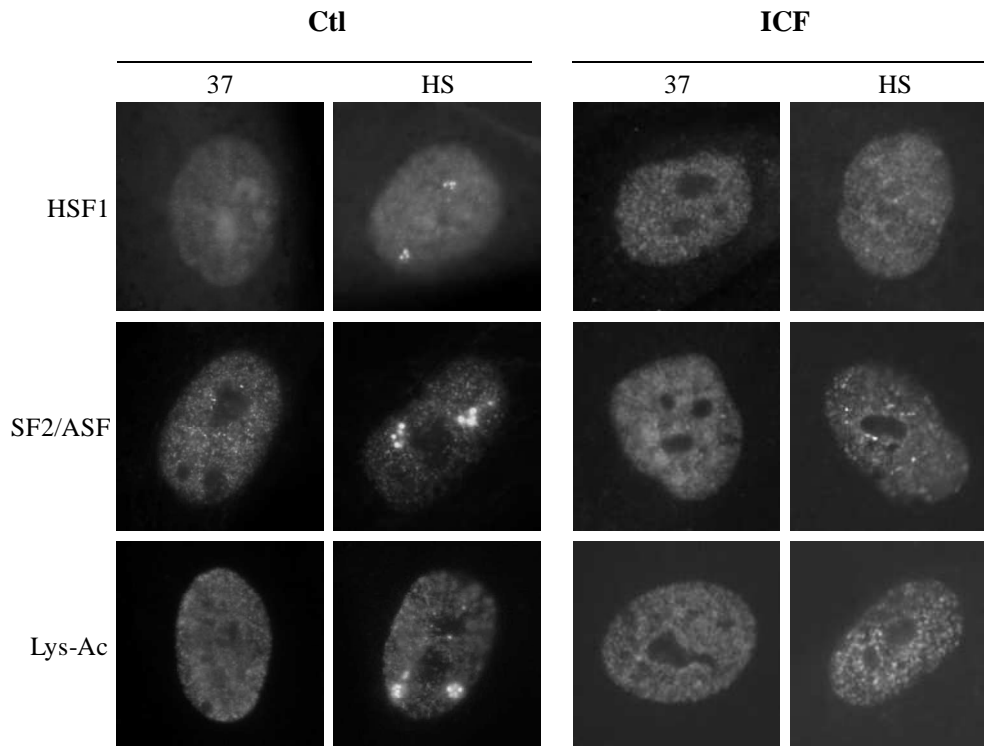
L'expérience a été réalisée selon le protocole décrit dans l'article n°2 sur des cellules HCT116 parentales ou déficientes en DNMT3b seule (3bKO) ou en combinaison avec DNMT1 (DKO).

### III.3 Résultats

#### 1. Absence de granule de stress dans les cellules ICF

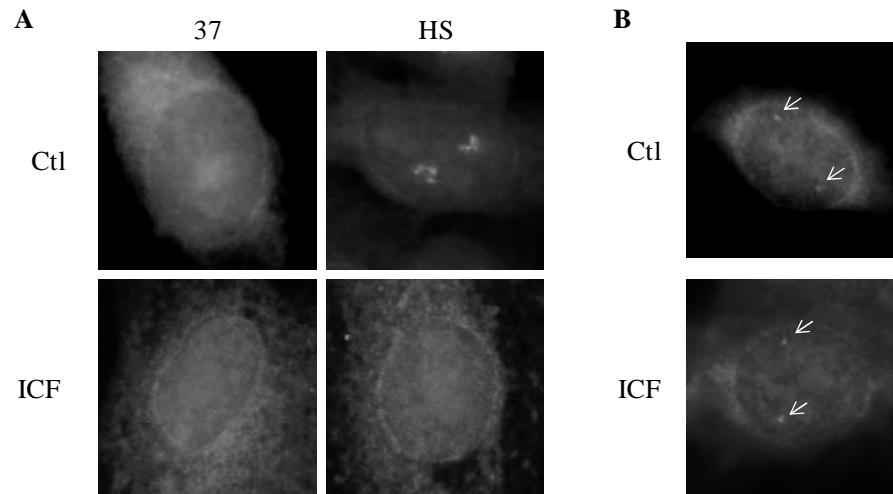
Afin d'évaluer l'impact de la méthylation de l'ADN sur l'expression des séquences satellites 3, nous avons dans un premier temps cherché à mettre en évidence la présence de granules de stress dans les cellules ICF. HSF1 a été détecté par immunofluorescence, dans des cellules ICF soumises (HS) ou non (37) à un choc thermique. La localisation d'HSF1 a été étudiée en fonction de celle de deux autres facteurs, connus pour s'accumuler au locus 9q12 dans les cellules soumises à un choc thermique, les histones acétylées et le facteur d'épissage SF2/ASF. La cinétique de formation des granules de stress et d'activation des séquences satellites 3 varie d'un type de cellule à l'autre au cours de la réponse au choc thermique. Ainsi plusieurs conditions de stress ont été testées dans les cellules ICF et les fibroblastes contrôles (Ctl) La condition optimale pour les deux lignées est la suivante : 1h de choc thermique à 45°C, suivie de 3h de récupération à 37°C (voir matériel et méthodes). Afin de simplifier les figures cette condition sera notée « HS » dans la suite des expériences, à l'exception de la Figure 79-C.

En absence de stress, les protéines HSF1, SF2/ASF et les lysines acétylées présentent toutes une distribution nucléaire homogène (Figure 77). Au cours du stress, deux foyers enrichis en HSF1, SF2/ASF et lysines acétylées sont aisément détectables dans les cellules contrôles. En revanche, ces foyers sont détectés dans seulement 1% des cellules ICF soumises à un choc thermique. Ces cellules pourraient correspondre aux cellules présentant un phénotype révertant fréquemment observé dans les cultures de cellules ICF. Ainsi, ces résultats indiquent non seulement une absence de formation des granules de stress dans les cellules ICF soumises à un choc thermique, mais suggèrent également un défaut de transcription des séquences satellites 3 au cours du stress.



**Figure 77 :** Les cellules ICF sont incapables de former des granules de stress. La détection d' HSF1, de SF2/ASF, et des lysines acétylées (Lys-Ac) a été réalisée par immunofluorescence dans des cellules ICF et contrôles (Ctl) soumises (HS) ou non (37) à un choc thermique. Après un choc thermique, les granules de stress, enrichies en HSF1, SF2/ASF et en lysines acétylées, sont détectées dans les cellules contrôles mais pas dans les cellules ICF.

Afin de confirmer l'absence de transcription des séquences satellites 3 au cours du stress dans les cellules ICF, les ARN satellites 3 ont été détectés par FISH ARN (Figure 78-A). En absence de stress, un marquage nucléaire diffus non spécifique est observé dans les fibroblastes contrôles, et dans les cellules ICF, suggérant qu'aucun ARN issus des séquences satellites 3 n'est présent. Ce résultat indique clairement que les cellules ICF ne sont pas capables d'induire la transcription constitutive des séquences satellites 3, même en présence d'une déméthylation du locus. Sous l'effet du stress, deux foyers nucléaires enrichis en ARN satellites 3 sont détectés dans l'ensemble des cellules contrôles. Par contre, seulement 1% des cellules ICF présentent des foyers d'ARN satellites 3.



**Figure 78** : Les cellules ICF sont incapables de transcrire les séquences satellites 3 sous l'effet d'un choc thermique. **A** Les ARN satellites 3 sont détectés par FISH ARN dans les cellules ICF et contrôles (Ctl). Sous l'effet du choc thermique, des foyers d'ARN satellites 3 sont observés dans les cellules contrôles mais pas dans les cellules ICF. **B** Au contraire, les ARN hsp70 sont détectés à la fois dans les cellules ICF et dans les cellules contrôles, soumises au stress thermique (flèches).

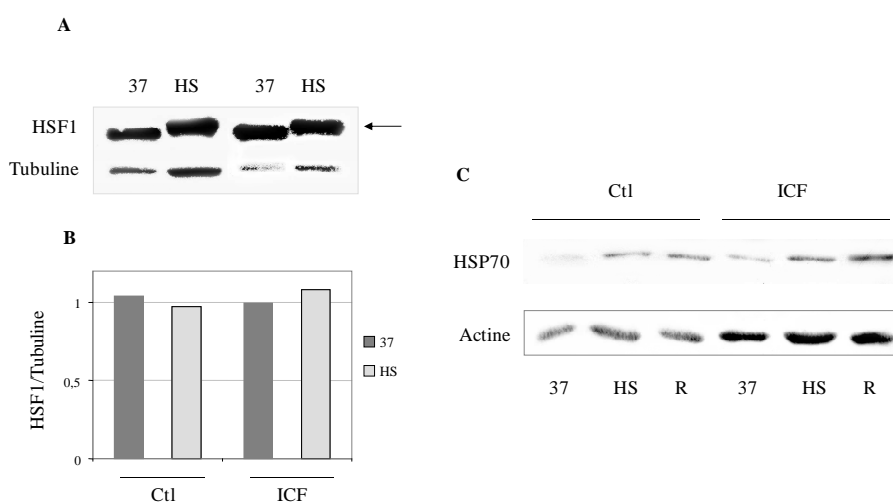
## 2. La réponse au stress est conservée

L'absence de formation des granules de stress ainsi que l'incapacité des cellules ICF à transcrire les satellites 3, laissait suggérer que ces cellules pourraient présenter une altération de la réponse au stress. La capacité des cellules ICF à transcrire le gène hsp70 a donc été testée. Comme on peut le voir dans la Figure 78-B deux foyers nucléaires enrichis en ARN hsp70 sont détectés à la fois dans les cellules ICF et dans les cellules contrôles, soumises au stress thermique. Ces résultats montrent donc que les cellules ICF présentent une altération de la transcription des séquences satellites 3 en dépit d'une transcription normale du gène hsp70.

Néanmoins, la possibilité que ces gènes hsp70 puissent être transcrits par des mécanismes indépendants d'HSF1 restait possible. L'activation d'HSF1 dans des cellules ICF soumises ou non au choc thermique a donc été examinée sur la base de sa mobilité sur gel d'acrylamide. En effet, lorsqu'il est activé, HSF1 présente une mobilité réduite due aux modifications post-traductionnelles qu'il subit au cours de son activation. Comme on peut le voir dans la Figure 79-A, une réduction de la mobilité d'HSF1 est observée, pour les extraits protéiques issus des cellules ICF et des cellules contrôles soumises au choc thermique. Ainsi, HSF1 est parfaitement activable dans les cellules ICF soumises au choc thermique. De plus, HSF1 semble être présent dans des quantités similaires dans ces deux lignées (Figure 79-B).



Enfin, la capacité des cellules ICF à répondre au choc thermique a également été évaluée en analysant l'expression de la protéine HSP70 dans des cellules soumises ou non au stress (Figure 79-C). Les résultats indiquent que les cellules ICF présentent la même capacité à induire l'expression de la protéine HSP70 que des cellules contrôles. Finalement, ces expériences montrent clairement que la réponse au stress « classique » n'est altérée dans les cellules ICF. Par contre la réponse au stress impliquant la formation des granules de stress et la transcription des séquences répétées est déficiente dans les cellules ICF.

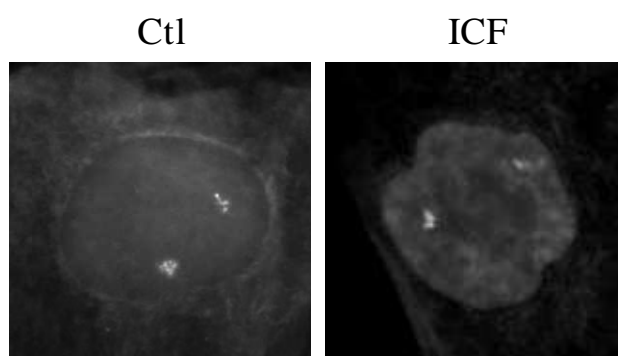


**Figure 79** : La réponse au stress « classique » est fonctionnelle dans les cellules ICF. **A** La mobilité électrophorétique d'HSF1 dans des extraits protéiques issus de cellules ICF et de cellules contrôles soumises (HS) ou non (37) au choc thermique a été analysée par western blot en utilisant un anticorps anti-HSF1. Le retard de migration d'HSF1 observé dans les extraits protéiques de cellules ICF et de cellules contrôles soumises au choc thermique reflète la présence de modifications post-traductionnelles de la forme active d'HSF1 (flèche). **B** La quantification de la protéine HSF1 est obtenue, grâce à ImageJ, en calculant le ratio HSF1/tubuline. **C** HSP70 est détectée par western blot dans les cellules ICF et dans les cellules contrôles soumises à une cinétique de choc thermique (37: cellules non stressées, HS: 1h de choc thermique à 45°C, R: 3h de récupération à 37°C après une heure de choc thermique à 45°C). HSP70 est exprimée au cours du stress, dans les cellules ICF et dans les cellules contrôles, d'une manière similaire.

### 3. La déméthylation de l'ADN n'empêche pas la formation des granules de stress

Nous nous sommes demandé si l'absence de relocalisation d' HSF1 au locus 9q12 ainsi que de l'absence de transcription des séquences répétées ne pourrait pas être due à une délétion de ces séquences satellites 3 dans les cellules ICF. Une expérience de FISH ADN a donc été effectuée dans des cellules ICF en utilisant une sonde spécifique des séquences satellites 3 du locus 9q12, la sonde pHuR98. Deux signaux

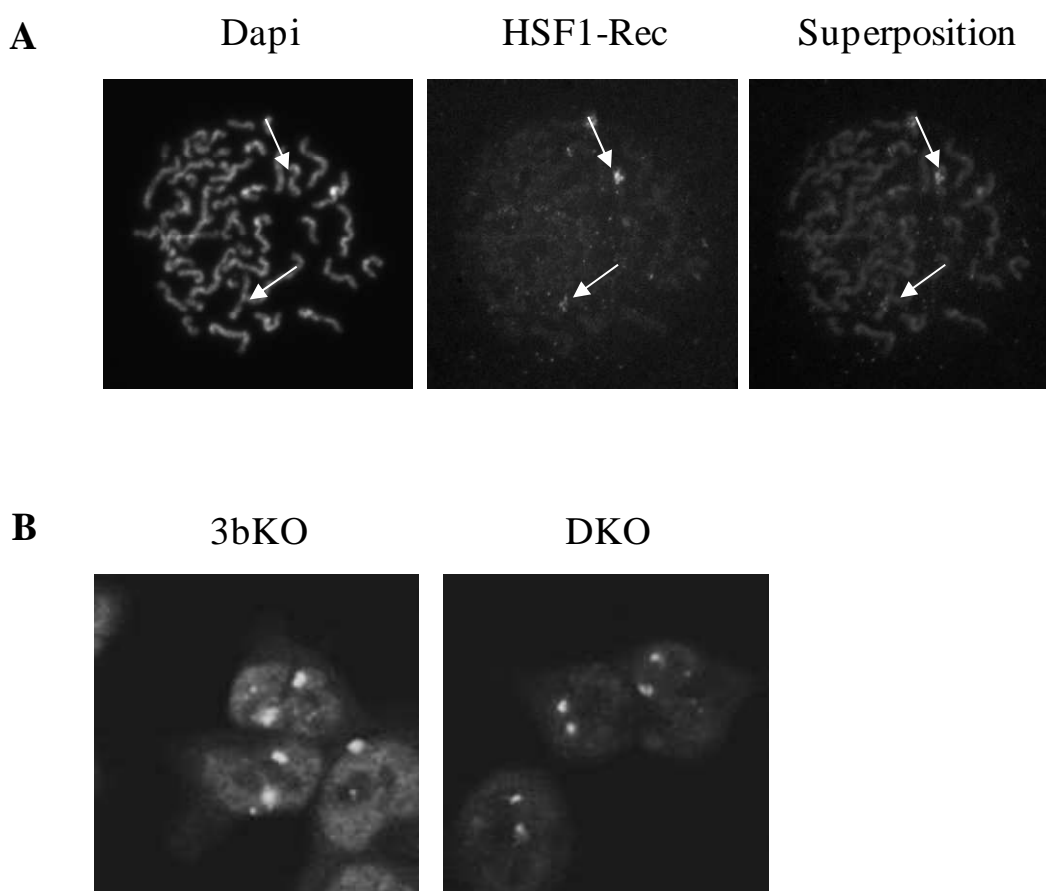
d'hybridation, correspondant aux deux loci 9q12, sont observés dans le noyau des cellules ICF (Figure 80). L'intensité et la taille de ces signaux est comparable à celles des cellules contrôles. Ce résultat suggère donc que l'absence d' HSF1 au locus 9q12 dans les cellules ICF n'est pas due à une perte des séquences satellites 3.



**Figure 80** : Les séquences génomiques satellites 3 du locus 9q12 sont présentes dans les cellules ICF. Les séquences satellites 3 présentes au sein du locus 9q12 sont détectées par FISH ADN à l'aide d'une sonde spécifique, pHuR98. La présence de deux signaux d'hybridation dans les cellules ICF et dans les cellules contrôles indique l'absence d'une délétion de ce locus dans les cellules ICF.

L'absence de granules de stress dans les cellules ICF soumises à un choc thermique pourrait s'expliquer par l'incapacité du facteur HSF1 à se lier sur des séquences déméthylées. Pour tester cette hypothèse, nous avons utilisé une technique de reconstitution de granules, décrite dans l'article n°1, qui consiste à incuber une protéine HSF1 recombinante sur des chromosomes métaphasiques provenant de cellules issues de patients atteints par le syndrome ICF. La protéine HSF1 est ensuite détectée par immunofluorescence classique. Comme on peut le voir sur la Figure 81-A, deux signaux fluorescents correspondant aux sites de liaison d'HSF1 sur les chromosomes 9 sont détectés dans les cellules ICF (flèches) Ces signaux sont comparables à ce que l'on peut obtenir avec des cellules normales (Jolly, Konecny et al. 2002). Ainsi, les séquences satellites 3 des cellules ICF, bien que déméthylées, sont parfaitement capables de lier la protéine HSF1 recombinante. Pour confirmer ces résultats, nous avons réalisé une immunofluorescence dirigée contre HSF1 dans des cellules déficientes en DNMT3b seule (3bKO) ou en combinaison avec la DNMT1 (DKO). En effet, l'absence de ces protéines conduit à la déméthylation des séquences satellites péricentromériques (article n°3 : Horard B, Eymery A et al, en préparation). Comme on peut le constater sur la Figure 81-B, la déméthylation de l'ADN au niveau du locus 9q12 n'empêche pas la formation des foyers HSF1. Ainsi, ces résultats excluent la possibilité que l'absence de relocalisation d'HSF1 au locus 9q12

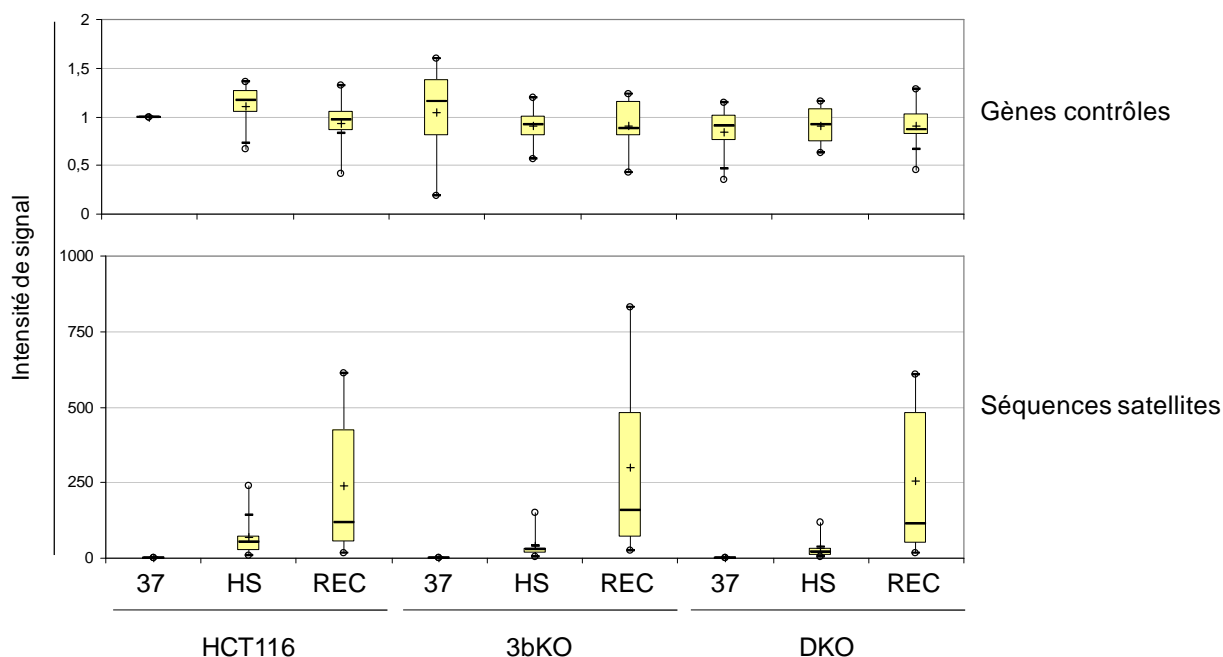
dans les cellules ICF puisse être la simple conséquence d'une déméthylation de ce locus.



**Figure 81** : Les séquences satellites 3 déméthylées sont capables de lier HSF1. **A** Les granules de stress ont été reconstituées *in vitro* en incubant une protéine HSF1 recombinante (HSF1-rec) avec des étalements de chromosomes métaphasiques provenant de cellules ICF. La protéine est ensuite détectée par immunofluorescence. Deux signaux sont observés sur les deux loci 9q12 (flèches), indiquant qu'HSF1 est parfaitement capable de se lier aux séquences satellites 3 déméthylées. **B** La protéine HSF1 est détectée par immunofluorescence dans des cellules déficientes en DNMT3b seule (3bKO) ou en combinaison avec la DNMT1 (DKO), soumises à un choc thermique. Des granules de stress sont observées dans ces deux lignées cellulaires soumises au choc thermique indiquant ainsi que la déméthylation des séquences satellites 3 n'empêche pas la liaison d'HSF1.

Enfin, une dernière interrogation demeurait quant à la capacité transcriptionnelle de ces séquences satellites déméthylées au cours du choc thermique. Pour répondre à cette question, nous avons évalué, par une approche transcriptomique, le taux d'ARN issus des séquences satellites péricentromériques (contenant les séquences satellites 3 du locus 9q12) au cours de la réponse au stress, dans des cellules 3bKO et DKO. Les résultats sont présentés dans la Figure 82. Ces cellules sont incapables, tout comme les cellules ICF de transcrire les séquences répétées en absence de stress. Cependant,

lorsqu'elles sont soumises à une cinétique de choc thermique, les séquences satellites péri-centromériques sont transcrites.



**Figure 82 :** Les séquences satellites déméthylées sont transcrites au cours du choc thermique. L'expression des séquences satellites, comprenant les séquences satellites 3 du locus 9q12, a été évaluée par approche transcriptomique dans des cellules HCT116 normales ou déficientes en DNMT3b (3bKO) ou en DNMT3b et DNMT1 (DKO). Les résultats ont été normalisés par rapport aux données obtenues dans les HCT116 normales (cellules dans lesquelles les délétions de *dnmt3b* et *dnmt1* ont été réalisées) non stressées. En absence de stress aucune de ces lignées ne présente de transcription des séquences satellites. Cependant au cours de la cinétique de choc thermique, des ARN issus des séquences satellites sont détectés pour les trois lignées. Ainsi, la déméthylation des séquences satellites n'empêche pas leur transcription au cours de la réponse au stress.

### III.4 Discussion

La relation entre la déméthylation des régions péri-centromériques et la transcription des séquences répétées qui les composent n'est pas clairement établie. En effet, nous avons précédemment montré que la déméthylation des séquences péri-centromériques sous l'effet de la 5 Aza-Cytidine pouvait conduire à leur expression. Cependant, des données contradictoires ont été obtenues avec les cellules HCT116 déficientes en DNMT3b seule ou en combinaison avec la DNMT1. Dans cet article, nous montrons que la déméthylation du locus 9q12 dans les cellules ICF ne

conduit pas à l'activation constitutive des séquences répétées de type satellites 3. L'absence de transcription constitutive des séquences satellites 3 dans les cellules ICF est en accord avec une étude montrant l'absence de transcription d'autres séquences péri-centromériques, les satellites 2, dans les cellules ICF (Alexiadis, Ballestas et al. 2007). Ainsi, conformément aux résultats présentés dans l'article n°2, cette observation suggère que la déméthylation de l'ADN n'est pas suffisante pour déréguler l'expression des séquences répétées, à moins que celle-ci soit drastique, comme c'est certainement le cas lors d'un traitement à la 5 Aza-Cytidine.

D'autre part, dans cet article, nous mettons pour la première fois en évidence un découplage de la réponse au stress chez l'Homme. Ainsi, alors que la voie « classique » de la réponse au stress (celle impliquant les gènes hsp) est parfaitement fonctionnelle dans les cellules ICF, ce n'est pas le cas pour celle impliquant les granules de stress. L'absence de localisation d'HSF1 au locus 9q12 dans les cellules ICF n'est pas directement liée au statut déméthylé de ces séquences. En effet, l'hypométhylation des séquences satellites 3 du locus 9q12 dans les cellules ICF n'empêche pas la fixation de la protéine HSF1 recombinante *in vitro*. De plus, les granules de stress sont détectés dans les cellules 3bKO et DKO. Les différences observées entre les cellules ICF et les cellules 3bKO et DKO peuvent s'expliquer par le fait que dans le cas de la pathologie la protéine est présente mais mutée et non totalement absente, ce qui pourrait laisser place à des mécanismes compensatoires.

Les mécanismes empêchant la fixation du facteur HSF1 et par conséquent la transcription des séquences satellites 3 n'ont pas encore été identifiées. Cependant, on peut imaginer que l'interaction d'HSF1 avec les séquences satellites 3 pourrait être modifiée par la présence (ou l'absence) de protéines particulières. Ces protéines pourraient empêcher la liaison d'HSF1 soit en agissant au niveau de la structure de la chromatine, soit en interférant avec des protéines connues pour jouer un rôle dans la structure de la chromatine péri-centromérique telles que les HDAC1, HP1 $\alpha$  et  $\beta$ , Suv39H, ou SNF2h. Il a notamment été proposé que SNF2h, le composant catalytique du complexe de remodelage SWI2/SNF2h, puisse en effet permettre le maintien d'une structure chromatiniennne ouverte au niveau du promoteur d'hsp70 *in vivo*, en absence de stress (Corey, Weirich et al. 2003). Dans ce contexte, il serait donc possible qu'à travers son interaction avec SNF2h, DNMT3b puisse faciliter l'ouverture chromatiniennne nécessaire à la liaison d'HSF1 au locus 9q12 (Geiman, Sankpal et al. 2004) Etant donné que le ciblage de SNF2h sur les promoteurs des gènes hsp est probablement indépendant de DNMT3b, les conséquences spécifiques liées à un défaut de localisation de SNF2h au niveau des séquences répétées pourraient constituer une

hypothèse attractive à tester afin d'expliquer ce découplage de la réponse au stress dans les cellules ICF.

### **III.5 Perspectives**

Concernant le syndrome ICF, les perspectives de ce travail sont assez limitées. En effet, le manque de matériel biologique, lié à la rareté du syndrome et au faible pouvoir de prolifération de ces cellules, rend les analyses difficiles. De plus, le maintien des cellules en culture aboutit souvent à la sélection d'une population indésirable de cellules révertantes.

Cependant, afin de déterminer si la DNMT3b est impliquée dans l'inhibition de la transcription des séquences satellites 3 au cours de la réponse au stress, il serait intéressant de pouvoir réaliser ces mêmes expériences dans des cellules issues de patients ICF de type 2. Il a en effet été montré que près de 40% des patients atteints par le syndrome ICF ne possèdent pas de mutations dans les exons de la DNMT3b (Jiang 2005). Ce sous-type de syndrome a été nommé type 2, par opposition au type 1 pour lequel des mutations dans le gène codant pour la DNMT3b ont été observées. Il est cependant envisageable que des mutations dans le promoteur ou dans d'autres éléments de contrôle de la transcription puissent affecter l'épissage ou la polyadénylation de l'ARNm. Mais dans certains cas, aucune altération du promoteur ou de l'épissage de l'ARNm n'a été détectée (Kubota, Furuumi et al. 2004; Kloeckener-Gruissem, Betts et al. 2005) Fait intéressant, la déméthylation des régions péricentromériques est également une caractéristique du syndrome ICF de type 2, malgré l'absence de mutations dans les régions codantes de la DNMT3b.

# Conclusion

Mon travail de thèse s'est focalisé sur l'étude de la transcription des séquences répétées centromériques et péricentromériques et en particulier les alpha satellites d'une part et les satellites 2 et 3 d'autre part.

Ainsi, j'ai pu identifier de nouvelles cibles chromosomiques du facteur de transcription HSF1 au cours de la réponse au stress au niveau des régions péricentromériques des chromosomes 1, 5, 7, 9, 10, 13, 14, 15, 16, 17, 21, 22 et Y. Or, ces régions sont connues pour être enrichies en séquences satellites 2 et 3. J'ai également montré que ces granules de stress secondaires correspondaient à des sites de transcription. Ainsi, la transcription des séquences répétées au cours de la réponse au stress ne se limite pas aux séquences satellites 3 du locus 9q12. De plus, à travers l'étude du syndrome ICF, j'ai mis en évidence un modèle de découplage de la réponse au stress.

Sur la base de ces observations, nous nous sommes intéressés au potentiel transcriptionnel de l'ensemble des séquences répétées centromériques et péricentromériques. Pour cela, j'ai développé, en collaboration avec les groupes des Dr Saadi Khochbin, Eric Gilson et François Berger, la première approche transcriptomique dédiée à l'analyse des séquences répétées, la RepChip. Cet outil a été validé grâce à la réponse cellulaire au stress.

Grâce à la RepChip, j'ai pu identifier deux mécanismes indépendants impliqués dans l'expression des séquences répétées péricentromériques : la réponse au stress à travers le facteur de transcription HSF1 et la déméthylation de l'ADN, bien que ce dernier cas semble répondre à des mécanismes complexes. De plus, j'ai mis en évidence la présence d'une activité transcriptionnelle au sein des séquences répétées de type alpha satellites. Finalement, des premiers éléments d'analyse indiquent une dérégulation de l'expression des séquences centromériques et péricentromériques dans certains cas de cancer.

Mais quelle pourrait être la fonction de ces ARNnc du génome humain ? Si cette transcription peut être perçue par certains comme une transcription parasite, inutile, d'autres pensent au contraire que ces ARN pourraient être des acteurs épigénétiques participant à la structure de l'hétérochromatine mais également à la régulation de l'expression des gènes. Cette hypothèse est étayée par le fait que la quantité d'ARNnc présente dans les cellules est proportionnelle à la complexité des organismes. De plus,



de nombreuses études s'accordent pour dire que même si la plupart de l'ADN humain, n'est pas codant, il est tout de même transcrit dans une grande majorité (Birney, Stamatoyannopoulos et al. 2007). Enfin, de plus en plus d'études décrivent de nouvelles fonctions assignées aux ARNnc.

Comme nous l'avons vu en introduction, certains longs ARNnc, tels que Xist ou Air, sont impliqués dans la répression transcriptionnelle des gènes, à travers des mécanismes de compensation de dose ou d'empreinte parentale. Ils favoriseraient le recrutement de complexes répresseurs permettant d'instaurer un statut épigénétique particulier. Des ARNnc sont également impliqués dans le développement en régulant par exemple l'expression des gènes HOX. Les gènes HOX participent à la mise en place de l'identité cellulaire le long de l'axe antéro-postérieur. La régulation de leur expression spatio-temporelle est indispensable au bon déroulement du développement. (Krumlauf 1994). Il a été récemment montré qu'un ARNnc était impliqué dans le contrôle de la transcription des gènes HOX. L'ARNnc HOTAIR (Hox Antisens Intergenic RNA) est transcrit à partir d'une séquence d'ADN intergénique localisée au niveau du complexe HOX-C sur le chromosome 12 (Rinn, Kertesz et al. 2007). HOTAIR, en recrutant le complexe répresseur polycomb qui contient l'HMT EZH2, permet ainsi la répression transcriptionnelle en trans des gènes du complexe HOX-D localisés sur le chromosome 2. Les pseudogènes peuvent également être transcrits en ARNnc. Ces derniers peuvent être impliqués dans la régulation de l'expression du gène codant homologue, en permettant notamment la stabilisation de son ARNm (Piehler, Hellum et al. 2008). Des petits ARNnc jouent également un rôle dans le contrôle de la transcription des gènes. C'est notamment le cas des miRNA dont certains sont transcrits à partir d'introns (Berezikov, Chung et al. 2007). En effet, ces miRNA d'environ 22 nucléotides de long peuvent réguler l'expression des gènes au niveau transcriptionnel et post transcriptionnel (Makeyev and Maniatis 2008; Place, Li et al. 2008). Certains de ces miRNA sont exprimés exclusivement dans un tissu particulier suggérant qu'ils pourraient contribuer à l'établissement et/ou au maintien d'une identité cellulaire. Enfin, la dérégulation de l'expression des miRNA peut être associée à certaines pathologies telles que la leucémie (Calin, Dumitru et al. 2002). Les ARNnc issus des séquences répétées dispersées peuvent également avoir un rôle sur l'expression des gènes. Il a en effet été récemment montré que les séquences Alu, qui appartiennent à la famille d'éléments répétés dispersés SINE, sont transcrites sous l'effet d'un choc thermique. Ces ARNnc se lient directement à l'ARN polymérase II, et répriment ainsi la transcription des gènes. Ces ARNnc Alu participeraient ainsi au mécanisme de répression transcriptionnelle qui a lieu au cours de la réponse au stress (Mariner, Walters et al. 2008).

Qu'en est-il de la fonction des ARNnc issus des séquences répétées en tandem ? Le fait que l'orientation de l'expression de ces ARNnc soit parfaitement contrôlée (en sens dans le cas de la réponse au stress, en antisens dans le testicule) ainsi que l'implication de certains facteurs de transcription dans cette transcription (HSF1, TonEBP) constituent des arguments forts en faveur d'un contrôle rigoureux de l'expression des séquences répétées et suggèrent donc un rôle précis pour ces ARNnc. Néanmoins, à l'heure actuelle aucune fonction n'a encore été formellement attribuée aux ARNnc issus des séquences centromériques et péri-centromériques humaines.

Depuis la découverte chez *S.pombe* de l'implication d'ARNnc dans la structure et le maintien de l'hétérochromatine, beaucoup d'études ont été réalisées dans le but de déterminer si cette même fonction était conservée chez l'Homme. Cependant, il n'y a aucune preuve directe de l'existence de petits ARN issus des séquences péri-centromériques, à l'exception des expériences réalisées dans les cellules hybrides Homme-poulet (Fukagawa, Nogami et al. 2004). En dépit d'une conservation générale de l'organisation des centromères et des péri-centromères au cours de l'évolution, certains mécanismes impliqués dans la formation de l'hétérochromatine ne sont pas conservés. C'est notamment le cas de la méthylation de l'ADN, absente chez *S.pombe*, ou du système RDRP dont aucun homologue n'a été retrouvé chez les mammifères. Le fait qu'aucun petit ARNnc issus des régions centromériques et/ou péri-centromériques n'ait été mis en évidence chez l'Homme ne signifie pas pour autant que la mise en place et/ou le maintien de l'hétérochromatine soient indépendants de ces ARNnc. Ces ARNnc pourraient en effet être uniquement transcrits en réponse à des événements conduisant à une déstabilisation de l'hétérochromatine dans le but de participer à sa reformation. Les ARNnc péri-centromériques, produits au cours du choc thermique, sont également de grande taille et restent localisés à proximité des péri-centromères dont ils sont issus. Ainsi, une des hypothèses concernant leur rôle serait donc de participer à la restructuration de l'hétérochromatine péri-centromérique qui est perturbée au cours de la réponse au stress. Par ailleurs, certaines séquences répétées en tandem ont récemment été décrites comme ayant un rôle dans la structure de l'hétérochromatine. En effet, les séquences télomériques qui appartiennent à la classe des minisatellites sont transcrites en ARNnc. Ceux-ci participent activement au maintien de la structure de l'hétérochromatine télomérique (Azzalin, Reichenbach et al. 2007; Schoeftner and Blasco 2008). A l'image des ARNnc ribosomiques ou de transfert (ARNr et ARNt), les ARNnc issus des séquences satellites péri-centromériques pourraient également intervenir dans la régulation de l'expression des gènes. Un seul exemple est décrit à l'heure actuelle chez l'Homme. Il s'agit des séquences satellites 3 du locus Yq12 qui sont impliquées dans des mécanismes de trans-épissage (Jehan, Vallinayagam et al. 2007). Ainsi, on peut tout à fait imaginer que les ARNnc induits au

cours de la réponse au stress participent également à la régulation de l'expression des gènes en permettant la séquestration des facteurs de transcription et d'épissage au sein des granules de stress.

Les fonctions des ARNnc sont encore loin d'être toutes élucidées. Cependant, de nombreux ARNnc sont exprimés de manière spécifique dans certaines pathologies, telles que des maladies cardiaques (Ishii, Ozaki et al. 2006). Ainsi, ils pourraient être utilisés comme marqueurs de diagnostic ou de pronostic (Lin, Maeda et al. 2007; Perez, Hoage et al. 2008). De plus, la découverte de l'existence de ces ARNnc et de certaines de leurs fonctions a contribué à la modification du Dogme Central établi par Crick « un gène - une protéine - une fonction » dans lequel l'ARN n'avait qu'un rôle de messenger. Ainsi le nouveau dogme intègre l'idée selon laquelle les ARN, et en particulier les ARNnc, peuvent également assurer certaines fonctions au sein de l'organisme. Dans ce contexte, quelle utilisation doit-on faire du terme « gène » qui sous-entend que le produit qu'il donne est fonctionnel ? Doit-on l'appliquer aux séquences non codantes ? Même si cette idée n'est pas à l'ordre du jour, il serait grand temps de rendre ses lettres de noblesse à cet ADN tant décrié.

Dans l'ère post-génomique, cet ADN non codant, longtemps considéré comme un déchet, est au centre de l'attention d'un grand nombre de chercheurs. Il est probable qu'ils soient un jour récompensés de leur efforts et que cet ADN non codant perde définitivement son attribut d'ADN poubelle pour se révéler dans le futur comme l'une des plus belles « merveilles » que le monde du vivant ait engendrée...



# Références

- Abad, J. P., M. Agudo, et al. (2000). "Pericentromeric regions containing 1.688 satellite DNA sequences show anti-kinetochore antibody staining in prometaphase chromosomes of *Drosophila melanogaster*." Mol Gen Genet **264**(4): 371-7.
- Abad, J. P., M. Carmena, et al. (1992). "Dodeca satellite: a conserved G+C-rich satellite from the centromeric heterochromatin of *Drosophila melanogaster*." Proc Natl Acad Sci U S A **89**(10): 4663-7.
- Abad, J. P. and A. Villasante (2000). "Searching for a common centromeric structural motif: *Drosophila* centromeric satellite DNAs show propensity to form telomeric-like unusual DNA structures." Genetica **109**(1-2): 71-5.
- Adams, M. D., S. E. Celniker, et al. (2000). "The genome sequence of *Drosophila melanogaster*." Science **287**(5461): 2185-95.
- Agelopoulos, M. and D. Thanos (2006). "Epigenetic determination of a cell-specific gene expression program by ATF-2 and the histone variant macroH2A." Embo J **25**(20): 4843-53.
- Agresti, A., R. Meneveri, et al. (1989). "Linkage in human heterochromatin between highly divergent Sau3A repeats and a new family of repeated DNA sequences (HaeIII family)." J Mol Biol **205**(4): 625-31.
- Agresti, A., G. Rainaldi, et al. (1987). "Chromosomal location by in situ hybridization of the human Sau3A family of DNA repeats." Hum Genet **75**(4): 326-32.
- Ahmad, K. and S. Henikoff (2001). "Centromeres are specialized replication domains in heterochromatin." J Cell Biol **153**(1): 101-10.
- Alastalo, T. P., M. Hellesuo, et al. (2003). "Formation of nuclear stress granules involves HSF2 and coincides with the nucleolar localization of Hsp70." J Cell Sci **116**(Pt 17): 3557-70.
- Alexiadis, V., M. E. Ballestas, et al. (2007). "RNAPol-ChIP analysis of transcription from FSHD-linked tandem repeats and satellite DNA." Biochim Biophys Acta **1769**(1): 29-40.
- Allfrey, V. G., R. Faulkner, et al. (1964). "Acetylation And Methylation Of Histones And Their Possible Role In The Regulation Of Rna Synthesis." Proc Natl Acad Sci U S A **51**: 786-94.
- Almeida, L. G., N. J. Sakabe, et al. (2008). "CTdatabase: a knowledge-base of high-throughput and curated data on cancer-testis antigens." Nucleic Acids Res.
- Almeida, R. and R. C. Allshire (2005). "RNA silencing and genome regulation." Trends Cell Biol **15**(5): 251-8.
- Amaral, P. P. and J. S. Mattick (2008). "Noncoding RNA in development." Mamm Genome.
- Amor, D. J., P. Kalitsis, et al. (2004). "Building the centromere: from foundation proteins to 3D organization." Trends Cell Biol **14**(7): 359-68.
- Angelov, D., A. Verdel, et al. (2004). "SWI/SNF remodeling and p300-dependent transcription of histone variant H2ABbd nucleosomal arrays." Embo J **23**(19): 3815-24.

- Annunziato, A. T., L. L. Frado, et al. (1988). "Treatment with sodium butyrate inhibits the complete condensation of interphase chromatin." *Chromosoma* **96**(2): 132-8.
- Antonacci, R., R. Marzella, et al. (1995). "A panel of subchromosomal painting libraries representing over 300 regions of the human genome." *Cytogenet Cell Genet* **68**(1-2): 25-32.
- Attwood, J. T., R. L. Yung, et al. (2002). "DNA methylation and the regulation of gene transcription." *Cell Mol Life Sci* **59**(2): 241-57.
- Avalos, J. L., K. M. Bever, et al. (2005). "Mechanism of sirtuin inhibition by nicotinamide: altering the NAD(+) cosubstrate specificity of a Sir2 enzyme." *Mol Cell* **17**(6): 855-68.
- Avner, P. and E. Heard (2001). "X-chromosome inactivation: counting, choice and initiation." *Nat Rev Genet* **2**(1): 59-67.
- Azzalin, C. M., P. Reichenbach, et al. (2007). "Telomeric repeat containing RNA and RNA surveillance factors at mammalian chromosome ends." *Science* **318**(5851): 798-801.
- Bachand, F. (2007). "Protein arginine methyltransferases: from unicellular eukaryotes to humans." *Eukaryot Cell* **6**(6): 889-98.
- Bacolla, A., S. Pradhan, et al. (1999). "Recombinant human DNA (cytosine-5) methyltransferase. II. Steady-state kinetics reveal allosteric activation by methylated dna." *J Biol Chem* **274**(46): 33011-9.
- Bannister, A. J., R. Schneider, et al. (2002). "Histone methylation: dynamic or static?" *Cell* **109**(7): 801-6.
- Bannister, A. J., P. Zegerman, et al. (2001). "Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain." *Nature* **410**(6824): 120-4.
- Barr, M. L. and E. G. Bertram (1949). "A morphological distinction between neurones of the male and female, and the behaviour of the nucleolar satellite during accelerated nucleoprotein synthesis." *Nature* **163**(4148): 676.
- Baum, M. and L. Clarke (2000). "Fission yeast homologs of human CENP-B have redundant functions affecting cell growth and chromosome segregation." *Mol Cell Biol* **20**(8): 2852-64.
- Baum, M., V. K. Ngan, et al. (1994). "The centromeric K-type repeat and the central core are together sufficient to establish a functional *Schizosaccharomyces pombe* centromere." *Mol Biol Cell* **5**(7): 747-61.
- Baxter, J., S. Sauer, et al. (2004). "Histone hypomethylation is an indicator of epigenetic plasticity in quiescent lymphocytes." *Embo J* **23**(22): 4462-72.
- Beard, C., E. Li, et al. (1995). "Loss of methylation activates Xist in somatic but not in embryonic cells." *Genes Dev* **9**(19): 2325-34.
- Becker, P. B. and W. Horz (2002). "ATP-dependent nucleosome remodeling." *Annu Rev Biochem* **71**: 247-73.
- Berezikov, E., W. J. Chung, et al. (2007). "Mammalian mirtron genes." *Mol Cell* **28**(2): 328-36.
- Bernard, P., J. F. Maure, et al. (2001). "Requirement of heterochromatin for cohesion at centromeres." *Science* **294**(5551): 2539-42.
- Bernstein, E., A. A. Caudy, et al. (2001). "Role for a bidentate ribonuclease in the initiation step of RNA interference." *Nature* **409**(6818): 363-6.

- Bertucci, F., K. Bernard, et al. (1999). "Sensitivity issues in DNA array-based expression measurements and performance of nylon microarrays for small samples." Hum Mol Genet **8**(9): 1715-22.
- Bestor, T. H. (1988). "Cloning of a mammalian DNA methyltransferase." Gene **74**(1): 9-12.
- Bestor, T. H. (2000). "The DNA methyltransferases of mammals." Hum Mol Genet **9**(16): 2395-402.
- Biamonti, G. (2004). "Nuclear stress bodies: a heterochromatin affair?" Nat Rev Mol Cell Biol **5**(6): 493-8.
- Bianchi, A. and D. Shore (2008). "How telomerase reaches its end: mechanism of telomerase regulation by the telomeric complex." Mol Cell **31**(2): 153-65.
- Billy, E., V. Brondani, et al. (2001). "Specific interference with gene expression induced by long, double-stranded RNA in mouse embryonal teratocarcinoma cell lines." Proc Natl Acad Sci U S A **98**(25): 14428-33.
- Bird, A. P. and A. P. Wolffe (1999). "Methylation-induced repression--belts, braces, and chromatin." Cell **99**(5): 451-4.
- Birney, E., J. A. Stamatoyannopoulos, et al. (2007). "Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project." Nature **447**(7146): 799-816.
- Blower, M. D. and G. H. Karpen (2001). "The role of Drosophila CID in kinetochore formation, cell-cycle progression and heterochromatin interactions." Nat Cell Biol **3**(8): 730-9.
- Blower, M. D., B. A. Sullivan, et al. (2002). "Conserved organization of centromeric chromatin in flies and humans." Dev Cell **2**(3): 319-30.
- Bobrow, M., K. Madan, et al. (1972). "Staining of some specific regions of human chromosomes, particularly the secondary constriction of No. 9." Nat New Biol **238**(82): 122-4.
- Bonaccorsi, S. and A. Lohe (1991). "Fine mapping of satellite DNA sequences along the Y chromosome of Drosophila melanogaster: relationships between satellite sequences and fertility factors." Genetics **129**(1): 177-89.
- Boulard, M., P. Bouvet, et al. (2007). "Histone variant nucleosomes: structure, function and implication in disease." Subcell Biochem **41**: 71-89.
- Boulard, M., T. Gautier, et al. (2006). "The NH2 tail of the novel histone variant H2BFWT exhibits properties distinct from conventional H2B with respect to the assembly of mitotic chromosomes." Mol Cell Biol **26**(4): 1518-26.
- Bourc'his, D., G. L. Xu, et al. (2001). "Dnmt3L and the establishment of maternal genomic imprints." Science **294**(5551): 2536-9.
- Bouzinba-Segard, H., A. Guais, et al. (2006). "Accumulation of small murine minor satellite transcripts leads to impaired centromeric architecture and function." Proc Natl Acad Sci U S A **103**(23): 8709-14.
- Bowater, R. P. and R. D. Wells (2001). "The intrinsically unstable life of DNA triplet repeats associated with human hereditary disorders." Prog Nucleic Acid Res Mol Biol **66**: 159-202.
- Boyer, L. A., R. R. Latek, et al. (2004). "The SANT domain: a unique histone-tail-binding module?" Nat Rev Mol Cell Biol **5**(2): 158-63.
- Brock, M. V., J. G. Herman, et al. (2007). "Cancer as a manifestation of aberrant chromatin structure." Cancer J **13**(1): 3-8.

- Brockdorff, N., A. Ashworth, et al. (1991). "Conservation of position and exclusive expression of mouse Xist from the inactive X chromosome." Nature **351**(6324): 329-31.
- Brown, C. J., A. Ballabio, et al. (1991). "A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome." Nature **349**(6304): 38-44.
- Brown, S. W. and H. S. Chandra (1973). "Inactivation system of the mammalian X chromosome." Proc Natl Acad Sci U S A **70**(1): 195-9.
- Buhler, M. and D. Moazed (2007). "Transcription and RNAi in heterochromatic gene silencing." Nat Struct Mol Biol **14**(11): 1041-1048.
- Burel, C., V. Mezger, et al. (1992). "Mammalian heat shock protein families. Expression and functions." Experientia **48**(7): 629-34.
- Cairns, B. R. (2004). "Around the world of DNA damage INO80 days." Cell **119**(6): 733-5.
- Calin, G. A., C. D. Dumitru, et al. (2002). "Frequent deletions and down-regulation of micro- RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia." Proc Natl Acad Sci U S A **99**(24): 15524-9.
- Cam, H. P., T. Sugiyama, et al. (2005). "Comprehensive analysis of heterochromatin- and RNAi-mediated epigenetic control of the fission yeast genome." Nat Genet **37**(8): 809-19.
- Celeste, A., O. Fernandez-Capetillo, et al. (2003). "Histone H2AX phosphorylation is dispensable for the initial recognition of DNA breaks." Nat Cell Biol **5**(7): 675-9.
- Cen, H., S. Zheng, et al. (2004). "Induction of HSF1 expression is associated with sporadic colorectal cancer." World J Gastroenterol **10**(21): 3122-6.
- Chadwick, B. P. and H. F. Willard (2001). "A novel chromatin protein, distantly related to histone H2A, is largely excluded from the inactive X chromosome." J Cell Biol **152**(2): 375-84.
- Chen, E. S., K. Zhang, et al. (2008). "Cell cycle control of centromeric repeat transcription and heterochromatin assembly." Nature **451**(7179): 734-7.
- Chendrimada, T. P., R. I. Gregory, et al. (2005). "TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing." Nature **436**(7051): 740-4.
- Chikashige, Y., N. Kinoshita, et al. (1989). "Composite motifs and repeat symmetry in *S. pombe* centromeres: direct analysis by integration of NotI restriction sites." Cell **57**(5): 739-51.
- Chiodi, I., M. Biggiogera, et al. (2000). "Structure and dynamics of hnRNP-labelled nuclear bodies induced by stress treatments." J Cell Sci **113** (Pt 22): 4043-53.
- Cho, K. S., L. I. Elizondo, et al. (2004). "Advances in chromatin remodeling and human disease." Curr Opin Genet Dev **14**(3): 308-15.
- Choo, K. H. (1997). "Centromere DNA dynamics: latent centromeres and neocentromere formation." Am J Hum Genet **61**(6): 1225-33.
- Choo, K. H. (2001). "Domain organization at the centromere and neocentromere." Dev Cell **1**(2): 165-77.
- Chow, C. M., A. Georgiou, et al. (2005). "Variant histone H3.3 marks promoters of transcriptionally active genes during mammalian cell division." EMBO Rep **6**(4): 354-60.
- Christians, E., A. A. Davis, et al. (2000). "Maternal effect of Hsf1 on reproductive success." Nature **407**(6805): 693-4.

- Churikov, D., J. Siino, et al. (2004). "Novel human testis-specific histone H2B encoded by the interrupted gene on the X chromosome." Genomics **84**(4): 745-56.
- Clarke, L. and M. P. Baum (1990). "Functional analysis of a centromere from fission yeast: a role for centromere-specific repeated DNA sequences." Mol Cell Biol **10**(5): 1863-72.
- Clarke, L. and J. Carbon (1985). "The structure and function of yeast centromeres." Annu Rev Genet **19**: 29-55.
- Clos, J., J. T. Westwood, et al. (1990). "Molecular cloning and expression of a hexameric Drosophila heat shock factor subject to negative regulation." Cell **63**(5): 1085-97.
- Cooper, K. F., R. B. Fisher, et al. (1992). "Structure of the pericentric long arm region of the human Y chromosome." J Mol Biol **228**(2): 421-32.
- Corey, L. L., C. S. Weirich, et al. (2003). "Localized recruitment of a chromatin-remodeling activity by an activator in vivo drives transcriptional elongation." Genes Dev **17**(11): 1392-401.
- Corneo, G., E. Ginelli, et al. (1967). "A satellite DNA isolated from human tissues." J Mol Biol **23**(3): 619-22.
- Corneo, G., E. Ginelli, et al. (1968). "Isolation of the complementary strands of a human satellite DNA." J Mol Biol **33**(1): 331-5.
- Corneo, G., E. Ginelli, et al. (1970). "Repeated sequences in human DNA." J Mol Biol **48**(2): 319-27.
- Corneo, G., E. Ginelli, et al. (1971). "Renaturation properties and localization in heterochromatin of human satellite DNA's." Biochim Biophys Acta **247**(4): 528-34.
- Costanzi, C. and J. R. Pehrson (1998). "Histone macroH2A1 is concentrated in the inactive X chromosome of female mammals." Nature **393**(6685): 599-601.
- Cotto, J., S. Fox, et al. (1997). "HSF1 granules: a novel stress-induced nuclear compartment of human cells." J Cell Sci **110** (Pt 23): 2925-34.
- Craig, J. M., W. C. Earnshaw, et al. (1999). "Mammalian centromeres: DNA sequence, protein composition, and role in cell cycle progression." Exp Cell Res **246**(2): 249-62.
- Craig, J. M., L. H. Wong, et al. (2003). "Centromeric chromatin pliability and memory at a human neocentromere." Embo J **22**(10): 2495-504.
- Cullen, B. R. (2004). "Derivation and function of small interfering RNAs and microRNAs." Virus Res **102**(1): 3-9.
- Cuthbert, G. L., S. Daujat, et al. (2004). "Histone deimination antagonizes arginine methylation." Cell **118**(5): 545-53.
- Davie, J. R. (2003). "Inhibition of histone deacetylase activity by butyrate." J Nutr **133**(7 Suppl): 2485S-2493S.
- de Capoa, A., F. Menendez, et al. (1996). "Cytological evidence for 5-azacytidine-induced demethylation of the heterochromatic regions of human chromosomes." Chromosome Res **4**(4): 271-6.
- de la Cruz, X., S. Lois, et al. (2005). "Do protein motifs read the histone code?" Bioessays **27**(2): 164-75.
- de la Serna, I. L., Y. Ohkawa, et al. (2006). "Chromatin remodelling in mammalian differentiation: lessons from ATP-dependent remodellers." Nat Rev Genet **7**(6): 461-73.



- Denegri, M., I. Chiodi, et al. (2001). "Stress-induced nuclear bodies are sites of accumulation of pre-mRNA processing factors." *Mol Biol Cell* **12**(11): 3502-14.
- Denegri, M., D. Moralli, et al. (2002). "Human chromosomes 9, 12, and 15 contain the nucleation sites of stress-induced nuclear bodies." *Mol Biol Cell* **13**(6): 2069-79.
- Dover, G. (1980). "Ignorant DNA?" *Nature* **285**(5767): 618-20.
- Doyen, C. M., W. An, et al. (2006). "Mechanism of polymerase II transcription repression by the histone variant macroH2A." *Mol Cell Biol* **26**(3): 1156-64.
- du Sart, D., M. R. Cancilla, et al. (1997). "A functional neo-centromere formed through activation of a latent human centromere and consisting of non-alpha-satellite DNA." *Nat Genet* **16**(2): 144-53.
- Ebert, A., S. Lein, et al. (2006). "Histone modification and the control of heterochromatic gene silencing in Drosophila." *Chromosome Res* **14**(4): 377-92.
- Ehrlich, M. (2002). "DNA methylation in cancer: too much, but also too little." *Oncogene* **21**(35): 5400-13.
- Ehrlich, M., C. Sanchez, et al. (2008). "ICF, an immunodeficiency syndrome: DNA methyltransferase 3B involvement, chromosome anomalies, and gene dysregulation." *Autoimmunity* **41**(4): 253-71.
- Eissenberg, J. C., T. C. James, et al. (1990). "Mutation in a heterochromatin-specific chromosomal protein is associated with suppression of position-effect variegation in Drosophila melanogaster." *Proc Natl Acad Sci U S A* **87**(24): 9923-7.
- Ekwall, K. (2007). "'Arc' escorts siRNAs in heterochromatin assembly." *Nat Struct Mol Biol* **14**(3): 178-9.
- Ekwall, K., J. P. Javerzat, et al. (1995). "The chromodomain protein Swi6: a key component at fission yeast centromeres." *Science* **269**(5229): 1429-31.
- Elgin, S. C. (1996). "Heterochromatin and gene regulation in Drosophila." *Curr Opin Genet Dev* **6**(2): 193-202.
- Enukashvily, N. I., R. Donev, et al. (2007). "Human chromosome 1 satellite 3 DNA is decondensed, demethylated and transcribed in senescent cells and in A431 epithelial carcinoma cells." *Cytogenet Genome Res* **118**(1): 42-54.
- Fagard, M., S. Boutet, et al. (2000). "AGO1, QDE-2, and RDE-1 are related proteins required for post-transcriptional gene silencing in plants, quelling in fungi, and RNA interference in animals." *Proc Natl Acad Sci U S A* **97**(21): 11650-4.
- Feinberg, A. P. and B. Tycko (2004). "The history of cancer epigenetics." *Nat Rev Cancer* **4**(2): 143-53.
- Filion, G. J., S. Zhenilo, et al. (2006). "A family of human zinc finger proteins that bind methylated DNA and repress transcription." *Mol Cell Biol* **26**(1): 169-81.
- Fire, A., S. Xu, et al. (1998). "Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans." *Nature* **391**(6669): 806-11.
- Fischle, W., B. S. Tseng, et al. (2005). "Regulation of HP1-chromatin binding by histone H3 methylation and phosphorylation." *Nature* **438**(7071): 1116-22.
- Fisher, A. G. and M. Merckenschlager (2002). "Gene silencing, cell fate and nuclear organisation." *Curr Opin Genet Dev* **12**(2): 193-7.
- Fitzgerald-Hayes, M., L. Clarke, et al. (1982). "Nucleotide sequence comparisons and functional analysis of yeast centromere DNAs." *Cell* **29**(1): 235-44.

- Folco, H. D., A. L. Pidoux, et al. (2008). "Heterochromatin and RNAi are required to establish CENP-A chromatin at centromeres." Science **319**(5859): 94-7.
- Fraga, M. F. and M. Esteller (2007). "Epigenetics and aging: the targets and the marks." Trends Genet **23**(8): 413-8.
- Franco, N., J. Lamartine, et al. (2005). "Low-dose exposure to gamma rays induces specific gene regulations in normal human keratinocytes." Radiat Res **163**(6): 623-35.
- Freimer, N. B. and M. Slatkin (1996). "Microsatellites: evolution and mutational processes." Ciba Found Symp **197**: 51-67; discussion 67-72.
- Frenster, J. H., V. G. Allfrey, et al. (1963). "Repressed And Active Chromatin Isolated From Interphase Lymphocytes." Proc Natl Acad Sci U S A **50**: 1026-32.
- Frommer, M., C. Paul, et al. (1988). "Localisation of satellite DNA sequences on human metaphase chromosomes using bromodeoxyuridine-labelled probes." Chromosoma **97**(1): 11-8.
- Frommer, M., J. Prosser, et al. (1982). "Simple repeated sequences in human satellite DNA." Nucleic Acids Res **10**(2): 547-63.
- Fukagawa, T. (2004). "Centromere DNA, proteins and kinetochore assembly in vertebrate cells." Chromosome Res **12**(6): 557-67.
- Fukagawa, T., Y. Mikami, et al. (2001). "CENP-H, a constitutive centromere component, is required for centromere targeting of CENP-C in vertebrate cells." Embo J **20**(16): 4603-17.
- Fukagawa, T., M. Nogami, et al. (2004). "Dicer is essential for formation of the heterochromatin structure in vertebrate cells." Nat Cell Biol **6**(8): 784-91.
- Fukagawa, T., C. Pendon, et al. (1999). "CENP-C is necessary but not sufficient to induce formation of a functional centromere." Embo J **18**(15): 4196-209.
- Gagne, R., C. Laberge, et al. (1973). "[Cytological aspect and intranuclear localization of the heterochromatic segments of C9 chromosomes in man]." Chromosoma **41**(2): 159-66.
- Gall, J. G., E. H. Cohen, et al. (1971). "Repetitive DNA sequences in drosophila." Chromosoma **33**(3): 319-44.
- Gao, L., M. A. Cueto, et al. (2002). "Cloning and functional characterization of HDAC11, a novel member of the human histone deacetylase family." J Biol Chem **277**(28): 25748-55.
- Geiman, T. M., U. T. Sankpal, et al. (2004). "DNMT3B interacts with hSNF2H chromatin remodeling enzyme, HDACs 1 and 2, and components of the histone methylation system." Biochem Biophys Res Commun **318**(2): 544-55.
- Gerbi, S. A. and A. K. Bielinsky (2002). "DNA replication and chromatin." Curr Opin Genet Dev **12**(2): 243-8.
- Gjerstorff, M. F., L. Harkness, et al. (2008). "Distinct GAGE and MAGE-A expression during early human development indicate specific roles in lineage differentiation." Hum Reprod **23**(10): 2194-201.
- Goll, M. G., F. Kirpekar, et al. (2006). "Methylation of tRNA<sup>Asp</sup> by the DNA methyltransferase homolog Dnmt2." Science **311**(5759): 395-8.
- Gosden, J. R., A. R. Mitchell, et al. (1975). "The location of four human satellite DNAs on human chromosomes." Exp Cell Res **92**(1): 148-58.
- Gottesman, S. (2005). "Micros for microbes: non-coding regulatory RNAs in bacteria." Trends Genet **21**(7): 399-404.

- Govin, J., E. Escoffier, et al. (2007). "Pericentric heterochromatin reprogramming by new histone variants during mouse spermiogenesis." J Cell Biol **176**(3): 283-94.
- Grady, D. L., R. L. Ratliff, et al. (1992). "Highly conserved repetitive DNA sequences are present at human centromeres." Proc Natl Acad Sci U S A **89**(5): 1695-9.
- Grau, E., S. Oltra, et al. (2008). "MAGE-A1 expression is associated with good prognosis in neuroblastoma tumors." J Cancer Res Clin Oncol.
- Gregory, T. R. (2005). "The C-value enigma in plants and animals: a review of parallels and an appeal for partnership." Ann Bot (Lond) **95**(1): 133-46.
- Gregory, T. R. (2005). "Synergy between sequence and size in large-scale genomics." Nat Rev Genet **6**(9): 699-708.
- Greider, C. W. and E. H. Blackburn (1985). "Identification of a specific telomere terminal transferase activity in Tetrahymena extracts." Cell **43**(2 Pt 1): 405-13.
- Grewal, S. I. and S. C. Elgin (2002). "Heterochromatin: new possibilities for the inheritance of structure." Curr Opin Genet Dev **12**(2): 178-87.
- Grewal, S. I. and S. C. Elgin (2007). "Transcription and RNA interference in the formation of heterochromatin." Nature **447**(7143): 399-406.
- Griffith, J. D., L. Comeau, et al. (1999). "Mammalian telomeres end in a large duplex loop." Cell **97**(4): 503-14.
- Gross, S. R. and T. G. Kinzy (2005). "Translation elongation factor 1A is essential for regulation of the actin cytoskeleton and cell morphology." Nat Struct Mol Biol **12**(9): 772-8.
- Guenatri, M., D. Bailly, et al. (2004). "Mouse centric and pericentric satellite repeats form distinct functional heterochromatin." J Cell Biol **166**(4): 493-505.
- Guettouche, T., F. Boellmann, et al. (2005). "Analysis of phosphorylation of human heat shock factor 1 in cells experiencing a stress." BMC Biochem **6**: 4.
- Guillemette, B., A. R. Bataille, et al. (2005). "Variant histone H2A.Z is globally localized to the promoters of inactive yeast genes and regulates nucleosome positioning." PLoS Biol **3**(12): e384.
- Haaf, T. and M. Schmid (2000). "Experimental condensation inhibition in constitutive and facultative heterochromatin of mammalian chromosomes." Cytogenet Cell Genet **91**(1-4): 113-23.
- Hadnagy, A., R. Beaulieu, et al. (2008). "Histone tail modifications and noncanonical functions of histones: perspectives in cancer epigenetics." Mol Cancer Ther **7**(4): 740-8.
- Hahnenberger, K. M., J. Carbon, et al. (1991). "Identification of DNA regions required for mitotic and meiotic functions within the centromere of *Schizosaccharomyces pombe* chromosome I." Mol Cell Biol **11**(4): 2206-15.
- Hall, I. M., K. Noma, et al. (2003). "RNA interference machinery regulates chromosome dynamics during mitosis and meiosis in fission yeast." Proc Natl Acad Sci U S A **100**(1): 193-8.
- Hall, I. M., G. D. Shankaranarayana, et al. (2002). "Establishment and maintenance of a heterochromatin domain." Science **297**(5590): 2232-7.
- Hammond, S. M., E. Bernstein, et al. (2000). "An RNA-directed nuclease mediates post-transcriptional gene silencing in *Drosophila* cells." Nature **404**(6775): 293-6.

- Hansen, R. S., C. Wijmenga, et al. (1999). "The DNMT3B DNA methyltransferase gene is mutated in the ICF immunodeficiency syndrome." Proc Natl Acad Sci U S A **96**(25): 14412-7.
- He, H. and N. Lehming (2003). "Global effects of histone modifications." Brief Funct Genomic Proteomic **2**(3): 234-43.
- Heard, E. (2005). "Delving into the diversity of facultative heterochromatin: the epigenetics of the inactive X chromosome." Curr Opin Genet Dev **15**(5): 482-9.
- Hediger, F. and S. M. Gasser (2006). "Heterochromatin protein 1: don't judge the book by its cover!" Curr Opin Genet Dev **16**(2): 143-50.
- Hegemann, J. H., J. H. Shero, et al. (1988). "Mutational analysis of centromere DNA from chromosome VI of *Saccharomyces cerevisiae*." Mol Cell Biol **8**(6): 2523-35.
- Hendzel, M. J., Y. Wei, et al. (1997). "Mitosis-specific phosphorylation of histone H3 initiates primarily within pericentromeric heterochromatin during G2 and spreads in an ordered fashion coincident with mitotic chromosome condensation." Chromosoma **106**(6): 348-60.
- Henikoff, S. and T. D. Dreesen (1989). "Trans-inactivation of the *Drosophila* brown gene: evidence for transcriptional repression and somatic pairing dependence." Proc Natl Acad Sci U S A **86**(17): 6704-8.
- Henikoff, S., J. M. Jackson, et al. (1995). "Distance and pairing effects on the brown dominant heterochromatic element in *Drosophila*." Genetics **140**(3): 1007-17.
- Hirota, T., J. J. Lipp, et al. (2005). "Histone H3 serine 10 phosphorylation by Aurora B causes HP1 dissociation from heterochromatin." Nature **438**(7071): 1176-80.
- Hoang, A. T., J. Huang, et al. (2000). "A novel association between the human heat shock transcription factor 1 (HSF1) and prostate adenocarcinoma." Am J Pathol **156**(3): 857-64.
- Hogan, C. and P. Varga-Weisz (2007). "The regulation of ATP-dependent nucleosome remodelling factors." Mutat Res **618**(1-2): 41-51.
- Holliday, R. and J. E. Pugh (1975). "DNA modification mechanisms and gene activity during development." Science **187**(4173): 226-32.
- Hollis, M. and J. Hindley (1988). "Satellite II DNA of human lymphocytes: tandem repeats of a simple sequence element." Nucleic Acids Res **16**(1): 363.
- Holmberg, C. I., V. Hietakangas, et al. (2001). "Phosphorylation of serine 230 promotes inducible transcriptional activity of heat shock factor 1." Embo J **20**(14): 3800-10.
- Holmberg, C. I., S. E. Tran, et al. (2002). "Multisite phosphorylation provides sophisticated regulation of transcription factors." Trends Biochem Sci **27**(12): 619-27.
- Hong, B., P. Reeves, et al. (2001). "Identification of an autoimmune serum containing antibodies against the Barr body." Proc Natl Acad Sci U S A **98**(15): 8703-8.
- Howell, C. Y., T. H. Bestor, et al. (2001). "Genomic imprinting disrupted by a maternal effect mutation in the *Dnmt1* gene." Cell **104**(6): 829-38.
- Howman, E. V., K. J. Fowler, et al. (2000). "Early disruption of centromeric chromatin organization in centromere protein A (*Cenpa*) null mice." Proc Natl Acad Sci U S A **97**(3): 1148-53.

- Hudson, D. F., K. J. Fowler, et al. (1998). "Centromere protein B null mice are mitotically and meiotically normal but have lower body and testis weights." J Cell Biol **141**(2): 309-19.
- Huttenhofer, A., P. Schattner, et al. (2005). "Non-coding RNAs: hope or hype?" Trends Genet **21**(5): 289-97.
- Inoue, A., J. Hyle, et al. (2008). "Perturbation of HP1 localization and chromatin binding ability causes defects in sister-chromatid cohesion." Mutat Res.
- Irelan, J. T., G. I. Gutkin, et al. (2001). "Functional redundancies, distinct localizations and interactions among three fission yeast homologs of centromere protein-B." Genetics **157**(3): 1191-203.
- Ishii, N., K. Ozaki, et al. (2006). "Identification of a novel non-coding RNA, MIAT, that confers risk of myocardial infarction." J Hum Genet **51**(12): 1087-99.
- Izawa, M., V. G. Allefrey, et al. (1963). "The relationship between RNA synthesis and loop structure in lampbrush chromosomes." Proc Natl Acad Sci U S A **49**: 544-51.
- Jacobs, S. A., S. D. Taverna, et al. (2001). "Specificity of the HP1 chromo domain for the methylated N-terminus of histone H3." Embo J **20**(18): 5232-41.
- James, T. C. and S. C. Elgin (1986). "Identification of a nonhistone chromosomal protein associated with heterochromatin in *Drosophila melanogaster* and its gene." Mol Cell Biol **6**(11): 3862-72.
- Jamrich, M., A. L. Greenleaf, et al. (1977). "Localization of RNA polymerase in polytene chromosomes of *Drosophila melanogaster*." Proc Natl Acad Sci U S A **74**(5): 2079-83.
- Jasinska, A. and W. J. Krzyzosiak (2004). "Repetitive sequences that shape the human transcriptome." FEBS Lett **567**(1): 136-41.
- Jedlicka, P., M. A. Mortin, et al. (1997). "Multiple functions of *Drosophila* heat shock transcription factor in vivo." Embo J **16**(9): 2452-62.
- Jehan, Z., S. Vallinayagam, et al. (2007). "Novel noncoding RNA from human Y distal heterochromatic block (Yq12) generates testis-specific chimeric CDC2L2." Genome Res **17**(4): 433-40.
- Jehn, B., R. Niedenthal, et al. (1991). "In vivo analysis of the *Saccharomyces cerevisiae* centromere CDEIII sequence: requirements for mitotic chromosome segregation." Mol Cell Biol **11**(10): 5212-21.
- Jenuwein, T. and C. D. Allis (2001). "Translating the histone code." Science **293**(5532): 1074-80.
- Jeppesen, P., A. Mitchell, et al. (1992). "Antibodies to defined histone epitopes reveal variations in chromatin conformation and underacetylation of centric heterochromatin in human metaphase chromosomes." Chromosoma **101**(5-6): 322-32.
- Johnstone, R. W. (2002). "Histone-deacetylase inhibitors: novel drugs for the treatment of cancer." Nat Rev Drug Discov **1**(4): 287-99.
- Jolly, C., L. Konecny, et al. (2002). "In vivo binding of active heat shock transcription factor 1 to human chromosome 9 heterochromatin during stress." J Cell Biol **156**(5): 775-81.
- Jolly, C. and S. C. Lakhota (2006). "Human sat III and *Drosophila* hsr omega transcripts: a common paradigm for regulation of nuclear RNA processing in stressed cells." Nucleic Acids Res **34**(19): 5508-14.

- Jolly, C., A. Metz, et al. (2004). "Stress-induced transcription of satellite III repeats." J Cell Biol **164**(1): 25-33.
- Jolly, C., R. Morimoto, et al. (1997). "HSF1 transcription factor concentrates in nuclear foci during heat shock: relationship with transcription sites." J Cell Sci **110 (Pt 23)**: 2935-41.
- Jolly, C., Y. Usson, et al. (1999). "Rapid and reversible relocalization of heat shock factor 1 within seconds to nuclear stress granules." Proc Natl Acad Sci U S A **96**(12): 6769-74.
- Jones, K. W., I. F. Purdom, et al. (1974). "The chromosomal localisation of human satellite DNA I." Chromosoma **49**(2): 161-71.
- Kalitsis, P., K. J. Fowler, et al. (1998). "Targeted disruption of mouse centromere protein C gene leads to mitotic disarray and early embryo death." Proc Natl Acad Sci U S A **95**(3): 1136-41.
- Kanellopoulou, C., S. A. Muljo, et al. (2005). "Dicer-deficient mouse embryonic stem cells are defective in differentiation and centromeric silencing." Genes Dev **19**(4): 489-501.
- Kantor, B., K. Makedonski, et al. (2003). "Expression and localization of components of the histone deacetylases multiprotein repressory complexes in the mouse preimplantation embryo." Gene Expr Patterns **3**(6): 697-702.
- Kasten, M., H. Szerlong, et al. (2004). "Tandem bromodomains in the chromatin remodeler RSC recognize acetylated histone H3 Lys14." Embo J **23**(6): 1348-59.
- Kellum, R. and B. M. Alberts (1995). "Heterochromatin protein 1 is required for correct chromosome segregation in Drosophila embryos." J Cell Sci **108 (Pt 4)**: 1419-31.
- Kim, N. W., M. A. Piatyszek, et al. (1994). "Specific association of human telomerase activity with immortal cells and cancer." Science **266**(5193): 2011-5.
- Kim, S. M., D. D. Dubey, et al. (2003). "Early-replicating heterochromatin." Genes Dev **17**(3): 330-5.
- Kloc, A., M. Zaratiegui, et al. (2008). "RNA interference guides histone modification during the S phase of chromosomal replication." Curr Biol **18**(7): 490-5.
- Kloeckener-Gruissem, B., D. R. Betts, et al. (2005). "A new and a reclassified ICF patient without mutations in DNMT3B and its interacting proteins SUMO-1 and UBC9." Am J Med Genet A **136**(1): 31-7.
- Klose, R. J. and A. P. Bird (2006). "Genomic DNA methylation: the mark and its mediators." Trends Biochem Sci **31**(2): 89-97.
- Klose, R. J. and Y. Zhang (2007). "Regulation of histone methylation by demethylimination and demethylation." Nat Rev Mol Cell Biol **8**(4): 307-18.
- Kohlmaier, A., F. Savarese, et al. (2004). "A chromosomal memory triggered by Xist regulates histone methylation in X inactivation." PLoS Biol **2**(7): E171.
- Kondo, T., M. P. Bobek, et al. (2000). "Whole-genome methylation scan in ICF syndrome: hypomethylation of non-satellite DNA repeats D4Z4 and NBL2." Hum Mol Genet **9**(4): 597-604.
- Korber, P. and W. Horz (2004). "SWRred not shaken; mixing the histones." Cell **117**(1): 5-7.
- Kotake, Y., R. Cao, et al. (2007). "pRB family proteins are required for H3K27 trimethylation and Polycomb repression complexes binding to and silencing p16INK4alpha tumor suppressor gene." Genes Dev **21**(1): 49-54.

- Kourmouli, N., P. Jeppesen, et al. (2004). "Heterochromatin and tri-methylated lysine 20 of histone H4 in animals." *J Cell Sci* **117**(Pt 12): 2491-501.
- Kroeger, P. E., K. D. Sarge, et al. (1993). "Mouse heat shock transcription factors 1 and 2 prefer a trimeric binding site but interact differently with the HSP70 heat shock element." *Mol Cell Biol* **13**(6): 3370-83.
- Krumlauf, R. (1994). "Hox genes in vertebrate development." *Cell* **78**(2): 191-201.
- Kubota, T., H. Furuumi, et al. (2004). "ICF syndrome in a girl with DNA hypomethylation but without detectable DNMT3B mutation." *Am J Med Genet A* **129A**(3): 290-3.
- Lachner, M., R. J. O'Sullivan, et al. (2003). "An epigenetic road map for histone lysine methylation." *J Cell Sci* **116**(Pt 11): 2117-24.
- Lander, E. S., L. M. Linton, et al. (2001). "Initial sequencing and analysis of the human genome." *Nature* **409**(6822): 860-921.
- Lee, C., X. Li, et al. (1995). "Human gamma X satellite DNA: an X chromosome specific centromeric DNA sequence." *Chromosoma* **104**(2): 103-12.
- Lee, C., R. Wevrick, et al. (1997). "Human centromeric DNAs." *Hum Genet* **100**(3-4): 291-304.
- Lee, G. J., A. M. Roseman, et al. (1997). "A small heat shock protein stably binds heat-denatured model substrates and can maintain a substrate in a folding-competent state." *Embo J* **16**(3): 659-71.
- Lee, J. T., L. S. Davidow, et al. (1999). "Tsix, a gene antisense to Xist at the X-inactivation centre." *Nat Genet* **21**(4): 400-4.
- Lehnertz, B., Y. Ueda, et al. (2003). "Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin." *Curr Biol* **13**(14): 1192-200.
- Leonhardt, H., A. W. Page, et al. (1992). "A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei." *Cell* **71**(5): 865-73.
- Li, E., T. H. Bestor, et al. (1992). "Targeted mutation of the DNA methyltransferase gene results in embryonic lethality." *Cell* **69**(6): 915-26.
- Li, Y. C., A. B. Korol, et al. (2004). "Microsatellites within genes: structure, function, and evolution." *Mol Biol Evol* **21**(6): 991-1007.
- Lin, C. C., R. Sasi, et al. (1993). "Isolation and identification of a novel tandemly repeated DNA sequence in the centromeric region of human chromosome 8." *Chromosoma* **102**(5): 333-9.
- Lin, R., S. Maeda, et al. (2007). "A large noncoding RNA is a marker for murine hepatocellular carcinomas and a spectrum of human carcinomas." *Oncogene* **26**(6): 851-8.
- Lindquist, S. (1986). "The heat-shock response." *Annu Rev Biochem* **55**: 1151-91.
- Liu, Y., E. J. Oakeley, et al. (1998). "Multiple domains are involved in the targeting of the mouse DNA methyltransferase to the DNA replication foci." *Nucleic Acids Res* **26**(4): 1038-45.
- Lohe, A. R., A. J. Hilliker, et al. (1993). "Mapping simple repeated DNA sequences in heterochromatin of *Drosophila melanogaster*." *Genetics* **134**(4): 1149-74.
- Lomberk, G., D. Bensi, et al. (2006). "Evidence for the existence of an HP1-mediated subcode within the histone code." *Nat Cell Biol* **8**(4): 407-15.

- Lomberk, G., L. Wallrath, et al. (2006). "The Heterochromatin Protein 1 family." Genome Biol **7**(7): 228.
- Losada, A. (2007). "Cohesin regulation: fashionable ways to wear a ring." Chromosoma **116**(4): 321-9.
- Lu, B. Y., P. C. Emtage, et al. (2000). "Heterochromatin protein 1 is required for the normal expression of two heterochromatin genes in Drosophila." Genetics **155**(2): 699-708.
- Lu, J. and D. M. Gilbert (2007). "Proliferation-dependent and cell cycle regulated transcription of mouse pericentric heterochromatin." J Cell Biol **179**(3): 411-21.
- Lu, J. and D. M. Gilbert (2008). "Cell cycle regulated transcription of heterochromatin in mammals vs. fission yeast: functional conservation or coincidence?" Cell Cycle **7**(13): 1907-10.
- Lukusa, T. and J. P. Fryns (2008). "Human chromosome fragility." Biochim Biophys Acta **1779**(1): 3-16.
- Lyon, M. F. (1961). "Gene action in the X-chromosome of the mouse (*Mus musculus* L.)." Nature **190**: 372-3.
- Lyon, M. F. (1962). "Sex chromatin and gene action in the mammalian X-chromosome." Am J Hum Genet **14**: 135-48.
- Maison, C., D. Bailly, et al. (2002). "Higher-order structure in pericentric heterochromatin involves a distinct pattern of histone modification and an RNA component." Nat Genet **30**(3): 329-34.
- Makeyev, E. V. and T. Maniatis (2008). "Multilevel regulation of gene expression by microRNAs." Science **319**(5871): 1789-90.
- Manuelidis, L. and J. C. Wu (1978). "Homology between human and simian repeated DNA." Nature **276**(5683): 92-4.
- Maraschio, P., O. Zuffardi, et al. (1988). "Immunodeficiency, centromeric heterochromatin instability of chromosomes 1, 9, and 16, and facial anomalies: the ICF syndrome." J Med Genet **25**(3): 173-80.
- Mariner, P. D., R. D. Walters, et al. (2008). "Human Alu RNA is a modular transacting repressor of mRNA transcription during heat shock." Mol Cell **29**(4): 499-509.
- Martens, J. H., R. J. O'Sullivan, et al. (2005). "The profile of repeat-associated histone lysine methylation states in the mouse epigenome." Embo J **24**(4): 800-12.
- Martin, M., R. Kettmann, et al. (2007). "Class IIa histone deacetylases: regulating the regulators." Oncogene **26**(37): 5450-67.
- Masumoto, H., H. Masukata, et al. (1989). "A human centromere antigen (CENP-B) interacts with a short specific sequence in alphoid DNA, a human centromeric satellite." J Cell Biol **109**(5): 1963-73.
- Masumoto, H., M. Nakano, et al. (2004). "The role of CENP-B and alpha-satellite DNA: de novo assembly and epigenetic maintenance of human centromeres." Chromosome Res **12**(6): 543-56.
- Mateescu, B., P. England, et al. (2004). "Tethering of HP1 proteins to chromatin is relieved by phosphoacetylation of histone H3." EMBO Rep **5**(5): 490-6.
- Meneveri, R., A. Agresti, et al. (1993). "Molecular organization and chromosomal location of human GC-rich heterochromatic blocks." Gene **123**(2): 227-34.
- Mermoud, J. E., C. Costanzi, et al. (1999). "Histone macroH2A1.2 relocates to the inactive X chromosome after initiation and propagation of X-inactivation." J Cell Biol **147**(7): 1399-408.



- Mertineit, C., J. A. Yoder, et al. (1998). "Sex-specific exons control DNA methyltransferase in mammalian germ cells." Development **125**(5): 889-97.
- Metz, A., J. Soret, et al. (2004). "A key role for stress-induced satellite III transcripts in the relocalization of splicing factors into nuclear stress granules." J Cell Sci **117**(Pt 19): 4551-8.
- Meyne, J., E. H. Goodwin, et al. (1994). "Chromosome localization and orientation of the simple sequence repeat of human satellite I DNA." Chromosoma **103**(2): 99-103.
- Meyne, J., R. L. Ratliff, et al. (1989). "Conservation of the human telomere sequence (TTAGGG)<sub>n</sub> among vertebrates." Proc Natl Acad Sci U S A **86**(18): 7049-53.
- Miller, O. J., W. Schnedl, et al. (1974). "5-Methylcytosine localised in mammalian constitutive heterochromatin." Nature **251**(5476): 636-7.
- Minc, E., J. C. Courvalin, et al. (2000). "HP1gamma associates with euchromatin and heterochromatin in mammalian nuclei and chromosomes." Cytogenet Cell Genet **90**(3-4): 279-84.
- Mirre, C. and A. Stahl (1978). "Peripheral RNA synthesis of fibrillar center in nucleoli of Japanese quail oocytes and somatic cells." J Ultrastruct Res **64**(3): 377-87.
- Mitchell, A. R., R. S. Beauchamp, et al. (1979). "A study of sequence homologies in four satellite DNAs of man." J Mol Biol **135**(1): 127-49.
- Momparler, R. L. (2003). "Cancer epigenetics." Oncogene **22**(42): 6479-83.
- Motamedi, M. R., A. Verdel, et al. (2004). "Two RNAi complexes, RITS and RDRC, physically interact and localize to noncoding centromeric RNAs." Cell **119**(6): 789-802.
- Muchardt, C., M. Guilleme, et al. (2002). "Coordinated methyl and RNA binding is required for heterochromatin localization of mammalian HP1alpha." EMBO Rep **3**(10): 975-81.
- Muchardt, C., J. C. Reyes, et al. (1996). "The hbrm and BRG-1 proteins, components of the human SNF/SWI complex, are phosphorylated and excluded from the condensed chromosomes during mitosis." Embo J **15**(13): 3394-402.
- Murphy, T. D. and G. H. Karpen (1995). "Localization of centromere function in a Drosophila minichromosome." Cell **82**(4): 599-609.
- Murray, J. I., M. L. Whitfield, et al. (2004). "Diverse and specific gene expression responses to stresses in cultured human cells." Mol Biol Cell **15**(5): 2361-74.
- Musacchio, A. and E. D. Salmon (2007). "The spindle-assembly checkpoint in space and time." Nat Rev Mol Cell Biol **8**(5): 379-93.
- Nakagawa, H., J. K. Lee, et al. (2002). "Fission yeast CENP-B homologs nucleate centromeric heterochromatin by promoting heterochromatin-specific histone tail modifications." Genes Dev **16**(14): 1766-78.
- Nakai, A., M. Tanabe, et al. (1997). "HSF4, a new member of the human heat shock factor family which lacks properties of a transcriptional activator." Mol Cell Biol **17**(1): 469-81.
- Nakaseko, Y., Y. Adachi, et al. (1986). "Chromosome walking shows a highly homologous repetitive sequence present in all the centromere regions of fission yeast." Embo J **5**(5): 1011-21.
- Nakaseko, Y., G. Goshima, et al. (2001). "M phase-specific kinetochore proteins in fission yeast: microtubule-associating Dis1 and Mtc1 display rapid separation and segregation during anaphase." Curr Biol **11**(8): 537-49.

- Nakaseko, Y., N. Kinoshita, et al. (1987). "A novel sequence common to the centromere regions of *Schizosaccharomyces pombe* chromosomes." Nucleic Acids Res **15**(12): 4705-15.
- Napoli, C., C. Lemieux, et al. (1990). "Introduction of a Chimeric Chalcone Synthase Gene into *Petunia* Results in Reversible Co-Suppression of Homologous Genes in trans." Plant Cell **2**(4): 279-289.
- Narayan, A., W. Ji, et al. (1998). "Hypomethylation of pericentromeric DNA in breast adenocarcinomas." Int J Cancer **77**(6): 833-8.
- Negrutskii, B. S. and A. V. El'skaya (1998). "Eukaryotic translation elongation factor 1 alpha: structure, expression, functions, and possible role in aminoacyl-tRNA channeling." Prog Nucleic Acid Res Mol Biol **60**: 47-78.
- Nguyen, C. T., D. J. Weisenberger, et al. (2002). "Histone H3-lysine 9 methylation is associated with aberrant gene silencing in cancer cells and is rapidly reversed by 5-aza-2'-deoxycytidine." Cancer Res **62**(22): 6456-61.
- Nguyen, V. T., M. Morange, et al. (1989). "Protein denaturation during heat shock and related stress. *Escherichia coli* beta-galactosidase and *Photinus pyralis* luciferase inactivation in mouse cells." J Biol Chem **264**(18): 10487-92.
- Nicolas, E., T. Yamada, et al. (2007). "Distinct roles of HDAC complexes in promoter silencing, antisense suppression and DNA damage protection." Nat Struct Mol Biol **14**(5): 372-80.
- Nishihashi, A., T. Haraguchi, et al. (2002). "CENP-I is essential for centromere function in vertebrate cells." Dev Cell **2**(4): 463-76.
- Noma, K., T. Sugiyama, et al. (2004). "RITS acts in cis to promote RNA interference-mediated transcriptional and post-transcriptional silencing." Nat Genet **36**(11): 1174-80.
- O'Keefe, R. T., S. C. Henderson, et al. (1992). "Dynamic organization of DNA replication in mammalian cell nuclei: spatially and temporally defined replication of chromosome-specific alpha-satellite DNA sequences." J Cell Biol **116**(5): 1095-110.
- Oh, H. J., X. Chen, et al. (1997). "Hsp110 protects heat-denatured proteins and confers cellular thermoresistance." J Biol Chem **272**(50): 31636-40.
- Ohhata, T., Y. Hoki, et al. (2008). "Crucial role of antisense transcription across the Xist promoter in Tsix-mediated Xist chromatin modification." Development **135**(2): 227-35.
- Ohno, S. (1972). "So much "junk" DNA in our genome." Brookhaven Symp Biol **23**: 366-70.
- Ohzeki, J., M. Nakano, et al. (2002). "CENP-B box is required for de novo centromere chromatin assembly on human alphoid DNA." J Cell Biol **159**(5): 765-75.
- Okano, M., D. W. Bell, et al. (1999). "DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development." Cell **99**(3): 247-57.
- Okano, M., S. Xie, et al. (1998). "Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases." Nat Genet **19**(3): 219-20.
- Okano, M., S. Xie, et al. (1998). "Dnmt2 is not required for de novo and maintenance methylation of viral DNA in embryonic stem cells." Nucleic Acids Res **26**(11): 2536-40.
- Ooi, S. K. and T. H. Bestor (2008). "The colorful history of active DNA demethylation." Cell **133**(7): 1145-8.

- Orgel, L. E. and F. H. Crick (1980). "Selfish DNA: the ultimate parasite." Nature **284**(5757): 604-7.
- Pal, S. and S. Sif (2007). "Interplay between chromatin remodelers and protein arginine methyltransferases." J Cell Physiol **213**(2): 306-15.
- Palomeque, T. and P. Lorite (2008). "Satellite DNA in insects: a review." Heredity **100**(6): 564-73.
- Panniers, R. (1994). "Translational control during heat shock." Biochimie **76**(8): 737-47.
- Panning, B., J. Dausman, et al. (1997). "X chromosome inactivation is mediated by Xist RNA stabilization." Cell **90**(5): 907-16.
- Papait, R., C. Pistore, et al. (2007). "Np95 is implicated in pericentromeric heterochromatin replication and in major satellite silencing." Mol Biol Cell **18**(3): 1098-106.
- Parker, K., J. Maxson, et al. (2007). "Class I histone deacetylase Thd1p promotes global chromatin condensation in *Tetrahymena thermophila*." Eukaryot Cell **6**(10): 1913-24.
- Partridge, J. F., K. S. Scott, et al. (2002). "cis-acting DNA from fission yeast centromeres mediates histone H3 methylation and recruitment of silencing factors and cohesin to an ectopic site." Curr Biol **12**(19): 1652-60.
- Pauler, F. M., M. V. Koerner, et al. (2007). "Silencing by imprinted noncoding RNAs: is transcription the answer?" Trends Genet **23**(6): 284-92.
- Pehrson, J. R. and V. A. Fried (1992). "MacroH2A, a core histone containing a large nonhistone region." Science **257**(5075): 1398-400.
- Perche, P. Y., C. Vourc'h, et al. (2000). "Higher concentrations of histone macroH2A in the Barr body are correlated with higher nucleosome density." Curr Biol **10**(23): 1531-4.
- Perez, D. S., T. R. Hoage, et al. (2008). "Long, abundantly expressed non-coding transcripts are altered in cancer." Hum Mol Genet **17**(5): 642-55.
- Perocchi, F., Z. Xu, et al. (2007). "Antisense artifacts in transcriptome microarray experiments are resolved by actinomycin D." Nucleic Acids Res **35**(19): e128.
- Peters, A. H., S. Kubicek, et al. (2003). "Partitioning and plasticity of repressive histone methylation states in mammalian chromatin." Mol Cell **12**(6): 1577-89.
- Peters, A. H., D. O'Carroll, et al. (2001). "Loss of the Suv39h histone methyltransferases impairs mammalian heterochromatin and genome stability." Cell **107**(3): 323-37.
- Peters, A. H. and D. Schubeler (2005). "Methylation of histones: playing memory with DNA." Curr Opin Cell Biol **17**(2): 230-8.
- Piehler, A. P., M. Hellum, et al. (2008). "The human ABC transporter pseudogene family: Evidence for transcription and gene-pseudogene interference." BMC Genomics **9**: 165.
- Pinto, I. and F. Winston (2000). "Histone H2A is required for normal centromere function in *Saccharomyces cerevisiae*." Embo J **19**(7): 1598-612.
- Pirkkala, L., P. Nykanen, et al. (2001). "Roles of the heat shock transcription factors in regulation of the heat shock response and beyond." Faseb J **15**(7): 1118-31.
- Place, R. F., L. C. Li, et al. (2008). "MicroRNA-373 induces expression of genes with complementary promoter sequences." Proc Natl Acad Sci U S A **105**(5): 1608-13.

- Plath, K., J. Fang, et al. (2003). "Role of histone H3 lysine 27 methylation in X inactivation." *Science* **300**(5616): 131-5.
- Poot, R. A., L. Bozhenok, et al. (2005). "Chromatin remodeling by WSTF-ISWI at the replication site: opening a window of opportunity for epigenetic inheritance?" *Cell Cycle* **4**(4): 543-6.
- Porcu, M. and A. Chiarugi (2005). "The emerging therapeutic potential of sirtuin-interacting drugs: from cell death to lifespan extension." *Trends Pharmacol Sci* **26**(2): 94-103.
- Pradhan, S., A. Bacolla, et al. (1999). "Recombinant human DNA (cytosine-5) methyltransferase. I. Expression, purification, and comparison of de novo and maintenance methylation." *J Biol Chem* **274**(46): 33002-10.
- Prasanth, K. V., P. A. Sacco-Bubulya, et al. (2003). "Sequential entry of components of the gene expression machinery into daughter nuclei." *Mol Biol Cell* **14**(3): 1043-57.
- Prescott, D. M. and M. A. Bender (1962). "Synthesis of RNA and protein during mitosis in mammalian tissue culture cells." *Exp Cell Res* **26**: 260-8.
- Probst, A. V. and G. Almouzni (2008). "Pericentric heterochromatin: dynamic organization during early development in mammals." *Differentiation* **76**(1): 15-23.
- Prosser, J., M. Frommer, et al. (1986). "Sequence relationships of three human satellite DNAs." *J Mol Biol* **187**(2): 145-55.
- Provost, P., R. A. Silverstein, et al. (2002). "Dicer is required for chromosome segregation and gene silencing in fission yeast cells." *Proc Natl Acad Sci U S A* **99**(26): 16648-53.
- Pruitt, K., R. L. Zinn, et al. (2006). "Inhibition of SIRT1 reactivates silenced cancer genes without loss of promoter DNA hypermethylation." *PLoS Genet* **2**(3): e40.
- Quivy, J. P., D. Roche, et al. (2004). "A CAF-1 dependent pool of HP1 during heterochromatin duplication." *Embo J* **23**(17): 3516-26.
- Rabindran, S. K., G. Giorgi, et al. (1991). "Molecular cloning and expression of a human heat shock factor, HSF1." *Proc Natl Acad Sci U S A* **88**(16): 6906-10.
- Rabindran, S. K., R. I. Haroun, et al. (1993). "Regulation of heat shock factor trimer formation: role of a conserved leucine zipper." *Science* **259**(5092): 230-4.
- Rae, M. M. and W. W. Franke (1972). "The interphase distribution of satellite DNA-containing heterochromatin in mouse nuclei." *Chromosoma* **39**(4): 443-56.
- Rao, B. J., S. K. Brahmachari, et al. (1983). "Structural organization of the meiotic prophase chromatin in the rat testis." *J Biol Chem* **258**(22): 13478-85.
- Rasmussen, T. P. (2003). "Embryonic stem cell differentiation: a chromatin perspective." *Reprod Biol Endocrinol* **1**: 100.
- Rea, S., F. Eisenhaber, et al. (2000). "Regulation of chromatin structure by site-specific histone H3 methyltransferases." *Nature* **406**(6796): 593-9.
- Rinn, J. L., M. Kertesz, et al. (2007). "Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs." *Cell* **129**(7): 1311-23.
- Rizzi, N., M. Denegri, et al. (2004). "Transcriptional activation of a constitutive heterochromatic domain of the human genome in response to heat shock." *Mol Biol Cell* **15**(2): 543-51.
- Robertson, K. D. (2002). "DNA methylation and chromatin - unraveling the tangled web." *Oncogene* **21**(35): 5361-79.

- Robertson, K. D., E. Uzvolgyi, et al. (1999). "The human DNA methyltransferases (DNMTs) 1, 3a and 3b: coordinate mRNA expression in normal tissues and overexpression in tumors." Nucleic Acids Res **27**(11): 2291-8.
- Rogakou, E. P., D. R. Pilch, et al. (1998). "DNA double-stranded breaks induce histone H2AX phosphorylation on serine 139." J Biol Chem **273**(10): 5858-68.
- Ronne, M., O. Andersen, et al. (1979). "Effect of colcemid exposure and methanol acetic acid fixation on human metaphase chromosome structure." Hereditas **90**(2): 195-201.
- Rudert, F., S. Bronner, et al. (1995). "Transcripts from opposite strands of gamma satellite DNA are differentially expressed during mouse development." Mamm Genome **6**(2): 76-83.
- Rudert, F. and H. Gronemeyer (1993). "Retinoic acid-response elements with a highly repetitive structure isolated by immuno-selection from genomic DNA." J Steroid Biochem Mol Biol **46**(2): 121-33.
- Sado, T., Z. Wang, et al. (2001). "Regulation of imprinted X-chromosome inactivation in mice by Tsix." Development **128**(8): 1275-86.
- Sangster, T. A., S. Lindquist, et al. (2004). "Under cover: causes, effects and implications of Hsp90-mediated genetic capacitance." Bioessays **26**(4): 348-62.
- Santos, F. and W. Dean (2004). "Epigenetic reprogramming during early development in mammals." Reproduction **127**(6): 643-51.
- Sarge, K. D., S. P. Murphy, et al. (1993). "Activation of heat shock gene transcription by heat shock factor 1 involves oligomerization, acquisition of DNA-binding activity, and nuclear localization and can occur in the absence of stress." Mol Cell Biol **13**(3): 1392-407.
- Sarge, K. D., V. Zimarino, et al. (1991). "Cloning and characterization of two mouse heat shock factors with distinct inducible and constitutive DNA-binding ability." Genes Dev **5**(10): 1902-11.
- Sassone-Corsi, P. (2002). "Unique chromatin remodeling and transcriptional regulation in spermatogenesis." Science **296**(5576): 2176-8.
- Saunders, L. R. and E. Verdin (2007). "Sirtuins: critical regulators at the crossroads between cancer and aging." Oncogene **26**(37): 5489-504.
- Saunders, M., M. Fitzgerald-Hayes, et al. (1988). "Chromatin structure of altered yeast centromeres." Proc Natl Acad Sci U S A **85**(1): 175-9.
- Scanlan, M. J., A. O. Gure, et al. (2002). "Cancer/testis antigens: an expanding family of targets for cancer immunotherapy." Immunol Rev **188**: 22-32.
- Schoeftner, S. and M. A. Blasco (2008). "Developmentally regulated transcription of mammalian telomeres by DNA-dependent RNA polymerase II." Nat Cell Biol **10**(2): 228-36.
- Schotta, G., A. Ebert, et al. (2002). "Central role of Drosophila SU(VAR)3-9 in histone H3-K9 methylation and heterochromatic gene silencing." Embo J **21**(5): 1121-31.
- Schubeler, D., D. Scalzo, et al. (2002). "Genome-wide DNA replication profile for Drosophila melanogaster: a link between transcription and replication timing." Nat Genet **32**(3): 438-42.
- Schuetz, T. J., G. J. Gallo, et al. (1991). "Isolation of a cDNA for HSF2: evidence for two heat shock factor genes in humans." Proc Natl Acad Sci U S A **88**(16): 6911-5.

- Sears, D. D., J. H. Hegemann, et al. (1995). "Cis-acting determinants affecting centromere function, sister-chromatid cohesion and reciprocal recombination during meiosis in *Saccharomyces cerevisiae*." *Genetics* **139**(3): 1159-73.
- Shamovsky, I., M. Ivannikov, et al. (2006). "RNA-mediated response to heat shock in mammalian cells." *Nature* **440**(7083): 556-60.
- Shamovsky, I. and E. Nudler (2008). "New insights into the mechanism of heat shock response activation." *Cell Mol Life Sci* **65**(6): 855-61.
- Sharma, P. C., A. Grover, et al. (2007). "Mining microsatellites in eukaryotic genomes." *Trends Biotechnol* **25**(11): 490-8.
- Sheardown, S. A., S. M. Duthie, et al. (1997). "Stabilization of Xist RNA mediates initiation of X chromosome inactivation." *Cell* **91**(1): 99-107.
- Shelby, R. D., K. Monier, et al. (2000). "Chromatin assembly at kinetochores is uncoupled from DNA replication." *J Cell Biol* **151**(5): 1113-8.
- Shi, Y., F. Lan, et al. (2004). "Histone demethylation mediated by the nuclear amine oxidase homolog LSD1." *Cell* **119**(7): 941-53.
- Shibata, S. and J. T. Lee (2003). "Characterization and quantitation of differential Tsix transcripts: implications for Tsix function." *Hum Mol Genet* **12**(2): 125-36.
- Shogren-Knaak, M., H. Ishii, et al. (2006). "Histone H4-K16 acetylation controls chromatin structure and protein interactions." *Science* **311**(5762): 844-7.
- Shumaker, D. K., T. Dechat, et al. (2006). "Mutant nuclear lamin A leads to progressive alterations of epigenetic control in premature aging." *Proc Natl Acad Sci U S A* **103**(23): 8703-8.
- Siedlecki, P. and P. Zielenkiewicz (2006). "Mammalian DNA methyltransferases." *Acta Biochim Pol* **53**(2): 245-56.
- Sieger, M., F. Pera, et al. (1970). "Genetic inactivity of heterochromatin and heteropycnosis in *Microtus agrestis*." *Chromosoma* **29**(3): 349-64.
- Sif, S., P. T. Stukenberg, et al. (1998). "Mitotic inactivation of a human SWI/SNF chromatin remodeling complex." *Genes Dev* **12**(18): 2842-51.
- Sims, R. J., 3rd, C. F. Chen, et al. (2005). "Human but not yeast CHD1 binds directly and selectively to histone H3 methylated at lysine 4 via its tandem chromodomains." *J Biol Chem* **280**(51): 41789-92.
- Sleutels, F., G. Tjon, et al. (2003). "Imprinted silencing of *Slc22a2* and *Slc22a3* does not need transcriptional overlap between *Igf2r* and *Air*." *Embo J* **22**(14): 3696-704.
- Sleutels, F., R. Zwart, et al. (2002). "The non-coding *Air* RNA is required for silencing autosomal imprinted genes." *Nature* **415**(6873): 810-3.
- Sorger, P. K. and H. R. Pelham (1988). "Yeast heat shock factor is an essential DNA-binding protein that exhibits temperature-dependent phosphorylation." *Cell* **54**(6): 855-64.
- Sparvoli, E., M. G. Galli, et al. (1977). "Perinuclear DNA and heterochromatin." *Exp Cell Res* **110**(2): 315-21.
- Spradling, A., S. Penman, et al. (1975). "Analysis of drosophila mRNA by in situ hybridization: sequences transcribed in normal and heat shocked cultured cells." *Cell* **4**(4): 395-404.
- Stahl, A., M. Hartung, et al. (1976). "Chromosomal constitution of nucleolus-associated chromatin in man." *Hum Genet* **35**(1): 27-34.

- Stoler, S., K. C. Keith, et al. (1995). "A mutation in CSE4, an essential gene encoding a novel chromatin-associated protein in yeast, causes chromosome nondisjunction and cell cycle arrest at mitosis." Genes Dev **9**(5): 573-86.
- Strahl, B. D. and C. D. Allis (2000). "The language of covalent histone modifications." Nature **403**(6765): 41-5.
- Suetake, I., F. Shinozaki, et al. (2004). "DNMT3L stimulates the DNA methylation activity of Dnmt3a and Dnmt3b through a direct interaction." J Biol Chem **279**(26): 27816-23.
- Sullivan, B. and G. Karpen (2001). "Centromere identity in *Drosophila* is not determined in vivo by replication timing." J Cell Biol **154**(4): 683-90.
- Sullivan, B. A. and G. H. Karpen (2004). "Centromeric chromatin exhibits a histone modification pattern that is distinct from both euchromatin and heterochromatin." Nat Struct Mol Biol **11**(11): 1076-83.
- Sun, F. L., M. H. Cuaycong, et al. (2001). "Long-range nucleosome ordering is associated with gene silencing in *Drosophila melanogaster* pericentric heterochromatin." Mol Cell Biol **21**(8): 2867-79.
- Sun, X., H. D. Le, et al. (2003). "Sequence analysis of a functional *Drosophila* centromere." Genome Res **13**(2): 182-94.
- Sun, X., J. Wahlstrom, et al. (1997). "Molecular structure of a functional *Drosophila* centromere." Cell **91**(7): 1007-19.
- Suzuki, M. M. and A. Bird (2008). "DNA methylation landscapes: provocative insights from epigenomics." Nat Rev Genet **9**(6): 465-76.
- Tagarro, I., A. M. Fernandez-Peralta, et al. (1993). "Digestion of centromeric DNA from each human metaphase chromosome by the 6 bp-restriction enzyme StuI." Histochemistry **99**(6): 453-6.
- Tagarro, I., A. M. Fernandez-Peralta, et al. (1994). "Chromosomal localization of human satellites 2 and 3 by a FISH method using oligonucleotides as probes." Hum Genet **93**(4): 383-8.
- Tagarro, I., J. Wiegant, et al. (1994). "Assignment of human satellite 1 DNA as revealed by fluorescent in situ hybridization with oligonucleotides." Hum Genet **93**(2): 125-8.
- Takahashi, K., S. Murakami, et al. (1992). "A low copy number central sequence with strict symmetry and unusual chromatin structure in fission yeast centromere." Mol Biol Cell **3**(7): 819-35.
- Takebayashi, S., M. Nakao, et al. (2001). "5-Aza-2'-deoxycytidine induces histone hyperacetylation of mouse centromeric heterochromatin by a mechanism independent of DNA demethylation." Biochem Biophys Res Commun **288**(4): 921-6.
- Talbert, P. B., C. D. LeCiel, et al. (1994). "Modification of the *Drosophila* heterochromatic mutation brownDominant by linkage alterations." Genetics **136**(2): 559-71.
- Taylor, J. H. (1960). "Asynchronous duplication of chromosomes in cultured cells of Chinese hamster." J Biophys Biochem Cytol **7**: 455-64.
- Ten Hagen, K. G., D. M. Gilbert, et al. (1990). "Replication timing of DNA sequences associated with human centromeres and telomeres." Mol Cell Biol **10**(12): 6348-55.
- Terranova, R., S. Sauer, et al. (2005). "The reorganisation of constitutive heterochromatin in differentiating muscle requires HDAC activity." Exp Cell Res **310**(2): 344-56.

- Tian, X. and J. Fang (2007). "Current perspectives on histone demethylases." Acta Biochim Biophys Sin (Shanghai) **39**(2): 81-8.
- Tiepolo, L., P. Maraschio, et al. (1979). "Multibranched chromosomes 1, 9, and 16 in a patient with combined IgA and IgE deficiency." Hum Genet **51**(2): 127-37.
- Tomilin, N. V. (2008). "Regulation of mammalian gene expression by retroelements and non-coding tandem repeats." Bioessays **30**(4): 338-48.
- Tomkiel, J., C. A. Cooke, et al. (1994). "CENP-C is required for maintaining proper kinetochore size and for a timely transition to anaphase." J Cell Biol **125**(3): 531-45.
- Trinklein, N. D., J. I. Murray, et al. (2004). "The role of heat shock transcription factor 1 in the genome-wide regulation of the mammalian heat shock response." Mol Biol Cell **15**(3): 1254-61.
- Tse, C., T. Sera, et al. (1998). "Disruption of higher-order folding by core histone acetylation dramatically enhances transcription of nucleosomal arrays by RNA polymerase III." Mol Cell Biol **18**(8): 4629-38.
- Tyler-Smith, C. and G. Florida (2000). "Many paths to the top of the mountain: diverse evolutionary solutions to centromere structure." Cell **102**(1): 5-8.
- Vagnarelli, P., S. A. Ribeiro, et al. (2008). "Centromeres: old tales and new tools." FEBS Lett **582**(14): 1950-9.
- Vakoc, C. R., S. A. Mandat, et al. (2005). "Histone H3 lysine 9 methylation and HP1gamma are associated with transcription elongation through mammalian chromatin." Mol Cell **19**(3): 381-91.
- Valdivia, M. M., J. Figueroa, et al. (1998). "A novel centromere monospecific serum to a human autoepitope on the histone H3-like protein CENP-A." FEBS Lett **422**(1): 5-9.
- Valgardsdottir, R., I. Chiodi, et al. (2008). "Transcription of Satellite III non-coding RNAs is a general stress response in human cells." Nucleic Acids Res **36**(2): 423-34.
- van Attikum, H., O. Fritsch, et al. (2004). "Recruitment of the INO80 complex by H2A phosphorylation links ATP-dependent chromatin remodeling with DNA double-strand break repair." Cell **119**(6): 777-88.
- Van Hooser, A. A., Ouspenski, II, et al. (2001). "Specification of kinetochore-forming chromatin by the histone H3 variant CENP-A." J Cell Sci **114**(Pt 19): 3529-42.
- van Roijen, H. J., M. P. Ooms, et al. (1998). "Immunoexpression of testis-specific histone 2B in human spermatozoa and testis tissue." Hum Reprod **13**(6): 1559-66.
- Verdel, A., S. Jia, et al. (2004). "RNAi-mediated targeting of heterochromatin by the RITS complex." Science **303**(5658): 672-6.
- Verdel, A. and D. Moazed (2005). "RNAi-directed assembly of heterochromatin in fission yeast." FEBS Lett **579**(26): 5872-8.
- Vissel, B. and K. H. Choo (1987). "Human alpha satellite DNA--consensus sequence and conserved regions." Nucleic Acids Res **15**(16): 6751-2.
- Voellmy, R. (2004). "On mechanisms that control heat shock transcription factor activity in metazoan cells." Cell Stress Chaperones **9**(2): 122-33.
- Volpe, T. A., C. Kidner, et al. (2002). "Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi." Science **297**(5588): 1833-7.



- Waizenegger, I. C., S. Hauf, et al. (2000). "Two distinct pathways remove mammalian cohesin from chromosome arms in prophase and from centromeres in anaphase." Cell **103**(3): 399-410.
- Walia, H., H. Y. Chen, et al. (1998). "Histone acetylation is required to maintain the unfolded nucleosome structure associated with transcribing DNA." J Biol Chem **273**(23): 14516-22.
- Wallrath, L. L. and S. C. Elgin (1995). "Position effect variegation in *Drosophila* is associated with an altered chromatin structure." Genes Dev **9**(10): 1263-77.
- Wang, Y., J. Wysocka, et al. (2004). "Human PAD4 regulates histone arginine methylation levels via demethyliminination." Science **306**(5694): 279-83.
- Warren, W. D., S. Steffensen, et al. (2000). "The *Drosophila* RAD21 cohesin persists at the centromere region in mitosis." Curr Biol **10**(22): 1463-6.
- Watt, F. and P. L. Molloy (1988). "Cytosine methylation prevents binding to DNA of a HeLa cell transcription factor required for optimal expression of the adenovirus major late promoter." Genes Dev **2**(9): 1136-43.
- Waye, J. S. and H. F. Willard (1989). "Human beta satellite DNA: genomic organization and sequence definition of a class of highly repetitive tandem DNA." Proc Natl Acad Sci U S A **86**(16): 6250-4.
- Weber, M., J. J. Davies, et al. (2005). "Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells." Nat Genet **37**(8): 853-62.
- Welch, W. J. and J. P. Suhan (1985). "Morphological study of the mammalian stress response: characterization of changes in cytoplasmic organelles, cytoskeleton, and nucleoli, and appearance of intranuclear actin filaments in rat fibroblasts after heat-shock treatment." J Cell Biol **101**(4): 1198-211.
- Wevrick, R. and H. F. Willard (1989). "Long-range organization of tandem arrays of alpha satellite DNA at the centromeres of human chromosomes: high-frequency array-length polymorphism and meiotic stability." Proc Natl Acad Sci U S A **86**(23): 9394-8.
- Wiederrecht, G., D. Seto, et al. (1988). "Isolation of the gene encoding the *S. cerevisiae* heat shock transcription factor." Cell **54**(6): 841-53.
- Wiens, G. R. and P. K. Sorger (1998). "Centromeric chromatin and epigenetic effects in kinetochore assembly." Cell **93**(3): 313-6.
- Wilson, A. S., B. E. Power, et al. (2007). "DNA hypomethylation and human diseases." Biochim Biophys Acta **1775**(1): 138-62.
- Witt, O., W. Albig, et al. (1996). "Testis-specific expression of a novel human H3 histone gene." Exp Cell Res **229**(2): 301-6.
- Wong, A. K. and J. B. Rattner (1988). "Sequence organization and cytological localization of the minor satellite of mouse." Nucleic Acids Res **16**(24): 11645-61.
- Wood, V., R. Gwilliam, et al. (2002). "The genome sequence of *Schizosaccharomyces pombe*." Nature **415**(6874): 871-80.
- Woodfine, K., H. Fiegler, et al. (2004). "Replication timing of the human genome." Hum Mol Genet **13**(2): 191-202.
- Workman, J. L. and R. E. Kingston (1998). "Alteration of nucleosome structure as a mechanism of transcriptional regulation." Annu Rev Biochem **67**: 545-79.
- Wright, W. E., V. M. Tesmer, et al. (1997). "Normal human chromosomes have long G-rich telomeric overhangs at one end." Genes Dev **11**(21): 2801-9.

- Wu, C. (1995). "Heat shock transcription factors: structure and regulation." Annu Rev Cell Dev Biol **11**: 441-69.
- Wu, R., P. B. Singh, et al. (2006). "Uncoupling global and fine-tuning replication timing determinants for mouse pericentric heterochromatin." J Cell Biol **174**(2): 185-94.
- Wutz, A. and R. Jaenisch (2000). "A shift from reversible to irreversible X inactivation is triggered during ES cell differentiation." Mol Cell **5**(4): 695-705.
- Xiao, X., X. Zuo, et al. (1999). "HSF1 is required for extra-embryonic development, postnatal growth and protection during inflammatory responses in mice." Embo J **18**(21): 5943-52.
- Xie, S., Z. Wang, et al. (1999). "Cloning, expression and chromosome locations of the human DNMT3 gene family." Gene **236**(1): 87-95.
- Xu, G. L., T. H. Bestor, et al. (1999). "Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene." Nature **402**(6758): 187-91.
- Yang, X. J. and E. Seto (2007). "HATs and HDACs: from structure, function and regulation to novel strategies for therapy and prevention." Oncogene **26**(37): 5310-8.
- Yazgan, O. and J. E. Krebs (2007). "Noncoding but nonexpendable: transcriptional regulation by large noncoding RNA in eukaryotes." Biochem Cell Biol **85**(4): 484-96.
- Yoshida, M., M. Kijima, et al. (1990). "Potent and specific inhibition of mammalian histone deacetylase both in vivo and in vitro by trichostatin A." J Biol Chem **265**(28): 17174-9.
- Zagradisnik, B. and N. Kokalj-Vokac (2000). "Hypomethylation of alphoid DNA and classical satellite DNA on chromosome 1, 9, 16 and Y in extraembryonic tissue." Pflugers Arch **440**(5 Suppl): R190-2.
- Zeitlin, S. G., R. D. Shelby, et al. (2001). "CENP-A is phosphorylated by Aurora B kinase and plays an unexpected role in completion of cytokinesis." J Cell Biol **155**(7): 1147-57.
- Zhao, W., H. Soejima, et al. (2005). "The essential role of histone H3 Lys9 dimethylation and MeCP2 binding in MGMT silencing with poor DNA methylation of the promoter CpG island." J Biochem **137**(3): 431-40.
- Zhimulev, I. F., E. S. Belyaeva, et al. (1988). "Cytogenetic and molecular aspects of position effect variegation in *Drosophila melanogaster*. II. Peculiarities of morphology and genetic activity of the 2B region in the T(1;2)dotvar7 chromosome in males." Chromosoma **96**(3): 255-61.

Le gène est défini comme une unité fonctionnelle de l'hérédité. En effet, les gènes, transcrits en ARN, codent pour des protéines qui assurent l'essentiel des fonctions cellulaires. Cependant, les gènes ne représentent que 2% du génome humain. Ainsi, notre génome est presque exclusivement composé d'ADN non codant (ADNnc) qui, en raison d'une absence apparente de fonction (c'est-à-dire de transcription et de traduction), est également connu sous le terme peu élogieux d'« ADN poubelle ». De manière surprenante, la quantité d'ADNnc présente dans les cellules est proportionnelle au degré de complexité des organismes, suggérant ainsi que cet ADNnc pourrait avoir une fonction. En particulier, les centromères et des péricentromères, qui contiennent des séquences d'ADNnc de type satellite, permettent la ségrégation correcte de chromatides sœurs lors de la mitose. De plus des études récentes réalisées chez la levure *S.pombe* montrent que les régions péricentromériques peuvent être transcrites. Les ARNnc ainsi produits participent à l'établissement et au maintien de l'hétérochromatine péricentromérique. Bien que les mécanismes impliqués dans ces processus soient en grande partie identifiés, très peu de choses sont connues quant à la capacité transcriptionnelle de ces séquences satellites chez d'autres espèces.

**Ainsi, l'objectif de ma thèse a été d'évaluer le potentiel transcriptionnel des séquences satellites du génome humain.**

A mon arrivée dans le laboratoire, l'équipe du Pr Claire Vourc'h était particulièrement engagée dans ce projet puisqu'elle venait de mettre en évidence la transcription des séquences satellites 3 du locus 9q12 au cours de la réponse au stress. Mon travail de thèse m'a conduit à approfondir cette étude en identifiant de nouvelles séquences satellites dont la transcription est induite par le stress thermique. Par ailleurs, j'ai développé la première approche transcriptomique dédiée à l'expression des séquences satellites du génome humain : la RepChip. Le modèle de la réponse au stress m'a permis de valider cet outil. Ainsi, j'ai pu identifier non seulement de nouveaux contextes cellulaires permettant l'expression de ces séquences mais également deux mécanismes indépendants impliqués dans leur transcription. En particulier, un lien entre la capacité transcriptionnelle de ces séquences et des modifications de l'épigénome a pu être établi. Finalement, j'ai identifié un modèle d'inhibition de la transcription des séquences satellites au cours du choc thermique, le syndrome ICF (Immunodeficiency, Centromeric instability and Facial anomalies).