



**HAL**  
open science

# Étude des “ Topologically Associated Domains (TADs) ” dans les anomalies du développement

Aurelien Trimouille

► **To cite this version:**

Aurelien Trimouille. Étude des “ Topologically Associated Domains (TADs) ” dans les anomalies du développement . Médecine humaine et pathologie. 2017. dumas-01628629

**HAL Id: dumas-01628629**

**<https://dumas.ccsd.cnrs.fr/dumas-01628629v1>**

Submitted on 3 Nov 2017

**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.

# Étude des “ Topologically Associated Domains (TADs) ” dans les anomalies du développement

Aurelien Trimouille

► **To cite this version:**

Aurelien Trimouille. Étude des “ Topologically Associated Domains (TADs) ” dans les anomalies du développement . Médecine humaine et pathologie. 2017. <dumas-01628629>

**HAL Id: dumas-01628629**

**<https://dumas.ccsd.cnrs.fr/dumas-01628629>**

Submitted on 3 Nov 2017

**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 de Bordeaux  
U.F.R. DES SCIENCES MEDICALES

Année 2017

Thèse n°3173

**Thèse pour l'obtention du  
DIPLOME d'ETAT de DOCTEUR EN MEDECINE**

Présentée et soutenue publiquement

Le 24/10/2017

Par Aurélien TRIMOUILLE

Né le 05/02/1988

**Étude des « Topologically Associated Domains (TADs) »  
dans les anomalies du développement**

**Directeur de thèse : Madame le Dr Caroline ROORYCK-THAMBO**

**Rapporteur de thèse : Madame le Pr Jeanne AMIEL**

**Membres du jury :**

Monsieur le Pr Benoit ARVEILER (président)

Monsieur le Pr Didier LACOMBE (juge)

Monsieur le Pr François MOREAU-GAUDRY (juge)

Madame le Dr Patricia FERGELOT (juge)

## Remerciements

Je remercie tous les membres du jury, ainsi que le rapporteur, Madame le Professeur Amiel, du temps consacré à la lecture de ce travail.

Je remercie le Professeur Benoit Arveiler d'avoir accepté la présidence de ce jury. Je le remercie également pour son accueil dès mon arrivée au sein du laboratoire de génétique moléculaire, ainsi que pour la confiance qu'il m'a accordée pour les différents projets réalisés durant mon internat.

Je remercie le Professeur Didier Lacombe pour son enseignement de la Génétique Clinique, et au delà, du contact et de la prise en charge globale des patients. Je le remercie également pour sa particulière bienveillance durant mon parcours d'interne.

Je remercie le Professeur François Moreau-Gaudry d'avoir accepté de faire partie de ce jury, et d'avoir pu se rendre disponible pour évaluer ce travail de thèse.

Je remercie le Docteur Patricia Fergelot pour ces très nombreux moments d'échanges, concernant de larges pans de la médecine, et pour tous ses conseils précieux.

Je remercie le Professeur Jeanne Amiel d'avoir accepté d'établir le rapport de ma thèse, ainsi que pour l'excellence de son enseignement durant mon stage réalisé à l'hôpital Necker.

Je remercie très chaleureusement le Docteur Caroline Rooryck-Thambo d'avoir accepté la direction de ma thèse d'exercice. Je la remercie particulièrement de m'avoir accordé mon autonomie pour le choix de ce sujet, tout en me guidant dans la réalisation de ce travail.

Un immense merci au Docteur Angela Tingaud-Sequeira pour son implication dans mon travail, pour tout ce temps qu'elle m'aura consacré, et pour son expertise sans laquelle les expériences de 3C n'auraient simplement pas été envisageables. Je la remercie également pour la relecture soignée qu'elle a pu faire de ce manuscrit.

Je remercie Mesdames Julie Bouron et Armelle Courrèges pour leur motivation, pour leur aide très précieuse et leurs avis éclairés sur les différentes expériences de CGH-array, de qPCR et de PCR Long-Range que nous avons pu réaliser.

Je remercie l'ensemble des équipes des trois écoles de la Génétique au sein desquelles j'ai eu la chance d'être formé : les services de Génétique du CHU de Bordeaux, de l'hôpital Robert-Debré et de l'hôpital Necker.

Je remercie en particulier le Docteur Eulalie Lasseaux pour son soutien et son amitié, et pour le travail que nous réalisons ensemble sur le séquençage de l'exome en diagnostic.

Je remercie également le Docteur Perrine Pennamen, le Docteur Julien Van-Gils, ainsi que mes co-internes de Génétique, Vincent, Henri, Audrey, Caroline, Ariane, Pauline, Benjamin et Yoann pour tous ces moments passés ensemble ces 5 dernières années, au CHU et en dehors, et sans lesquels mon internat n'aurait pas été le même !

## Sommaire

I.	Introduction .....	9
A.	La chromatine .....	9
B.	Organisation spatiale du génome.....	14
1.	Territoires chromosomiques .....	14
2.	Compartiments chromosomiques.....	15
3.	Topologically associated domain (TAD).....	19
a)	Propriétés des TADs .....	22
b)	Formation des TADs.....	23
c)	Fonction des TADs .....	25
4.	Études des TADs par Chromosome Conformation Capture (3C) et techniques derives .....	29
a)	Technique de Chromosome Conformation Capture (3C).....	29
b)	Techniques de 4C et de 5C .....	32
c)	Techniques de Hi-C et ChIA-PET .....	34
5.	TADS et anomalies du développement : études issues de la littérature .....	40
a)	Remaniements chromosomiques et altérations des TADs.....	40
(1)	Délétions.....	40
(2)	Duplications.....	42
(3)	Translocation et inversion .....	44
b)	Défaut de mise en place des TADs .....	49
II.	Matériel et Méthodes.....	52
A.	Ré-analyse de la cohorte hospitalière de patients sans CNV pathogène identifié par ACPA .....	52
1.	Description de la cohorte .....	52
2.	Méthode de tri .....	53
3.	Visualisation et interprétation des données.....	55

B.	Mise au point de la technique de 3C.....	57
1.	Choix du locus CIITA.....	57
2.	Protocole de 3C.....	59
a)	Cross-linking.....	59
b)	Lyse des cellules et digestion par NcoI.....	60
c)	Ligation.....	60
d)	Purification de l'ADN.....	60
e)	Détection des produits de ligation.....	61
III.	Résultats.....	61
A.	Ré-analyse de la cohorte hospitalière de patients sans CNV pathogène identifié par ACPA.....	61
1.	Délétion au locus <i>WNT6/IHH/EPHA4/PAX3</i> .....	61
a)	Description foetopathologique et moléculaire.....	61
b)	Description du locus <i>WNT6/IHH/EPHA4/PAX3</i> .....	64
c)	Variants structuraux au locus <i>WNT6/IHH/EPHA4/PAX3</i> : modèle de Lupiáñez et collaborateurs.....	66
d)	Application du modèle au cas fœtal.....	69
2.	Délétion au locus <i>ID4/E2F3/SOX4</i> .....	70
a)	Description Clinique et moléculaire.....	70
b)	Etudes complémentaires.....	73
c)	CNV au locus <i>ID4/E2F3/SOX4</i> dans la littérature.....	74
d)	Application du modèle de Flöttmann.....	76
3.	Délétion au locus <i>FMN2</i> .....	77
4.	Délétion au locus d' <i>UBE2A</i> .....	80
5.	Délétion au locus <i>MBD5</i> .....	84
6.	Duplication au locus <i>WASHC5</i> .....	88
7.	Duplication au locus d' <i>ANKRD11</i> .....	92

B.	Mise au point de la technique de 3C.....	94
a) b) c)	Courbes de fusion des produits de ligation pIV/enhancer -8kb d'un échantillon, comparés à : .....	95
IV.	Discussion et conclusion .....	96
A.	Méthode de ré-analyse de la cohorte analysée par ACPA : jusqu'où pousser l'automatisation du pipeline ? .....	96
B.	Taux de cas positifs : influence de la sélection des CNVs et de la technique utilisée	98
C.	Remaniements, intervalles critiques et gènes causaux : vers une reconsidération de certains syndromes ? .....	99
D.	Prédictions <i>in silico</i> , preuves <i>in vitro</i> ?.....	100
E.	Mise au point de la technique de 3C.....	101
F.	Défaut de mise en place des TADs : un des mécanismes des cohésinopathies ?.....	101
G.	L'organisation chromatinienne : une aide à l'interprétation du séquençage du génome complet ?.....	102
H.	Conclusion .....	103
V.	Annexes .....	104
A.	Détail de la méthode de tri des CNV .....	104
B.	Légende de la segmentation de la chromatine ChromHMM(90) visualisée sur UCSC Genome Browser.....	106
C.	Liste de 44 gènes impliqués dans la déficience intellectuelle. ....	107
VI.	Références .....	108



Figure 1 : Walther Flemming, Zellsubstanz, Kern und Zellteilung(1).....	9
Figure 2 : Modèle classique de repliement hiérarchique de l'ADN.....	11
Figure 3 : Noyaux de cellules épithéliales des voies respiratoires humaines (SAEC) visualisés en ChromEMT.....	12
Figure 4 : Modèle d'organisation de la chromatine d'après les données de ChromEMT.....	13
Figure 5 : FISH 3D de l'ensemble des chromosomes de fibroblastes humains en phase G0 ..	15
Figure 6 : Données de Hi-C obtenues par Liebermann-Aiden et collaborateurs(25).....	17
Figure 7 : Globule à l'équilibre et globule fractal.....	18
Figure 8 : TADs et compartiments.....	20
Figure 9 : Boucle locale au sein d'un TAD.....	21
Figure 10 : Modèle Strings and Binders Switch (SBS).....	22
Figure 11 : Modèle de formation des TADs par extrusion de l'ADN.....	25
Figure 12 : Fonctions d'un TAD sur la régulation transcriptionnelle.....	26
Figure 13 : Régulation de la transcription au locus <i>HoxD</i> chez la souris.....	28
Figure 14 : Schéma général d'une expérience de 3C.....	30
Figure 15 : Expériences de 3C au locus <i>CFTR</i> .....	31
Figure 16 : Résumé des méthodes dérivées de Chromosome Conformation Capture.....	32
Figure 17 : Données de 4C.....	33
Figure 18 : Données de 5C.....	34
Figure 19 : Données de Hi-C au locus <i>Hoxa</i> chez la souris(30).....	37
Figure 20 : Comparaison de différents algorithmes de détection des TADs.....	38
Figure 21 : Données de ChIAPet.....	39
Figure 22 : Leucodystrophie autosomique dominante de l'adulte par délétion d'une barrière de TAD.....	41
Figure 23 : Conséquences des différentes duplications en tandem au locus de <i>KCNJ2</i> et <i>SOX9</i> .....	43
Figure 24 : Points de cassure de translocations au locus de <i>MEF2C</i> .....	46

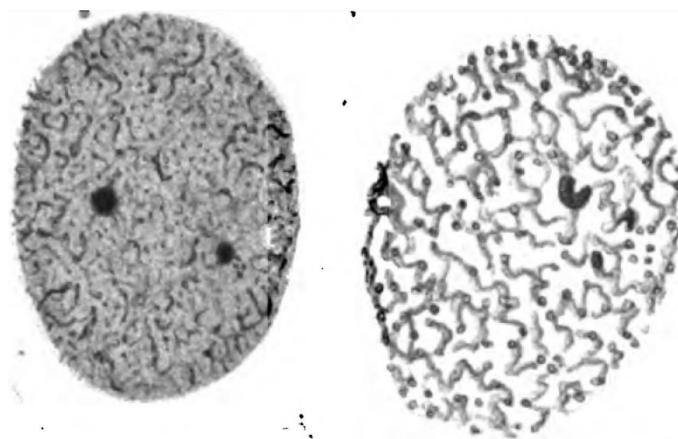
Figure 25 : Remaniements chromosomiques et TAD : Résumé .....	48
Figure 26 : Données de Hi-C au locus cPcdh des souris <i>Setdb1</i> <sup>2lox/2lox</sup> (44) .....	50
Figure 27 : Fichiers .bed et outils BEDtools .....	54
Figure 28 : Interprétation des données de Hi-C .....	57
Figure 29 : Données de 3C au locus <i>CIITA</i> , d'après Ni et collaborateurs(114). .....	58
Figure 30 : Délétion au locus <i>WNT6/IHH/EPHA4/PAX3</i> .....	63
Figure 31 : Position des fragments amplifiés par qPCR chez le fœtus par rapport à <i>IHH</i> .....	64
Figure 32 : Résumé des remaniements au locus <i>WNT6/IHH/EPHA4/PAX3</i> .....	69
Figure 33 : Délétion au locus <i>ID4/E2F3/SOX4</i> .....	71
Figure 34 : Etude d'expression de <i>SOX4</i> .....	74
Figure 35 : Délétion au locus <i>FMN2</i> .....	78
Figure 36 : Délétion au locus d' <i>UBE2A</i> .....	82
Figure 37 : Délétion au locus <i>MBD5</i> .....	85
Figure 38 : Phénotypes neurodéveloppementaux liés à <i>MBD5</i> : un gène, différents mécanismes .....	87
Figure 39 : Duplication au locus <i>WASHC5 (KIAA0196)</i> .....	91
Figure 40 : Duplication au locus d' <i>ANKRD11</i> .....	93
Figure 41 : Résultat des expériences de 3C au locus <i>CIITA</i> (n=3) .....	95

## I. Introduction

### A. La chromatine

La généralisation des examens tels que l'Analyse Chromosomique sur Puces à ADN (ACPA), et le séquençage par technologie Next-Generation Sequencing (NGS), dans le cadre du diagnostic des anomalies du développement, a amené à la détection très fréquente de variants pour lesquelles une interprétation simple n'est pas possible. Ces difficultés d'interprétation concernent notamment les variants structuraux détectés par ACPA pour lesquels, bien souvent, seul le contenu en gènes est considéré. Ces variants structuraux sont en premier lieu une désorganisation de larges régions chromatiniennes par rapport à un génome de référence. L'étude des éléments constituant la chromatine, ainsi que de leur organisation, pourrait donc permettre de mieux comprendre l'impact fonctionnel de ces variants de structure.

Seulement 3 ans après l'identification de la chromatine par Walther Flemming en 1882(1) (Figure 1), l'étude de son organisation au sein du noyau durant le cycle cellulaire fait déjà l'objet de recherche par Carl Rabl(2). Celui-ci observe par microscopie optique, dans les cellules épithéliales de larves de salamandre, une organisation non aléatoire et polarisée de la chromatine, dénommée pseudo-Rabl. Cette organisation correspond à une agglomération des centromères en amas proche du corps polaire du fuseau, tandis que les télomères sont eux situés en périphérie, et liés à l'enveloppe du noyau.



**Figure 1** : Walther Flemming, Zellsubstanz, Kern und Zellteilung(1)

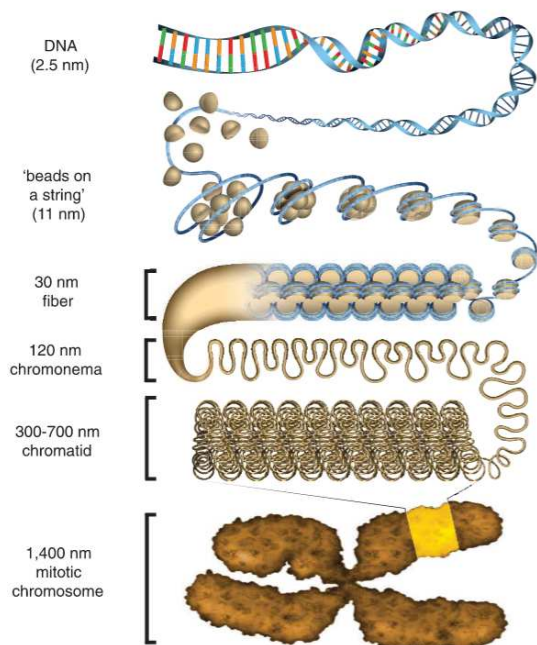
« Noyaux dans leur premier stade de remaniement en forme de pelote »

Les principes généraux de la compaction de la double hélice d'ADN au sein de la chromatine sont désormais connus à différentes échelles. Cette organisation tridimensionnelle n'a pas qu'un rôle structural : elle est essentielle pour des fonctions biologiques critiques telles que la régulation de la transcription, la réparation de l'ADN et la réplication(3).

Le premier niveau de repliement de l'ADN est effectué par les protéines histones. Il existe cinq classes principales d'histones : 4 histones « de cœur », H2A, H2B, H3 et H4, et une histone « de liaison » H1. Les histones de cœur sont caractérisées par la présence d'un domaine C-terminal globulaire très conservé permettant leur dimérisation, servant à l'assemblage du nucléosome. Celui-ci est constitué d'un octamère contenant deux dimères H2A-H2B et deux dimères H3-H4(4), chargé positivement, autour duquel s'enroule une boucle d'ADN d'environ 146 nucléotides, suivant une hélice gauche d'un tour et 65 centièmes. Ces boucles d'ADN autour des nucléosomes, formant une structure dite en « perles sur un fil » (beads-on-a-string) se répètent tous les 200 nucléotides environ. Les modifications biochimiques des queues des histones, comme les méthylations ou les acétylations, constituent un mécanisme épigénétique majeur. Elles vont avoir pour conséquence de modifier la structure chromatinienne, permettant ainsi un changement de l'accessibilité de l'ADN, et donc une régulation de la transcription.

Afin de déterminer le mode de repliement de l'ADN à une échelle supérieure à celle des nucléosomes, de nombreuses études ont été réalisées sur des modèles *in vitro*, à partir d'ADN et d'histones purifiés, ou de noyaux perméabilisés dont les autres composants ont été extraits. Ces études(5–10), par microscopie électronique et par cristallographie à rayons X, ont permis d'établir un modèle classique de repliement hiérarchique de l'ADN au sein du noyau (Figure 2). Selon ce modèle, les nucléosomes s'assemblent d'abord en une fibre d'environ 30nm, selon deux formes possibles, en solénoïde de 33nm, ou en zig-zag de 27,2 à 29,9nm. Cette fibre de 30nm s'assemble par la suite en chromonéma de 120nm, puis en chromatide de 300 à 700nm, avant de former le chromosome individualisé, tel qu'il peut être identifié lors de la métaphase.

Cependant, les données issues d'études par d'autres techniques plus poussées (cryo-microscopie électronique(11), diffusion des rayons X(12) et spectroscopie électronique(13)) ne sont pas en faveur de l'existence de ce modèle, bien que les limitations de chacune de ces technologies ne permettent pas à elles seules de déterminer l'organisation tridimensionnelle de la chromatine à une large échelle.



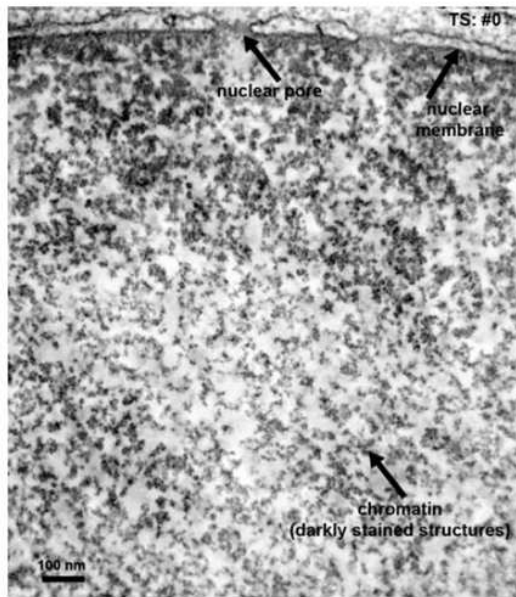
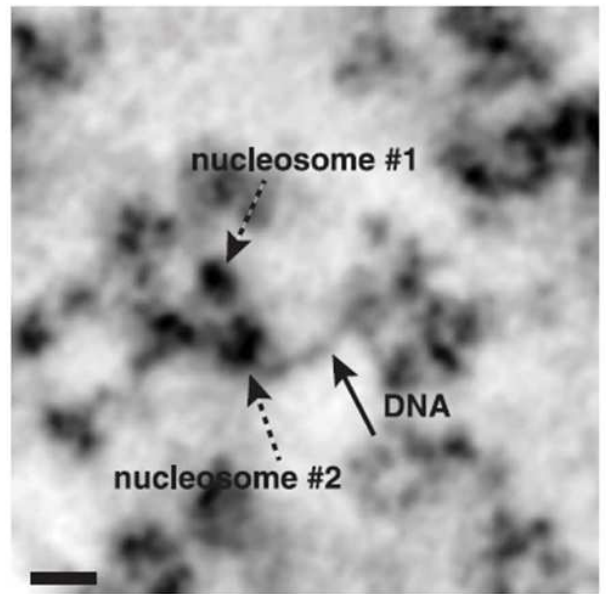
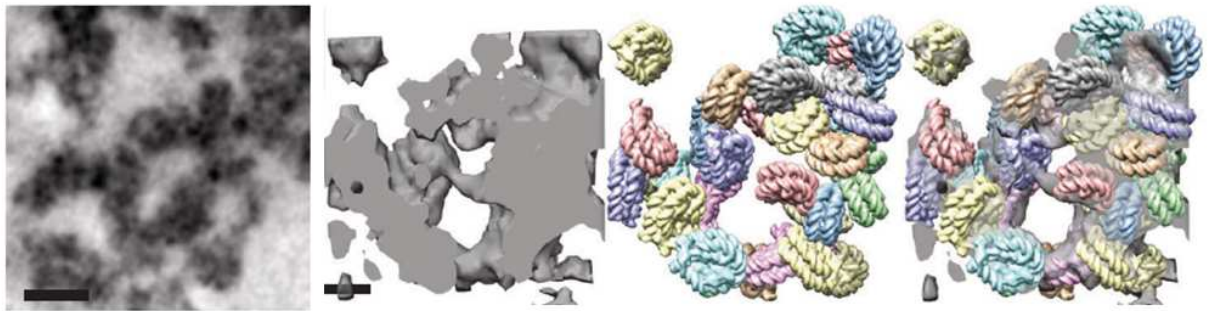
**Figure 2 :** Modèle classique de repliement hiérarchique de l'ADN.

Adapté de Ou et collaborateurs(14)

Les nucléosomes en « perles sur un fil » s'assemblent en une fibre de 30nm, puis en chromonéma de 120nm, en chromatide de 300 à 700 nm, avant de former le chromosome identifiable durant la métaphase.

Par ailleurs, une technique très récente, la ChromEMT(14), permettant de visualiser la chromatine et son organisation, *in situ*, et ce à l'échelle du noyau (Figure 3). Son principe repose sur une fixation de fluorophores sur l'ADN ayant la propriété de catalyser le dépôt de polymères de diaminobenzidine à sa surface, permettant une visualisation en microscopie électronique après fixation par tétraoxyde d'osmium.

Les données de ChromEMT ne sont pas en faveur de l'existence du modèle hiérarchique précédemment admis. En effet, aucun des niveaux de repliement décrits, à commencer par la fibre de 30nm, n'est visualisable sur les images obtenues. Un autre modèle d'organisation de la chromatine a été ainsi proposé : la chromatine formerait une chaîne de 5 à 24nm de diamètre d'apparence désordonnée, et présente à différentes concentrations dans le noyau (Figure 4).

**A****B****C**

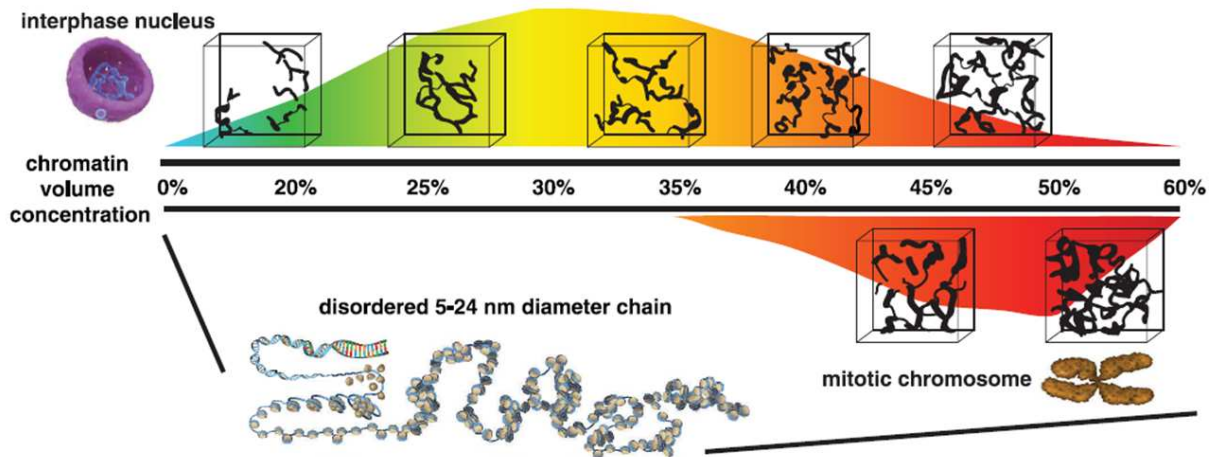
**Figure 3 :** Noyaux de cellules épithéliales des voies respiratoires humaines (SAEC) visualisés en ChromEMT.

Adapté de Ou et collaborateurs(14).

A : Au sein du noyau, la chromatine apparait comme une chaîne désordonnée, agrégée à différentes concentrations, et notamment plus concentrée autour de la lamina.

B : Visualisation de nucléosomes individualisables. La barre d'échelle est de 20nm.

C : Visualisation d'une boucle de chromatine. L'algorithme Sculptor(15) permet à partir des données de densités de microscopie électronique en ChromEMT de prédire l'emplacement des nucléosomes.



**Figure 4** : Modèle d'organisation de la chromatine d'après les données de ChromEMT.

Adapté de Ou et collaborateurs(14)

La chromatine formerait une chaîne flexible d'apparence désordonnée, de 5 à 24nm de diamètre, présente à différentes concentrations dans le noyau durant l'interphase et la mitose.

Si la technique de ChromEMT apparaît comme prometteuse pour identifier les principes de bases du repliement de la chromatine au sein du noyau en partant de l'échelle du nucléosome, elle ne permet pas encore l'étude des interactions entre locus distants spécifiques au sein du génome.

Or l'étude des interactions à distance est fondamentale, car ce sont ces interactions qui vont permettre la régulation de la transcription de tous les gènes, via l'implication de locus non codants, comme les promoteurs et les enhanceurs, et la fixation de facteurs de transcription.

Ces dernières années ont vu le développement de techniques permettant de mettre en évidence ces contacts physiques au sein du génome, en partant de l'échelle la plus grande, le noyau et les chromosomes, pour gagner peu à peu en résolution. Ces techniques, initialement l'Hybridation In Situ en Fluorescence (FISH) puis les techniques dérivées de 3C, ont permis de découvrir au sein du noyau l'existence de territoires chromosomiques, puis de

compartiments chromatiniens, pour enfin mettre en lumière à l'échelle d'une mégabase les Topologically Associated Domains (TAD), ou domaines topologiquement associés. En contrôlant la régulation des gènes, les TADs relient la structure du génome à sa fonction.

Comme tout élément critique du génome, les altérations des TADs, du fait d'un remaniement chromosomique ou d'un défaut de mise en place, sont susceptibles d'entraîner une anomalie du développement d'origine génétique. Ce mécanisme physiopathologique de découverte récente pourrait concerner un nombre significatif de diagnostics en génétique médicale, et l'étude des TADs apparaît désormais comme nécessaire dans le champ des anomalies du développement. Le travail réalisé dans le cadre de cette thèse a donc pour but d'appréhender la méthodologie nécessaire à cette étude, ses résultats possibles et leurs limites, ainsi que les perspectives offertes par la connaissance de l'organisation tridimensionnelle du génome dans le diagnostic des anomalies du développement.

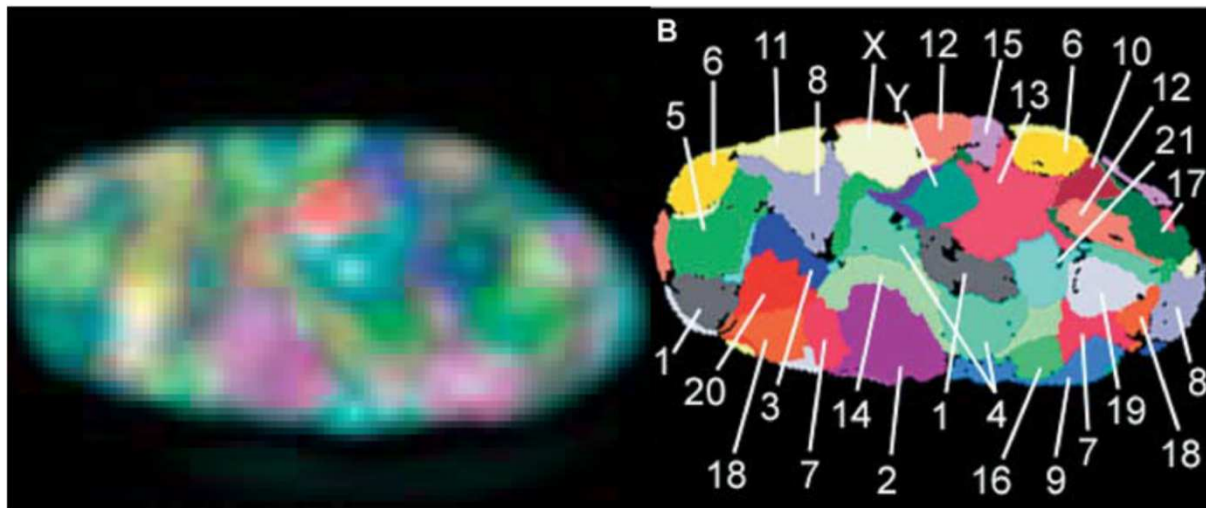
## B. Organisation spatiale du génome

### 1. Territoires chromosomiques

La répartition des chromosomes dans le noyau durant l'interphase a pu faire l'objet d'études depuis la fin des années 1970(16–19). Une des expériences décisives dans ce domaine a été réalisée par Zorn et collaborateurs(17). En induisant des dommages de l'ADN par l'utilisation d'un laser sur une région très limitée du noyau, puis en étudiant par l'incorporation de nucléotides marqués par radioactivité ces régions génomiques au cours de leur réparation, les auteurs de cette étude(17) ont pu mettre en évidence une répartition non aléatoire des régions chromosomiques au sein du noyau.

Les études ultérieures utilisant la FISH(20) ont permis de confirmer l'existence de territoires chromosomiques, délimitant l'espace occupé par un chromosome donné dans le noyau, ainsi que ces rapports avec les autres chromosomes (Figure 5). Ces différents territoires chromosomiques apparaissent comme spécifiques d'un type cellulaire et d'un tissu, et influent sur l'expression génique.





**Figure 5** : FISH 3D de l'ensemble des chromosomes de fibroblastes humains en phase G0

Adapté de Bolzer et collaborateurs(20)

A gauche, image en couleurs RGB obtenue par microscopie. A droite, identification des différents territoires chromosomiques en fausses couleurs.

A une échelle inférieure, environ 1000 à 1500 régions chromosomiques ont été identifiées comme se localisant à proximité de la lamina, bordant la membrane nucléaire(21,22). Ces domaines associés à la lamina (LADs), d'une taille allant de 100kb à 10Mb et représentant environ 35% du génome, sont caractérisés par une faible expression des gènes et une réplication tardive durant la phase S, et correspondent à des régions d'hétérochromatine.

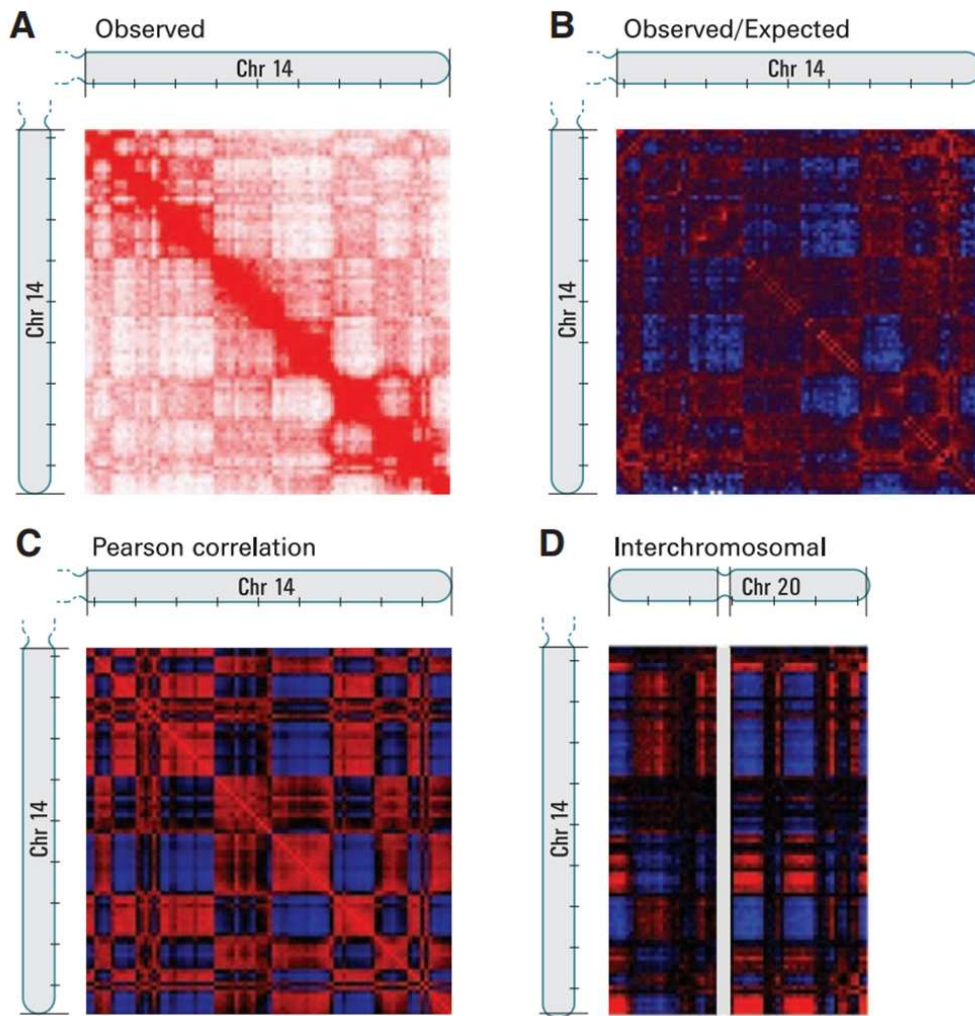
Certaines régions sont par ailleurs associées au nucléole, sous-compartiment nucléaire permettant la transcription des ARN ribosomiques. Ces domaines associés au nucléole, ou NADs, se recoupent partiellement avec les LADs, et représentent eux 4% du génome(23,24).

## 2. Compartiments chromosomiques

En 2009, grâce à la mise au point de la technique de Hi-C par Lieberman-Aiden et collaborateurs(25), des découvertes fondamentales sur l'organisation tridimensionnelle du génome à une échelle inférieure vont être possibles. Cette technique permet d'étudier de façon pangénomique l'ensemble des interactions physiques de l'ADN, intrachromosomiques et interchromosomiques, à une échelle allant de 1Mb à 40 kb, voir 1kb. Les données issues de cette technologie retrouvent la présence de contacts interchromosomiques correspondant aux territoires chromosomiques précédemment identifiés par FISH(19). La première découverte

faite grâce à la technologie de Hi-C appliquée au génome humain est que toutes les régions du génome peuvent être classées comme appartenant à 2 compartiments différents, dénommés arbitrairement compartiments A et B (Figure 6). Les régions appartenant à un compartiment donné vont interagir préférentiellement avec les autres régions du même compartiment situées sur le même chromosome, mais également sur les autres chromosomes.

En étudiant la corrélation entre les 2 compartiments A et B et les données concernant l'épigénétique ainsi que l'expression des gènes, il a été montré que le compartiment A est caractérisé par une forte expression des gènes, une chromatine accessible, une réplication précoce durant la phase S, et une déplétion en LADs et NADs, comparativement au compartiment B. Le compartiment A peut donc être défini comme l'ensemble des régions chromatiniennes actives, tandis que le compartiment B regroupe les régions inactives(25). De fait, le compartiment B constitue l'union de tous les domaines correspondant à l'hétérochromatine, associés à la lamina, au nucléole, ainsi qu'au centromère. Les données de Hi-C de meilleure résolution publiée par Rao et collaborateurs(26) a permis de mettre en évidence l'existence de sous-catégories de compartiments, présentant des caractéristiques génétiques et épigénétiques différentes. Le compartiment A se divise en deux sous compartiments A1 et A2, le sous compartiment A2 étant caractérisé par une présence accrue de marques H3K9me, un contenu plus faible en GC, et contenant des gènes plus longs que le sous compartiment A1. Le compartiment B se divise lui en sous compartiments B1, qui correspondrait à des régions d'hétérochromatine facultative, B2, contenant 62% de l'hétérochromatine péricentromérique, et B3, particulièrement enrichis en LADs. Un sixième sous-compartiment B4 a pu être mis en évidence. Celui-ci n'est présent que sur 11Mb du chromosome 19, et possède une signature épigénétique unique : il correspondrait aux locus des gènes de la superfamille KRAB-ZNF, comprenant 278 protéines dont 130 sur le chromosome 19. Il est important de noter que ces compartiments ne sont pas conservés pas entre les lignées cellulaires.



**Figure 6** : Données de Hi-C obtenues par Liebermann-Aiden et collaborateurs(25)

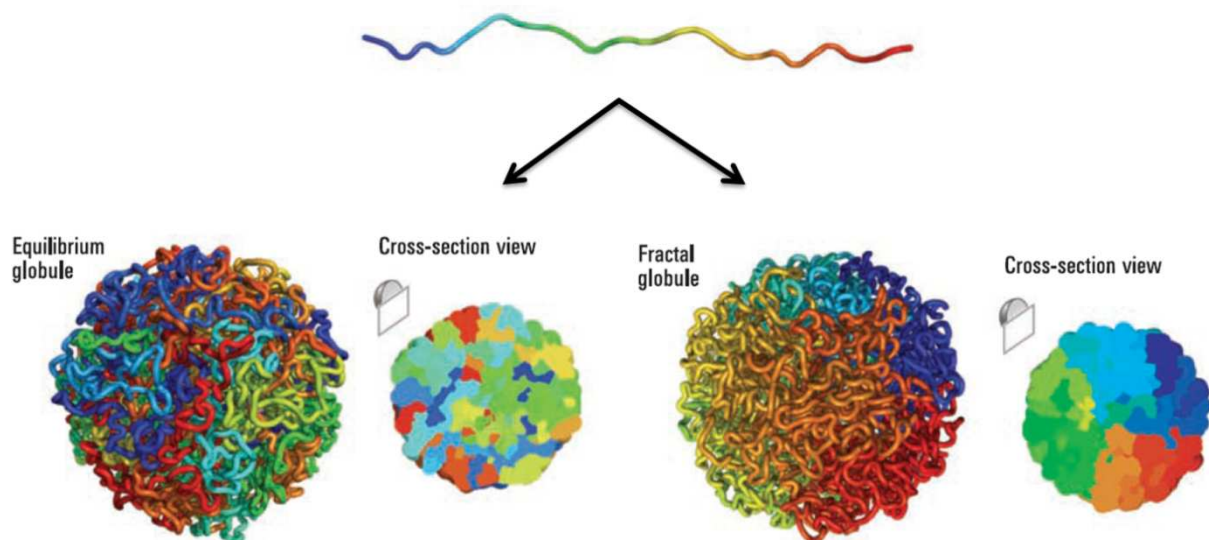
A. En rouge, matrice des interactions intrachromosomiques du chromosome 14, à la résolution de 1 Mb. Chaque pixel rouge correspond à une interaction entre deux locus.

B. Matrice des interactions normalisée par le calcul du ratio du signal observé/attendu ( $>1$  en rouge,  $<1$  en bleu). Le signal attendu correspond à la probabilité d'interaction entre 2 locus du fait de leur distance génomique.

C. Matrice de corrélation de Pearson de ces mêmes données (rouge :  $+1$ , bleu :  $-1$ ). Chaque point correspond à la valeur de la corrélation des données de la ligne et de la colonne correspondantes. A cette échelle, on peut voir clairement une répartition « en tartan », ou « en plaid », correspondant aux interactions privilégiées au sein des compartiments A et B.

D. Matrice de corrélation de Pearson des interactions interchromosomiques entre les chromosomes 14 et 20. La répartition « en plaid » montre que les compartiments A et B interagissent également préférentiellement entre les chromosomes.

Un autre résultat de l'étude de Lieberman-Aiden et collaborateurs(25) concerne la forme globale du repliement de la chromatine au sein d'un compartiment, en étudiant la probabilité d'une interaction entre 2 locus en fonction de leur distance génomique. Des études précédentes basées sur les modélisations mathématiques utilisées en physique des polymères avaient proposé un modèle de repliement appelé globule à l'équilibre (equilibrium globule), correspondant à une chromatine fortement entremêlée(27). La fonction de la probabilité d'une interaction entre 2 locus distants d'après les données de Hi-C obtenues n'est pas compatible avec ce modèle. Au contraire, celle-ci soutient un modèle alternatif appelé globule fractal, proposé en 1988 par Grosberg et collaborateurs(28,29), qui apparaît comme non entremêlé, avec l'existence de parties isolées les unes des autres (Figure 7). Une structure en globule fractal permettrait ainsi à la chromatine de s'étendre et se replier facilement du fait de l'absence de nœud, et d'isoler les différents domaines les uns des autres.



**Figure 7** : Globule à l'équilibre et globule fractal

Adapté de Lieberman-Aiden et collaborateurs(25).

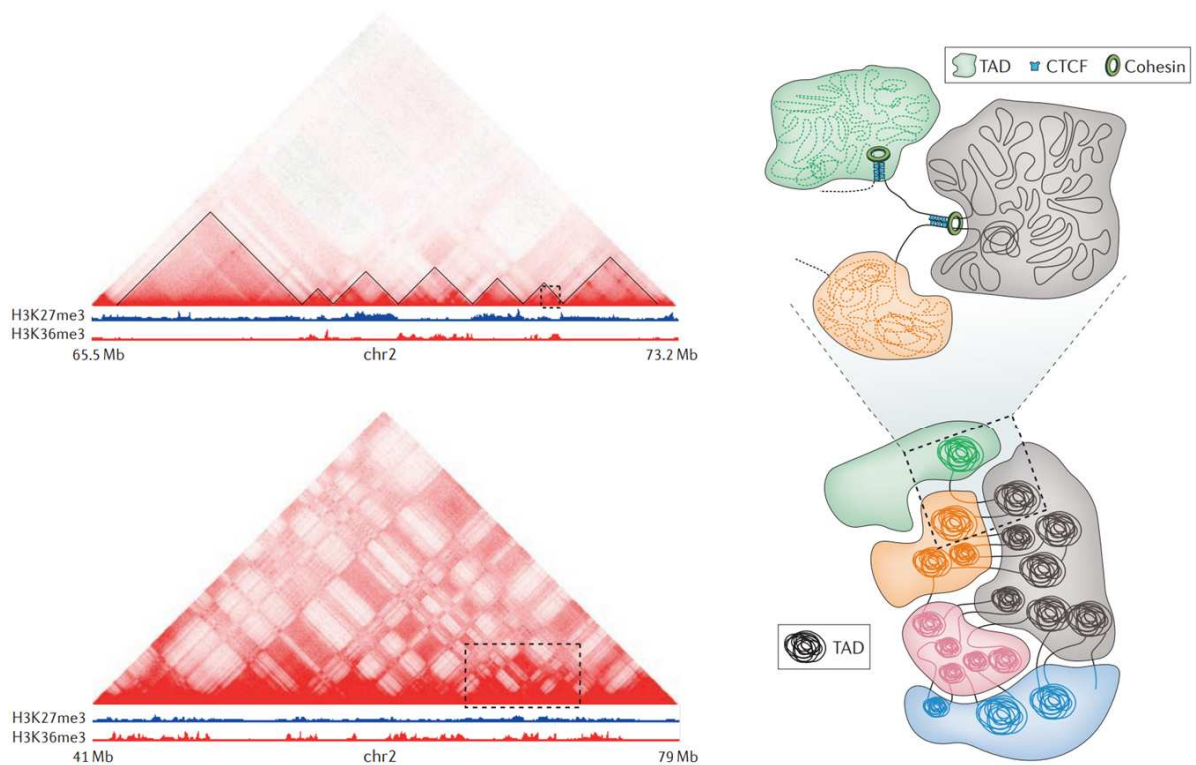
La compaction d'un polymère non replié (en haut) peut se faire selon ces 2 modèles. Le globule à l'équilibre apparaît comme fortement entremêlé, à la différence du globule fractal, où les différentes portions du polymère, en couleurs, sont isolées les unes des autres. Le globule fractal représenterait le mode de repliement de la chromatine à l'échelle de plusieurs mégabases.

### 3. Topologically associated domain (TAD)

A une échelle inférieure de celle des compartiments, aux alentours d'un mégabase, les données de Hi-C mettent en lumière l'existence de structures le plus souvent désignées comme topologically associated domain (TAD), ou encore topologically associating domain (Figure 8).

Un TAD peut être défini comme une région génomique, de plusieurs centaines de kilobases à quelques mégabases, au sein de laquelle les interactions physiques de l'ADN sont fréquentes, tandis que les interactions entre une position au sein du TAD et une située en dehors sont rares(30). Chaque chromosome est ainsi compartimenté en de multiples TADs. Ceux-ci sont stables au cours de plusieurs divisions cellulaires, conservés entre différents types cellulaires, ainsi qu'entre différentes espèces au cours de l'évolution. En raison de ce haut degré de conservation, les TADs sont considérés comme une unité structurelle fondamentale(31). Si les TADs sont conservés entre les lignées cellulaires, les compartiments A et B ne le sont pas. En effet, un TAD peut basculer d'un compartiment à un autre, passant ainsi d'un état de la chromatine inactif à actif par exemple, permettant ainsi une régulation de la transcription des gènes compris dans le TAD. On peut noter par ailleurs qu'il existe de nombreux exemples où un TAD va se superposer entièrement à une région définie comme LAD, même si cette observation n'est pas systématique.

La première étude ayant permis la réalisation d'une cartographie pangénomique des TADs chez l'homme a été réalisée en 2012 par Dixon et collaborateurs(30). Les TADs ainsi détectés ont une taille moyenne de 880kb. Cette étude comprend la réalisation d'expériences de Hi-C sur 3 lignées cellulaires différentes : des cellules embryonnaires murines, humaines, et des fibroblastes humains IMR90, ainsi que l'exploitation de données précédemment obtenues à partir de cortex murins. Les données de Hi-C d'une résolution allant jusqu'à 1kb publiées par Rao et collaborateurs(26) ont permis de préciser cette cartographie des TADs. La taille moyenne des TADs suggérée par cette étude serait inférieure, proche de 185kb, et pouvant aller de 40kb à 3Mb. Cette publication par Rao et collaborateurs(26) s'est particulièrement concentrée sur un sous-groupe de TADs, caractérisés par la présence à leurs bornes d'un signal de Hi-C (« peak ») se démarquant très nettement de bruit de fond aléatoire, synonyme de la présence d'une boucle. De tels TADs sont dénommés domaines à boucle (loop-domain). Ceux-ci représentent environ 39% de l'ensemble des domaines détectés, et pourraient correspondre à un ensemble de TAD particulièrement stables(3).



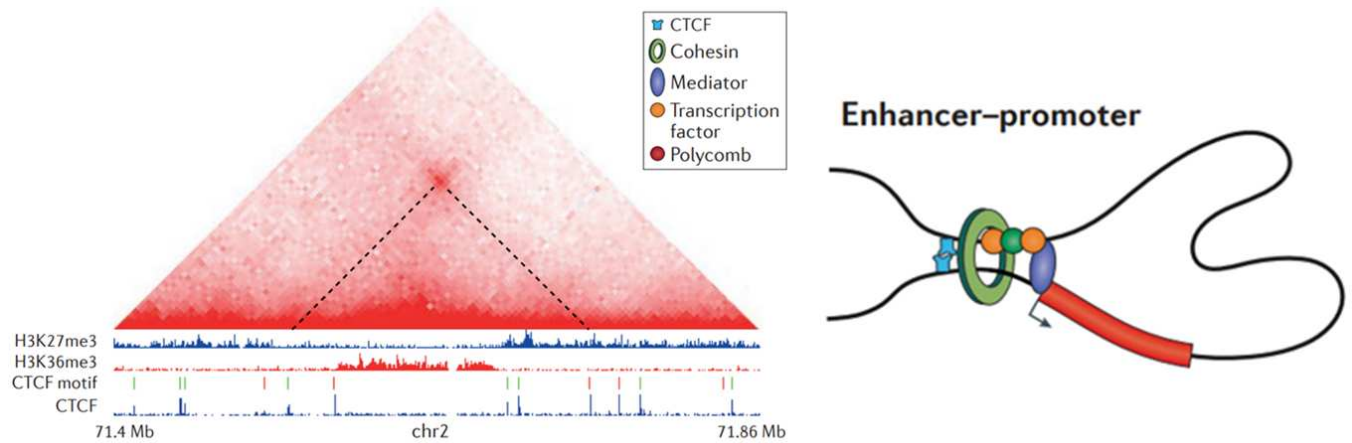
**Figure 8** : TADs et compartiments

Adapté de Bonev et Cavalli(3).

En haut : à gauche, données de Hi-C montrant la répartition d'une région du chromosome 2 en TADs (triangles surlignés en noir) d'une taille proche d'une mégabase. Une matrice d'interactions interchromosomiques d'Hi-C étant « en miroir », centrée sur une diagonale, seule une des deux moitiés est visualisée ici. A droite, illustration de 3 TADs consécutifs, fermés par des boucles via la fixation de cohésine et de protéine CTCF (CCCTC-binding factor) pour deux d'entre eux.

En bas : à gauche, données de Hi-C et illustration à une échelle plus large, montrant la répartition des TADs en compartiments.





**Figure 9** : Boucle locale au sein d'un TAD

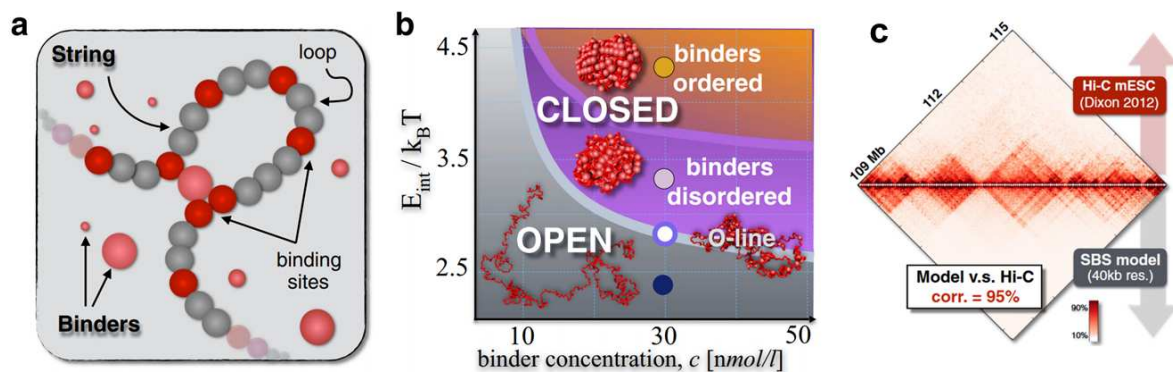
Adapté de Bonev et Cavalli(3).

A gauche, données de Hi-C mettant en évidence une interaction entre deux locus distants d'une centaine de kilobases, au sein d'un TAD plus large. Noter la corrélation des deux points de cette boucle avec la présence d'une paire de motifs CTCF d'orientation convergente, en vert en sens, en rouge en anti-sens. A droite, illustration de cette boucle, permettant par exemple une interaction entre un enhancer et un promoteur.

Les données de Hi-C « classiques » correspondent à l'ensemble des interactions se produisant en moyenne dans une population cellulaire, et non pas dans cellule unique. Afin de déterminer si les structures chromatiniennes se retrouvent dans l'ensemble des cellules à l'échelle individuelle, des protocoles de single-cell Hi-C ont été développés(32,33). L'un des résultats de ces expériences est que la position des TADs est en réalité très variable d'une cellule à l'autre. Cependant, en réalisant la moyenne de données de single-cell Hi-C de différentes cellules, on retrouve les TADs identifiés par les expériences de Hi-C conventionnelles. Un TAD donné n'existe donc pas en tant que tel dans chaque cellule, mais représente plutôt le repliement moyen au sein d'une population cellulaire(34). L'un des éléments ayant notamment été identifié comme influençant les TADs à l'échelle de chaque cellule est la phase du cycle cellulaire dans laquelle elle se trouve(35).

### a) Propriétés des TADs

Ce qui caractérise fondamentalement un TAD est donc une interaction préférentielle de l'ADN au sein du TAD, et une faible probabilité d'interaction avec les TADs adjacents. L'organisation des TADs repose donc sur deux propriétés fondamentales : les régions au sein d'un TAD sont capables de s'auto-associer, tandis que les régions bordant une barrière de TAD sont capables de s'isoler l'une de l'autre. Il est admis que les TADs sont le reflet de la formation de boucles au sein du génome. De façon notable, les modélisations utilisées en physique des polymères de type « Strings and Binders Switch model » (SBS) retrouvent ces deux propriétés d'auto-association et d'isolement comme conséquences de la formation de boucles entre deux monomères, au sein d'un polymère(36)(Figure 10). Ces modélisations permettent notamment de simuler le repliement de la chromatine à un locus donné, avec une corrélation de près de 95% avec les données issues de Hi-C.



**Figure 10** : Modèle Strings and Binders Switch (SBS)

a) Le principe de ce modèle de dynamique moléculaire repose sur la simulation d'un polymère, représentant la chromatine. Ce polymère est composé de monomères (bille grise), dont certains contiennent des sites de fixation (billes rouges, binding sites) pour un élément « binder » (billes roses)

b) En faisant varier lors de la simulation la concentration de binders  $c$ , ainsi que sa capacité de fixation  $E_{int}$  par rapport à la température  $T$ , on obtient différents états du polymère, en chaîne ouverte, en globule désordonnée, ainsi qu'en structure ordonnée pseudo cristalline.

c) Corrélation entre les simulations du modèle SBS et les données de Hi-C obtenues par Dixon et collaborateurs au locus de *Sox9*(30). On observe une corrélation de près de 95% entre les données *in silico* et *in vitro*.



Un des éléments empêchant une identification univoque des TADs d'un locus donné est leur nature intriquée. Plusieurs TADs de tailles différentes peuvent se chevaucher partiellement. Les bornes des TADs peuvent également être dites « nichées » (« nested »), avec de grands meta-TADs regroupant plusieurs TADs, divisés en sous-TADs de quelques kilobases, eux-mêmes présentant plusieurs boucles locales(26). Ces boucles locales peuvent par exemple correspondre à une interaction entre le promoteur d'un gène et un enhancer distant de quelques kilobases (**Figure 9**). A chacune des échelles de taille considérées, les 2 propriétés d'auto-association et d'isolement sont conservées. Une question fondamentale dans ce champ d'étude est de savoir si les TADs d'une part, et ces sous-TADs et boucles locales d'autre part, correspondent aux mêmes mécanismes d'organisation, et ne différeraient que par leur taille, ou s'ils sont fondamentalement différents dans leur structure et dans leurs conséquences fonctionnelles. Un des éléments de réponse est le degré de conservation de ces différents éléments : les TADs, à l'échelle d'un mégabase, sont très conservés entre différentes lignées cellulaires, alors que les sous-TADs et boucles peuvent être spécifiques d'un type cellulaire. Les TADs constitueraient donc une structure fondamentale et constante au sein d'un génome donné, alors que les sous-TADs et boucles pourraient être impliqués dans la différenciation cellulaire, via une régulation spécifique de la transcription génique.

#### b) Formation des TADs

Des études de ChIP-seq(26) ont également pu mettre en évidence de façon très significative la fixation de la protéine CTCF, et des sous-unités de la cohésine RAD21 et SMC3, aux frontières entre les TADs. Cette fixation est également retrouvée à une échelle inférieure au TAD, au sein des boucles locales.

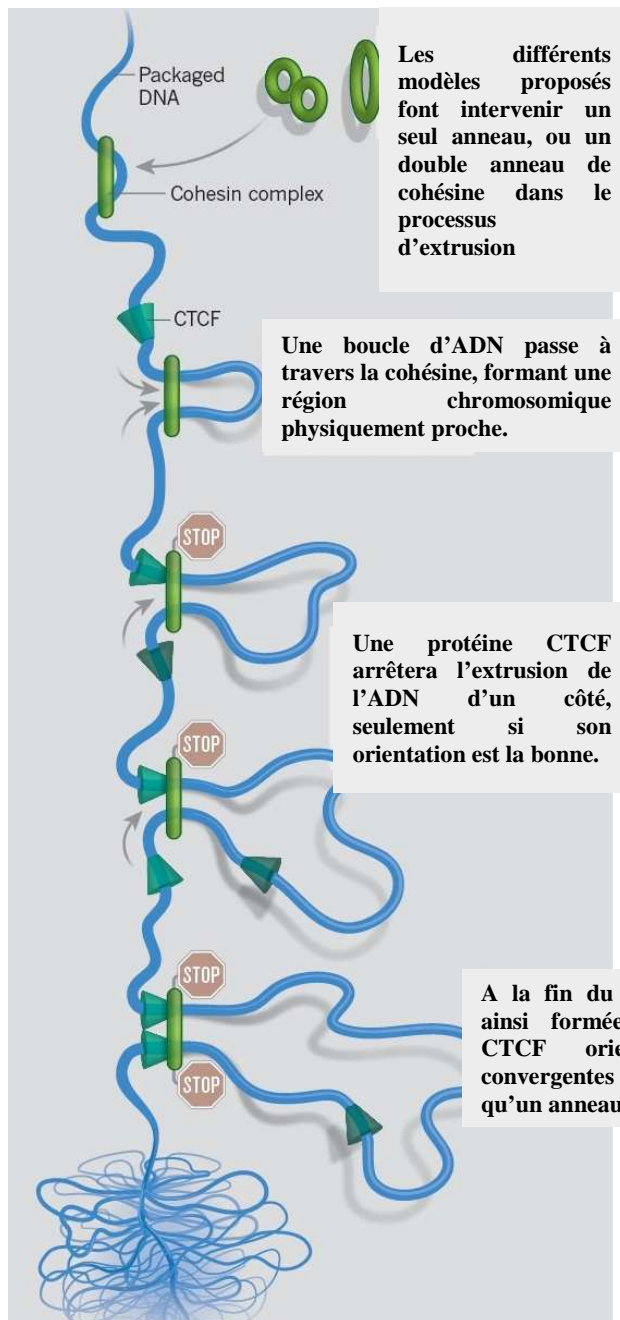
Ce résultat est concordant avec les nombreuses études établissant le rôle de CTCF et de la cohésine dans la formation de boucle chromatinienne(37). La protéine CTCF, pour CCCTC-binding factor, est la principale protéine décrite comme insulatrice chez les vertébrés. Cette protéine a initialement été décrite comme un facteur de transcription capable d'activer et d'inactiver l'expression des gènes, avant que sa faculté à interrompre les interactions entre promoteur et enhancers ne soit mise en évidence(38). La cohésine est un complexe protéique constitué de 4 sous unités (RAD21, SMC1, SMC3 et STAG1 ou STAG2), conservé depuis la levure, et ayant la propriété de lier deux segments d'ADN dans une structure en forme

d'anneau(39). Elle joue un rôle essentiel dans la liaison entre les deux chromatides sœurs d'un chromosome durant la mitose, la méiose, et la réparation par recombinaison homologue.

Le motif de fixation à l'ADN de CTCF est la séquence 5'-CCACNAGGTGGCAG-3'. Cette séquence non palindromique impose donc une orientation aux motifs de fixation, qui peuvent être en sens ou en anti-sens. Les deux frontières d'un TAD donné comportent donc une paire de motifs CTCF, pouvant être dans 4 orientations possibles : les 2 motifs en sens, les 2 motifs en anti-sens, convergente ou divergente. En examinant les bornes des TADs détectés par Hi-C, 92% des paires de motifs CTCF apparaissent comme convergentes(26).

L'existence d'une fixation de paires de CTCF orientées de façon convergente associée à la présence de cohésine a permis l'émergence d'un modèle de formation des TADs dit « par extrusion » de la chromatine. Un anneau de chromatine contiendrait deux brins d'ADN, formant une boucle d'ADN à l'un des côtés. La chromatine se déviderait pour agrandir la boucle, jusqu'à ce que la cohésine rencontre l'un puis l'autre CTCF, fixés de façon convergente(40,41) (**Figure 11**). On observe néanmoins que sur l'ensemble des TADs potentiels du fait de la présence d'une paire de motifs CTCF convergents, seul un petit nombre se forme réellement. D'autres facteurs de régulation, qu'il reste à identifier, doivent donc intervenir dans la détermination de cette organisation chromatinienne.

L'importance des motifs de fixation de CTCF dans la formation des TADs a été vérifiée par des expériences de CRISPR/Cas9 sur des cellules humaines HAP1, réalisant des délétions, ou des inversions de ces motifs(40). La réalisation d'expériences de Hi-C sur ces cellules met en évidence le fait que la perte d'une paire de CTCF orientée de façon convergente entraîne la disparition du TAD qu'elle définissait. De la même façon, l'insertion d'une base dans un motif CTCF (chr5:31,581,788) entraîne la disparition du TAD correspondant(40). Néanmoins, il faut noter qu'il n'a pas encore été établi qu'un tel variant ponctuel entraîne une modification de l'architecture génomique de façon constante.



**Figure 11** : Modèle de formation des TADs par extrusion de l'ADN.

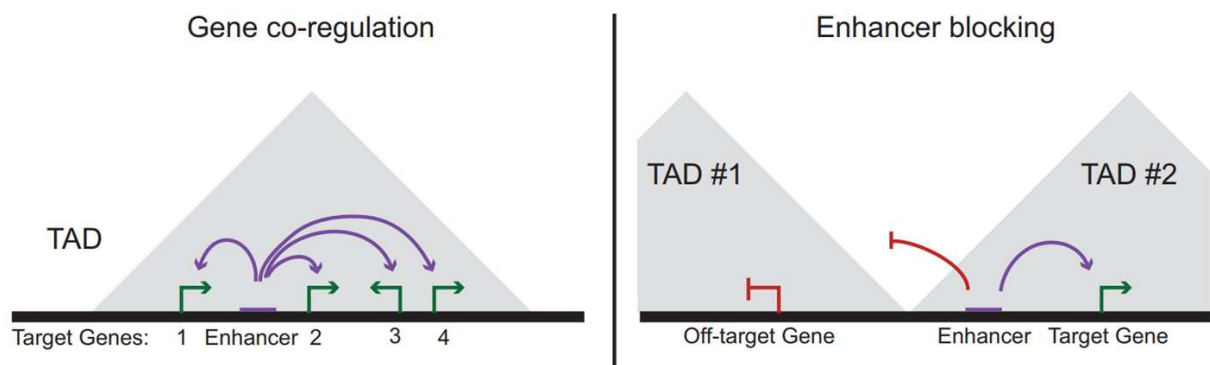
Extrait de Dolgin, « DNA's secret weapon against knots and tangles. », Nature 2017.(42)

Ce modèle fait intervenir la cohésine comme effecteur de la formation des boucles d'interaction au sein du génome, ainsi que des paires de CTCF orientée de façon convergente comme site de fixation.

### c) Fonction des TADs

La compartimentation du génome en TADs va structurer l'ensemble des interactions entre gènes et éléments régulateurs. Au sein d'un TAD, les gènes pourront être régulés par les différents enhanceurs présents, étant donné qu'une interaction via la présence de facteur de transcription sera possible (**Figure 12**). Un certain degré de co-régulation des gènes situés dans un même TAD est possible : il a été montré que différents gènes d'un même TAD peuvent présenter un profil d'expression similaire entre différents types cellulaires(43). Il

existe ainsi plusieurs clusters de gènes paralogues, comme le cluster des protocadhérines(44), ou celui des récepteurs olfactifs, se situant au sein d'un TAD unique, et permettant ainsi une co-régulation de ces différents gènes. Cependant, un TAD n'est pas l'équivalent d'un opéron bactérien : au sein d'un TAD, tous les gènes peuvent ne pas être exprimés simultanément, et cette co-régulation peut n'en concerner qu'une partie seulement(43). A l'inverse, la présence d'une frontière entre 2 TADs rendra impossible l'interaction d'un enhancer avec un gène situé dans un TAD adjacent.



**Figure 12** : Fonctions d'un TAD sur la régulation transcriptionnelle.

Adapté de Dixon et collaborateurs(31).

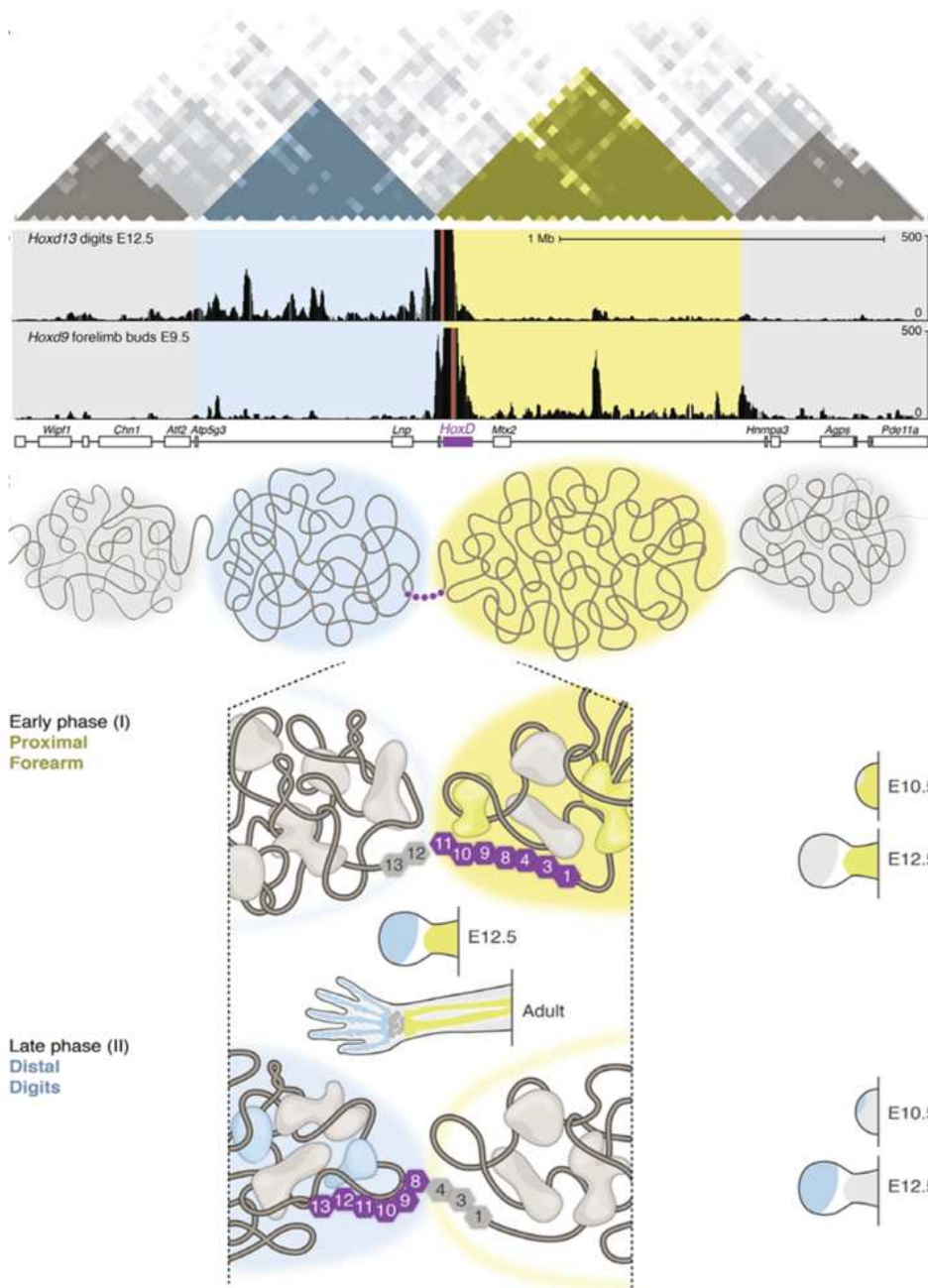
Au sein d'un TAD, les interactions physiques entre locus vont aboutir à une régulation des différents gènes par les enhancers présents, avec un possible effet de co-régulation des gènes par un même enhancer. A l'inverse, un enhancer situé dans un TAD ne pourra pas réguler les gènes situés dans un TAD adjacent.

Ces phénomènes de co-régulation et d'isolement entre les enhancers et les promoteurs vont permettre une régulation spatiale et temporelle de la transcription. Un locus pour lequel cette régulation a été particulièrement étudiée chez les vertébrés est le locus des gènes *HoxD*, appartenant à la famille des gènes homéotiques Hox. Le locus *HoxD* comprend 9 gènes (*HoxD1*, 3, 4, 8, 9, 10, 11, 12, 13), exprimés de manière colinéaire durant le développement des membres chez les vertébrés, en deux phases distinctes. Durant une première phase, les gènes de *HoxD1* à *HoxD11* sont exprimés dans le bourgeon en développement, permettant la formation de la partie proximale des membres, comprenant le bras et l'avant-bras pour le membre supérieur. Durant une seconde phase, les gènes de *HoxD8* à *HoxD13* sont exprimés dans la partie distale du bourgeon, et sont nécessaires à la formation des doigts.

L'étude de l'organisation chromatinienne et des éléments régulateurs de ce locus montre qu'il est organisé en 2 TADs distincts, chacun de ces TADs comprenant plusieurs enhancers situés dans des déserts géniques(45,46)(**Figure 13**). Les gènes *HoxD* sont eux situés exactement à la frontière entre ces deux TADs, *HoxD1* étant proche du TAD télomérique, et *HoxD13*, à l'autre bout du cluster de gènes, proche du TAD centromérique.

Durant la première phase, la chromatine située dans le TAD télomérique présente une conformation active, permettant la transcription des gènes de *HoxD1* à *HoxD11*, sous le contrôle des enhancers situés dans ce TAD. Le TAD centromérique est durant cette phase dans une conformation inactive. Durant la seconde phase, la conformation de la chromatine s'inverse entre les deux TADs : le TAD centromérique devient actif, et permet la transcription des gènes *HoxD8* à *HoxD13* sous le contrôle de ses enhancers, tandis que le TAD télomérique est inactif. On remarque que les gènes centraux du cluster, principalement *HoxD9*, *10* et *11*, sont exprimés durant les deux phases, mais sous la dépendance d'enhancers différents, appartenant successivement à l'un puis à l'autre TAD. Ce phénomène de bascule entre deux TADs d'un même gène situé à leur frontière ajoute donc un niveau de complexité supplémentaire au modèle de régulation de la transcription précédemment décrit.

En plus de leur rôle dans la régulation de l'expression génique, les TADs sont également des unités de régulation temporelle de la réplication. Le génome des eucaryotes se réplique durant la phase S selon un programme spatio-temporel strictement régulé(47). Dès 1958, il a été mis en évidence par Taylor l'existence de domaines de réplication, définis comme des segments chromosomiques incorporant de la thymidine- $H^3$  de façon asynchrone durant la phase S(48). Ces différents domaines contiennent eux-mêmes plusieurs réplicons, correspondant à un segment génomique se répliquant à partir d'une seule origine de réplication. Certains domaines de réplication sont caractérisés par une réplication précoce durant la phase S, tandis que d'autres présentent une initiation tardive de la réplication. Pope et collaborateurs(49) ont pu démontrer que les TADs se superposent à ces domaines de réplication. Ces propriétés de réplication précoce ou tardive sont à relier à l'appartenance d'un TAD donné au compartiment A ou B pour le type cellulaire étudié.



**Figure 13** : Régulation de la transcription au locus HoxD chez la souris

D'après Lonfat et Duboule(46)

En haut, données de Hi-C au locus Hoxd, et de 4C centrée sur *Hoxd13* et sur *Hoxd9*.

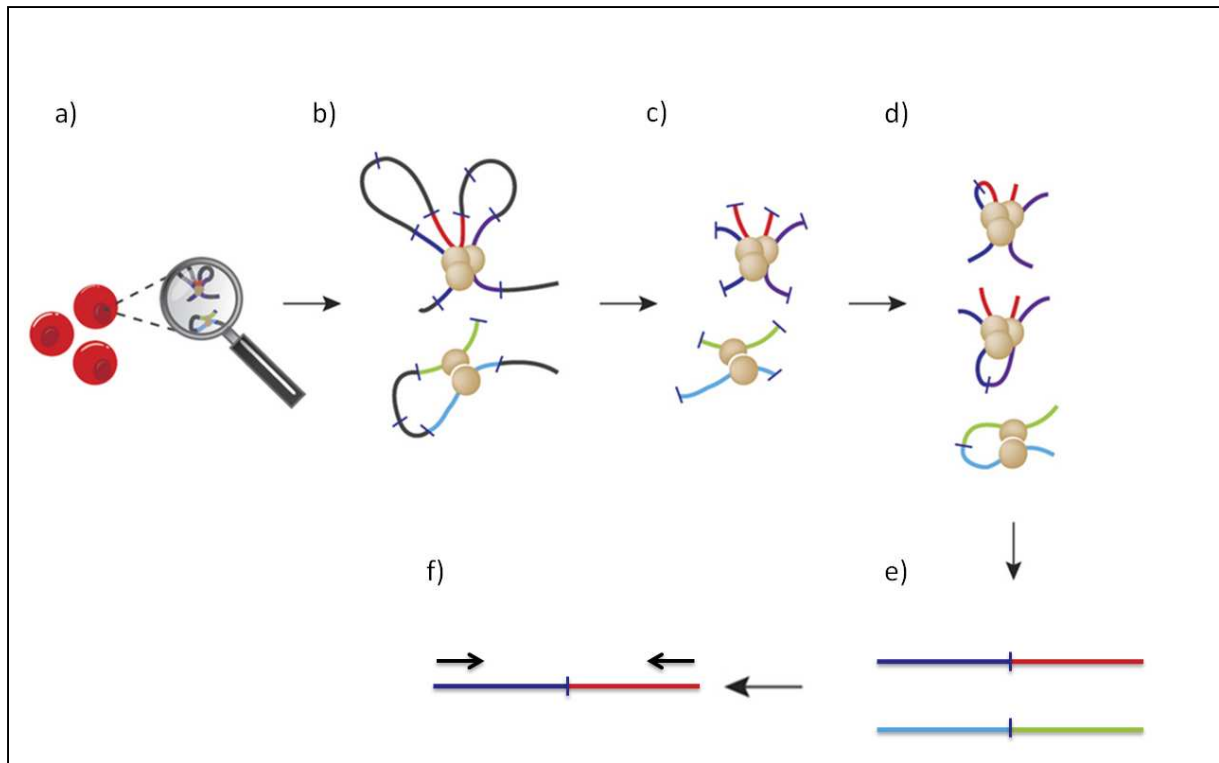
En bas, représentation de la conformation possible de la chromatine à ce locus. Durant la première phase, le TAD télomérique, en jaune, est actif, et permet une interaction entre ses enhancers et les gènes de *Hoxd1* à *Hoxd11*. Cette phase permet la formation du bras et de l'avant-bras. Durant la seconde phase, le TAD centromérique, en bleu, devient actif, et permet l'expression des gènes de *Hoxd8* à *Hoxd13*, et la formation des doigts.

#### 4. Études des TADs par Chromosome Conformation Capture (3C) et techniques dérivées

##### a) Technique de Chromosome Conformation Capture (3C)

La première mise au point d'une technique permettant de mettre en évidence une interaction physique entre un promoteur et un enhancer distant a été réalisée en 1993, avec l'étude de la régulation du gène de prolactine chez le rat(50). Basée sur une digestion enzymatique et la ligation de fragments physiquement proches, elle a par la suite été améliorée en 2002 pour aboutir à la technique actuelle de Chromosome Conformation Capture, ou 3C(51). Cette première étude par Dekker et collaborateurs a eu pour but de déterminer l'organisation spatiale du chromosome III chez la levure.

Les expériences de 3C sont basées sur 4 étapes successives (**Figure 14**). La première étape consiste en une fixation *in vivo* de cellules au formaldéhyde, permettant de fixer les interactions entre régions génomiques spatialement proches, via la création de liaisons dites « cross-links ». On procède ensuite à la digestion de l'ADN par une enzyme de restriction : celle-ci est choisie en fonction du locus d'intérêt, et des fragments de restrictions qui seront générés. Après digestion, on réalise une ligation des extrémités des fragments. Les produits de ligation vont ensuite être amplifiés par PCR spécifique, pour être détecté après migration sur gel d'agarose. La réalisation d'une quantification par qPCR TaqMan(52), est également possible.



**Figure 14** : Schéma général d'une expérience de 3C

Adapté de Stadhouders et collaborateurs(53)

**a) et b)** Cross-link, et lyse des cellules      **c)** Digestion par enzyme de restriction      **d)** Ligation des extrémités des fragments      **e)** Purification de l'ADN      **f)** Détection des fragments de ligation par PCR spécifique

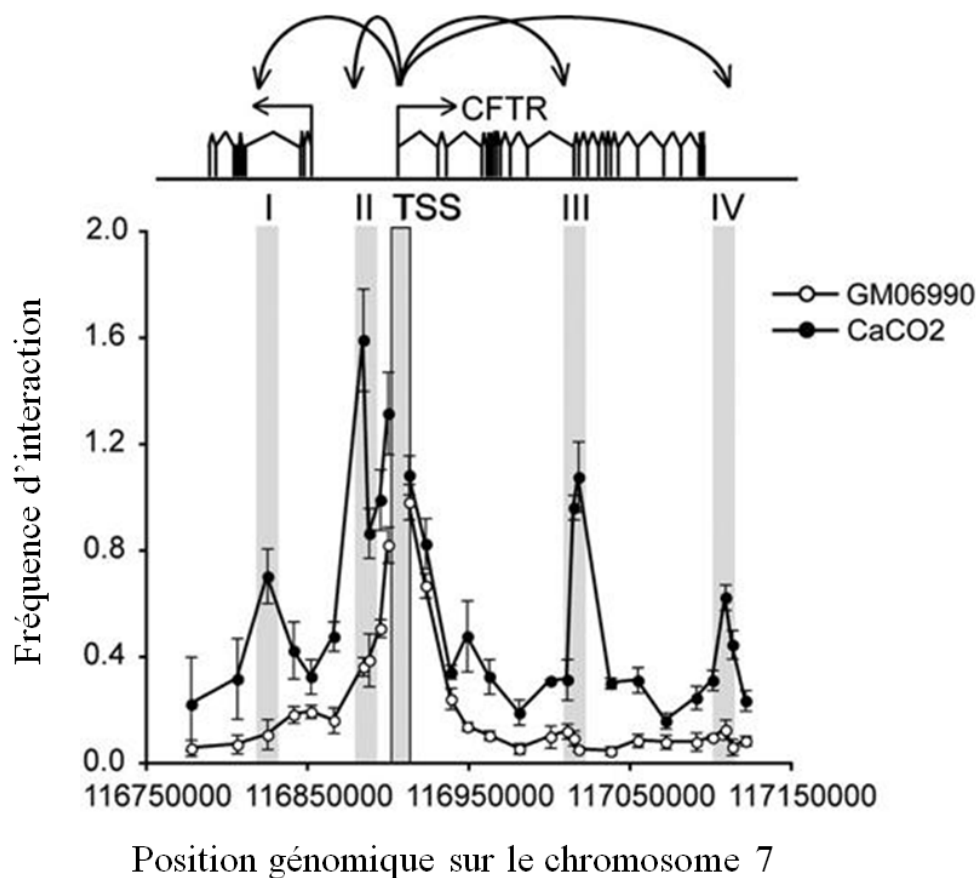
Une expérience de 3C permet donc de mettre en évidence une proximité physique entre deux régions génomiques distantes, déterminées à l'avance. L'analyse quantitative des interactions entre un promoteur et différents enhanceurs potentiels nécessite donc de réaliser plusieurs expériences de 3C successives (**Figure 15**). Par ailleurs, afin de permettre une comparaison entre ces différentes expériences comportant des PCRs d'efficacité variable, des expériences contrôles doivent être réalisées. Celles-ci utilisent généralement le chromosome bactérien artificiel (BAC) correspondant au locus d'intérêt. Les BAC n'ayant pas de conformation tridimensionnelle préférentielle, les expériences de 3C réalisées à partir de ceux-ci permettent d'obtenir une quantification de l'efficacité de chaque PCR, qui servira à la normalisation des données obtenues(54).

Par ailleurs, il faut noter qu'il se produit des interactions physiques aléatoires entre les régions génomiques d'un même brin, avec une fréquence d'interaction inversement proportionnelle à



la distance génomique entre ces 2 régions. Avant de conclure à la détection d'une interaction spécifique entre 2 loci par 3C, il faut donc s'assurer que ces 2 loci interagissent d'avantage entre eux qu'avec les régions génomiques voisines.

La technique de 3C est particulièrement bien adaptée pour étudier des interactions entre loci génomiques distants de quelques kilobases, même si des études ont été réalisées sur des distances plus importantes, jusqu'à 1 Mb(55,56). Cette limite de taille, ainsi que la nécessité de réaliser de nombreuses expériences indépendantes afin d'étudier toutes les interactions possibles au sein d'un locus, a motivé la mise aux point de techniques plus performantes, dérivées de la 3C.








**Figure 15** : Expériences de 3C au locus CFTR

Ces différentes expériences permettent de visualiser la fréquence d'interaction entre le site d'initiation de la transcription de *CFTR* et différents fragments d'une région de 460kb, sur 2 lignées cellulaires(57). On peut noter que 4 fragments présentent une interaction privilégiée avec le site d'initiation de la transcription. L'étude de ces 4 sites a révélée qu'il s'agit de sites d'hypersensibilité à la DNase I et de modifications d'histones, caractéristiques de la présence d'enhancers.

## b) Techniques de 4C et de 5C

Toutes les techniques dérivées de la 3C ont en commun d'utiliser les 3 premières étapes de la 3C, à savoir le cross-linking, la digestion et la ligation(58) (**Figure 16**).

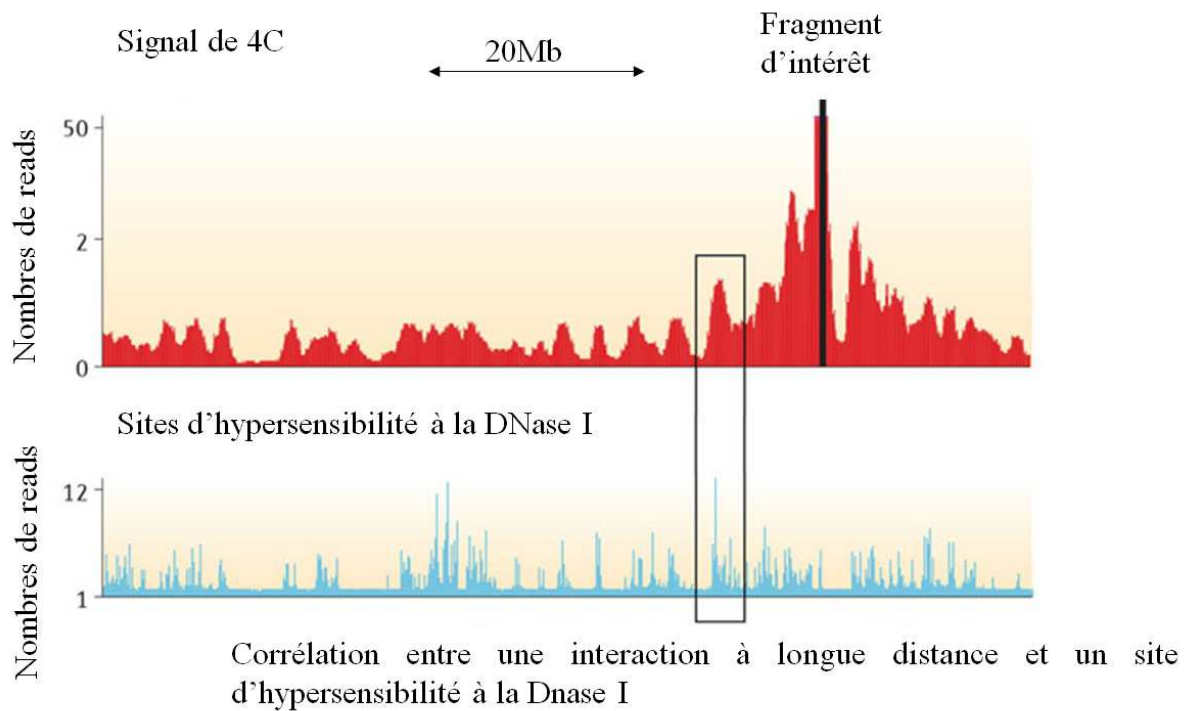
Technique	Fragments étudiés	Méthode de détection privilégiée
3C	Un contre un	qPCR 
4C	Un contre tous	iPCR et séquençage 
5C	Beaucoup contre beaucoup	Multiplex LMA et séquençage 
Hi-C	Tous contre tous	Séquençage 
ChIA-PET	Beaucoup contre beaucoup, via un FT	ChIP-seq 

**Figure 16** : Résumé des méthodes dérivées de Chromosome Conformation Capture.

Adapté de Krijger et collaborateurs(59)

Ces différentes méthodes diffèrent sur le nombre de fragments étudiés, et sur la méthode de détection des produits de ligation.

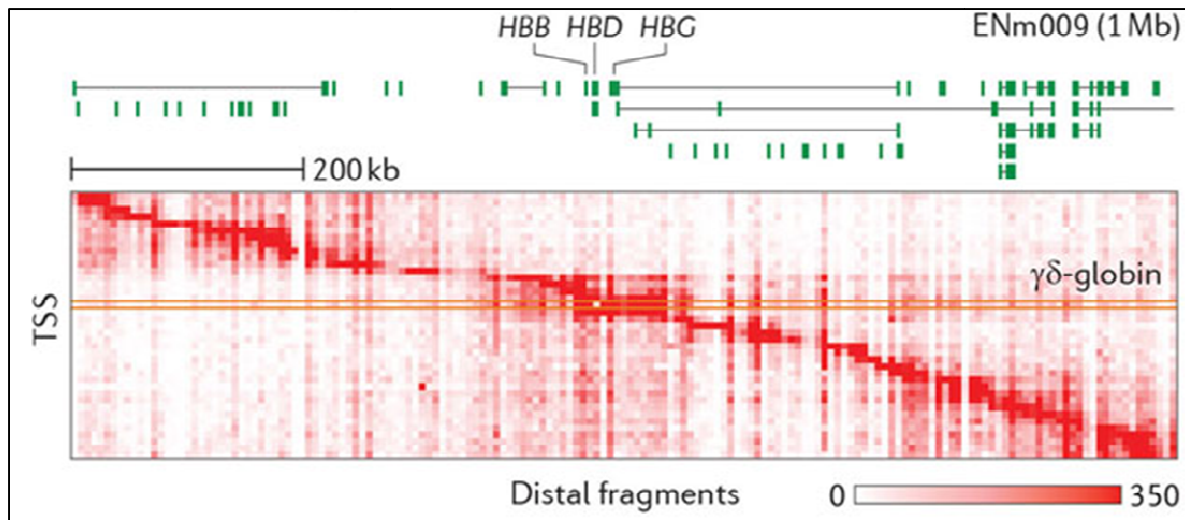
La technique de Chromosome Conformation Capture-on-Chip, ou 4C, permet de détecter en une seule expérience l'ensemble des régions interagissant physiquement avec un fragment de restriction unique(60,61) (**Figure 17**). La 4C permet de détecter à la fois les contacts intra-chromosomiques, mais également inter-chromosomiques. La 4C est basée sur l'obtention de produits de ligation circulaires, et leur amplification par PCR inverse utilisant une seule paire de primers spécifiques du fragment d'intérêt. Les fragments amplifiés sont par la suite caractérisés par puce à ADN, ou séquençage nouvelle génération. Cette technique a été initialement validée par l'étude du locus de la beta-globine(60) chez la souris.



**Figure 17** : Données de 4C

Adaptée de Dekker et collaborateurs(58). Ces données sont issues du génome murin, simulées à partir des données d'Imakaev et collaborateurs(62). Le signal de 4C est représentatif des interactions entre le fragment d'intérêt, et les différents fragments de la région génomique l'entourant.

La technique de Chromosome Conformation Capture Carbon Copy, ou 5C, est l'équivalent de multiples expériences de 3C multiplexées(63). Les premières étapes consistent en la fabrication d'une librairie de 3C classique. Grâce à l'utilisation d'une amplification multiplexe dépendante d'une ligation, ou LMA, les produits de ligation correspondant aux régions d'intérêt sont spécifiquement amplifiés. La détection des interactions entre ces différentes régions se fait par la suite via l'utilisation de puce à ADN, ou plus communément par technique de séquençage nouvelle génération. Une seule expérience de 5C permet donc de définir tous les contacts possibles entre les différents fragments de restriction constituant un locus donné (**Figure 18**). Le contrôle de la validité de cette technique a été initialement fait par la détection d'interactions précédemment identifiées au sein du locus de la beta-globine chez l'humain(63).



**Figure 18** : Données de 5C

Adaptée de Dekker et collaborateurs(58).

Ces données permettent de visualiser les interactions entre différents sites d'initiation de la transcription, et différents fragments, au sein du locus de la beta-globine(64). Chaque ligne correspond au profil d'interaction entre un site d'initiation de la transcription et l'ensemble du locus.

### c) Techniques de Hi-C et ChIA-PET

La technique de Hi-C, développée à partir de 2009 par Lieberman-Aiden et collaborateurs(25), est issue de l'utilisation des technologies de séquençage à haut débit appliquées aux techniques de Chromosome Capture Conformation. Elle permet l'identification de l'ensemble des interactions chromatinienne, intra et inter-chromosomiques, et ce à l'échelle du génome entier. Il s'agit à l'heure actuelle de la technique de référence d'étude de l'organisation spatiale du génome à l'échelle des TADs, ou à une échelle supérieure(65). Si les expériences de Hi-C sont généralement réalisées sur une population cellulaire, des protocoles modifiés ont été publiés, permettant d'utiliser cette même technique sur cellule unique (single-cell Hi-C)(32,34,35). Ces différentes méthodes reposent sur l'ajout d'une étape de cytométrie en flux, de barcoding ou d'isolement manuel, permettant d'obtenir des données sur les noyaux à l'échelle individuelle.

Après réalisation du cross-link et de la digestion, étapes communes à la 3C, les extrémités des fragments de restriction sont biotynilées. Après une étape de ligation de ces extrémités, les fragments biotynylés sont capturés via l'utilisation de billes magnétiques recouvertes de streptavidine. L'ensemble de ces fragments sont ensuite séquencés par technologie Next-Generation Sequencing (NGS) « paired-end ». La taille, la complexité du génome étudié et le nombre de reads obtenus vont influencer la résolution à laquelle l'étude des interactions chromatinienne sera possible. Ainsi, l'étude du génome humain à une résolution de 40kb nécessite 100 millions de reads(66).

Dans les suites du séquençage à haut-débit de la librairie de Hi-C, une analyse bio-informatique complexe des données est nécessaire. La première étape consiste en un alignement des reads obtenus sur le génome de référence. Les différents aligneurs habituels peuvent être utilisés (Bowtie, BWA-MEM, etc...), à condition de ne pas utiliser le mode d'alignement paired-end. En effet, ce mode d'alignement suppose que les 2 reads d'une paire se situent sur des segments contigus du génome, ce qui n'est pas le cas pour les produits de ligation obtenus dans ce type d'expérience.

Chacun des reads mappés est ensuite assigné au fragment de restriction auquel il correspond, en fonction de l'enzyme de restriction utilisée dans le protocole. Pour une paire de reads donnée, il existe ainsi deux possibilités : soit les deux reads ont été alignés sur le même fragment de restriction, soit ils se situent sur des fragments différents. Les paires de reads ainsi situées sur le même fragment de restriction sont filtrées. Ils représentent soit des fragments non issus d'une ligation, soit d'une ligation circulaire : dans ces deux cas, ils ne sont pas informatifs. On procède également à un filtrage des duplicats de PCR sur les reads restants.

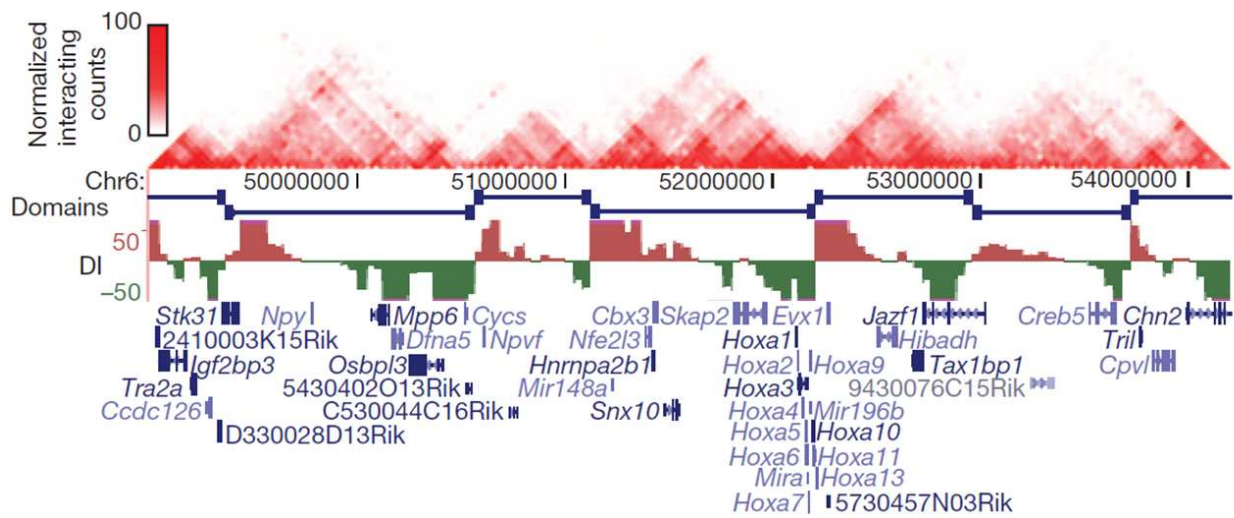
A ce stade, chaque paire de reads est donc le reflet d'une proximité physique entre les deux fragments de restriction sur lesquels ils ont été alignés. Cependant, le nombre de reads générés par les actuelles technologies de séquençage ne permettent pas d'obtenir une résolution correspondant à la taille de chacun des fragments de restriction. Le génome de référence est donc divisé en régions de taille égale, ou bin (intervalle), et les reads de chaque fragment de restriction sont assignés au bin correspondant. Cette étape de binning permet de réduire la complexité des données, et d'augmenter le ratio signal sur bruit. La taille d'un bin

varie généralement de 40kb à 1 Mb, et les bins ayant un signal trop faible sont filtrés. Différentes normalisations doivent également être appliquées, afin de corriger l'effet des régions répétées du génome et du contenu en GC. On procède ensuite à la construction d'une matrice correspondant à l'ensemble des interactions bin à bin.

Chaque point correspond à la fréquence d'interaction entre 2 bins de la région étudiée. Les interactions inter-chromosomiques sont également visualisables.

Dans le cas de l'analyse des interactions intra chromosomiques, cette matrice est en miroir centrée sur une diagonale. Le plus souvent, seule la moitié supérieure droite est représentée.

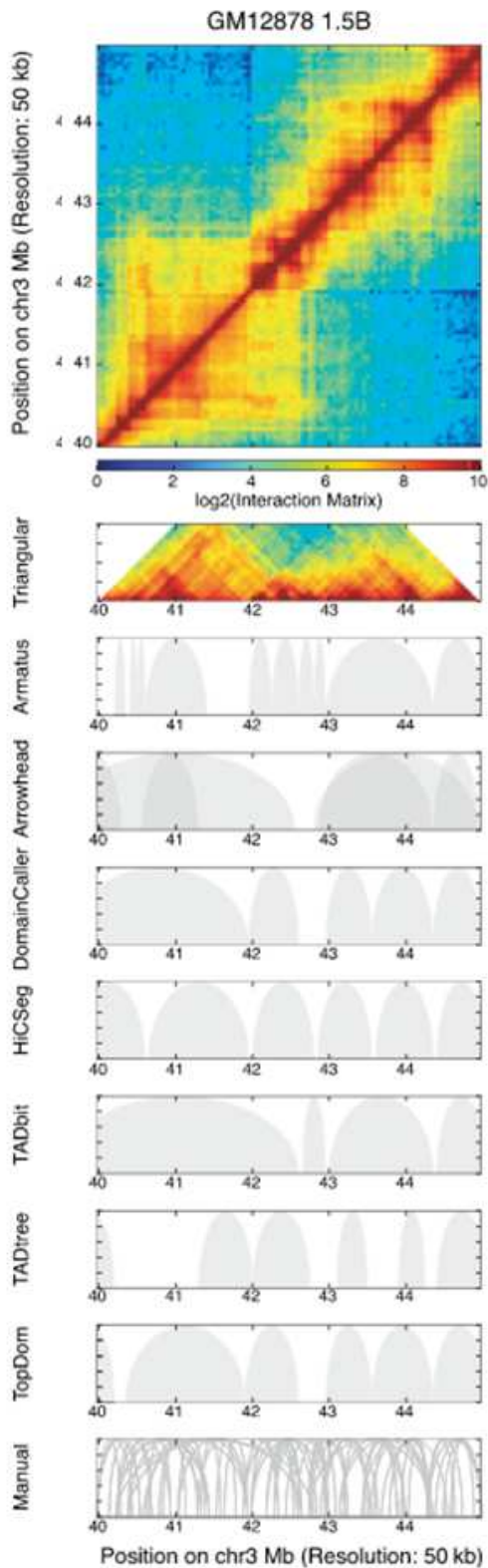
A l'échelle d'une région génomique de quelques mégabases, les TADs apparaissent sur les données de Hi-C comme des régions contiguës au sein desquelles les bins interagissent plus les uns avec les autres qu'avec les bins situés en dehors (**Figure 19**). De nombreux algorithmes permettent de réaliser une annotation des TADs. Cependant, comme précédemment expliqué, la nature intriquée des TADs sur différentes échelles rend leur annotation peu reproductible entre 2 algorithmes. Deux évaluations récentes des différents outils disponibles par Dali et collaborateurs(67) et par Forcato et collaborateurs(68) concluent toutes deux à la non supériorité d'un outil sur un autre (**Figure 20**). De plus, l'annotation manuelle d'une région par l'analyse visuelle d'une matrice de données d'Hi-C n'est pas forcément inférieure à l'utilisation d'algorithmes.



**Figure 19** : Données de Hi-C au locus *Hoxa* chez la souris(30)

DI : index directionnel, définit comme la différence entre les interactions en aval et en amont d'un locus donné. Un index directionnel élevé correspond à une interaction fréquente d'un bin avec les bins situés en aval, et une absence d'interaction avec les bins situés en amont. L'alternance d'un bin à index directionnel négatif et d'un bin à index directionnel positif signe donc un isolement entre ces deux bins, et donc potentiellement à une barrière de TAD.

En bleu foncé, les différents TAD annotés, via l'application d'un modèle de Markov caché sur les valeurs de l'index directionnel.



**Figure 20** : Comparaison de différents algorithmes de détection des TADs

Adapté de Dali et collaborateurs(67).

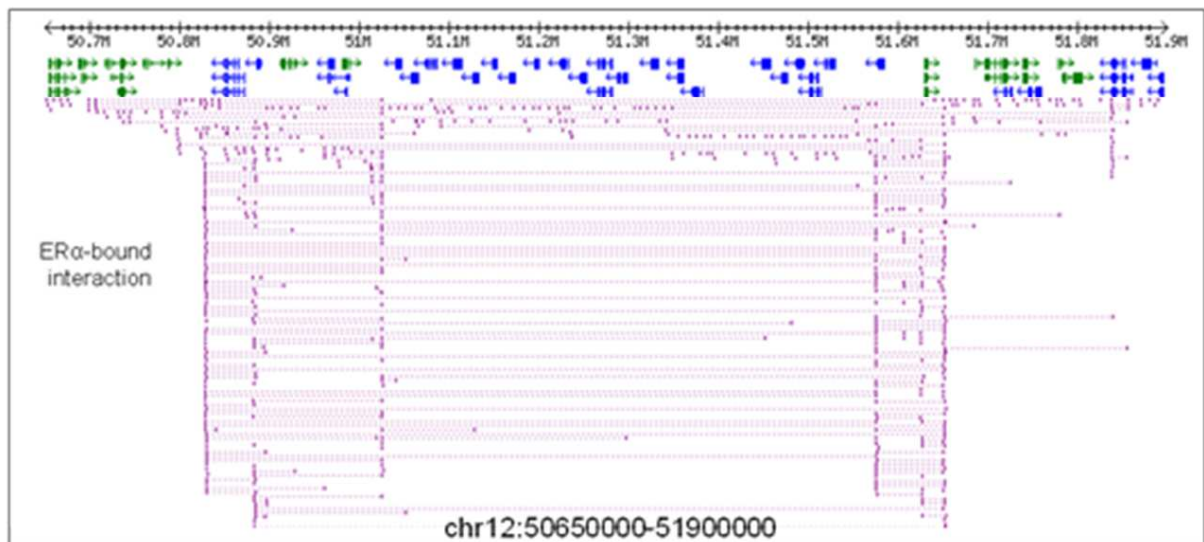
Comparaison de la détection des TADs entre 7 algorithmes différents, et une annotation manuelle(67). Les données de Hi-C utilisées sont issues de la publication de Rao et collaborateurs(26).

Pour exemple, l’algorithme initialement développé dans cette publication est Arrowhead, tandis que l’algorithme développé par Dixon et collaborateurs(30) basé sur l’index directionnel est DomainCaller.

Chaque algorithme a des présupposés différents sur la nature des TADs, concernant leur taille, le signal de Hi-C correspondant, la possibilité de chevauchements, la possibilité d’intrication ou « nesting », etc... expliquant les différences observées.



La technique de ChIA-PET est dérivée de celle de Hi-C, et permet d'analyser les interactions chromatinienne médiées par la présence d'un facteur de transcription d'intérêt(69). Après fixation de la chromatine par le formaldéhyde, l'ADN est fragmenté par sonication. Via l'utilisation d'un anticorps spécifique, les complexes d'interactions liés au facteur de transcription d'intérêt sont enrichis dans l'échantillon par réalisation d'une immunoprécipitation de la chromatine. Après ligation et biotinylation, les fragments obtenus sont séquencés selon la même méthode que la Hi-C. Cette technique a été initialement utilisée afin d'étudier les interactions chromatinienne médiées par une fixation du récepteur aux œstrogènes alpha (**Figure 21**).



**Figure 21** : Données de ChIAPet

Ces données correspondent à une étude des interactions médiées par le récepteur aux œstrogènes alpha au locus des gènes de la kératine(69).

En bleu et vert, les différents gènes du locus. En rose les paires de locus interagissant sont en traits épais, rejoints par des pointillés.

## 5. TADS et anomalies du développement : études issues de la littérature

### a) Remaniements chromosomiques et altérations des TADs

Si les TADs sont conservés entre les espèces et entre les lignées cellulaires, les remaniements chromosomiques peuvent altérer cette organisation de la chromatine à un locus donné, et ainsi avoir des conséquences fonctionnelles pouvant aboutir à la survenue d'anomalies du développement.

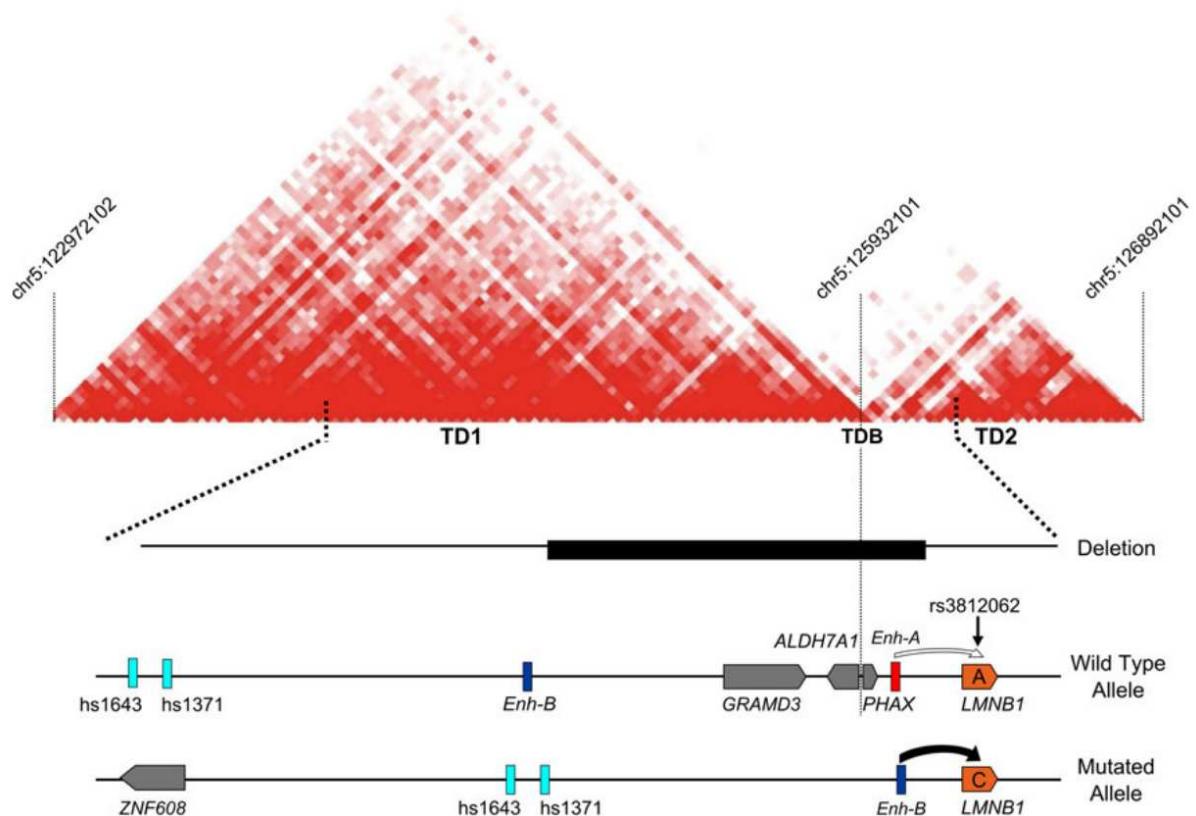
Tous les types de remaniements peuvent potentiellement induire une modification de l'organisation chromatinienne : les délétions, les duplications, les translocations, ainsi que les inversions. Dans tous les cas, l'élément déterminant permettant de prédire l'impact d'un variant de structure sur les TADs et donc sur la régulation des gènes du locus va être l'implication d'une des barrières entre 2 TADs.

#### (1) Délétions

Une délétion dont les bornes sont entièrement comprises au sein d'un TAD, soit une délétion intra-TAD, ne modifiera pas l'architecture globale du locus. L'effet potentiel d'une telle délétion est simplement une haplo-insuffisance des gènes délétés. Néanmoins, si une délétion comprend une barrière entre 2 TADs, l'effet potentiel est une fusion de ces 2 TADs formant un néo-TAD, où des enhancers et des promoteurs de gènes, précédemment isolés les uns des autres, vont pouvoir interagir de façon ectopique(70). L'effet pathogène de telles délétions inter-TADs a été mis en évidence dans différentes pathologies.

Un exemple de délétion inter-TADs pathogène a été rapporté par Giorgio et collaborateurs(71), avec une étude portant sur une grande famille italienne multigénérationnelle dans laquelle ségrège une leucodystrophie autosomique dominante de l'adulte. La cause moléculaire la plus fréquente de cette pathologie est une duplication hétérozygote du gène *LMNB1*, codant pour la lamine B-1. Cette duplication aboutit à une surexpression de *LMNB1*, et entraîne une réduction de l'expression de *PLP*, codant pour la protéine protéolipide, composant primaire de la myéline. Les CGH-arrays réalisées chez les membres atteints de cette famille italienne n'ont pas retrouvé la duplication classique de *LMNB1*. Cependant, une délétion de 660kb située à 66kb en amont de *LMNB1* a été mise en évidence. L'analyse des données de Dixon et collaborateurs(30) montre que ce locus est

organisé en deux TADs, avec une barrière située à 145kb en amont de *LMNB1* (**Figure 22**). En supprimant la barrière entre les 2 TADs, la délétion présente chez cette famille a eu pour conséquence l'apparition d'une interaction ectopique entre un enhancer, dénommé *Enh-B*, et le promoteur de *LMNB1*, entraînant une surexpression.



**Figure 22** : Leucodystrophie autosomique dominante de l'adulte par délétion d'une barrière de TAD

Adapté de Giorgio et collaborateurs(71).

En haut, données de Hi-C au locus de *LMNB1*, organisé en 2 TADs : TD1 et TD2, séparés par une barrière (TDB). En dessous, position de la délétion pathogène par rapport à l'allèle Wild Type, et structure de l'allèle muté en résultant. La délétion de la barrière entre les 2 TADs entraîne l'interaction de l'enhancer *Enh-B* avec *LMNB1*.

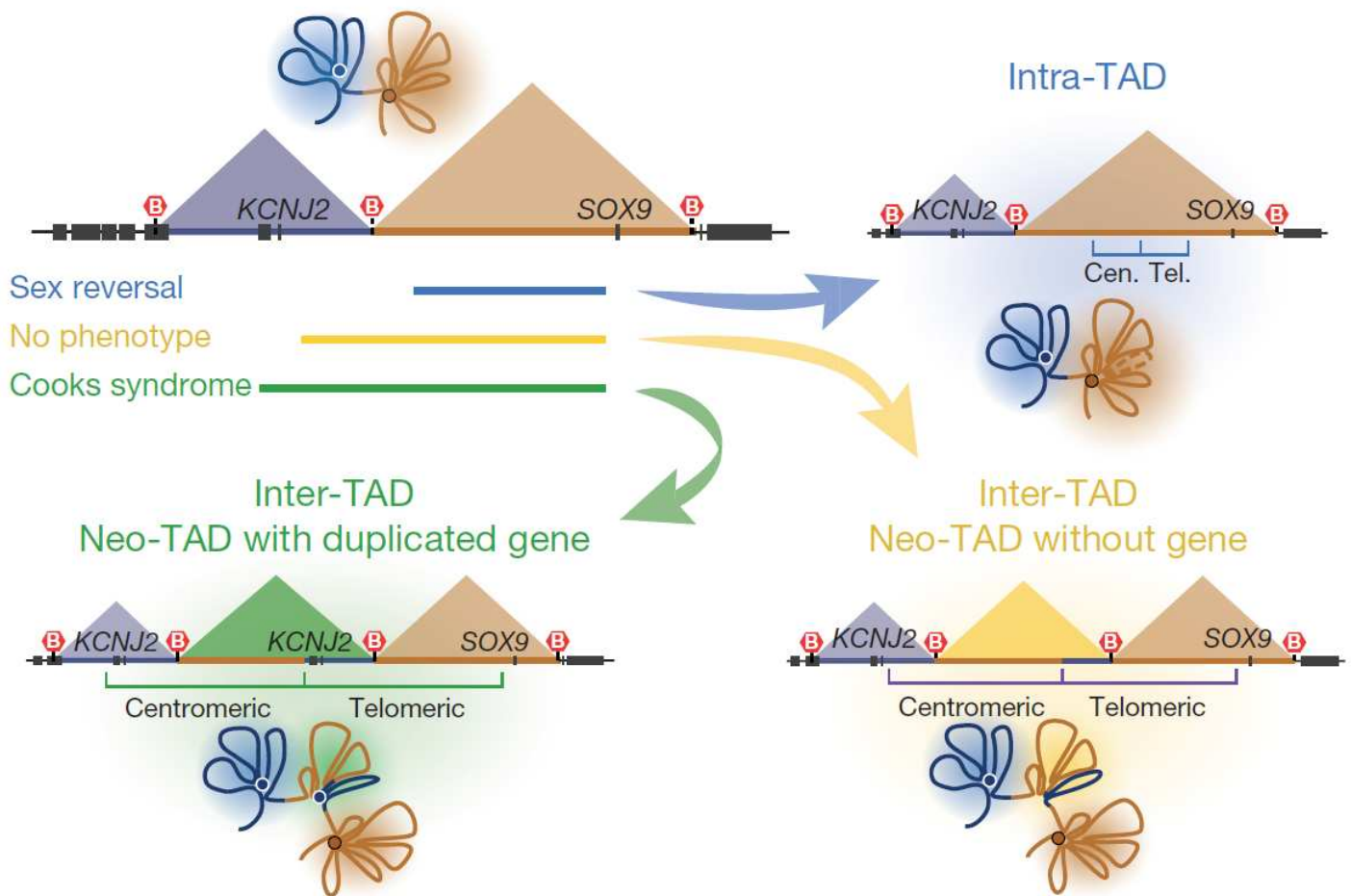
## (2) Duplications

Comme les délétions, les duplications peuvent être à l'origine d'altérations de la structure des TADs, via l'implication d'une barrière et la création de néo-TADs. Deux mécanismes différents sont possibles : soit un enhancer est compris dans la duplication, et la copie surnuméraire est positionnée à proximité d'un gène dans le néo-TAD, soit c'est un gène qui est dupliqué, et dont la copie surnuméraire est mise en contact avec un enhancer. Dans ces deux cas cependant, l'orientation du fragment dupliqué rendra possible ou non ces interactions ectopiques.

L'un des locus les plus étudiés concerné par de telles duplications est le locus du gène *SOX9*, facteur de transcription impliqué dans la différenciation des chondrocytes et la détermination sexuelle. Ce locus est organisé en 2 TADs : un comprenant *SOX9*, et un comprenant 2 gènes codant pour des canaux potassiques, *KCNJ2* et *KCNJ16* (**Figure 23**). Le TAD de *SOX9* contient de nombreux éléments régulateurs, contrôlant son expression dans différents tissus. Des délétions emportant des éléments régulateurs de l'expression cranio-faciale de *SOX9* ont notamment été décrites chez des patients présentant des séquences de Pierre-Robin non syndromique(72). Des duplications comprenant des éléments régulateurs de son expression gonadique, dans une région de 500kb en amont de *SOX9*, ont elles été décrites comme entraînant une réversion de sexe femelle en sexe mâle(73). Cependant, les duplications incluant ces éléments mais s'étendant plus loin en amont, et comprenant *KCNJ2*, n'entraîneront pas d'anomalie du développement sexuel, mais un syndrome de Cooks, caractérisé par des malformations des membres, des brachydactylies et une aplasie des ongles(74). Un troisième type de duplication a été retrouvé dans une famille, comprenant les éléments régulateurs de *SOX9*, s'étendant en amont, mais sans inclure *KCNJ2* et *KCNJ16*. Les porteurs de cette duplication n'ont aucun phénotype.

L'effet fonctionnel de ces duplications peut être déduit par l'étude de leur impact sur l'organisation des 2 TADs du locus. Les duplications intra-TAD incluant uniquement les éléments régulateurs de *SOX9* ne modifient pas la position des TADs : le mécanisme physiopathologique est une dérégulation de *SOX9* par ces éléments dupliqués. Les duplications comprenant *KCNJ2* vont inclure la barrière située entre les 2 TADs. Si elles sont en tandem, elles entraînent la formation d'un néo-TAD comprenant *KCNJ2* et une copie dupliquée des éléments régulateurs de *SOX9*, isolée du TAD de *SOX9*. Une interaction ectopique entre les éléments régulateurs et *KCNJ2* sera donc possible, avec pour conséquence l'apparition d'un syndrome de Cooks. Le troisième type de duplication, comprenant les

éléments régulateurs de *SOX9* sans comprendre *KCNJ2* entraîne l'apparition d'un néo-TAD dépourvu de gène, et où la copie mutée des éléments régulateurs est isolée de *SOX9* : il n'existe pas d'interaction ectopique, et donc le phénotype résultant est normal. Les conséquences fonctionnelles prédites de ces différentes duplications ont été prouvées par la réalisation de différentes expériences de 4-C sur des cellules de patients, ainsi que de Hi-C sur des modèles murins obtenus par technologie CRISPR-Cas9(75).



**Figure 23 :** Conséquences des différentes duplications en tandem au locus de *KCNJ2* et *SOX9*

Adapté de Franke et collaborateurs(75).

Selon que la duplication soit intra-TAD, ou inter-TAD incluant ou non *KCNJ2*, les modifications de l'organisation chromatinienne entraîneront un phénotype différent, voir une absence de phénotype.

### (3) Translocation et inversion

Comme les variations du nombre de copies, les translocations équilibrées et les inversions peuvent entraîner la désorganisation des TADs d'un locus, via la formation de néo-TADs. Ce phénomène, correspondant à un effet de position, était auparavant soupçonné chez les patients présentant des remaniements équilibrés *de novo*, avec des points de cassure n'entraînant pas l'interruption d'un gène codant. Les données disponibles sur l'organisation chromatinienne permettent désormais d'établir le lien causal entre ces remaniements et le phénotype des patients.

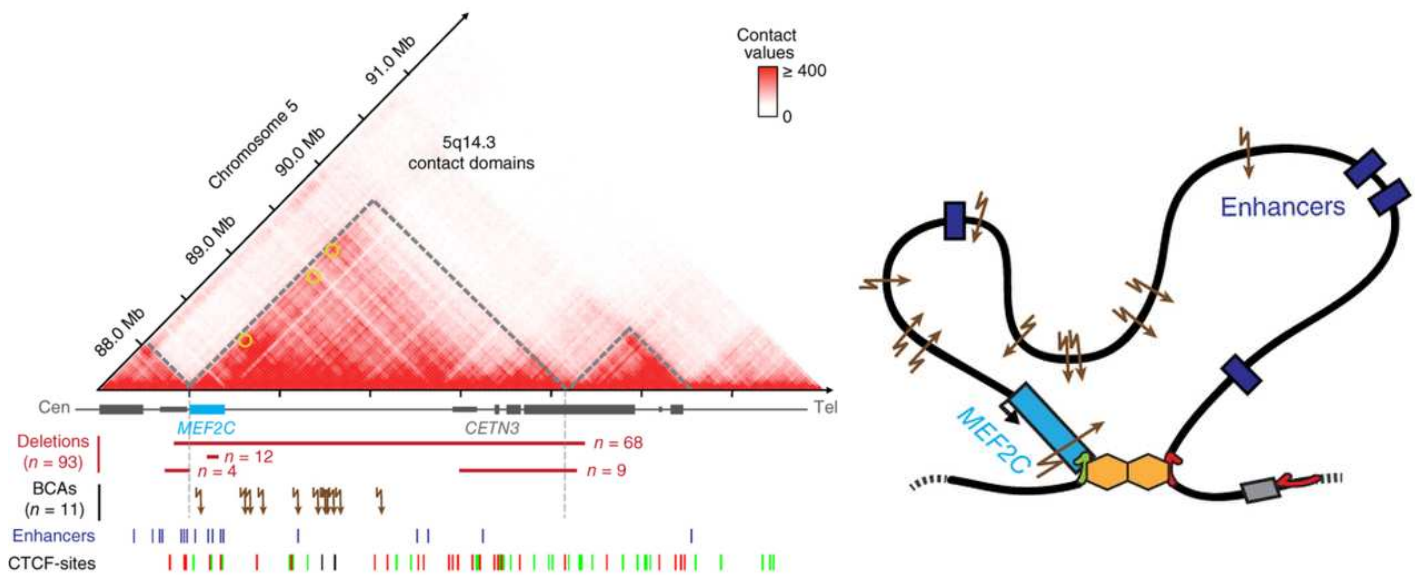
En 2016, Redin et collaborateurs(76) ont publié une étude portant sur la pathogénicité des translocations équilibrées présentes chez 273 patients présentant différentes anomalies du développement. Parmi celles-ci, 72,5 % étaient survenues *de novo*, ou ségrégeaient avec un phénotype pathogène dans la famille, l'étude complète de la ségrégation n'ayant pas pu être réalisée pour les 27,5% restant.

Un séquençage du génome complet de ces 273 patients a été réalisé, permettant d'obtenir au nucléotide près les points de cassure de 248 translocations (90,8 %), avec un total de 876 points de cassures identifiés. 26% des réarrangements se sont avérés complexes, avec 3 points de cassures ou plus, et 5% étaient compatibles avec des phénomènes de chromothripsis ou de chromoplexie. 67% des points de cassures interrompaient un gène, ce qui correspond à ce qui est attendu de façon aléatoire : bien que la séquence codante des gènes ne représente que 2% du génome humain, l'ensemble des régions transcrites en représente 60%. L'ensemble de ces gènes est fortement enrichi en gènes associés à des maladies mendéliennes, ou exprimés durant le développement embryonnaire. 66 patients de cette cohorte présentaient une translocation équilibrée pouvant être considérée comme un variant pathogène car interrompant un gène précédemment associé à une anomalie de développement de transmission autosomique dominante (par exemple : *SATB2*, *MBD5*, *EHMT1*, *NFIA* et *ZBTB20*). Vingt-cinq patients présentaient également des points de cassures interrompant des gènes candidats d'anomalies du développement. Au total, 36,7% des patients présentaient donc une translocation dont le mécanisme pathogène était probablement l'interruption d'un gène codant.

En analysant les données des autres patients, présentant des points de cassures situés en région intergénique, les auteurs de cette étude ont constaté que 7 de ces points de cassures se situaient en 5q14.3 (**Figure 24**). Cette région comprend le gène *MEF2C*, codant pour un facteur de transcription jouant un rôle majeur dans la myogénèse, mais également dans la neurogénèse, le développement cranio-facial et la migration des cellules dérivées des crêtes neurales, via une induction de la voie des endothélines(77). Des variants nucléotidiques délétères hétérozygotes *de novo* de *MEF2C* ont été identifiés chez des patients présentant une déficience intellectuelle sévère(78). Au moins un de ces patients présente une oreille en point d'interrogation, habituellement caractéristique du syndrome auriculocondytaire, dont les différents gènes responsables se situent sur la voie des endothélines(79). Les délétions de *MEF2C* sont également responsables du phénotype dans le syndrome microdélétionnel 5q14.3(80). Dans la littérature, trois autres patients atteints de déficience intellectuelle présentent des translocations équilibrées impliquant des points de cassures proches de *MEF2C*(81–83).

En reprenant les données de Hi-C publiées par Dixon et collaborateurs(30), il apparaît que les points de cassure de ces translocations équilibrées se situent à l'intérieur du TAD comprenant *MEF2C*. Par ailleurs, treize patients présentant des délétions en 5q14.3 ne comprenant pas *MEF2C* ont été rapportés : ces délétions sont situées sur les barrières de ce même TAD, en position centromérique ou télomérique. Redin et collaborateurs(76) ont donc fait l'hypothèse que ces différents réarrangements entraînent une perturbation de l'architecture de ce TAD, aboutissant à une dérégulation de *MEF2C*.

Cette hypothèse a été vérifiée par la réalisation d'une étude de l'expression de *MEF2C*, sur ARN extrait de lignées lymphoblastiques de 4 des patients avec une translocation équilibrée, montrant une expression significativement réduite par rapport aux contrôles.



**Figure 24** : Points de cassure de translocations au locus de *MEF2C*

Adapté de Redin et collaborateurs(76)

A gauche : Données de Hi-C au locus de *MEF2C*(30). En rouge, les délétions issues de la littérature. Les flèches représentent les points de cassure identifiés chez 7 patients de cette étude, et chez 3 patients de la littérature. Un patient supplémentaire a été rapporté comme présentant un point de cassure interrompant directement *MEF2C*.

A droite : Représentation d'un modèle de repliement de la chromatine à ce locus.

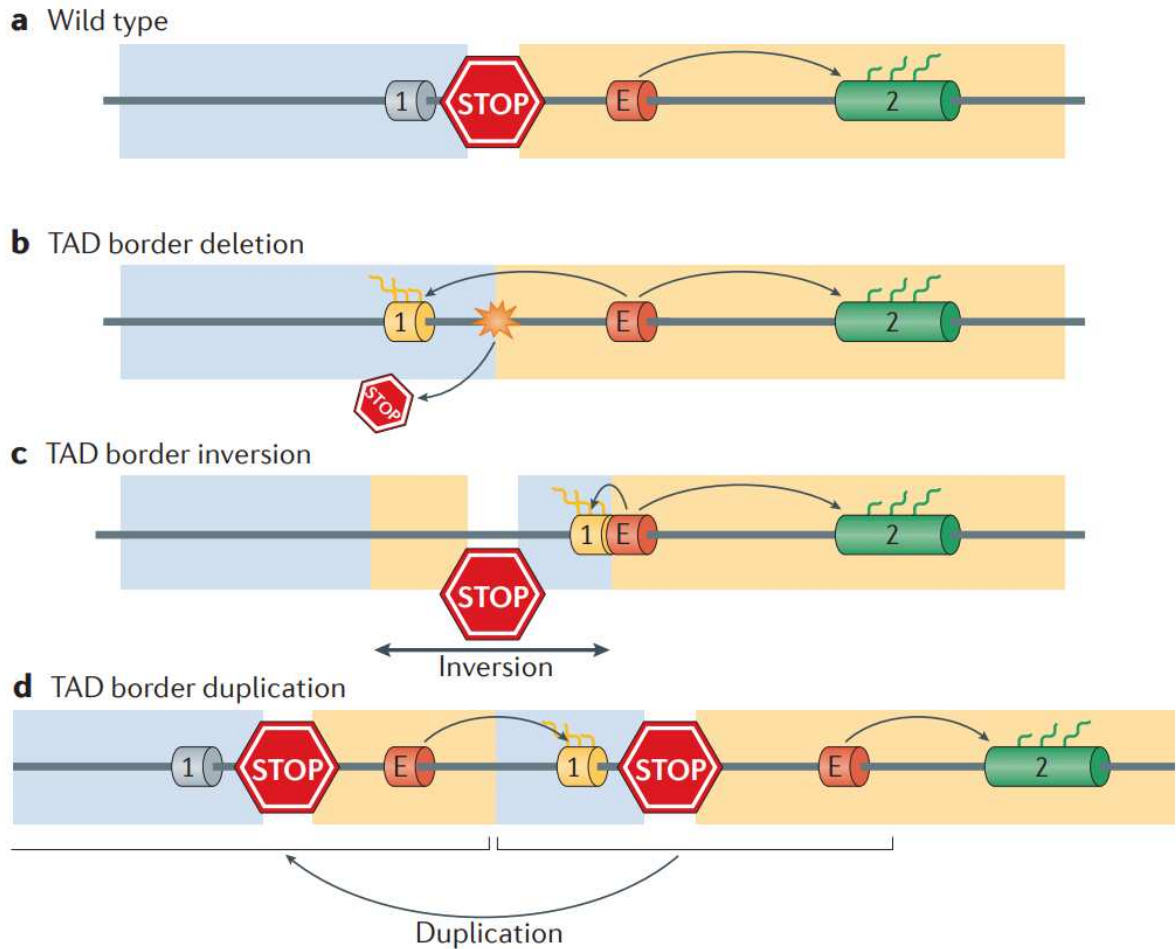
Dans l'étude réalisée par Redin et collaborateurs(76), des points de cassures situés dans d'autres loci sont également suspects d'entraîner une désorganisation des TADs de gènes connus comme responsables d'anomalies du développement. Trois patients présentent des points de cassures en 2q33.1, situés dans le TAD du gène *SATB2* impliqué dans une déficience intellectuelle syndromique, le syndrome de Glass, de transmission autosomique dominante(84). Quatre patients présentent des points de cassures compris dans le TAD de *FOXG1*, impliqué dans le syndrome de Rett atypique(85). Quatre autres patients présentent des points de cassure à proximité de *SYNCRIP*, dont des variants perte de fonction *de novo* ont été rapportés chez 2 patients présentant une déficience intellectuelle non syndromique(86). De plus, quatre patients présentent différents points de cassure situés dans des TAD de quatre gènes impliqués dans des d'anomalies du développement (*PITX2*, *SLC2A1*, *SOX9*, et *SRCAP*).



Au total, 7,3% des patients de cette cohorte(76) présentent une translocation équilibrée entraînant la survenue d'un phénotype par la désorganisation du TAD d'un gène impliqué dans une maladie mendélienne.

Ce mécanisme d'interruption de TAD serait donc dans cette cohorte 5 fois moins fréquent que l'interruption directe de la séquence transcrite d'un gène. Cependant, l'interprétation des données d'Hi-C reste difficile, et plusieurs interruptions de TAD ont pu ne pas être considérées. Il est tout à fait notable que 7 points de cassure se concentrent sur le TAD de *MEF2C*. Ce mécanisme moléculaire semble donc être impliqué de façon non négligeable dans les anomalies du développement liées à *MEF2C*, et il en est probablement de même pour d'autres syndromes.

Une seconde étude sur l'interprétation des réarrangements équilibrés à la lumière des données de Hi-C a été réalisée par Zepeda-Mendoza(87). Celle-ci porte sur 17 patients, issus du Developmental Genome Anatomy Project (DGAP)(88), porteurs de 15 translocations et de 2 inversions sans perte de matériel génomique, dont les points de cassure sont situés en région intergénique, et qui sont de survenue *de novo*. Le séquençage du génome de ces patients a été réalisé, permettant d'obtenir les coordonnées de ces remaniements au nucléotide près. Différentes données publiques ont été utilisées pour permettre une interprétation des conséquences fonctionnelles de ces anomalies. Les données de Hi-C de Dixon et collaborateurs(30), ainsi que de Rao et collaborateurs(26) ont permis de définir les coordonnées des TADs. La présence d'éléments régulateurs a été évaluée par l'utilisation de la base de données d'enhancers VISTA(89), des données d'annotation de la chromatine Broad ChromHMM(90), ainsi que des données concernant les sites d'hypersensibilité à la DNase I, issues du projet ENCODE. Le phénotype des patients, traduit en termes Human Phenotype Ontology (HPO)(91), a été comparé aux phénotypes associés aux gènes situés jusqu'à 3 Mb de distance des points de cassure, via le calcul du score Phenomatch(92). Sur ces 17 patients, 11 présentent des réarrangements impliquant potentiellement une dérégulation de 16 gènes différents, pouvant expliquer en partie leur phénotype. Cependant, les auteurs soulignent la difficulté à établir des prédictions concernant ces effets de positions, et la nécessité d'une analyse rigoureuse des différents gènes candidats.



**Figure 25** : Remaniements chromosomiques et TAD : Résumé

Adapté de Krijger et collaborateurs(59).

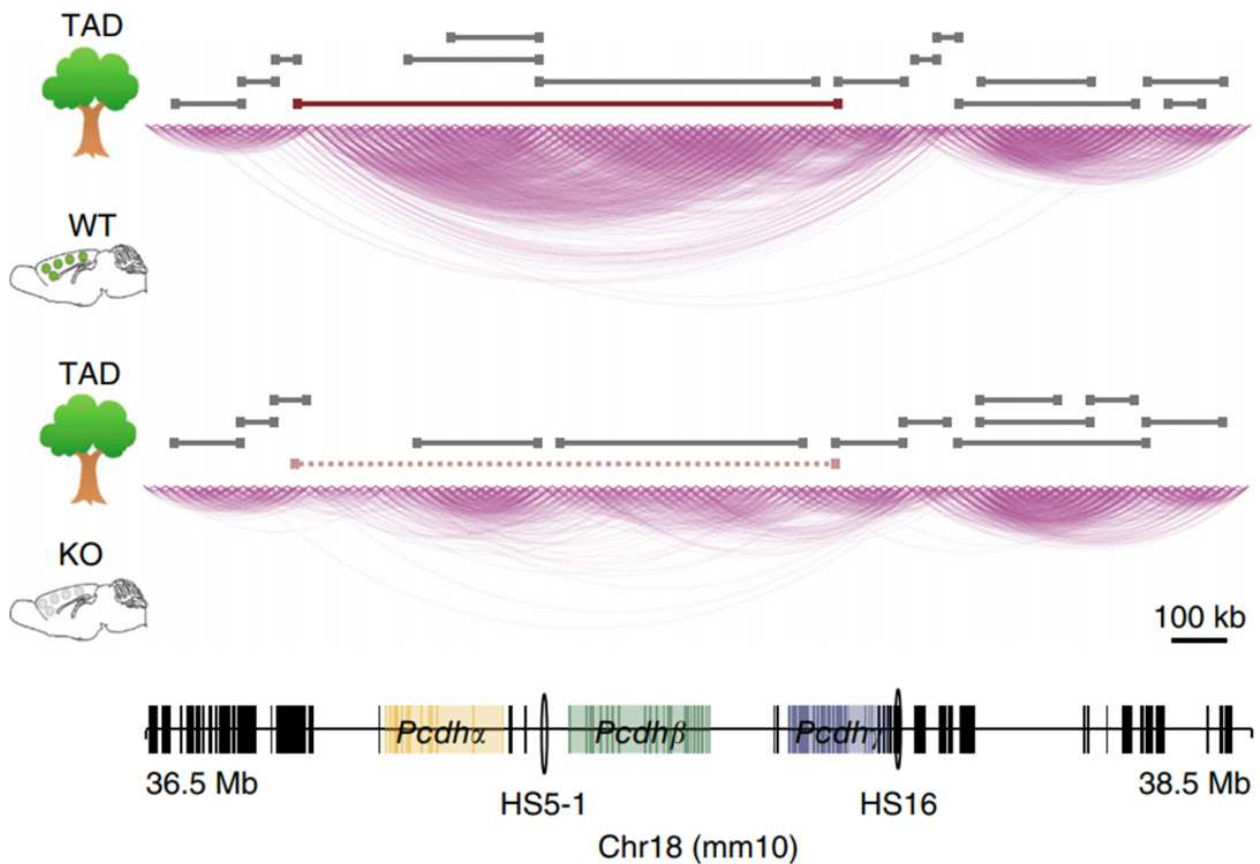
Ces différents remaniements peuvent aboutir à une interaction ectopique entre un enhancer et un gène, via une désorganisation des TADs d'un locus. En bleu et jaune, deux TADs, séparés par une barrière (STOP). Sur l'allèle sauvage, l'enhancer (E) régule le gène (2), et n'interagit pas avec le gène (1). Les différents remaniements aboutissent à une interaction de l'enhancer (E) avec le gène (1).

## b) Défaut de mise en place des TADs

Récemment, des études de l'organisation spatiale du génome ont été réalisées sur un modèle murin KO conditionnel pour le gène *Setdb1*(44). Ce gène code pour l'histone H3Lys9 méthyltransférase, intervenant dans la répression de la transcription, et a été identifié comme fréquemment dérégulé dans différents cancers(93). Par ailleurs, des variants ponctuels de *SETDB1* ont été rapportés chez plusieurs patients présentant un trouble du spectre autistique(94), et la présence d'une délétion hétérozygote *de novo* comprenant *SETDB1* chez un autre patient présentant le même phénotype a également été rapportée(95).

Un modèle de souris Cre/Lox a été généré, la Cre recombinase étant sous le contrôle du promoteur du gène *Camk2a*, et permettant d'obtenir des délétions de l'exon 3 de *Setdb1* uniquement dans les neurones du prosencéphale, en post-natal. Des cellules neuronales de cortex de souris ainsi porteuses d'une délétion homozygote de l'exon 3 de *Setdb1*, *Setdb1*<sup>2lox/2lox</sup>, ont été isolées par technique de cytométrie en flux, et des expériences de Hi-C ont été réalisées sur celles-ci.

Les cellules neuronales *Setdb1*<sup>2lox/2lox</sup> ne présentent pas de désorganisation généralisée des interactions détectées par Hi-C. Cependant, 110 boucles de contact se trouvent affectées chez ce modèle murin. Parmi celles-ci, un hot-spot se démarque particulièrement sur le chromosome 18, et correspond au locus des protocadherines (cPcdh) (**Figure 26**). Ce locus comprend 77 gènes, dont 58 codant pour des protéines impliquées dans l'adhésion cellulaire, répartis en 3 clusters Pcdh- $\alpha$ , Pcdh- $\beta$  et Pcdh- $\gamma$ . Les protocadhérines sont majoritairement exprimées dans le système nerveux, et interviennent dans de nombreux processus cellulaires durant le développement, comme le guidage axonal et la ramification dendritique(96). La dérégulation de l'expression de ces protéines a par ailleurs été retrouvée chez différents modèles de pathologies neuropsychiatriques(97). Chez la souris sauvage, les données de Hi-C montrent une organisation du locus cPcdh en de multiples sous-TAD d'environ 100kb, compris au sein d'un super-TAD de 1,2Mb. Dans les neurones de la souris *Setdb1*<sup>2lox/2lox</sup>, ce super-TAD n'est pas présent.



**Figure 26** : Données de Hi-C au locus cPcdh des souris *Setdb1*<sup>2lox/2lox</sup> (44)

Cette expérience a été réalisée sur les cellules neuronales des souris *Setdb1*<sup>2lox/2lox</sup> et sauvages. Les interactions sont représentées par les lignes courbes violettes. En gris, les TADs appelés grâce à l’algorithme TADtree. Le super-TAD présent chez la souris sauvage, mais absent sur le modèle murin, est représenté en rouge.

Afin d’expliquer cette perturbation de l’organisation de la chromatine induite par la perte de fonction de *Setdb1*, des expériences de CHIP-seq ont été réalisées sur les cellules neuronales *Setdb1*<sup>2lox/2lox</sup>, montrant une diminution pangénomique du nombre de marques H3K9me3, normalement produites par l’activité de *Setdb1*. Le locus cPcdh constitue l’un des locus les plus hypométhylés pour cette marque chez le modèle murin. Une analyse des motifs de séquence les plus spécifiquement touchés par cette hypométhylation a été réalisée. Parmi les 5 motifs les plus concernés, 3 constituent des motifs de fixation de CTCF. Les auteurs ont donc

fait l'hypothèse que l'hypométhylation de ces motifs, pour la plupart cryptiques, conduit à une fixation de CTCF sur ceux-ci. Cette hypothèse a pu être vérifiée par la réalisation d'expériences de ChIP-seq étudiant la présence de CTCF, et montrant l'apparition d'une fixation de CTCF sur ces motifs cryptiques chez le modèle murin *Setdb1*<sup>2lox/2lox</sup>. Là encore, le locus *cPcdh* est l'un des plus concerné par cette hyperfixation de CTCF.

Le modèle proposé est donc que l'absence de *Setdb1* conduit à une diminution des marques H3K9me3 sur les motifs de fixation de CTCF, aboutissant à une hyperfixation de CTCF au locus *cPcdh*. Cette hyperfixation conduit à une modification de l'architecture tridimensionnelle de la chromatine à ce locus, avec une conservation des multiples sous-TADs, et une disparition du super-TAD.

L'étude de l'expression des gènes du locus *cPcdh* montre que l'ablation de *Setdb1* chez le modèle murin conduit à une augmentation de l'expression des gènes codants pour les protocadhérines. Afin d'expliquer comment la disparition du super-TAD entraîne cette surexpression, les auteurs(44) ont étudié les différents éléments régulateurs et marques épigénétiques présents à ce locus, et notamment sur les régions mises en contact par les interactions au sein du super-TAD. L'analyse de ces éléments, précédemment identifiés(98), montre que ce super-TAD met en contact chez la souris sauvage, deux enhanceurs, HS16 et HS5-1 en contact avec deux régions, R1 et R2, comportant des marques épigénétiques de répression de la chromatine. Chez la souris sauvage, l'activité de ces enhanceurs seraient donc inhibée par un contact avec les régions R1 et R2. Chez le modèle *Setdb1*<sup>2lox/2lox</sup>, l'absence de ces boucles de répression conduirait à une interaction des 2 enhanceurs avec les gènes du locus, et donc à leur surexpression. Cette surexpression des différentes protocadhérines est potentiellement impliquée dans le phénotype des patients présentant une perte de fonction de SETDB1.

Il s'agit d'un exemple de défaut d'une des protéines régulant la mise en place des TADs aboutissant à une modification de l'architecture de la chromatine. De façon très surprenante, alors que *Setdb1* à une action sur l'ensemble du génome, seul le locus *cPcdh* semble radicalement désorganisé.

## II. Matériel et Méthodes

### A. Ré-analyse de la cohorte hospitalière de patients sans CNV pathogène identifié par ACPA

#### 1. Description de la cohorte

Afin d'identifier les patients porteurs de remaniements chromosomiques dont le mécanisme physiopathologique est l'altération de barrières de TADs, nous avons décidé d'exploiter les données disponibles pour l'ensemble des patients pour lesquels une CGH-array pangénomique ou ACPA a été réalisée au CHU de Bordeaux.

Au sein du laboratoire de génétique moléculaire du CHU de Bordeaux, la technique de CGH-array est utilisée dans le cadre du diagnostic des anomalies du développement, afin de mettre en évidence des délétions ou duplications pathogènes. Les indications pour lesquelles ces CGH-arrays sont réalisées regroupent l'ensemble de ce champ diagnostique, avec en majorité l'exploration de patients présentant une déficience intellectuelle syndromique ou non, ou un syndrome polymalformatif.

Ces analyses sont réalisées à la fois dans un contexte post-natal, majoritairement sur ADN extrait de lymphocytes circulant, mais également en période anténatale, par prélèvement invasif de villosités choriales, ou de liquide amniotique. La technique utilise des lames à oligonucléotides pangénomiques Agilent Technologies, comportant initialement 44000 sondes, avant qu'un changement ne soit effectué au profit de puces plus résolutive comportant 600000 sondes, avec une distance moyenne de 58 kb entre deux sondes.

L'hybridation de l'ADN des patients est effectuée selon la méthode du trio en « color swap » : chaque ADN est hybridé deux fois, en cyanine 3 et en cyanine 5, et comparé à chaque fois à l'ADN d'un autre patient, hybridé avec le fluorochrome inverse. L'hybridation est considérée comme ayant une qualité suffisante à l'interprétation des données si le score DLRS est inférieur à 0,3.

Les données obtenues après lecture du signal d'intensité de fluorescence sont analysées sur le logiciel Genomic Workbench Lite Edition 6.0, avec utilisation de l'algorithme ADM-2 permettant la détection des Variants du nombre de copies ou CNVs. Un remaniement sera considéré si 3 sondes consécutives dévient des valeurs de Log<sub>2</sub> ratio patient/référence attendues.

Les remaniements génomiques détectés sont ensuite interprétés grâce au logiciel Cartagena Bench Lab CNV. Ce logiciel permet une interrogation simultanée de plusieurs bases de données de CNV pathogènes ou non, notamment DGV(99), ISCA(100), et DECIPHER(101), des données du réseau français Achropuce, et des données des patients locaux déjà analysés. Par ailleurs, les informations sur les gènes contenus dans les CNVs sont obtenues via l'utilisation des données des navigateurs UCSC Genome Browser(102), Ensembl(103), et de la base de données OMIM(104).

Grâce à l'ensemble de ces données permettant l'interprétation, les remaniements détectés sont étiquetés comme appartenant à différentes catégories : pathogènes, bénins, de signification inconnue, facteurs de prédisposition, etc.

## 2. Méthode de tri

Au sein du laboratoire de génétique moléculaire du CHU de Bordeaux, le logiciel Cartagena Bench Lab CNV comporte actuellement les données de 3271 patients, correspondant à l'ensemble des ACPA réalisées.

Une des difficultés rencontrées lors du tri de ces données a été l'utilisation au cours du développement de cette technique de différents systèmes d'étiquetage des CNV, avec plusieurs catégories synonymes. Par ailleurs, le système de filtre du logiciel Cartagena Bench Lab CNV fonctionne en filtrant non pas les CNVs de façon individuelle, mais les patients, porteur d'au moins un CNV répondant aux critères choisis. Une fois ces patients identifiés, l'utilisation d'un tri par feuille de calcul Excel est nécessaire, afin de ne considérer que les CNVs répondant aux critères choisis initialement.

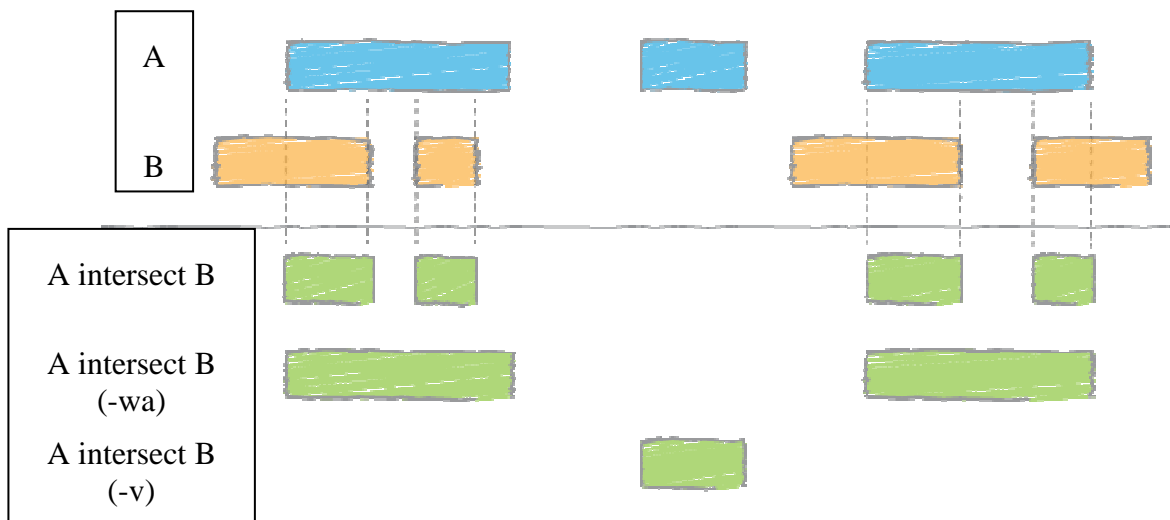
Le détail de la méthode utilisée afin de trier les CNV est précisé en annexe. Seuls les CNV initialement étiquetés comme de signification inconnue ou probablement bénins, non formellement identifiés comme hérités, et non portés par un patient ayant par ailleurs un CNV pathogène ou probablement pathogène ont été considérés. A noter que les CNV du chromosome Y ont été éliminés de l'analyse, les données de Hi-C disponibles ne permettant pas d'interprétation des TADs sur ce chromosome.

Au total, 1045 CNVs, dont 649 duplications, et 396 délétions ont été retenus.

Cette répartition entre délétions et duplications peut-être expliquée par le fait que la pathogénicité des duplications est souvent plus difficile à établir que pour les délétions, et que ces duplications sont donc plus souvent étiquetées comme variants de signification inconnue.

Afin d'éliminer les CNVs se trouvant dans des locus dont les données de Hi-C ne sont pas assez informatives, les coordonnées des CNVs ont été intersectées avec les coordonnées des TADs d'après l'étude de Rao et collaborateurs(26). Les CNVs ne se superposant à aucun TAD ont été éliminés. Cette intersection a été réalisée sur les différents fichiers de coordonnées convertis au format .bed, en utilisant l'outil Intersect du package BEDtools(105), via la plateforme d'outils bioinformatiques bio-informatiques Galaxy(106) (**Figure 27**).

```
chr7 127471196 127472363 Pos1 10 + 127471196 127472363 255,0,0
chr7 127472363 127473530 Pos2 20 + 127472363 127473530 255,0,0
chr7 127473530 127474697 Pos3 35 + 127473530 127474697 255,0,0
chr7 127474697 127475864 Pos4 20 + 127474697 127475864 255,0,0
chr7 127475864 127477031 Neg1 12 - 127475864 127477031 0,0,255
chr7 127477031 127478198 Neg2 30 - 127477031 127478198 0,0,255
```



**Figure 27** : Fichiers .bed et outils BEDtools

En haut, format de fichier .bed. Chaque ligne correspond à un intervalle différent. Le fichier comporte 3 colonnes obligatoires, correspondant au chromosome considéré du génome de référence, une coordonnée de début et une coordonnée de fin. Les autres colonnes, jusqu'à 9, sont optionnelles et servent à préciser un nom, un score et un brin à chaque intervalle, ainsi que des options d'affichages (trait fin/épais, jonction intron-exon, couleur).

En bas, outil Intersect du package BEDtools(105). A partir de deux fichiers comportant chacun plusieurs intervalles au format .bed, ici A et B, cet outil bio-informatique permet d'obtenir différents recouvrements (par exemple, options `-wa` et `-v`).

Après cette étape, 735 CNVs, dont 498 duplications et 237 délétions, ont été conservées pour analyse.



### 3. Visualisation et interprétation des données

Une interprétation manuelle de l'impact de ces CNVs sur les TADs, utilisant les mêmes jeux de données publiques que l'étude de Zepeda-Mendoza(87), a pu être ainsi réalisée.

Un fichier .bed (**Figure 27**) correspondant à l'ensemble des coordonnées des duplications et des délétions a été chargé en tant que piste personnalisée dans le visualisateur de données génomiques UCSC Genome Browser(102). Cette session personnalisée a ensuite été superposée au sein du visualisateur de données de Hi-C 3D Genome Browser(107) (**Figure 28**).

Une première visualisation a été effectuée sur un nombre restreint de locus. Les régions pour lesquelles une altération de TAD a déjà été identifiée comme pathogène d'après les données de la littérature(75,76,108–110) ont été analysées préférentiellement, de même que les locus de 44 gènes impliqués dans la déficience intellectuelle, indication majoritaire dans notre cohorte (liste en annexe).

Par ailleurs, l'information sur le caractère *de novo* ou hérité des CNVs étaient connus pour 40 d'entre eux, avec 19 délétions et 21 duplications *de novo*. Ces CNVs ont été analysés en priorité, car plus susceptibles d'être pathogènes.

Une seconde analyse a ensuite été réalisée en visualisant de façon systématique une région de +/- 1 Mb autour de chacun des 735 CNV.

Dans le cas où une barrière de TAD est impactée par un CNV, avec un possible effet sur un gène connue en pathologie humaine ou dont les données de la littérature sont compatibles avec le phénotype du patient, les éléments régulateurs du locus ont été recherchés. La visualisation des données issues du projet ENCODE (ChromHMM, H3K27Ac et DNaseI), de la base de données VISTA(89), et de l'atlas publiés par Andersson et collaborateurs(111) permettent d'identifier de potentiels enhancers. Cependant, en l'absence d'étude fonctionnelle, ces prédictions ne peuvent pas amener à des conclusions définitives quant à la présence, ou à l'absence d'enhancers dans un locus donné.

Données de Hi-C d'après Rao, à une résolution de 25kb

TADs

Délétions et duplications

Design des sondes de CGH-array

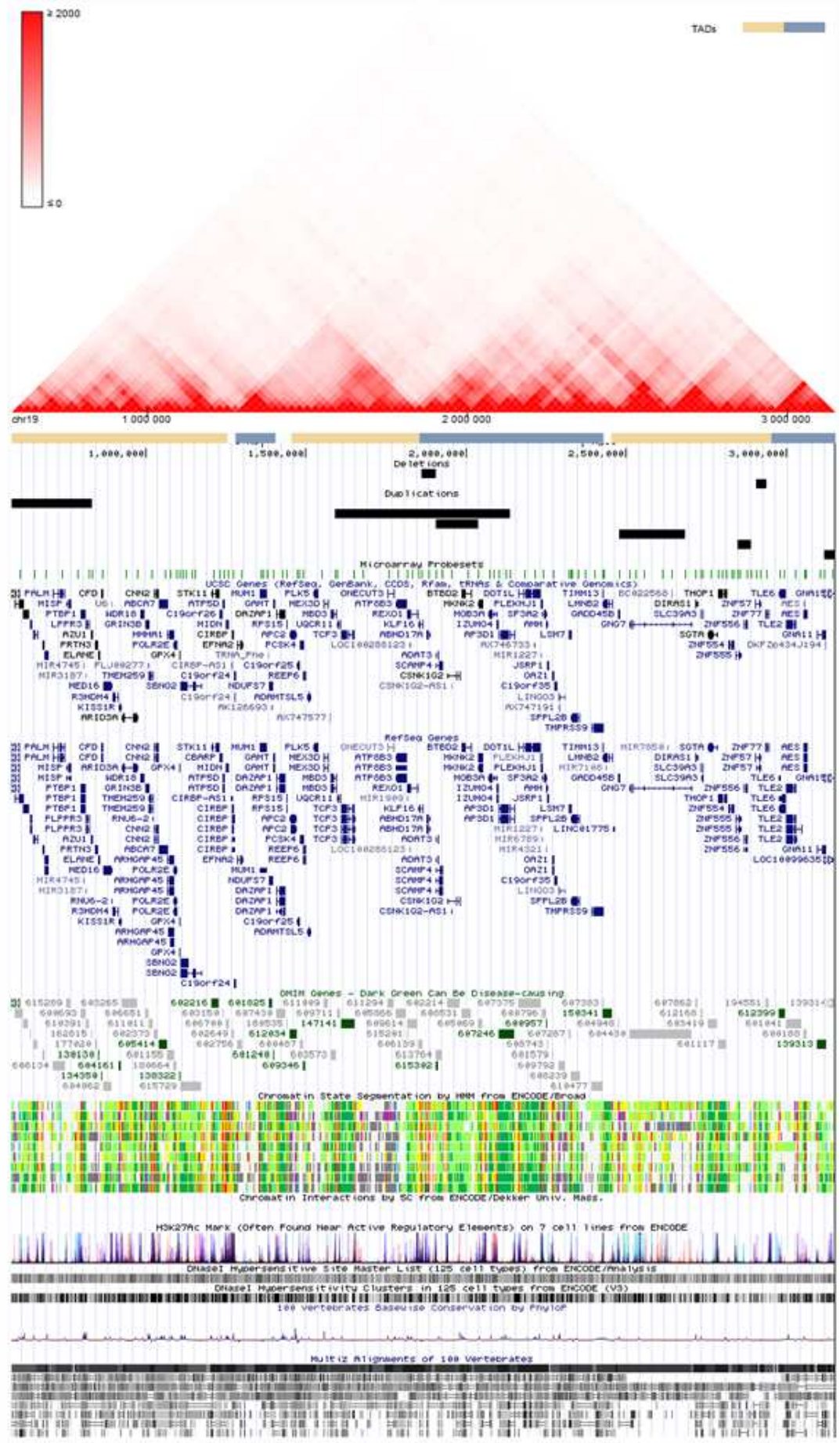
Gènes UCSC et RefSeq

Gènes OMIM

Segmentation de la chromatine ChromHMM

ENCODE H3K27Ac et DNaseI

Données de conservation



**Figure 28** : Interprétation des données de Hi-C

Visualisation des données sur 3D Genome Browser(107), couplée à une session personnalisée de l'UCSC Genome Browser(102). Les données de Hi-C et d'annotation des barrières de TADs sont celles issues de la publication de Rao et collaborateurs(26). Grâce à cette visualisation, on peut identifier quel CNV est susceptible d'altérer une barrière de TADs d'après les modèles de mécanismes moléculaires connus (voir **I.B.5.a**). La segmentation de chromatine ChromHMM(90) (légende complète en annexe) permet de vérifier que la barrière de TAD annotée est bien enrichie en motif insulateur, notamment CTCF, sur différentes lignées cellulaires (en bleu ciel.) Par la suite, on peut identifier quel gène OMIM déjà connu en pathologie humaine peut être impacté. Les données issues du projet ENCODE (ChromHMM, H3K27Ac et DNaseI) permettent ensuite de repérer d'éventuels éléments régulateurs.

## B. Mise au point de la technique de 3C

En parallèle de la réanalyse des résultats d'ACPA de notre cohorte, nous avons souhaité développer un test fonctionnel approprié pour étudier l'impact des remaniements chromosomiques sur l'architecture chromatinienne. Si elle permet d'explorer les interactions physiques entre 2 locus, la technique de 3C n'est pas en soi la technique à privilégier pour analyser les modifications architecturales d'un locus à l'échelle des TADs. En effet, la distance entre 2 locus qu'il est possible d'étudier de façon courante n'est que de quelques kilobases. Cependant, toutes les autres techniques dérivées de la 3C, et notamment la Hi-C, méthode de référence, ont en commun les étapes initiales de la 3C : le cross-linking, la digestion et la ligation. Mettre au point ces étapes par 3C sur un locus de référence nous a donc paru comme un préalable essentiel à la réalisation de techniques plus sophistiquées dans le futur.

### 1. Choix du locus *CIITA*

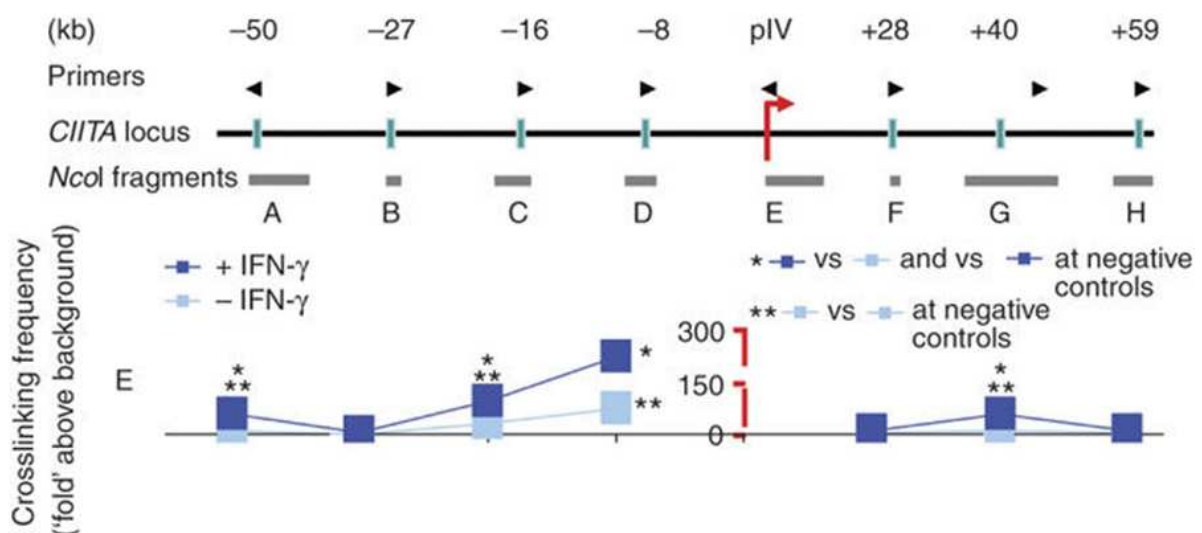
Afin de mettre au point un protocole de 3C, nous avons fait le choix d'étudier les interactions entre le promoteur pIV du gène *CIITA*, et deux enhanceurs situés en amont, à -8kb et -16kb.

Le gène *CIITA*, ou class II transactivator, code pour le régulateur central de la transcription des gènes du complexe d'histocompatibilité de classe II. La transcription du gène *CIITA* est

elle-même contrôlée par différents mécanismes épigénétiques(112), et notamment par un remodelage chromatinien induit par l'interféron-gamma, et dépendant de SMARCA4 (BRG1)(113).

L'étude réalisée par Ni et collaborateurs(114) a eu pour but de démontrer que de multiples enhancers distaux sont nécessaires à la régulation de la transcription de *CIITA* induite par l'interféron-gamma. Plusieurs expériences de 3C ont été réalisées lors de cette étude, afin d'identifier les contacts existant entre le promoteur pIV de *CIITA* et ces éléments distaux, à l'état basal, et après exposition à l'interféron-gamma.

Il a été notamment montré qu'il existe un contact à l'état basal entre pIV et un enhancer situé à -8kb, ainsi que de façon plus modérée entre pIV et un enhancer situé à -16kb (**Figure 29**). Ce sont ces interactions que nous avons choisi d'étudier afin de valider notre protocole.



**Figure 29** : Données de 3C au locus *CIITA*, d'après Ni et collaborateurs(114).

En haut, positions des différents fragments de restriction de *NcoI* au locus *CIITA*.

En dessous, données de 3C : le fragment de référence dont les interactions avec les autres locus sont considérées est le fragment E, soit celui du promoteur pIV.

On observe qu'à l'état basal, il existe une interaction entre le promoteur pIV et les fragments situés à -8kb et -16kb. Ces interactions sont renforcées après incubation des cellules par interféron gamma.

Le protocole détaillé dans cette publication présente plusieurs avantages techniques. L'étude est réalisée sur des cellules HeLa, que nous avons à notre disposition au laboratoire. La séquence des primers, l'enzyme de restriction et le BAC utilisé ont été publiés. De plus, l'expérience montre une interaction à l'état basal entre le promoteur pIV et l'enhancer situé à -8kb. L'existence d'un autre enhancer à -16kb nous permet de vérifier que les interactions mises en évidence entre pIV et l'enhancer situé à -8kb ne sont pas dépendantes de la distance génomique, par la vérification de l'absence d'interaction entre les deux enhanceurs, situés eux-mêmes à 8kb de distance.

Par ailleurs, l'étude du locus de *CIITA* a également été réalisée par une autre équipe, dans le but de mettre point une technique de quantification des interactions par 3C par réalisation d'une courbe de fusion utilisant du SYBR Green(115). Cette étude précise notamment la température de fusion du produit de ligation des fragments correspondant à pIV et à l'enhancer situé à -8kb, soit 81,3°C +/- 0,7.

## 2. Protocole de 3C

Après adaptation du protocole publié par Ni et collaborateurs(114), voici les différentes étapes du protocole de 3C que nous avons pu tester, en prenant comme locus d'intérêt le locus *CIITA*.

### a) Cross-linking

Des cellules HeLa cultivées en flasques T75 ont été utilisées. Ainsi, des cellules témoins issues de cultures réalisées dans des conditions standards sont utilisées (DMEM, 10% SVF, Pénicilline/streptomycine). En parallèle, des cellules HeLa ont été traitées à l'interféron gamma IFN- $\gamma$  pendant 6h à la concentration de 0,1  $\mu\text{g}/\text{mL}$ . Pour chaque flasque contenant 5mL de milieu de culture, 140 $\mu\text{L}$  de Formaldéhyde à 37% sont ajoutés, afin d'obtenir une fixation à 1% de Formaldéhyde. Après 10 minutes d'incubation à température ambiante, 280  $\mu\text{L}$  de Glycine à une concentration initiale de 2,5M sont ajoutés, afin de stopper le processus de cross-linking. Après incubation durant 5 minutes à température ambiante, puis 15 minutes sur glace, les cellules sont transférées par grattage dans des tubes de 15mL.

Après centrifugation à 800g, à 4°C durant 10 minutes, le surnageant est éliminé. Les culots sont ressuspendus dans 1mL et transférés dans des microtubes de 1,5mL. Ces tubes sont de nouveau centrifugés à 800g, à 4°C durant 10 minutes. Le surnageant est éliminé, et le culot congelé à -80°C, avec possibilité de stockage à ce stade.

#### b) Lyse des cellules et digestion par NcoI

Après décongélation les culots cellulaires sont resuspendus et vortexés dans une solution de lyse comprenant 2 $\mu$ L de NP40, 20 $\mu$ L de Tris-HCl pH 8 à 0,5M, 10 $\mu$ L de NaCl à 1 M, 10 $\mu$ L de Protease Inhibitor Mix 100X, complétés par de l'eau pour un volume total de 1mL. Après incubation de 5min à 65°C, les tubes sont centrifugés à 2000g durant 5min, à température ambiante, et le surnageant éliminé.

Les culots sont rincés à 2 reprises dans 500 $\mu$ l de tampon de restriction 3.1 à 1X, avec centrifugation à 2000g durant 5min à température ambiante et élimination du surnageant. Les culots sont finalement resuspendus dans un mix de restriction comprenant 140 $\mu$ l de tampon de restriction 3.1 à 1X et 8 $\mu$ L, soit 400U, d'enzyme de restriction NcoI suivi d'une incubation à 37°C durant 2h minimum. L'arrêt du processus de digestion peut se faire selon 2 conditions différentes : chauffage à 80°C durant 20min, ou ajout de 30 $\mu$ L de SDS 10%.

#### c) Ligation

La ligation se fait par ajout d'un mix de ligation comprenant : 2,5  $\mu$ L (soit 5000U) de T4 DNA Ligase, 60 $\mu$ L de tampon de ligation 10X, et 389,5  $\mu$ L d'eau. Le mix réactionnel est incubé à 16°C durant 1h. A ce stade, 30  $\mu$ L de Protéinase K à 10mg/mL sont ajoutés, et les tubes sont incubés à 65°C durant 4h minimum. Le processus est arrêté par chauffage à 95°C durant 5min. Finalement, 5  $\mu$ L de RNase A sont ajoutés, et les tubes incubés à 37°C durant 30 minutes.

#### d) Purification de l'ADN

L'ADN présent est purifié via l'utilisation de billes magnétiques. Ainsi, 650  $\mu$ L de billes magnétiques sont ajoutés et 15 aspiration/refoulement successifs sont réalisés afin d'homogénéiser le mélange. Les tubes sont incubés 1h sur portoir standard, puis 15 minutes sur support magnétique, à température ambiante. Le surnageant est éliminé. On procède à 2 lavages par ajout d'1 mL d'éthanol 80% avec une incubation durant 1min sur support magnétique, et élimination du surnageant. L'ADN fixé sur les billes est ensuite élué dans 200  $\mu$ L d'eau, avec réalisation de 15 aspiration/refoulements successifs, et incubation 5min sur portoir standard puis 10min sur support magnétique. Le surnageant est ensuite transféré dans un nouveau tube. On réalise ensuite le dosage de l'ADN par technique spectrophotométrique NanoDrop (ThermoFisher).

#### e) Détection des produits de ligation

La détection des produits de ligation présents dans les bibliothèques de 3C a été réalisée par PCR, via l'utilisation d'un mix commercial : l'iQ SYBR Green Supermix (BioRad, CA, USA). Les amorces de chaque couple (pIV/enhancer -8kb, pIV/enhancer -16kb, enhancer -8kb/enhancer -16kb) sont ajoutées à une concentration 0,2 µM. Le mix est distribué en plaque, et 2 µl de la banque issue 3C diluée à 25ng/µL est ajouté à chaque puit. La plaque est installée dans un thermocycleur et la réaction se déroulera selon le protocole : Dénaturation initiale (95°C, 3 min), puis 40 cycles de dénaturation (95°C, 30 sec) et hybridation (55°C, 30 sec). Une courbe de fusion est ensuite réalisée, par chauffage de 55 à 95°C. La visualisation et l'analyse des données se fait par le système CFX96 Touch™ Real-Time PCR Detection (BioRad, CA, USA).

### III. Résultats

#### A. Ré-analyse de la cohorte hospitalière de patients sans CNV pathogène identifié par ACPA

##### 1. Délétion au locus *WNT6/IHH/EPHA4/PAX3*

##### a) Description foetopathologique et moléculaire

Nous décrivons ici le cas d'un fœtus de sexe féminin dont la grossesse a été interrompue à 15 semaines d'aménorrhée pour un syndrome polymalformatif associant un volumineux hygrome kystique, des polydactylies, une phocomélie, des reins hyperéchogènes, et une méga grande citerne.

Un examen foetopathologique a pu être réalisé. Le fœtus était macéré, avec des plaques d'épidermolyse représentant 80% de la surface corporelle. On note cependant qu'il est hypotrophique. L'examen externe retrouve une double fente palatine, et des oreilles très

basses implantées. Les organes génitaux externes sont ambigus avec une impression de défaut de fusion des bourrelets urétraux, le terme précoce limitant cependant l'examen.

Il existe une micromélie globale à prédominance rhizomélique, avec des ptérygiums scapulaires. On observe une polydactylie des 4 extrémités. Aux membres supérieurs, on compte 8 doigts à chaque main, avec pour la main droite une impression de main en miroir, avec duplication du premier rayon. Aux membres inférieurs, il existe 7 orteils au pied gauche et 6 au pied droit dont un hallux élargi.

L'examen viscéral macroscopique retrouve différentes anomalies. La vésicule biliaire n'est pas retrouvée. On observe une polysplénie. Les organes génitaux internes sont féminins, anormaux, avec un utérus mal visible. La dissection segmentaire du cœur et des gros vaisseaux met en évidence un seul gros vaisseau. Les deux oreillettes sont visibles avec un situs normal. L'aorte est à cheval sur les deux ventricules avec une large communication interventriculaire. Aucun tronc artériel pulmonaire n'est identifiable. Il n'y a pas d'artère pulmonaire issue de ce tronc artériel unique. L'examen de la crosse permet de mettre en évidence une possible collatérale aorto-pulmonaire issue de l'isthme et rejoignant le poumon droit et une autre issue de l'aorte descendant vers le poumon gauche.

Des CGH-array 60K, puis 180K ont été réalisées chez le fœtus. Celles-ci ont mis en évidence la présence d'une délétion interstitielle hétérozygote *de novo* d'une taille minimale de 972 kb en 2q35 : chr2:g.(219912758\_219928949)\_(220900731\_220931554)del(hg19) (**Figure 30**).

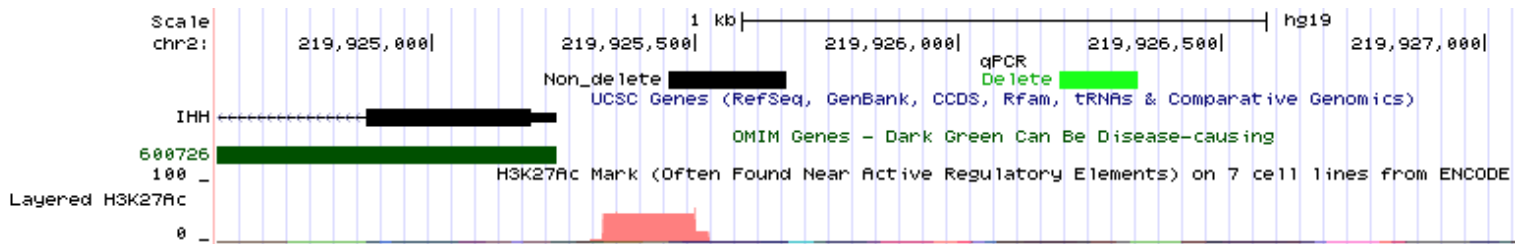
Afin de préciser la position du point de cassure centromérique, nous avons réalisées plusieurs qPCR successives dans la zone d'incertitude (**Figure 31**). Le point de cassure se trouve entre 400 et 1000 nucléotides en position télomérique par rapport au gène *IHH*, sans l'inclure dans la délétion.





**Figure 30** : Délétion au locus *WNT6/IHH/EPHA4/PAX3*

On observe sur la visualisation des données du 3D Genome Browser que la délétion (flèche rouge) emporte la barrière entre le TAD d'*EPHA4* (cercle bleu) et le TAD adjacent, comprenant *IHH* (cercle rouge).



**Figure 31** : Position des fragments amplifiés par qPCR chez le fœtus par rapport à *IHH*.

En noir, locus sauvage. En vert, locus emporté par la délétion. La borne de la délétion est située en dehors d'*IHH*, et en dehors d'une région identifiée comme marque H3K27Ac, pouvant correspondre au promoteur d'*IHH*.

Cette délétion comprend 25 gènes *Refseq* dont 5 gènes *OMIM Morbid* (*NHEJ1*, *ABCB6*, *DNAJB2*, *DES* et *OBSL1*). Cependant, l'haplo-insuffisance d'aucun de ces gènes ne semble pouvoir expliquer la survenue de ce tableau polymalformatif, d'après les données de la littérature.

#### b) Description du locus *WNT6/IHH/EPHA4/PAX3*

Cette délétion est incluse dans une région de 4 Mb du chromosome 2, comprenant notamment les gènes *WNT6*, *IHH*, *EPHA4* et *PAX3*.

*IHH* code pour une protéine faisant partie de la famille des protéines intervenant dans la voie de signalisation *hedgehog* comprenant *SHH*, *IHH* et *DHH*, impliquées dans de nombreux processus du développement embryonnaire(116). *IHH* intervient en particulier dans la régulation de l'ostéogénèse et de la chondrogénèse, et est partiellement redondant avec *SHH*. Des variants faux-sens d'*IHH* sont responsables de 2 phénotypes distincts : la brachydactylie de type A1, de transmission dominante(117), et la dysplasie acrocapitofémorale, de transmission récessive(118). Par ailleurs, des duplications en tandem en amont d'*IHH*, incluant des éléments de régulation distant, ont été rapportées dans 4 familles, présentant un phénotype comprenant une syndactylie et une craniosténose(119,120). *IHH* apparaît donc comme un gène majeur impliqué dans le développement des membres, et dont la dérégulation de l'expression entraîne des phénotypes variables.

Les 2 autres gènes de la famille *hedgehog* sont également impliqués dans des maladies mendéliennes. DHH est responsable d'une forme de dysgénésie gonadique associée à une polyneuropathie(121). Des variants entraînant la perte de fonction de SHH sont eux responsables d'une forme autosomique dominante d'holoprosencéphalie(122). Nous pouvons également noter que les gènes codant pour des protéines situées en aval de la cascade de signalisation de SHH sont également impliqués dans d'autres formes d'anomalies du développement : *PTCH1* dans le syndrome de Gorlin, et *GLI3* dans le syndrome de Pallister-Hall(123), le syndrome de céphalopolysyndactylie de Greig(124) et dans des formes autosomiques dominantes de polydactylies(125).

*WNT6* code pour une protéine de signalisation de la famille WNT, impliquées de manière fondamentale dans les mécanismes de prolifération, de polarité, et de destin cellulaire durant l'embryogénèse. WNT6 a notamment été identifiée comme impliquée dans l'induction de la formation de la crête neurale(126), et dans la répression de la chondrogénèse chez le poulet(127).

*EPHA4* code pour un récepteur à activité tyrosine kinase de la sous famille des éphrines. Il est particulièrement impliqué dans la régulation de la guidance axonale durant le développement de la moelle épinière(128). Chez la souris, il a été montré qu'*EPHA4* est exprimé durant le développement des membres, et est nécessaire à leur innervation. Cependant son inactivation n'entraîne pas de phénotype malformatif(129). Chez l'homme, il a par ailleurs été mis en évidence qu'*EPHA4* joue un rôle dans la sévérité du phénotype dans la sclérose latérale amyotrophique(130).

*PAX3* code pour un facteur de transcription de la famille PAX, et joue un rôle critique durant le développement fœtal, notamment par l'induction du développement de la crête neurale(131). Des variants hétérozygotes entraînant la perte de fonction de *PAX3* sont responsables du syndrome de Waardenburg de type 1(132), caractérisé par une surdité congénitale, des anomalies pigmentaires touchant l'iris, les cheveux et la peau, et une dysmorphie caractéristique avec dystopie des canthis. Par ailleurs, des variants homozygotes ont été rapportés chez des patients présentant un syndrome de Waardenburg de type 3, ou Klein-Waardenburg, qui associe à un phénotype de Waardenburg un phénotype musculo-squelettique avec une hypoplasie ou des contractures des membres, une fusion des os du carpe, ou une syndactylie(133).

Le locus *WNT6/IHH/EPHA4/PAX3* comprend donc au moins 3 gènes, *WNT6*, *IHH* et *PAX3*, intervenant de façon essentielle durant le développement embryonnaire et foetal, et dont les deux derniers sont connus pour être impliqués dans des maladies mendéliennes avec anomalies des membres.

Par ailleurs, l'organisation spatiale du locus *WNT6/IHH/EPHA4/PAX3* a été étudiée(30,108) via la réalisation d'expériences de Hi-C et de 4C dans des cellules souches embryonnaires humaines et chez la souris. Il a pu être déterminé que cette région comprenait 3 TADs : un TAD centromérique incluant les gènes *WNT6* et *IHH*, un TAD médian comprenant *EPHA4*, et un TAD télomérique comprenant le gène *PAX3*.

c) Variants structuraux au locus *WNT6/IHH/EPHA4/PAX3* : modèle de Lupiáñez et collaborateurs.

En étudiant une cohorte de patients présentant des anomalies des membres, Mundlos et collaborateurs ont pu mettre en évidence des variants structuraux de ce locus chez des patients présentant 3 phénotypes distincts : une brachydactylie, un syndrome F, ou une polydactylie(108).

Dans chacun des cas, la simple haploinsuffisance ou surexpression d'un ou des gènes compris dans le variant structural ne permettait pas d'expliquer la survenue du phénotype.

Chez 3 familles présentant une brachydactylie touchant les 1er et 2e rayons, les CGH-array de haute résolution identifient des délétions hétérozygotes de 1,75 à 1,9 Mb. Ces délétions incluent la barrière située entre le TAD d'*EPHA4* et le TAD de *PAX3*, et peuvent donc entraîner une interaction entre les deux gènes.

Deux familles où ségrège un syndrome F ont été étudiées. Ce phénotype est caractérisé par une syndactylie sévère et complexe, touchant les 1<sup>er</sup> et 2<sup>e</sup> rayons aux mains, associée avec une polydactylie des orteils(134). La réalisation d'un séquençage de l'exome des membres de ces familles n'a pas permis d'identifier de variant pouvant expliquer la pathologie. Néanmoins, la réalisation du séquençage de génomes complets a permis de mettre en évidence une inversion hétérozygote de 1,1 Mb dans la première famille, et une duplication hétérozygote en tandem de 1,4 Mb dans la seconde. Ces deux réarrangements ont pour conséquence de mettre en relation la partie centromérique du TAD d'*EPHA4* et le gène *WNT6*(108).

Un patient porteur d'une duplication hétérozygote de 900 kb comprenant *IHH* a été rapporté comme présentant une déficience intellectuelle sévère associée à une polysyndactylie des mains et des pieds, des anomalies craniofaciales avec macrocéphalie, et une agénésie du corps calleux(135). Une seconde grossesse de la mère de ce patient, avec un autre partenaire, a été interrompue à 18 semaines d'aménorrhée devant la récurrence d'un phénotype similaire. Ce fœtus portait également cette duplication, qui n'a cependant pas pu être mise en évidence chez la mère, correspondant à une probable mosaïque germinale. Les auteurs rapportant ces cas ont rapproché cette présentation clinique du modèle murin *doublefoot*. Celui-ci présente notamment une polydactylie des 4 membres, associée à des anomalies cranio-faciales, et est dû à une délétion de 600 kb située en 5' du gène *IHH*. Dans le cas du patient porteur de la duplication d'*IHH* et dans celui du modèle murin *doublefoot* porteur lui d'une délétion en aval du gène, les remaniements chromosomiques responsables du phénotype comprennent la barrière située entre le TAD *WNT6/IHH* et le TAD d'*EPHA4*, permettant une interaction entre ces 2 TADs.

Tous les variants structuraux retrouvés ont donc pour conséquence une possible interaction entre un gène du locus, et le TAD d'*EPHA4* (**Figure 32**).

Une étude d'expression sur des souris porteuses des réarrangements décrits, obtenues via l'utilisation de la technologie CRISPR-Cas9, a été réalisée. Il a été montré que les gènes mis en relation avec le TAD d'*EPHA4* acquièrent un profil d'expression dans les membres correspondant normalement à celui d'*EPHA4*, principalement dans le mésoderme distal antérieur(108).

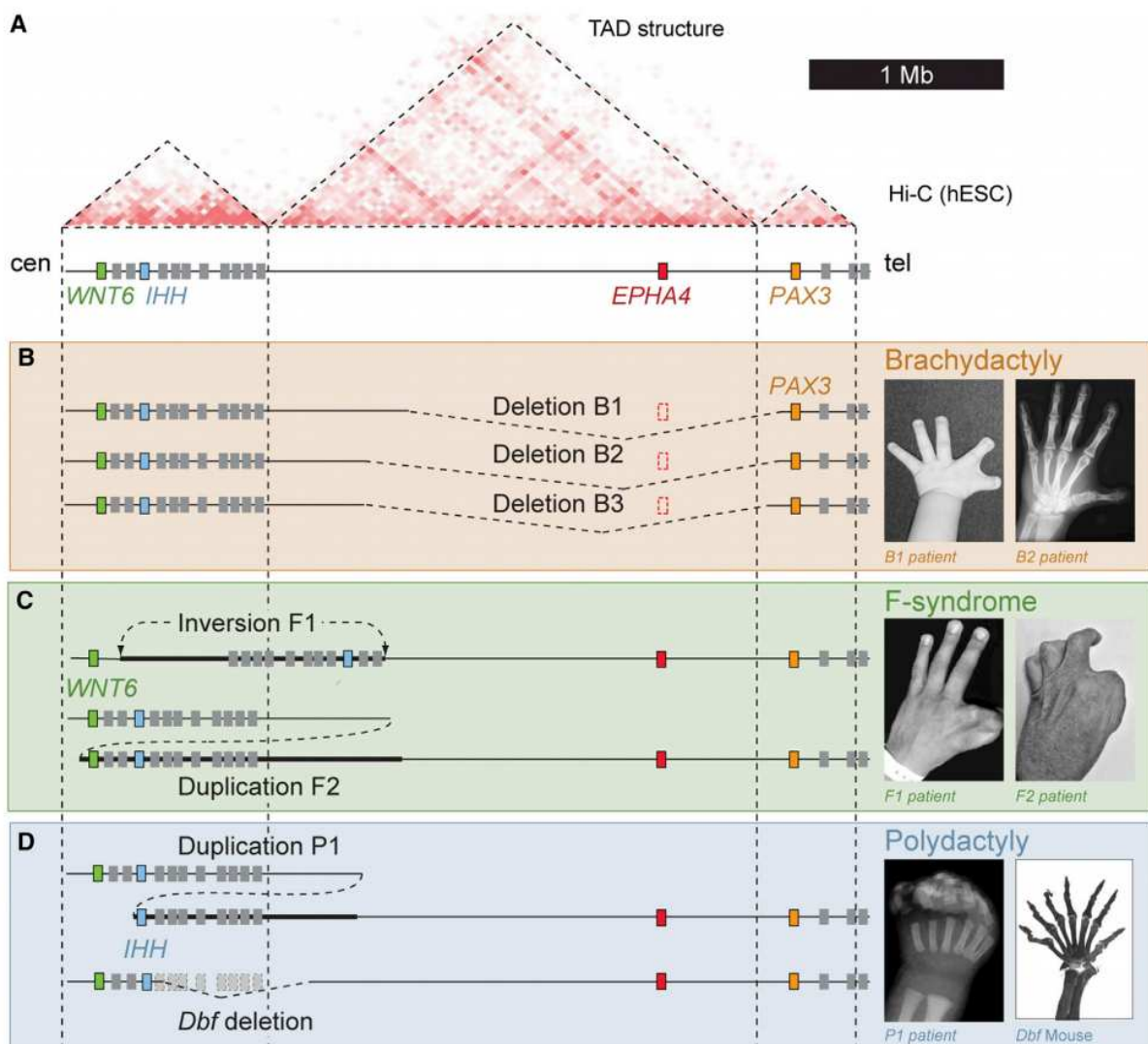
Par ailleurs, des expériences de 4C chez les mêmes modèles murins ciblant les promoteurs de *WNT6*, d'*IHH* et de *PAX3* ont pu être réalisées. Celles-ci montrent la présence chez les souris modèles d'un signal dans le TAD d'*EPHA4*, non présent chez la souris sauvage, et prouvant l'existence d'interactions ectopiques entre les différents gènes du locus et le TAD d'*EPHA4* résultant des variants structuraux.

Ces mêmes expériences de 4C ont été réalisées sur des fibroblastes des patients présentant des anomalies des membres, et retrouvent des résultats similaires.

L'ensemble de ces éléments suggère que le mécanisme physiopathologique de ces variants structuraux est le contrôle de l'expression de *WNT6*, *IHH* ou *PAX3* par des enhancers régulant normalement l'expression d'*EPHA4*. Ce contrôle est permis par la disparition des barrières

situées entre les TADs, permettant des interactions ectopiques entre les promoteurs de ces 3 gènes, et ces possibles enhancers situés dans le TAD d'*EPHA4*.

Afin d'identifier ces enhancers, les données publiées de ChIP-Seq, d'hyperméthylation à la DNase et de conservation ont été utilisées. Sur les 5 enhancers candidats les plus forts, 2 étaient connus et préalablement étudiés dans la base de données d'enhancers VISTA(89). Les 3 autres ont été étudiés chez des souris transgéniques en utilisant le gène rapporteur LacZ. Quatre de ces 5 enhancers montrent une activité LacZ au niveau des membres durant leur développement, dont 3 montrant un haut degré de superposition avec l'expression endogène connue d'*EPHA4*.



**Figure 32** : Résumé des remaniements au locus *WNT6/IHH/EPHA4/PAX3*

Adapté de Lupiáñez et collaborateurs(108)

En A), données de Hi-C. On observe que le locus est organisé en 3 TADs distincts.

En B), C) et D), remaniements aboutissant aux trois phénotypes différents : brachydactylie, F-syndrome et polydactylie. Dans chacun des cas, l'implication de la barrière d'un TAD dans le remaniement aboutit à une interaction ectopique entre un gène et un cluster d'enhancers situés dans le TAD d'*EPHA4*.

A noter que la délétion que présente le fœtus que nous rapportons correspond à la délétion *Dbf*, retrouvée chez le modèle murin *doublefoot*.

d) Application du modèle au cas fœtal

Si l'on applique ce modèle au cas de notre fœtus, on observe que la délétion minimum de 972 kb inclut uniquement la barrière entre le TAD *WNT6/IHH* et le TAD d'*EPHA4*, et correspond à la délétion présente chez le modèle murin *doublefoot*. La conséquence de cette délétion chez le fœtus serait donc l'interaction entre le promoteur d'*IHH* et le cluster d'enhancers d'*EPHA4*.

Le phénotype présenté par le fœtus est par ailleurs très similaire à celui du modèle murin *doublefoot*, ainsi qu'à celui du patient porteur d'une duplication d'*IHH*, avec notamment dans ces 3 cas la présence d'une polydactylie en miroir. On peut noter cependant que le fœtus présente d'autres anomalies sévères non retrouvées chez le patient porteur de la duplication, comme la phocomélie, ou les anomalies génitales. Une des hypothèses pourrait être que l'absence de copie normalement exprimée d'*IHH* sur l'un des allèles dans le cas de la délétion aggrave le phénotype. Une autre hypothèse que l'on ne peut pas exclure est l'implication dans le phénotype des gènes inclus dans la délétion.

Il existe donc des arguments très forts étayant l'hypothèse d'une altération des TADS par cette délétion au locus *WNT6/IHH/EPHA4/PAX3*. Ce fœtus constituerait le premier cas humain correspondant réellement au modèle murin *doublefoot*, et un nouvel exemple de remaniement pathogène à ce locus.

L'exploitation des données de notre cohorte de CGH-array nous a permis également de mettre en lumière les CNVs décrits ci-dessous comme candidats potentiels à une altération de TAD. Néanmoins, en raison de la difficulté d'interprétation des conséquences de ces CNVs sur l'architecture de la chromatine, et en l'absence d'étude fonctionnelle, il convient de rester prudent quant aux conclusions à en tirer.

## 2. Délétion au locus *ID4/E2F3/SOX4*

### a) Description Clinique et moléculaire

Nous décrivons ici le cas d'un garçon de 4 ans, né d'un couple non apparenté. Il s'agit du premier enfant du couple. L'histoire familiale n'est pas informative. Il a une sœur cadette, en bonne santé. La grossesse s'est déroulée sans complication. Il est né à terme, avec un poids de 2,415 kg, pour une taille à 48 cm et un périmètre crânien à 34 cm. Il présente une fente palatine postérieure avec luette bifide, traitée chirurgicalement.

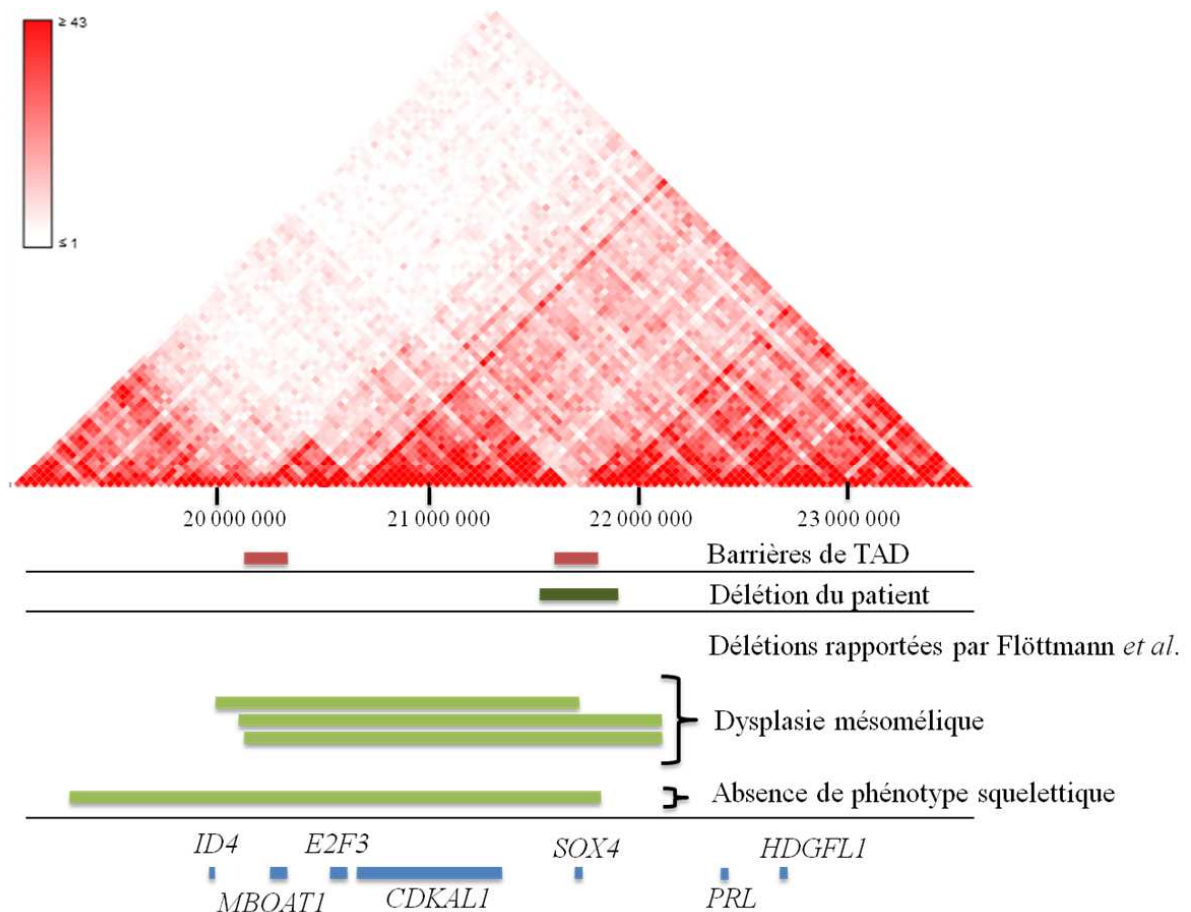
A l'âge d'un an, des examens d'imagerie sont réalisés devant une augmentation de son périmètre abdominal, et retrouvent une tumeur intrarénale droite, avec métastases pulmonaires. L'examen anatomo-pathologique conclut à un néphroblastome de sous type épithélial.

Il évolue par ailleurs avec un décalage modéré des acquisitions, avec une marche acquise à 18 mois, et un retard de langage avec quelques mots à l'âge de 4 ans. A noter qu'après réévaluation à distance, l'évolution sur le plan cognitif est favorable, et ce décalage a pu être le fait d'un hospitalisme.

A l'examen clinique on note plusieurs critères physiques mineurs, à savoir des bosses frontales, une racine du nez large, un philtrum long avec lèvre supérieure fine, et un pli palmaire transverse unique gauche.

Une CGH-array 60K a été réalisée, et retrouve la présence d'une délétion interstitielle à l'état hétérozygote *de novo* de 389 kb au locus 6p22.3 : chr6:g.(21390094)\_(21778878)(hg19)del. Cette délétion inclut 2 gènes codants *RefSeq* : *CASC15* et *SOX4* (**Figure 33**).





**Figure 33** : Délétion au locus *ID4/E2F3/SOX4*

En vert sombre, la délétion présentée par notre patient en 6p22.3 [arr 6p22.3(21,390,094-21,778,878)](Array-CGH 60K, hg19). Au-dessus, données de Hi-C publiées par Dixon et collaborateurs(30).

*CASC15* est inclus dans un locus de prédisposition au neuroblastome(136), et est associé à la progression du mélanome(137). Il n'existe cependant pas d'autre argument pour relier la délétion de ce gène au phénotype présenté par notre patient, bien qu'il puisse participer au mécanisme de survenu du néphroblastome.

*SOX4* code pour un facteur de transcription de la famille SOX, et fait partie du sous-groupe SOX C, avec *SOX11* et *SOX12*(138). Ces 3 facteurs de transcription présentent un haut degré d'homologie, ont des patterns d'expression proches, et des fonctions se recoupant partiellement. Par ailleurs, des délétions hétérozygotes et des variants hétérozygotes *de novo* entraînant la perte de fonction de *SOX11* ont été rapportés chez des patients présentant un phénotype correspondant au syndrome de Coffin-Siris(139).

*SOX4* est connu pour avoir un rôle critique durant l'embryogénèse(140), et en particulier dans la neurogénèse(141–143). Par ailleurs, il a été montré que *SOX4* est nécessaire à la néphrogénèse, et est une cible de *WT1*(144). *WT1* est une protéine à doigt de zinc pouvant agir comme activateur ou répresseur de la transcription, en fonction du contexte cellulaire. *WT1* est impliqué dans différentes formes syndromiques de prédisposition à la survenue d'un néphroblastome comme le syndrome WAGR(145), le syndrome de Denys-Drash(146) et le syndrome de Frasier(147). Des mutations somatiques de *WT1* sont par ailleurs retrouvées dans 5 à 10% des néphroblastomes de survenue sporadique(148). Il joue par ailleurs un rôle majeur dans le développement du système uro-génital(149). La famille de facteurs de transcription SOX-C interagit également avec la voie Wnt/ $\beta$ -catenin durant la formation du cartilage de croissance(150). Par ailleurs, il a été montré que *SOX4* est également impliqué dans la fusion et l'extension du palais durant l'embryogénèse(151). Ce rôle dans la formation du palais peut être rapproché de l'existence d'une fente palatine chez notre patient.

Des modèles animaux KO pour *SOX4* ont été réalisés. La souris KO pour *SOX4* n'est pas viable, avec une mort fœtale à E14. Elle présente des malformations cardiaques, dont une communication interventriculaire et un développement incomplet de la valve pulmonaire(152). Le modèle murin avec double délétion *SOX4-SOX11* spécifique du cortex présente une réduction de la taille des hémisphères cérébraux et des bulbes olfactifs(142). La souris avec délétion de *SOX4* spécifique des néphrons présente un nombre réduit de néphrons(144).

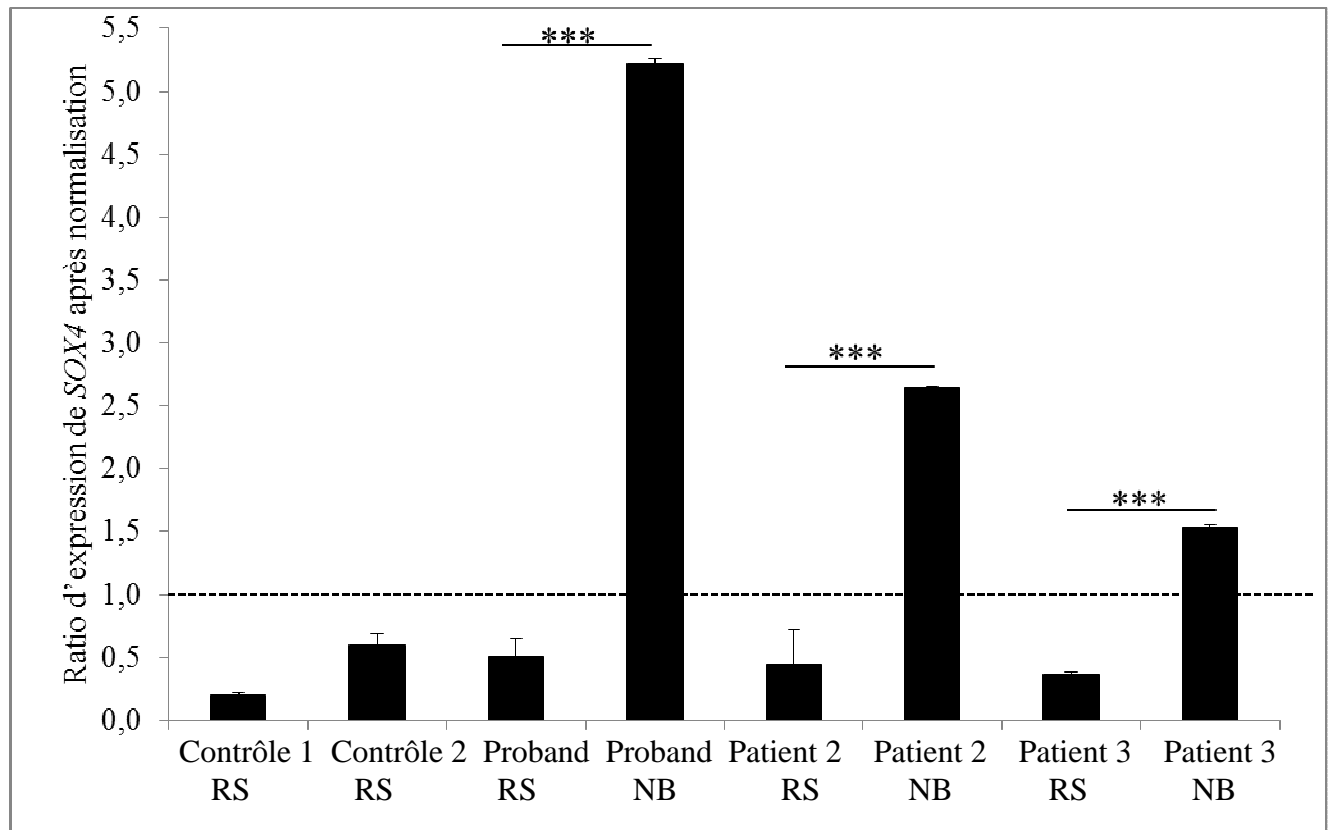
*SOX4* est également impliqué dans la tumorigénèse de nombreux cancers. Il a été montré que *SOX4* peut être considéré comme un biomarqueur de mauvais pronostic dans le carcinome de la vésicule biliaire(153), le cholangiocarcinome(154), le cancer de la prostate(155), et le cancer du poumon non à petites cellules(156). Un des mécanismes physiopathologiques possibles expliquant cela est un rôle anti-apoptotique de *SOX4*, comme suggéré par des expériences d'ARN interférent réalisées sur des lignées cellulaires dérivées de carcinome adénoïde kystique(157).

## b) Etudes complémentaires

Dans l'hypothèse d'un modèle "two-hit" de Knudson, nous avons réalisé le séquençage de *SOX4* par technique de Sanger à partir d'ADN extrait du néphroblastome de notre patient, à la recherche d'un second évènement. Nous n'avons retrouvé aucun variant délétère dans la séquence codante, ni dans les jonctions introns-exons. Nous avons également réalisé ce séquençage sur de l'ADN issu de néphroblastomes de 3 autres patients, présentant des formes familiales de néphroblastomes. Nous n'avons pas non plus mis en évidence de variant chez ces autres patients.

Nous avons réalisé des expériences de RT-PCR à partir d'ARN extrait du rein sain et du néphroblastome de notre patient, de 2 patients avec néphroblastome familial, ainsi que de reins sains de 2 contrôles (**Figure 34**). Nous avons mis en évidence une surexpression de *SOX4* dans le néphroblastome du patient, comparativement à son rein sain, avec un ratio de 10 (n=4). Chez les 2 autres patients, on retrouve également une surexpression comparativement au rein sain, avec des ratios de 4 et de 6 respectivement.

*SOX4* est un gène fréquemment surexprimé dans de nombreux cancers(158). Cependant, une surexpression apparaît comme paradoxale chez notre patient porteur d'une délétion hétérozygote de *SOX4*. Cette surexpression pourrait néanmoins être le fait des différents remaniements génomiques tardifs se produisant dans une tumeur, et n'apporte pas d'information sur le niveau d'expression de *SOX4* durant les premières phases de la tumorigénèse du néphroblastome, durant la vie fœtale.



**Figure 34** : Etude d'expression de *SOX4*

Cette étude d'expression de *SOX4* a été réalisée dans du tissu rénal sain (RS) et dans les néphroblastomes (NB) du cas index, de 2 autres patients, et de 2 contrôles. Le niveau d'expression est normalisé par rapport à 2 gènes de référence, *B2M* et *GUSB*, et la significativité a été testé par un test de Student (\*\*\*)  $p < 0,001$

c) CNV au locus *ID4/E2F3/SOX4* dans la littérature

Les données sur les différents rôles biologiques de *SOX4* suggèreraient donc un lien causal entre la délétion hétérozygote de *SOX4* présentée par notre patient, et la survenue d'un néphroblastome. Cependant, l'étude de l'ARNm n'a pas montré de baisse d'expression de *SOX4*, dans l'hypothèse où le mécanisme physiopathologique serait une simple haploinsuffisance de *SOX4*. Le séquençage de l'allèle en *trans* de la délétion n'a pas non plus mis en évidence la survenue d'un second évènement somatique.

De plus, Flöttmann et collaborateurs(109) ont rapporté plusieurs patients présentant des délétions au locus 6p22.3, comprenant le gène *SOX4*. Or aucun cas de néphroblastome ni d'autres cancers n'a été décrits chez eux, démontrant que la seule délétion de *SOX4* n'est pas suffisante pour induire la survenue d'un néphroblastome (**Figure 33**).

Cette étude a été réalisée dans le but d'identifier les causes moléculaires conduisant à la survenue de la dysplasie mésomélique de type Savarirayan. Cette maladie est caractérisée par une absence de fibula et des tibias triangulaires et hypoplasiques, associés à un raccourcissement mésomélique modéré des membres supérieurs, avec élargissement proximal des ulnas(159). Chez trois patients présentant ce phénotype, Flöttmann et collaborateurs ont retrouvé la présence de délétions hétérozygotes *de novo* au locus 6p22.3, incluant 4 gènes : *MBOAT1*, *E2F3*, *CDKAL1* et *SOX4*. Un patient précédemment décrit dans la littérature présentait également une délétion similaire(160).

Les auteurs rapportent également un autre patient présentant une déficience intellectuelle, sans signe de dysplasie osseuse, porteur d'une délétion hétérozygote de ce même locus, et incluant un gène supplémentaire en plus des 4 pré-cités, le gène *ID4*. Le cas de cet autre patient montre que l'haplo-insuffisance de *MBOAT1*, *E2F3*, *CDKAL1* ou *SOX4* ne peut pas être le mécanisme physiopathologique de la survenue de la dysplasie mésomélique de type Savarirayan.

D'après les données de Hi-C de Dixon et collaborateurs(30), ce locus est constitué de 3 TADs : un TAD télomérique comprenant *ID4*, un TAD médian comprenant *MBOAT1*, *E2F3*, et *CDKAL1*, et un TAD centromérique. *SOX4* se situe au niveau de la barrière entre le TAD médian et le TAD centromérique.

Par ailleurs, les données de ChIP-Seq publiées par Cotney *et al*(161) permettent d'identifier deux enhancers potentiels spécifiques des membres, inclus dans le TAD centromérique de ce locus.

Les délétions des 3 patients présentant une dysplasie mésomélique emportent la barrière entre les TADs télomérique et médian, le TAD médian dans son intégralité et la barrière entre le TAD médian et le TAD centromérique. On peut donc prédire que la conséquence sur l'architecture du locus sera la création d'un néo-TAD comprenant *ID4* et les 2 enhancers spécifiques des membres, permettant une interaction entre eux, et donc une expression ectopique d'*ID4* dans les membres. Il a été montré qu'*ID4* a une fonction dans la différenciation des ostéoblastes, et que le modèle murin *Id4*<sup>-/-</sup> n'est pas viable, et présente un

retard de croissance sévère(162). Une expression ectopique d'*ID4* dans les membres pourrait donc expliquer la survenue d'une dysplasie mésomélique chez ces patients.

Ce modèle explique également l'absence de phénotype squelettique chez le patient présentant une déficience intellectuelle : la délétion dont il est porteur inclue le gène *ID4*, qui ne peut donc pas être dérégulé.

#### d) Application du modèle de Flöttmann

Si l'on applique ce modèle physiopathologique à la délétion que présente le patient que nous rapportons, nous pouvons observer que la délétion qu'il présente inclut seulement la barrière entre les TADs médian et centromérique, située à proximité de *SOX4*.

Les 3 autres gènes délétés chez les patients présentant une dysplasie mésomélique, *MBOAT1*, *E2F3* et *CDKAL1*, sont donc potentiellement dérégulés par une mise en contact avec des enhancers situés dans le TAD centromérique.

*MBOAT1* appartient à la super famille des O-acétyltransférase membranaire(163), et *CDKAL1* est une méthylthiotransférase(164). Il n'existe pas d'argument dans la littérature permettant de relier une dérégulation de ces gènes à la survenue d'un néphroblastome.

*E2F3* est lui un facteur de transcription oncogène connu et précédemment étudié(165–167), faisant partie de la famille des protéines E2F, interagissant avec pRB, et impliquées de façon critique dans le contrôle du cycle cellulaire. De plus, un lien entre l'expression d'*E2F3* et l'agressivité des tumeurs de Wilms a été mis en évidence(168). On peut donc faire l'hypothèse qu'une surexpression d'*E2F3*, due à sa position dans un néo-TAD comprenant un cluster d'enhancers, a pu intervenir dans la survenue d'un néphroblastome chez notre patient.

Un argument qui renforcerait cette hypothèse serait la présence d'un enhancer actif dans les tissus rénaux en formation à proximité du cluster d'enhancers spécifiques des membres. Cette présence d'enhancer peut être supposée : il s'agit d'une région inter-génique pauvre en gène, et montrant une zone annotée comme enhancer fort sur plusieurs lignées cellulaire à partir des données issues du projet ENCODE(169). Il n'existe néanmoins pas de données obtenues à partir de cellules embryonnaires rénales.

Le phénotype de notre patient, associant une fente palatine, des critères physiques mineurs et la survenue d'un néphroblastome pourrait être donc dû à la conjugaison d'une haplo-insuffisance de *SOX4*, et de la dérégulation d'*E2F3*. Ce cas a pu faire l'objet d'une publication(170).

### 3. Délétion au locus *FMN2*

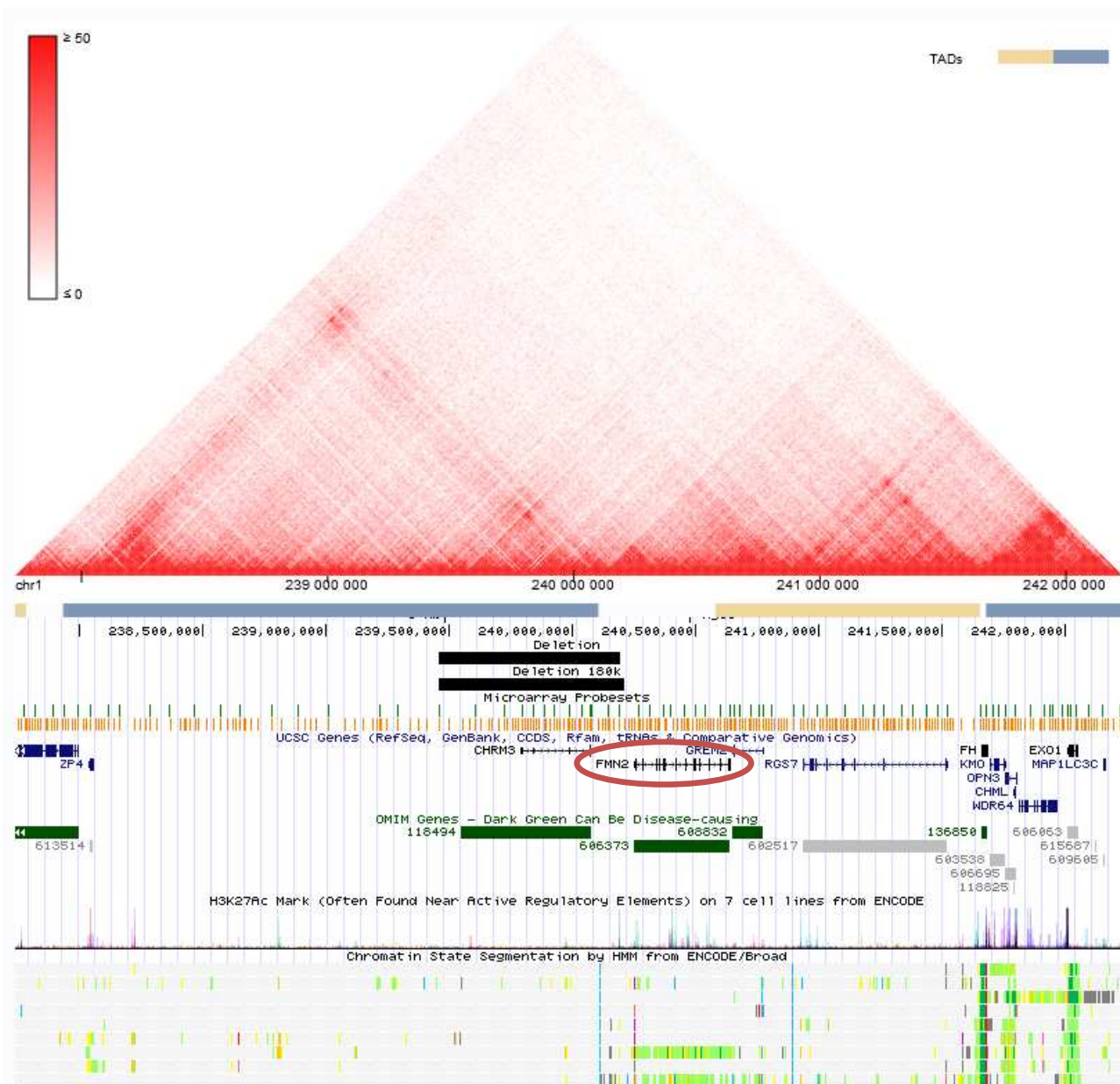
Nous décrivons ici le cas d'un enfant de 16 ans, adressé en consultation de génétique pour bilan étiologique d'un retard des acquisitions et d'une obésité. Il s'agit d'un enfant unique, issu d'un couple non apparenté. A l'interrogatoire, on retient l'existence de troubles des apprentissages chez sa mère, qui a arrêté sa scolarité en CM1-CM2, ainsi qu'un surpoids chez son père. Le reste de l'histoire familiale n'est pas informative. La grossesse pour cet enfant s'est déroulée sans complication. La naissance a eu lieu par césarienne pour bassin étroit à 39 semaines d'aménorrhée. A la naissance, l'APGAR était à 10/10. Le poids était de 3,280kg pour une taille à 50cm et un PC à 35cm. Par la suite, il a évolué avec un retard prédominant sur le langage, sans retard sur le plan moteur, avec une marche acquise à l'âge de 10 mois. Il présente notamment des troubles de la prononciation et une difficulté dans la construction des phrases. Il est actuellement scolarisé en 3<sup>ème</sup> ULIS, et bénéficie d'un suivi orthophonique. Dans ses antécédents, on note une amygdalectomie en 2011, et un syndrome d'apnées du sommeil appareillé. Sur le plan du comportement, il est décrit comme plutôt sociable et ne présente pas d'agressivité, ni d'intolérance à la frustration.

A l'examen clinique, il pèse 97,5kg (+4,6DS) pour une taille à 170,5cm (+0,3DS) et un PC à 57,5cm (+0,9DS). L'IMC est calculé à 33,7kg/m<sup>2</sup>. L'examen clinique montre une obésité faciotronculaire avec une bosse de bison. Il existe un épicanthus bilatéral. L'examen des extrémités met en évidence une hyperlaxité des petites articulations, une syndactylie cutanée partielle II-III. Il existe un acanthosis nigricans discret au niveau du cou. La puberté est à un stade normal pour son âge G4P4.

Dans le cadre des explorations à visée étiologique, un dosage du taux de leptine a été effectué, et retrouvé normal. Un caryotype et une FISH 22q11.2 ont été réalisés, également normaux.

Des CGH-arrays 60K, puis 180K identifient une délétion hétérozygote en 1q43 d'une taille maximale de 819kb, emportant un gène codant, *CHRM3*, et située à 25kb en amont du gène *FMN2*. Les parents de ce patient n'ont pas souhaité réaliser l'étude de la ségrégation de ce variant.

L'étude des données publiées de Hi-C(30) montre que cette délétion emporte une barrière du TAD comprenant *CHRM3*, situé en position centromérique par rapport à *FMN2* (**Figure 35**).



**Figure 35** : Délétion au locus *FMN2*

La visualisation des données permet d'observer que la délétion se situe en amont de *FMN2*, et interrompt la barrière entre deux TADs adjacents. Afin de préciser les bornes de la délétion, une CGH-array 180K a été réalisée, dont le design des sondes est visible en orange. Les coordonnées de la délétion détectée par cette CGH-array, d'une taille minimale de 751kb, et maximale de 819kb, sont précisées sous l'étiquette « Deletion 180K » : on observe que celle-ci n'inclut pas *FMN2*. Par ailleurs, on peut noter la correspondance entre la barrière entre les deux TADs, et la présence d'une fixation d'insulateur sur toutes les lignées cellulaires, sur la piste de la segmentation de la chromatide ChromHMM.



*CHRM3* code pour le récepteur muscarinique à l'acétylcholine M3. Ce récepteur est responsable de la neurotransmission du sphincter pupillaire de l'iris et du muscle détrusor.

Dans la littérature, on retrouve le cas de plusieurs patients présentant un variant perte de fonction homozygote du gène *CHRM3*(171,172). Ces patients présentent une atteinte de la vessie, allant d'une incapacité à réaliser une vidange complète par acontractilité du détrusor à un prune-belly-like syndrome. Ces patients présentent par ailleurs une mydriase aréactive. Ces symptômes sont donc en rapport avec la perte de fonction connue de *CHRM3*, et très similaires à ceux que présentent le modèle murin *Chrm3<sup>-/-</sup>*(173). Cependant, il n'existe pas dans la littérature d'arguments laissant à penser que la délétion hétérozygote de ce gène soit impliquée dans le phénotype de notre patient.

*FMN2* code lui pour la formine 2, protéine faisant partie des 15 protéines formines identifiées dans l'espèce humaine. Ces protéines sont caractérisées par la présence d'un domaine catalytique formin homology 2 (FH2), et sont toutes impliquées dans la régulation du cytosquelette d'actine(174). Via son action sur la dynamique du réseau d'actine, *FMN2* coordonne la relocalisation du fuseau mitotique des ovocytes murins durant la métaphase I, permettant l'établissement de la polarité cellulaire et l'élimination du globule polaire(175). A noter que les femelles du modèle murin *Fmn2<sup>-/-</sup>* présentent une hypofertilité, liée à un défaut de cette fonction.

*FMN2* joue également un rôle dans le système nerveux central, et est majoritairement exprimé dans le cerveau en développement, ainsi qu'à l'âge d'adulte(176). Il s'agit notamment d'un gène identifié comme étant régulé par l'apprentissage chez la souris, avec la présence d'une acétylation des histones H4K12 après réalisation de tests de conditionnement à la peur(177). De plus, le modèle murin *Fmn2<sup>-/-</sup>* présente un défaut d'apprentissage par association, avec sur le plan neuro-anatomique une densité d'épines dendritiques diminuée au niveau des hippocampes(178).

Chez l'humain, des délétions hétérozygotes *de novo* de *FMN2* ont été rapportées chez 2 patients présentant une déficience intellectuelle(179,180). Par ailleurs, des variants bi-alléliques entraînant la perte de fonction de *FMN2* ont été rapportés chez 5 individus issus de 2 familles consanguines différentes et présentant une déficience intellectuelle non syndromique(178). Les cellules neurales dérivées de cellules pluripotentes de ces patients présentent une densité synaptique réduite. Ces résultats, concordant avec les anomalies neuro-anatomiques présentées par le modèle murin KO, sont en faveur de l'hypothèse que *FMN2* est

nécessaire à la régulation du réseau d'actine durant le développement, la maturation et le remodelage synaptique.

Si l'haplo-insuffisance de *FMN2* entraîne un phénotype, il a également été montré que la surexpression de *FMN2* dans des modèles cellulaires NIH/3T3 entraîne d'importantes modifications des réseaux d'actine et de microtubules, avec une diminution du nombre de fibres de stress(181).

La dérégulation spatiale ou temporelle de l'expression de *FMN2* due à la disparition d'une des barrières du TAD où il se situe pourrait donc jouer un rôle dans le phénotype neurocomportemental de notre patient.

#### 4. Délétion au locus d'*UBE2A*

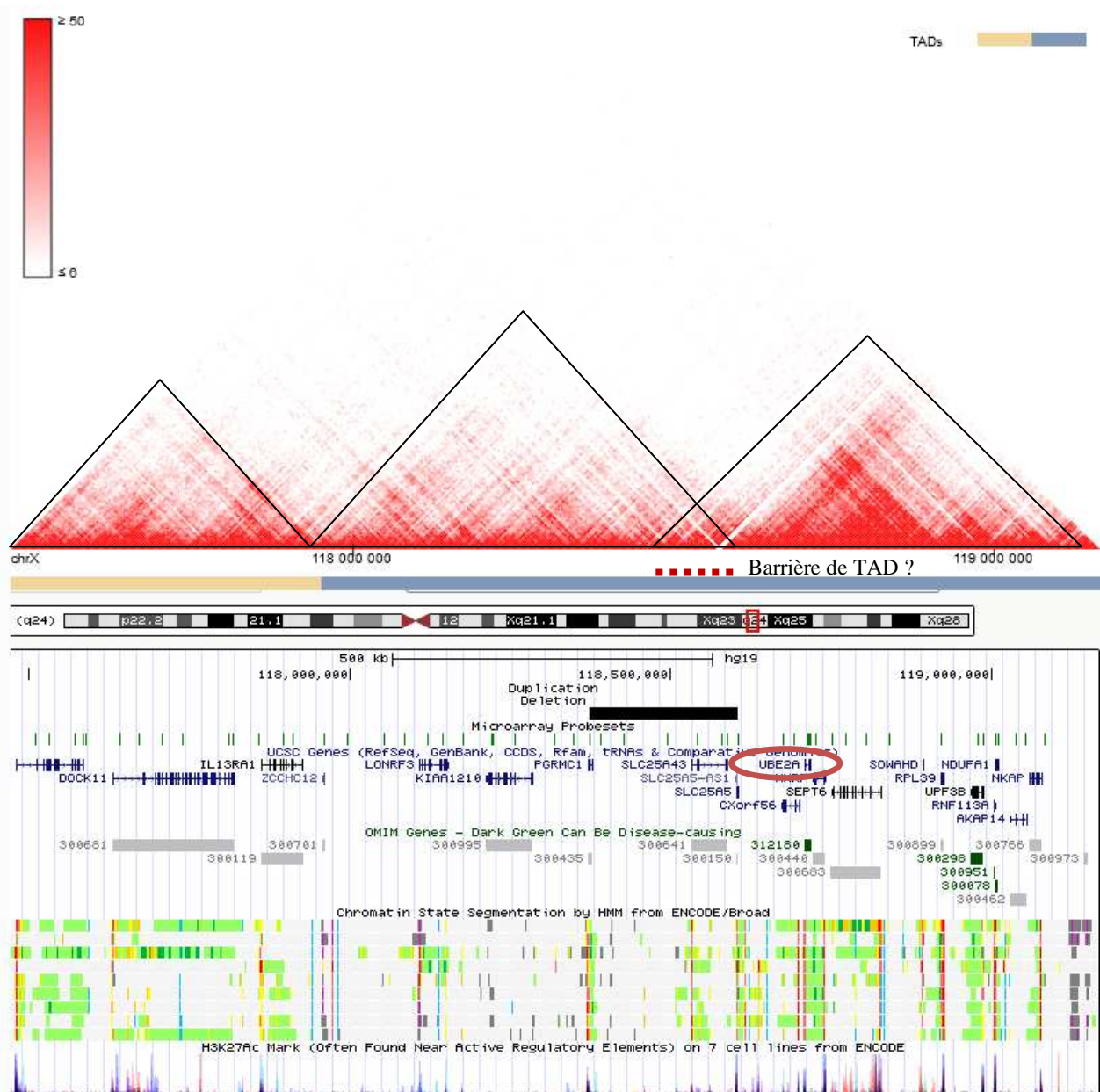
Nous décrivons ici le cas d'un patient âgé de 24 ans, et suivi dans le cadre d'une encéphalopathie épileptique. Il s'agit du premier enfant d'une fratrie de deux, issu d'un couple non apparenté. Il a une sœur puînée, en bonne santé. La grossesse s'est déroulée sans particularité, et il est né à 37 semaines d'aménorrhée par césarienne, sans notion de souffrance néonatale, avec des mensurations normales. Il a par la suite évolué avec un retard de langage, ainsi qu'un léger retard sur le plan moteur, avec une station assise acquise à 7 mois et demi, et une marche à 21 mois. Il présente par ailleurs d'importants troubles du comportement, ainsi qu'une épilepsie à type d'absences, contrôlée sous DEPAKINE. A l'examen clinique, on ne retrouve pas de critères morphologiques notables, et il évolue sur la courbe des +2 DS pour le poids, +2,5 DS pour la taille et +2,5 DS pour le PC. Il a été pris en charge jusqu'à l'âge de 20 ans en IME, puis a été orienté vers un foyer occupationnel. Sur le plan des examens complémentaires, une IRM cérébrale a été réalisée et retrouvée normale. Le caryotype standard réalisé est normal, et le séquençage d'un panel de gènes impliqués dans les encéphalopathies épileptiques par technologie NGS n'a pas mis en évidence de variant pathogène. La réalisation d'une CGH-array 60k a cependant mis en évidence la présence d'une délétion de 230kb située en Xq24, et comprenant 3 gènes codants : *PGRMC1*, *SLC25A43*, et *SLC25A5*. Cette délétion est par ailleurs située à 108kb en amont du gène *UBE2A*. Les données de Hi-C montrent que cette délétion pourrait emporter l'une des barrières du TAD d'*UBE2A* (**Figure 36**). Cependant, il faut noter que si cette barrière est visualisable sur la matrice d'interaction, et qu'elle se colocalise à des marques annotées

comme insulateur par ChromHMM, elle n'est pas annotée comme barrière par l'algorithme Arrowhead. De plus, la zone d'incertitude de la CGH-array est de 68kb à cette borne, ne permettant pas de conclure sans autre analyse plus résolutive quant aux conséquences de cette délétion sur l'architecture du locus.

*UBE2A* code pour une E2 ubiquitine ligase, impliquée comme les autres protéines de cette famille dans la dégradation protéique par le protéasome. Initialement identifié comme l'homologue de la protéine Rad6 chez la levure, *UBE2A* est impliquée dans la réparation de l'ADN et la répression de la transcription, et joue un rôle dans l'apprentissage et la mémoire(182). Des variants entraînant la perte de fonction d'*UBE2A* ont été mis en évidence chez des patients présentant une déficience intellectuelle syndromique lié à l'X, dite de type Nascimento(183), caractérisée par une macrocéphalie et par des traits dysmorphiques comprenant notamment un synophris, des rebords orbitaux proéminents et des yeux enfoncés dans les orbites, une macrostomie, un hirsutisme, une implantation anormale des cheveux avec tourbillons anormaux, un micropénis et une onychodystrophie. Les mères porteuses de ces variants délétères ont une intelligence normale, mais peuvent présenter une légère dysmorphie(184).

Des délétions emportant *UBE2A* ont également été rapportées chez des patients présentant un phénotype similaire(185), renforçant l'hypothèse que le mécanisme physiopathologique de cette maladie est une haploinsuffisance d'*UBE2A*. Cinq de ces patients porteurs d'une délétion en Xq24 présentent également une communication interventriculaire, qui n'est pas rapportée chez les patients porteurs d'un variant ponctuel d'*UBE2A*. La présence de cette cardiopathie congénitale a été attribuée par les auteurs(185) au fait que ces délétions emportaient également *SLC25A5*, gène codant pour un translocateur mitochondrial de nucléotides à adénine, et dont le modèle murin KO présenterait des malformations cardiaques similaires.

La délétion portée par notre patient pourrait donc aboutir à une dérégulation d'*UBE2A*, qui jouerait un rôle dans la survenue de l'encéphalopathie qu'il présente.



**Figure 36** : Délétion au locus d'*UBE2A*

La visualisation des données de Hi-C semble montrer que cette région est organisée en 3 TADs, surlignés en noir. Cependant, ces données présentent un bruit de fond important, gênant une identification non équivoque des TADs. La barrière entre le TAD médian et le TAD télomérique n'a pas été annotée comme telle par l'algorithme Arrowhead (pointillés rouges), mais semble bien présente, et corrélée avec la présence de motifs de fixation de CTCF (en bleu sur la segmentation de la chromatine ChromHMM). Par ailleurs, le design des sondes de CGH-array 80K (piste Microarray Probesets, en vert) laisse une zone d'incertitude de 68kb pour la borne télomérique.

En 2013, des patients présentant des délétions similaires ont été rapportés dans la littérature par Vandewalle et collaborateurs(186). Ces patients issus de 3 familles différentes présentent une déficience intellectuelle, sans critères dysmorphiques notables. Les auteurs font une hypothèse différente que celle de la dérégulation d'*UBE2A*. La région minimale critique des délétions de ces patients comprend 2 gènes, *SLC25A43* et *SLC25A5*, et une délétion emportant uniquement *SLC25A43* a été mise en évidence chez un garçon asymptomatique. Par ailleurs, il a été montré que *SLC25A5* a une expression cérébrale, notamment dans le cortex et l'hippocampe. Vandewalle et collaborateurs(186) proposent donc que seule l'haploinsuffisance de *SLC25A5* est responsable du phénotype de leurs patients. Un potentiel effet de position sur *UBE2A* a néanmoins été envisagé, et une étude de l'expression d'*UBE2A* par RT-qPCR sur lignée lymphoblastoïde a été réalisée, ne montrant pas de différence d'expression par rapport aux contrôles.

Il est tout à fait possible que l'haploinsuffisance de *SLC25A5* joue un rôle dans le phénotype de ces patients. Néanmoins, des critiques peuvent être émises par rapport à cette hypothèse. Le phénotype des patients porteurs d'une délétion emportant *UBE2A* et *SLC25A5* ne semble pas être plus grave que celui des patients porteurs d'un variant ponctuel d'*UBE2A*, comme attendu si *SLC25A5* était à lui seul responsable d'une déficience intellectuelle. Par ailleurs, la dérégulation de l'expression d'un gène par une rupture de barrière de TAD peut être à la fois tissu spécifique et propre à un stade de développement. L'absence de différence de l'expression d'*UBE2A* par rapport aux contrôles dans les lignées lymphoblastoïdes indique que ces délétions n'entraînent pas de dérégulation massive et ubiquitaire, mais n'apporte pas d'élément quant à une possible dérégulation durant la neurogénèse. De la même façon, l'absence de dysmorphie comparable entre les patients rapportés par Vandewalle et collaborateurs(186) et les traits caractéristiques de la déficience intellectuelle de Nascimento montrent que cette potentielle dérégulation d'*UBE2A* n'est pas une simple haplo-insuffisance, mais peut-être des changements plus subtils dans le schéma d'expression.

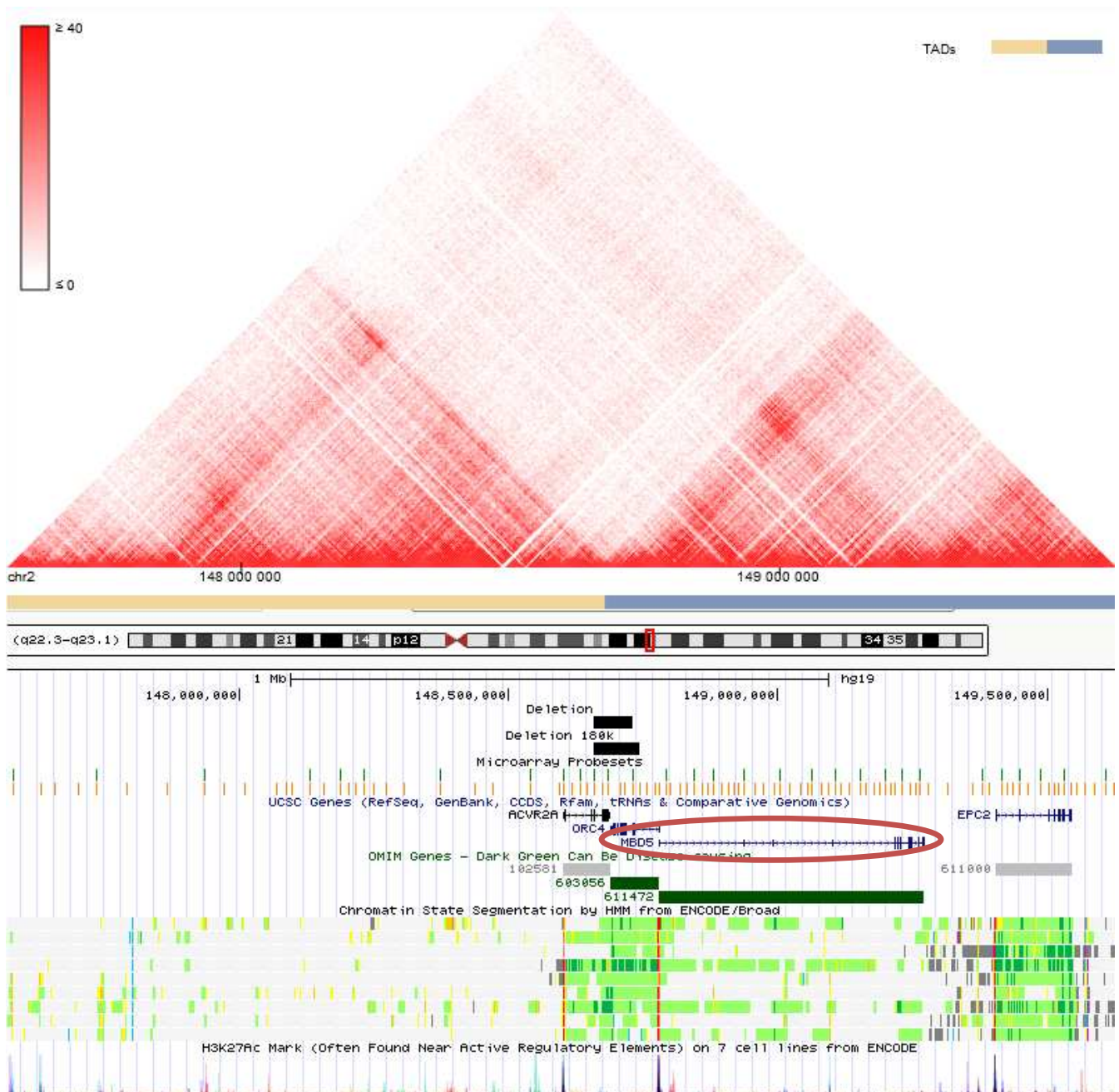
En l'absence d'études fonctionnelles supplémentaires, l'effet exact des délétions en Xq24 sur les différents gènes du locus ne peut être affirmé, et il est possible qu'un effet combiné sur *SLC25A5* et *UBE2A* soit responsable du phénotype. Dans tous les cas, cette délétion semble constituer l'anomalie causale chez notre patient. La vérification de la ségrégation de ce variant dans la famille sera prochainement réalisée.

## 5. Délétion au locus *MBD5*

Ce patient né en 2010 est pris en charge dans un CAMSP extérieur au CHU de Bordeaux dans le cadre du suivi d'un retard des acquisitions. Il présente une déficience intellectuelle modérée isolée, sans élément retenu comme notable à l'examen des critères physiques mineurs. Dans les antécédents familiaux, on note de possibles difficultés d'apprentissage chez son père, sans autre exploration. Afin de compléter les éléments cliniques disponibles pour ce patient, il sera prochainement revu au sein d'un service de génétique médical extérieur.

La réalisation d'une CGH-array 60k a permis de mettre en évidence une délétion de 74 kb en 2q22.3 (chr2:148657042-148731085), incluant deux gènes codant, *ACVR2A* et *ORC4*, et dont la zone d'incertitude comprend une partie de la séquence codante d'un 3<sup>ème</sup> gène, *MBD5*. La réalisation d'une CGH-array 180k a permis de mieux préciser les coordonnées de cette délétion : celle-ci est alors estimée à 87kb (chr2:148657042-148743939), et en définitive n'inclut pas *MBD5*. Une étude de la ségrégation a été réalisée, et ce variant est hérité du père du patient.

La visualisation des données sur le 3D Genome Browser permet de voir que ce locus est organisé en 2 TADs distincts, dont la frontière se situe entre *ACVR2A* et *ORC4* (**Figure 37**). La délétion emporte exactement la région prédite comme étant la barrière entre ces 2 TADs.



**Figure 37** : Délétion au locus *MBD5*

On observe que cette délétion comprend la barrière de deux TADs adjacents.

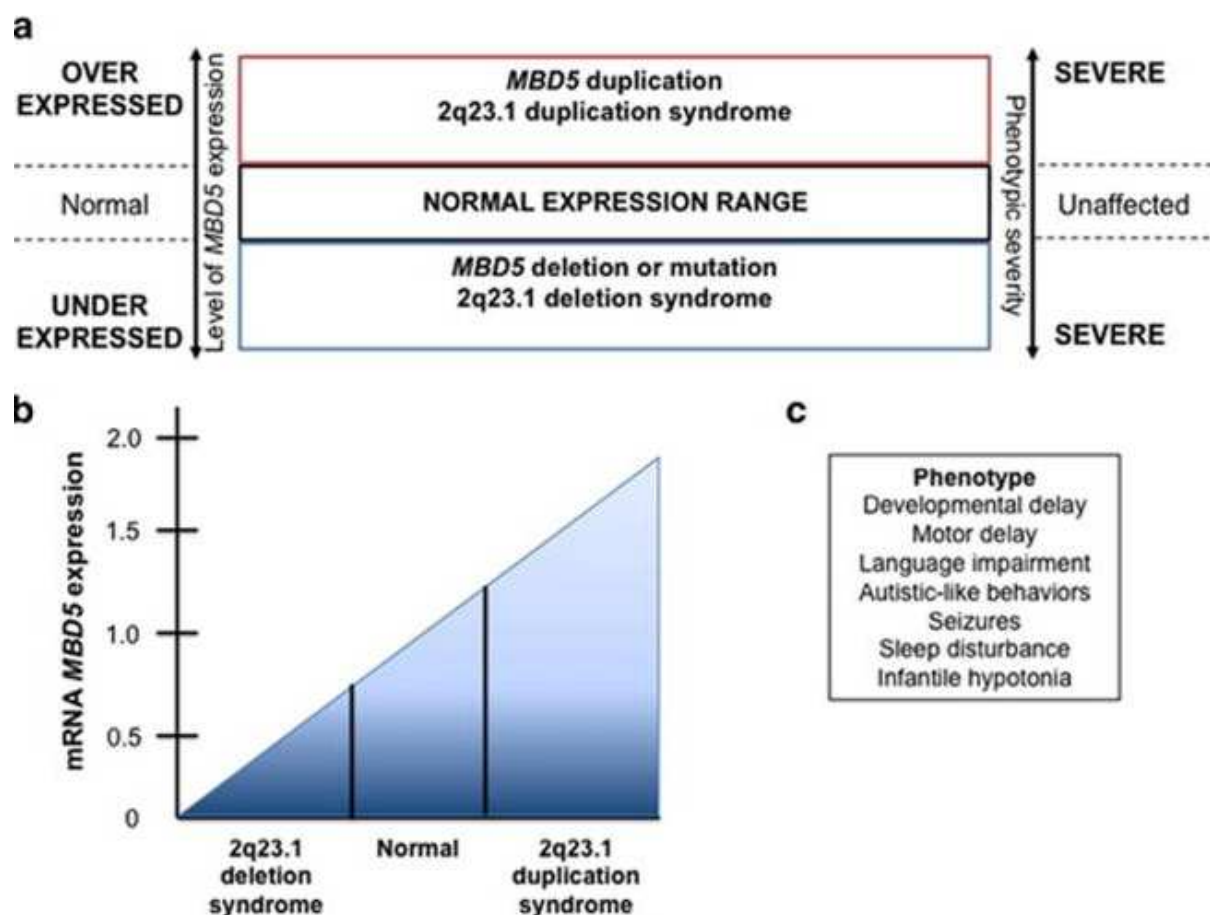
Afin de préciser les bornes de la délétion, une CGH-array 180K a été réalisée, dont le design des sondes est visible en orange. Les coordonnées de la délétion détectée par cette CGH-array sont précisées sous l'étiquette « Deletion 180K » : on observe que celle-ci n'inclut pas le gène *MBD5*.

*ACVR2A* code pour l'un des deux types de récepteur à l'activine, et jouerait un rôle de régulation autocrine ou paracrine dans le placenta(187), ainsi que dans l'axe gonadotrope(188). Il n'existe cependant pas d'argument dans la littérature permettant de relier une délétion hétérozygote d'*ACVR2A* au phénotype de notre patient.

*ORC4* code pour un homologue humain d'une sous-unité de complexe de reconnaissance des origines de réplication chez la levure. Des variants bi-alléliques de ce gène ont été identifiés comme responsables du syndrome de Meier-Gorlin, forme de nanisme primordial microcéphalique(189) de transmission autosomique récessive. Outre le retard de croissance et la microcéphalie, ce syndrome est caractérisé par une dysmorphie incluant des petites oreilles, et une absence ou une hypoplasie des rotules. On ne peut pas exclure qu'une délétion hétérozygote de ce gène ne participe pas au phénotype de notre patient. Cependant, celui-ci ne présente pas les symptômes spécifiques du syndrome de Meier-Gorlin, et les variants hétérozygotes d'*ORC4* n'ont pas été associés à un phénotype dans la littérature. De plus, 62 individus de la base de données GnomeAD(190) sont porteurs d'un variant frameshift ou non-sens à l'état hétérozygote dans *ORC4*.

*MBD5* code pour une protéine appartenant à la famille des protéines à domaine de fixation aux méthyl-CpG, et jouerait un rôle dans la régulation de la transcription. Ce gène a été associé à la survenue de déficience intellectuelle de transmission autosomique dominante, et ce via différents mécanismes moléculaires possibles. Des délétions en 2q23.1 incluant *MBD5* ont d'abord été identifiées chez des patients présentant une déficience intellectuelle, associées à des troubles du spectre autistique (191,192). Puis des variants hétérozygotes perte de fonction(193), ainsi que des duplications de *MBD5* entraînant une surexpression(194), ont été rapportés comme associés à des phénotypes similaires. *MBD5* apparaît donc comme un gène dont la dérégulation, quelle qu'elle soit, entraîne un phénotype comprenant une déficience intellectuelle et/ou un trouble du spectre autistique (**Figure 38**). A noter que 3 études(94,193,195) ont rapporté des patients dont l'anomalie de *MBD5* a été transmise d'un parent sain, suggérant une expressivité variable.





**Figure 38** : Phénotypes neurodéveloppementaux liés à *MBD5* : un gène, différents mécanismes

Adapté de Mullegama et Elsea(192)

La sévérité du phénotype lié à *MBD5* est fonction du degré de dérégulation de l'expression induit par l'anomalie moléculaire présentée par le patient. Néanmoins, haploinsuffisance et surexpression induisent le même phénotype, comprenant un retard global, des traits autistiques, et une épilepsie.

La délétion que présente notre patient pourrait induire une dérégulation de *MBD5*, via la disparition d'une barrière de TADs. Un autre mécanisme physiopathologique est néanmoins possible : la délétion pourrait inclure un élément régulateur de *MBD5*, dont la disparition expliquerait à elle seule la dérégulation. Cependant, les données disponibles ne mettent pas en évidence l'existence de tels éléments régulateurs.

Le fait que cette délétion soit héritée du père n'est pas un élément en faveur d'une pathogénicité certaine. Cependant, celui-ci présenterait des troubles des apprentissages, et il a

été montré que les phénotypes neuro-développementaux associés à une dérégulation de *MBD5* sont d'expressivité variable, avec des variants pathogènes parfois hérités de parents sains.

Des études de l'expression de *MBD5* par RT-qPCR sur lymphocytes chez notre patient seront prochainement réalisées, afin de mettre en évidence cette possible dérégulation.

## 6. Duplication au locus *WASHC5*

Nous décrivons ici le cas d'un patient né en 2006, et adressé en consultation de génétique pour une association malformative. Dans ces antécédents familiaux, on note une surdité chez le grand-père paternel et des troubles cardiaques nécessitant une chirurgie chez l'oncle paternel. Du côté maternel, les parents, un frère et une sœur sont atteints de trouble visuel, sans autre précision.

La grossesse s'est déroulée sans particularité et l'accouchement a eu lieu à 38 semaines d'aménorrhée par voie basse. Il n'y a pas eu de souffrance néonatale, et le poids de naissance était de 2,027 kg. Durant la période néonatale, l'examen clinique retrouve une fossette sacrée, ainsi qu'une microtie bilatérale asymétrique, prédominant à gauche. L'exploration ORL retrouve une surdité de transmission bilatérale. Le scanner des rochers a mis en évidence un aspect dysmorphique des deux enclumes et une mauvaise visualisation des étriers ainsi qu'une absence de contact entre les osselets et l'oreille interne.

Une échographie cardiaque met en évidence une communication inter-auriculaire, qui sera opérée à l'âge de 7 ans. Les radiographies du rachis montrent par ailleurs une hypoplasie C4-C5 associée à un pincement discal postérieur.

Sur le plan psychomoteur, la marche a été acquise à 16 mois, et un retard de langage a été constaté en raison de son atteinte auditive. Depuis la mise en place d'un appareillage auditif, l'évolution sur le plan psychomoteur est bonne, avec une scolarisation en milieu ordinaire.

À l'examen clinique à 7 ½ ans, il pèse 20 kg (-1 DS) pour une taille à 122 cm (Médiane) et un périmètre crânien à 48 cm (-3,5 DS). On note une asymétrie faciale discrète, une microtie bilatérale plus marquée à gauche, des cils longs, un nez rectiligne, une microstomie, un aspect anormal des dents pouvant être en rapport avec des caries notamment des canines supérieures, un pli palmaire transverse unique à droite et une fossette sacro-coccygienne profonde borgne. Il existe une hyperlaxité des coudes et des genoux, des pieds plats prédominants à gauche et une discrète exostose claviculaire.

La CGH-array 60k réalisée a mis en évidence la présence d'une duplication de 174 kb en 8q24.13 (ch8:125989614-126163472), comprenant entièrement 3 gènes codants, *ZNF572*, *SQLE* et *KIAA0196* (**Figure 39**). Un 4<sup>ème</sup> gène est lui compris partiellement dans la duplication, *NSMCE2*.

L'analyse des données sur le 3D Genome Browser montre que ce locus est composé de 2 TADs distincts. L'annotation des TADs par l'algorithme Arrowhead situe la frontière au niveau des premiers exons de *KIAA0196*, mais l'analyse visuelle de la matrice d'interaction et des marques de fixation d'insulateurs de ChromHMM laissent à penser que la frontière se situe plutôt proche du deuxième exon de *NSMCE2*.

*ZNF752* code pour une protéine à doigts de zinc de fonction inconnue, et *SQLE* code pour la squalène epoxidase, enzyme catalysant la première étape d'oxygénation dans la biosynthèse des stérols. Il n'existe pas d'argument dans la littérature permettant de relier une surexpression ou une dérégulation de *ZNF752* ou de *SQLE* à la survenue d'une pathologie humaine.

*NSMCE2* code pour une SUMO ligase, jouant un rôle dans la réparation de l'ADN, et identifiée comme l'un des gènes responsables du syndrome de Seckel, forme de nanisme microcéphalique primordial de transmission autosomique récessive(196). *NSMCE2* n'étant que partiellement compris dans la duplication, soit celle-ci est en miroir, entraînant une interruption du gène, soit elle est en tandem, n'entraînant pas d'interruption, mais n'entraînant pas non plus d'interaction ectopique possible des enhancers et la copie complète du gène.

*KIAA0196*, également identifié comme *WASHC5*, code pour la strumpelline, protéine exprimée de façon ubiquitaire, et localisée dans le cytosol et le réticulum endoplasmique(197). Cette protéine est une des sous unités du complexe WASH, activateur du complexe Arp2/3 et donc de la polymérisation de l'actine(198). Deux maladies distinctes, et de transmissions différentes, ont été associées à ce gène.

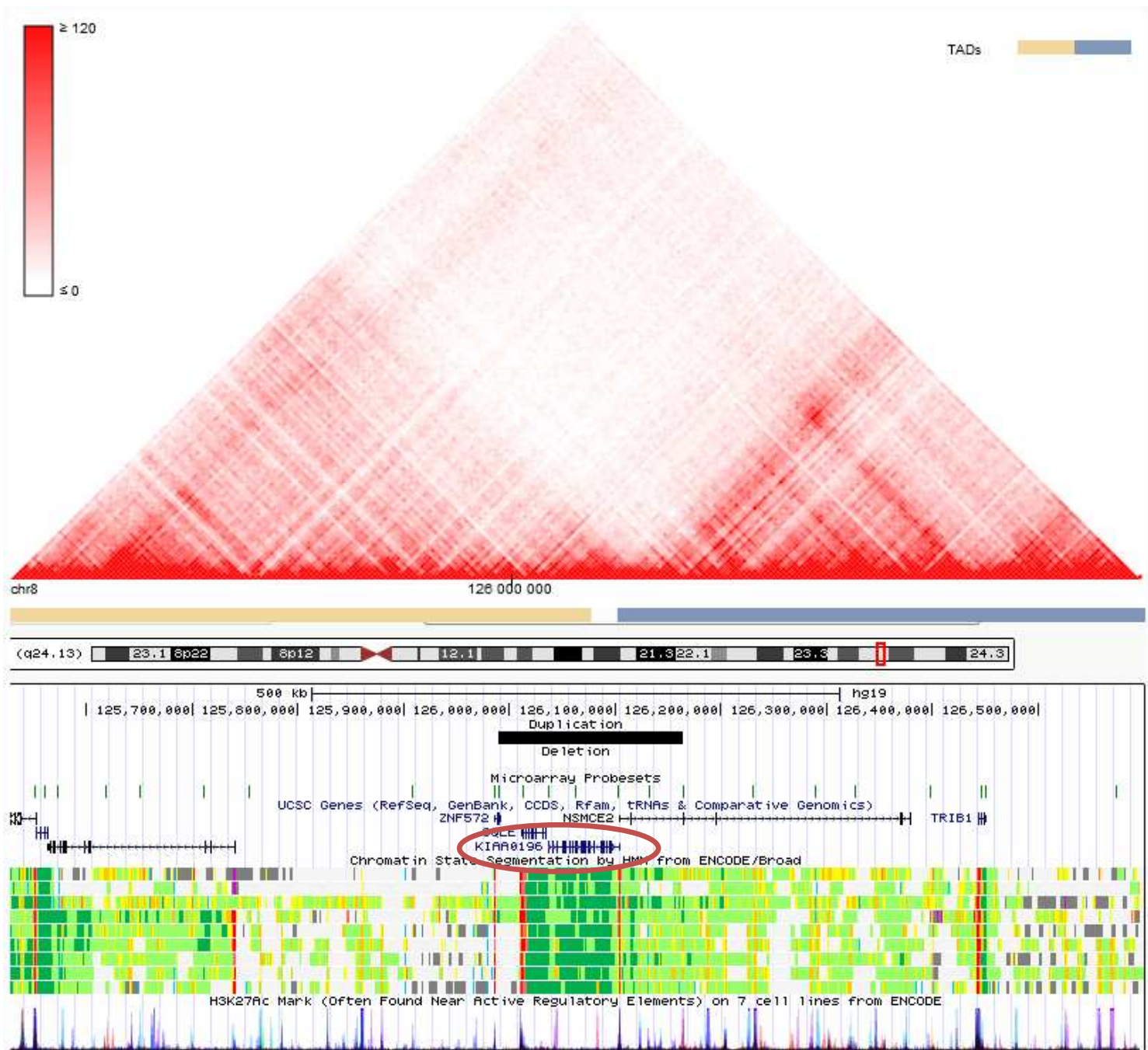
Des variants faux-sens hétérozygotes ont été rapportés dans la littérature comme responsable de paraplégie spastique héréditaire de type 8, de transmission autosomique dominante(199). Il a été montré que le mécanisme physiopathologique le plus probable de ces variants est un effet gain de fonction toxique(200).

D'autre part, Elliott et collaborateurs(201) ont rapporté la présence d'un variant d'épissage homozygote de *WASHC5*, entraînant la production d'une protéine tronquée, exprimée à un taux réduit de 60% par rapport au contrôle, chez 8 patients présentant un syndrome de

Ritscher-Schinzel, ou syndrome 3C, pour cranio-cérébello-cardiaque. Ces patients présentent une déficience intellectuelle sévère. Les anomalies craniofaciales incluent des fentes palatines, des colobomes oculaires et une dysmorphie caractéristique, avec occiput et front proéminent, oreilles basses implantées, fentes palpébrales orientées vers le bas, hypoplasie des os propres du nez, et micrognathie. Il existe des anomalies de la fosse postérieure, dont des malformations de Dandy-Walker et des hypoplasies vermiennes. Enfin, ces patients présentent des cardiopathies congénitales de type cono-troncal. Les 8 patients décrits par Elliott et collaborateurs(201) appartiennent tous à la communauté amérindienne de la région Nord du Manitoba, au Canada.

La visualisation des données de Hi-C montre qu'en incluant entièrement *WASHC5* (*KIAA0196*), ainsi que la frontière entre les 2 TADs, cette duplication, si elle est en tandem, peut entraîner une interaction ectopique entre des éléments régulateurs et *WASHC5*.

De par son rôle ubiquitaire, et l'existence d'anomalies crâniofaciales dans le syndrome de Ritscher-Schinzel, nous pouvons faire l'hypothèse qu'une dérégulation de *WASHC5* pourrait contribuer au syndrome malformatif que présente notre patient. On ne peut cependant pas écarter que la seule surexpression de *WASHC5*, induite par cette duplication qui n'a jamais été décrite dans la littérature, ne soit en définitive responsable du phénotype. L'étude de la ségrégation de cette duplication sera prochainement réalisée, et dans le cas où celle-ci serait *de novo*, pourrait apporter un élément supplémentaire en faveur de sa pathogénicité.



**Figure 39** : Duplication au locus *WASHC5* (*KIAA0196*)

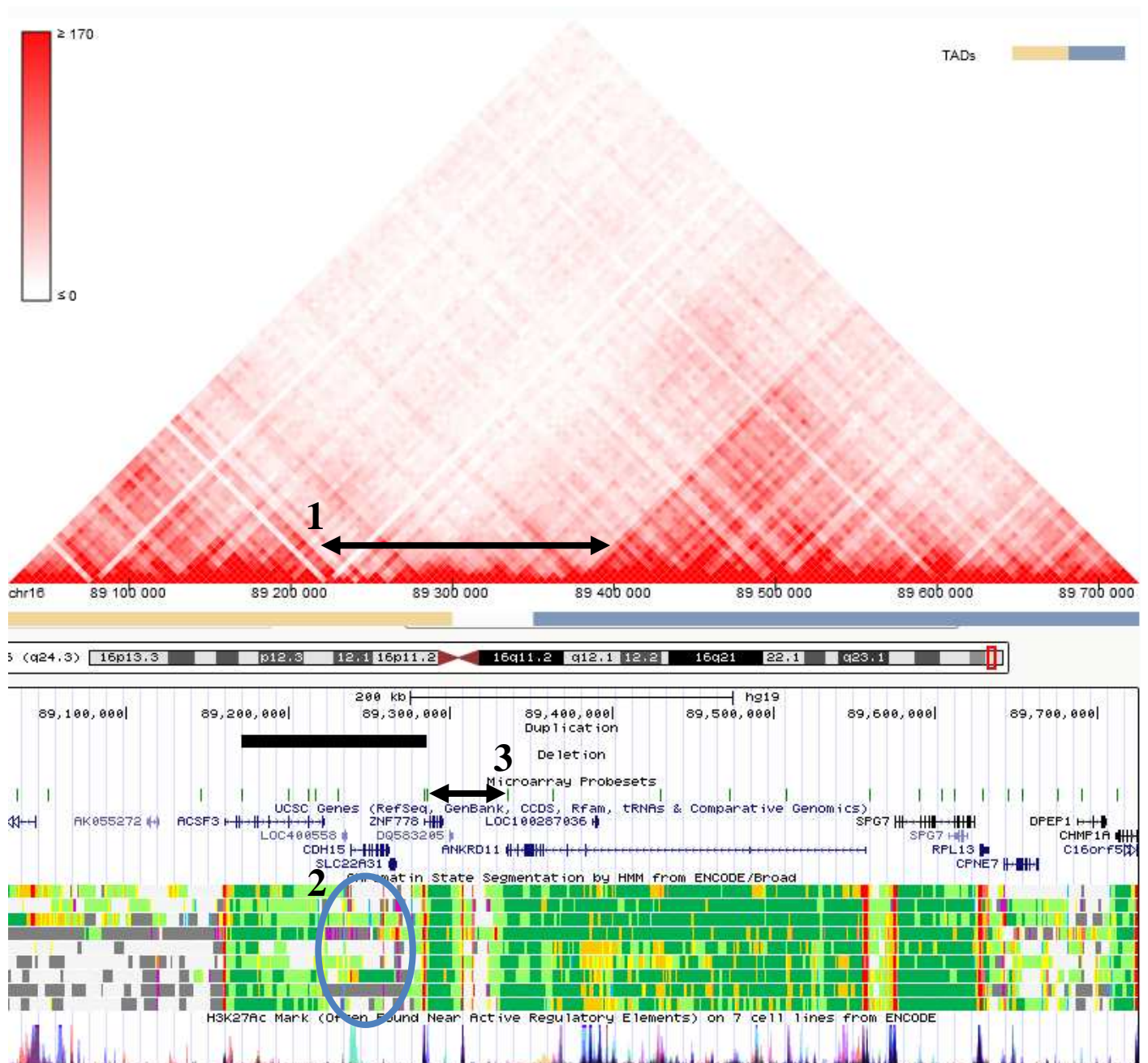
La duplication comprend entièrement la séquence codante de *WASHC5* (ou *KIAA0196*), ainsi que la barrière entre deux TADs consécutifs. Par ailleurs, de nombreuses marques H3K27Ac, pouvant correspondre à des enhanceurs, sont incluses dans la duplication.

## 7. Duplication au locus d'*ANKRD11*

Ce patient âgé de 20 ans a été adressé en consultation de génétique pour bilan d'une déficience intellectuelle avec troubles du comportement. Il s'agit du premier enfant d'une fratrie de 2, issu d'un couple non apparenté. Il a une demi-sœur côté maternelle, et une sœur, toutes deux en bonne santé. La grossesse comme l'accouchement se sont déroulés sans particularité. Il est né à terme à 40 semaines d'aménorrhée, avec un score d'Apgar à 9/10, un poids de naissance à 4070g, une taille à 59cm et un PC à 38 cm. Il a par la suite évolué avec un retard de langage, sans retard sur le plan moteur, avec une marche acquise à 12 mois. L'examen clinique retrouve une lèvre inférieure éversée, et une langue protruse, sans autre élément dysmorphique notable. A l'âge de 20 ans, il mesure 1,80 m pour 95kg, avec un périmètre crânien de 59 cm. En l'absence de photographie du patient, nous n'avons pas pu réaliser une étude a posteriori des critères physiques mineurs de ce patient.

La CGH-array réalisée retrouve la présence d'une duplication de 115kb, située sur le chromosome 16 (chr16:89170643-89285603). L'étude de la ségrégation de ce variant est en cours de réalisation. Cette duplication est située à 48kb du gène *ANKRD11*, codant pour une protéine membre de la famille des cofacteurs à répétition ankirine. Les variants entraînant la perte de fonction d'*ANKRD11* ainsi que des délétions hétérozygotes sont responsables du syndrome KBG, forme syndromique de déficience intellectuelle, associée à une petite taille, des anomalies squelettiques et une dysmorphie caractéristique(202). La visualisation des différentes données sur le 3D Genome Browser permet de constater que la barrière centromérique du TAD d'*ANKRD11* est potentiellement incluse dans la duplication. Si l'orientation de celle-ci est en miroir, des éléments régulateurs peuvent interagir de façon ectopique avec *ANKRD11*, et déréguler son expression, jouant ainsi un rôle dans la déficience intellectuelle de ce patient. Cependant, cette interprétation se heurte à plusieurs obstacles. Les données de Hi-C ne permettent pas d'établir clairement où se situe la barrière entre les deux TADs (**Figure 40**, flèche 1). L'annotation des TADs est elle-même peu précise, indiquant 2 TADs non contiguës, séparés par un espace non défini. Cette barrière pourrait se situer plus en amont de cette annotation, et être totalement incluse dans la duplication, comme indiquée par la présence de nombreuses marques annotées comme insulateur dans la segmentation ChromHMM (**Figure 40**, cercle bleu 2). Enfin, la région d'incertitude due à la résolution de la CGH-array complique également cette interprétation (**Figure 40**, flèche 3).





**Figure 40** : Duplication au locus d'*ANKRD11*

Flèche 1 : intervalle dans lequel la barrière entre les TADs pourrait se situer, d'après une interprétation visuelle des données de Hi-C

Cercle bleu 2 : à noter la présence de nombreuses bandes bleues, correspondant à une annotation en tant qu'insulateur, habituellement retrouvée comme co-localisé au niveau des barrières de TADs.

Flèche 3 : zone d'incertitude de la CGH-array 60k

Les données cliniques à notre disposition concernant notre patient ne semblent pas être en faveur d'un phénotype correspondant à un syndrome KBG tel que classiquement décrit. Cependant, il est possible là-encore que cette potentielle dérégulation d'*ANKRD11* soit spécifique d'un tissu ou d'une période du développement, et n'entraîne pas les mêmes conséquences qu'une perte de fonction germinale.

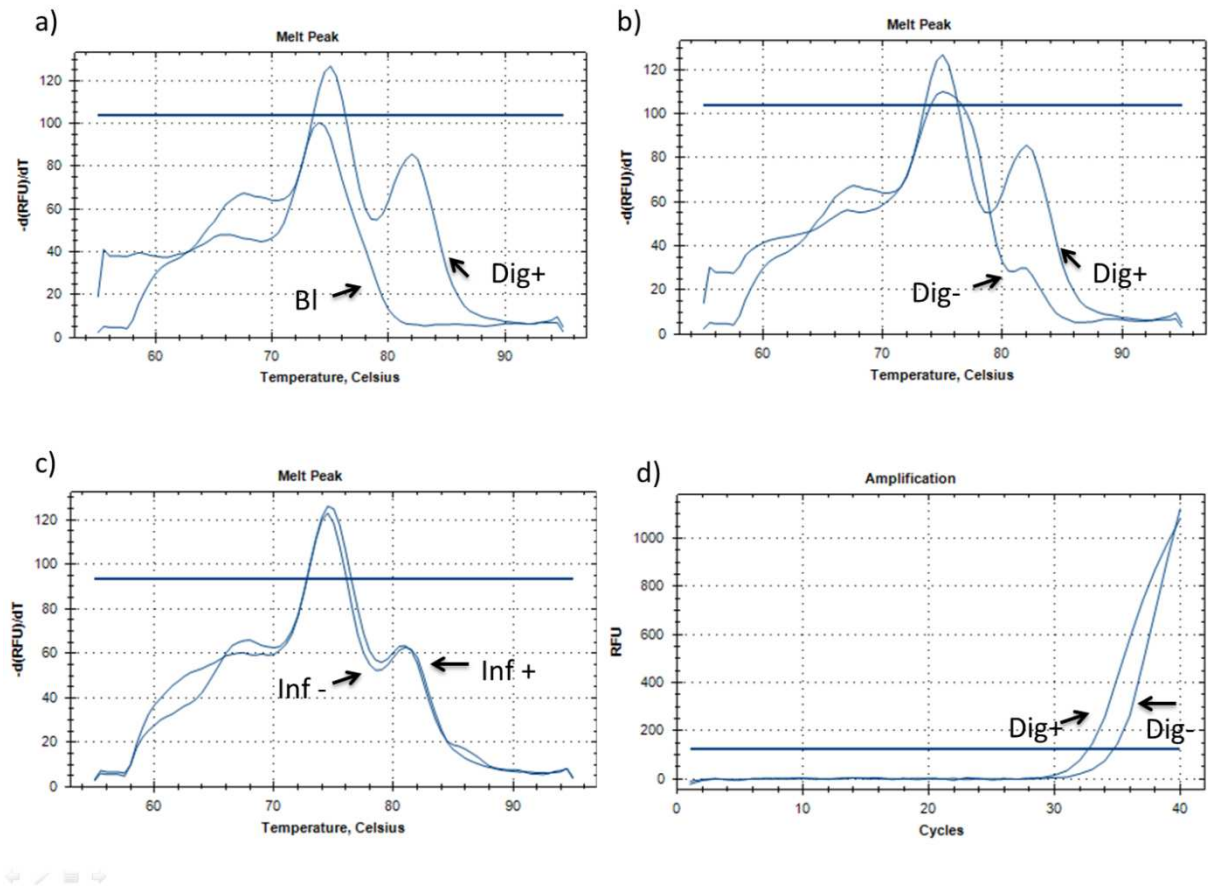
Ce cas illustre bien les difficultés d'interprétation auxquelles nous avons été confrontés lors de l'analyse de la cohorte : en l'absence d'étude de la ségrégation et de l'orientation du fragment, de données de CGH-array et d'Hi-C de plus haute résolution, ainsi que de la réalisation d'études fonctionnelles, on ne peut pas conclure formellement quant à la pathogénicité de cette duplication.

## B. Mise au point de la technique de 3C

Notre protocole de 3C, établi à partir des différents protocoles publiés dans la littérature, a été expérimenté sur des cellules Hela afin d'étudier les éléments régulateurs du locus du gène *CIITA*. Après obtention des librairies de 3C, 3 qPCR en temps réel avec courbes de fusion ont été réalisées, afin d'amplifier 3 produits de ligation possibles correspondant aux interactions entre le promoteur pIV et l'enhancer situé à -8kb, entre pIV et l'enhancer situé à -16kb, et entre ces 2 enhancers. Des échantillons témoins pour lesquels l'enzyme de restriction n'a pas été ajoutée durant le protocole ont également été analysés, afin de s'assurer de la spécificité des résultats obtenus.

On observe que pour la qPCR correspondant aux interactions entre pIV et l'enhancer à -8kb, la courbe de fusion présente une distribution bimodale (**Figure 41**, a) et b)). Un premier pic à un T<sub>m</sub> de 75°C est présent sur les échantillons, mais également sur les blancs d'échantillon, et les témoins non digérés : ce signal correspond probablement aux dimères d'amorces. Un second pic est présent à un T<sub>m</sub> de 82°C, soit le T<sub>m</sub> attendu pour le produit de ligation d'après la littérature(115) : ce pic n'est pas présent dans les blancs d'échantillon, ni chez le témoin non digéré. Ce signal pourrait correspondre à l'amplification du produit de ligation spécifique correspondant à une interaction entre le promoteur pIV et l'enhancer situé à -8kb.





**Figure 41** : Résultat des expériences de 3C au locus *CIITA* (n=3)

a) b) c) : Courbes de fusion des produits de ligation pIV/enhancer -8kb d'un échantillon, comparés à :

a) Blanc d'échantillon b) Témoin non digéré c) Echantillon obtenu après stimulation des cellules par interféron gamma.

d) Courbe d'amplification des produits de ligation pIV/enhancer -8kb, comparés au témoin non digéré.

Cependant, après incubation des cellules par de l'interféron gamma, nous ne constatons pas de différence entre les échantillons avec ou sans incubation, comme attendu d'après les données de la littérature (**Figure 41**, c)). De plus, cette amplification est trop faible pour nous permettre de réaliser une quantification (**Figure 41**, d)). Les qPCR réalisées pour les couples d'amorces correspondant aux interactions entre le promoteur pIV et l'enhancer situé à -16kb, ainsi qu'entre les enhancer situés à -8 et -16kb n'ont pas mis en évidence de signaux spécifiques, les courbes de fusion des échantillons et des témoins de digestion étant superposables, avec ou sans incubation à l'interféron gamma.

## IV. Discussion et conclusion

### A. Méthode de ré-analyse de la cohorte analysée par ACPA : jusqu'où pousser l'automatisation du pipeline ?

La ré-analyse de la cohorte de patients porteurs de remaniements chromosomiques de signification inconnue ou considérés comme probablement bénins nous a permis de mettre en évidence une très probable altération des TADs chez un cas fœtal, ainsi que chez 6 autres patients dont les CNVs impliquent potentiellement ce même mécanisme.

Afin de ré-analyser ces CNVs d'une façon efficace, une méthode automatisée comme celle proposée par Zepeda-Mendoza et collaborateurs(87) aurait pu être utilisée. Le principe de celle-ci repose sur une intersection des coordonnées des remaniements avec les coordonnées des TADs annotés dans des précédentes publications(26,30), d'une identification des éléments régulateurs des TADs impliqués via l'utilisation des bases de données publiques ; notamment issues du projet ENCODE, puis d'une priorisation des gènes à considérer comme potentiellement dérégulés via l'application d'un score basé sur les termes HPO correspondant aux phénotypes des patients.

Cependant, une telle méthode est fortement dépendante des données utilisées. Comme le conclue la revue de Dali et collaborateurs(67), les outils de prédiction des TADs utilisés par les différentes études ne sont pas reproductibles, et aucun n'est supérieur à un autre, ou à une annotation manuelle. Les données issues du projet ENCODE sont très informatives concernant la présence d'enhancers à l'échelle pangénomique, sans toutefois donner la spécificité temporelle et spatiale de ces enhancers. La base de données d'enhancers VISTA(89) repose elle sur des preuves fonctionnelles, mais de nombreux locus ne sont pas représentés. Par ailleurs, cette méthode automatique suppose une résolution des coordonnées des remaniements au nucléotide près par séquençage du génome complet, et une traduction en terme HPO du phénotype des patients.

De plus, les remaniements considérés sont différents dans notre cas, et dans l'étude de Zepeda-Mendoza(87). L'hérédité de la plupart des CNVs de notre cohorte n'a pas été étudiée, et on peut faire l'hypothèse qu'une grande majorité est non pathogène, et héritée d'un des deux parents. Même dans le cas où les CNVs sont *de novo*, il existe encore la probabilité que

l'effet pathogène soit l'haploinsuffisance pour une délétion, ou une surexpression pour une duplication, d'un des gènes compris dans le CNV, mais non encore décrits dans la littérature, plutôt qu'une perturbation de l'organisation des TADs. Les 17 remaniements étudiés par Zepeda-Mendoza(87) sont tous *de novo*, et n'interrompent pas de gènes, rendant très probable un effet de position, et donc bien plus rentable cette méthode de prédiction de l'effet sur les TADs.

Différents essais ont néanmoins été réalisés afin de reproduire partiellement cette méthode, en ne conservant pour analyse que les CNV incluant une barrière de TAD telle qu'annotée par les études de Dixon(30) (hESC HindIII) ou de Rao(26) (in situ MboI primary and replicate), puis en établissant la liste des gènes décrits dans la base de donnée OMIM(104) situés dans les TADs directement en amont et en aval de cette barrière. Ces analyses ont été réalisées sur la plateforme d'outils bio-informatiques Galaxy(106) : après traduction des coordonnées des CNVs, des barrières de TADs et des gènes OMIM au format .bed, les différents fichiers .bed sont intersectés par l'outil intersect du package BEDtools(105) (**Figure 27**).

Cette méthode n'a pas été concluante. L'étude par Rao et collaborateurs(26) utilisant l'algorithme Arrowhead est basée sur des données de Hi-C de meilleure résolution que les données de Dixon et collaborateurs(30), et devrait donc être préférentiellement utilisée afin de définir les barrières de TADs. Cependant, les TADs ainsi annotés ne sont pas forcément consécutifs, et peuvent être intriqués, empêchant une intersection simple avec les coordonnées des remaniements.

Par ailleurs, la priorisation des gènes présents dans les TADs par la base de données OMIM n'est pas pertinente : avec près de 3426 gènes associés à une maladie humaine dans cette base de donnée, la plupart des TADs comporte au moins un de ces gènes. Cependant, le phénotype auquel ils sont associés n'est que très rarement en rapport avec le phénotype du patient. De plus, le décalage entre la mise à jour de cette base de données et les dernières données de la littérature oblige à considérer également les autres gènes.

Si la méthode détaillée par Zepeda-Mendoza et collaborateurs(87) est efficace pour mettre en évidence l'effet de position spécifique d'un remaniement fortement candidat à ce type de mécanisme physiopathologique, elle semble ne pas être adaptée pour identifier ces remaniements au sein d'une large cohorte de CNVs « tout venant ». Dans ce cas, notre méthode reposant sur une interprétation manuelle de CNV à partir des différentes bases de données publiques semble être une approche plus pertinente, bien que fastidieuse.

## B. Taux de cas positifs : influence de la sélection des CNVs et de la technique utilisée

Sur notre cohorte, on peut constater un faible nombre de CNVs concernés potentiellement par une altération de TAD, soit 7, par rapport aux 735 CNVs ré analysés. Ce chiffre peut être comparé aux 7,3% de translocations équilibrées probablement pathogènes via une altération de TAD retrouvée dans l'étude réalisée par Redin et collaborateurs(76). Cependant, 72,5% de ces translocations se trouvent être *de novo*, ou d'une ségrégation compatible avec le phénotype, contrairement aux 735 CNVs que nous avons réanalysés, pour lesquels le caractère *de novo* ou hérité n'était connu que pour 40 d'entre eux. Il est probable qu'une ré-analyse ne portant que sur les CNVs suspects d'être pathogènes car *de novo* augmenterait significativement le taux d'altération de TAD retrouvé.

Une autre différence entre notre étude et celles portant sur des remaniements équilibrés(76,87) concerne la technologie utilisée, avec l'utilisation pour ces deux études de séquençage du génome complet, permettant la détermination des bornes des remaniements au nucléotide près. Les zones d'incertitude des CGH-arrays 60K utilisées, concernant généralement des zones intergéniques, compliquent également l'interprétation de nombreux CNVs, comme c'est le cas pour les patients porteurs d'une délétion au locus d'*UBE2A* (III.A.4) et d'une duplication au locus d'*ANKRD11* (III.A.7). L'utilisation de puces plus résolutive, comme les puces 180K utilisées pour les patients porteurs de délétions au locus *WNT6/IHH/EPHA4/PAX3* (III.A.1), *FMN2* (III.A.3) et *MBD5* (III.A.5), permet d'améliorer ce degré d'incertitude. La possible utilisation du séquençage de l'exome comme moyen courant d'exploration des remaniements chromosomiques ne répondrait que partiellement à ce problème, la résolution obtenue dépendant du kit de capture utilisé, qui peut inclure ou non des régions intergéniques. La technologie de référence à utiliser est en réalité le séquençage du génome complet : en plus d'une résolution au nucléotide près, celui-ci permettrait la détection des remaniements équilibrés tels que les translocations et les inversions, susceptibles d'altérer l'architecture chromatinienne.

Dans la littérature, il n'existe à ce jour qu'une seule étude, publiée par Di Gregorio et collaborateurs(110), portant sur l'analyse d'une large cohorte de patients par CGH-array, et ayant rapporté l'altération de TADs comme mécanisme physiopathologique. Les cas de 1015 patients présentant une déficience intellectuelle syndromique ou non sont rapportés dans cette

étude. Parmi eux, un patient présente une délétion interrompant la barrière du TAD de *CTNND2*, gène impliqué dans une forme de déficience intellectuelle avec traits autistiques, et dont la dérégulation pourrait expliquer le phénotype du patient. Ce rapport également faible, de 1 patient pour 1015, est en définitive du même ordre de grandeur que ce que nous avons retrouvé dans notre cohorte, avec 7 cas détectés parmi les 3271 ACPA réalisées au CHU de Bordeaux. Nous pouvons donc faire l'hypothèse que si l'interruption de TAD est un mécanisme physiopathologique privilégié pour certaines pathologies, comme les phénotypes associés au locus *KNCJ2/SOX9*(75), ou au gène *MEF2C*(76), il s'agit d'un phénomène rarement retrouvé au sein d'une cohorte de patients présentant des anomalies du développement. L'existence de ces cas, bien que rares, met néanmoins en lumière la nécessité de prendre en compte l'existence des TADs dans l'interprétation des variants de structure dans le cadre du diagnostic des anomalies du développement.

### C. Remaniements, intervalles critiques et gènes causaux : vers une reconsidération de certains syndromes ?

Cette étude de Di Gregorio et collaborateurs(110) rapporte également une patiente porteuse d'une large délétion du chromosome 8, qui ne présente pas, d'après la littérature, le phénotype associé aux gènes présents dans la délétion, à savoir *SULF1* et *SLCO5A1*. En effet, ces deux gènes sont compris dans l'intervalle minimale critique de délétions portées par des patients présentant un syndrome mésomélie-synostoses(MIM#600383)(203). Di Gregorio et collaborateurs(110) font l'hypothèse que ce syndrome précédemment décrit n'est pas lié à l'haploinsuffisance de *SULF1* et *SLCO5A1*, mais à une dérégulation impliquant le TAD contenant ces gènes, et dont l'une des barrières est comprise dans les délétions rapportées. L'intégration des données concernant les TADs dans l'analyse des remaniements chromosomiques permet donc de reconsidérer les bases moléculaires admises pour certains syndromes. Un gène dont l'haploinsuffisance est supposée être responsable d'une pathologie pourrait finalement ne pas être en cause : le mécanisme physiopathologique peut être l'effet de l'abolition d'une barrière d'un TAD sur un gène situé à distance. C'est cette hypothèse que nous soulevons, malgré l'absence de preuve fonctionnelle, pour le patient porteur d'une délétion proche d'*UBE2A* (III.A.4), alors que Vandewalle et collaborateurs(186) considéraient *SLC25A5* comme gène responsable du phénotype pour des délétions similaires.

Le mécanisme physiopathologique de variants structuraux entraînant la survenue d'un phénotype ne peut donc pas se déduire des seuls gènes présents dans l'intervalle critique identifié dans une cohorte de patients. L'ensemble des éléments participant à l'organisation tridimensionnelle du génome doit désormais être pris en compte.

#### D. Prédications *in silico*, preuves *in vitro* ?

Dans plusieurs cas positifs de notre cohorte, nous nous sommes heurtés aux limites des prédictions faisables à partir des données publiques disponibles. Ces données, que ce soit celles de Hi-C ou des marques épigénétiques, ne sont pas exhaustives. De plus, elles ne donnent pas toujours d'information sur le contexte fonctionnel de l'élément considéré, comme la différence entre boucle locale et TAD, ou la spécificité tissulaire d'un enhancer. Le seul moyen de tester les hypothèses que nous avons émises est la réalisation d'études fonctionnelles : c'est dans ce but que nous avons cherché à mettre au point la technique de 3C, qui permettrait d'affirmer l'existence d'interactions ectopiques chez nos patients. Les autres types d'études, sur l'ARNm, ou des modèles animaux sont à envisager en fonction de l'hypothèse soulevée, mais dans tous les cas, ces tests sont indispensables pour permettre de conclure sur la pathogénicité d'un variant.

Cette problématique est commune à l'étude des variants ponctuels des régions non codantes, tels que mis en évidence par le séquençage du génome complet.

Une méthode *in silico* permettrait néanmoins de contourner la réalisation de techniques dérivées de 3C pour l'étude des remaniements touchant les TADs : il s'agit de la réalisation de simulations en dynamique moléculaire, comme proposée par Chiariello et collaborateurs(36). Il est possible d'introduire des remaniements dans ces modèles bioinformatiques, et ainsi d'obtenir une prédiction fiable des données d'expériences de Hi-C. Il faut cependant noter que ces simulations nécessitent une puissance et un temps de calcul qui les rendent, pour l'instant, hors de portée du diagnostic en routine.

## E. Mise au point de la technique de 3C

Les résultats des expériences de 3C que nous avons réalisées ne nous permettent pas d'affirmer la détection d'une interaction au locus *CIITA* entre le promoteur pIV et l'enhancer à -8kb. Le pic présent sur la courbe de fusion est au Tm attendu, présent sur l'échantillon et absent du contrôle, mais n'est pas assez élevé pour conclure à une interaction réellement spécifique. De plus, l'absence d'effet de l'incubation par l'interféron gamma n'est pas en faveur de cette hypothèse.

Le protocole que nous avons établi va prochainement être modifié, grâce à la publication d'un protocole optimisé par Ea, Court et Forné(205). Un des principaux atouts de ce protocole est la réalisation d'une première étape d'isolement des noyaux, via l'utilisation de différentes solutions tampons permettant de garder intact l'architecture chromatinienne, avant de réaliser le cross-linking par formaldéhyde. La fixation des interactions est donc bien plus efficace, car elle n'est pas réalisée sur l'ensemble du volume cellulaire.

## F. Défaut de mise en place des TADs : un des mécanismes des cohésinopathies ?

Jiang et collaborateurs(44) ont montré que l'haploinsuffisance de la méthyltransférase *SETDB1* entraîne une perturbation de l'organisation chromatinienne localisée au locus des protocadhérines. *SETDB1* est par ailleurs connu comme gène impliqué dans la survenue de déficience intellectuelle(94,95). C'est le cas de nombreux autres gènes interagissant avec la chromatine, et notamment des gènes codant pour les protéines du complexe cohésine, directement responsable de la mise en place des TADs, et dont les variants délétères entraînent les différents syndromes de « cohésinopathies ». On peut par ailleurs noter que *CTCF*, autre gène codant pour un élément indispensable à la mise en place des TADs est également responsable d'une forme syndromique de déficience intellectuelle(205).

L'exemple de *SETDB1* peut donc amener à formuler une hypothèse plus large : la perturbation d'un ou plusieurs TADs est-elle un mécanisme physiopathologique de ces différentes maladies, et notamment des cohésinopathies ?

De par leur rôle essentiel, il paraît en effet plausible que l'haploinsuffisance de CTCF ou des gènes du complexe cohésine ait un impact sur la formation des TADs. Il est par ailleurs intéressant de noter que différents syndromes du spectre des « cohésinopathies », ou des « transcriptomopathies » (syndrome de Cornelia de Lange, de Wiedemann -Steinert, de Rubinstein-Taybi, KBG, etc...), dont les gènes responsables modifient la chromatine, présentent des anomalies des extrémités spécifiques, étant donné le rôle déterminant de l'organisation des TADs dans la formation des membres.

En dehors des remaniements chromosomiques, l'étude de l'architecture chromatinienne apparaît donc comme une piste possible à explorer pour expliquer la survenue de la symptomatologie de ces syndromes.

#### G. L'organisation chromatinienne : une aide à l'interprétation du séquençage du génome complet ?

En apportant une information essentielle sur les interactions possibles entre les séquences non codantes et les gènes, les données concernant l'architecture de la chromatine pourraient être utilisées dans le cadre de l'interprétation du séquençage du génome complet. Le séquençage du génome complet est une technique désormais régulièrement réalisée dans un cadre de recherche, et est appelée à l'être dans un futur proche dans un cadre diagnostique. Cependant, l'interprétation des variants situés en région non codante se heurte à l'absence de prédiction simple de leurs effets, comme il est possible de le faire pour les variants exoniques. Une méthode permettant l'interprétation des variants ponctuels considérés comme associés à un phénotype par les études de type GWAS a été proposée par Krijger et de Laat(59), et fait appel aux données disponibles sur les interactions chromatiniennes. En utilisant les informations disponibles sur les marques épigénétiques présentes sur le génome, on peut identifier quel variant se situe sur un élément régulateur. A partir des données de Hi-C identifiant TADs et boucles promoteur-enhancer, on peut prédire quel enhancer interagit avec quel gène. Le lien entre un variant d'une région non codante et un gène potentiellement causal est donc facilité par l'étude de l'organisation chromatinienne. Des études fonctionnelles sont néanmoins nécessaires par la suite pour étudier l'effet exact sur l'expression. Si cette méthode a été proposée pour les variants issus des études GWAS, elle pourrait également être appliquée à des variants identifiés par séquençage du génome complet, et dont la ségrégation serait compatible avec le phénotype.



## H. Conclusion

L'étude de l'organisation spatiale du génome, et en particulier des TADs, ouvre une nouvelle ère dans l'analyse des données issues des examens génétiques. Ces connaissances sont dès à présent applicables à l'interprétation des remaniements chromosomiques détectés par ACPA, et le seront prochainement à ceux détectés par séquençage du génome complet.

L'exploration des régions délétées ou dupliquées, et la définition d'intervalles critiques responsables de syndromes, ne doivent plus se limiter au seul contenu en gènes, mais doivent prendre en compte l'ensemble des régions environnantes, codantes et non codantes, et leur contexte chromatinien d'un point de vue tridimensionnel. La réanalyse des données d'ACPA semble indispensable, tout comme le développement de pipelines bioinformatiques robustes permettant d'identifier un potentiel impact sur un TAD, et la mise au point de tests fonctionnels le confirmant.

L'arrivée prochaine du séquençage du génome complet dans le diagnostic des anomalies du développement amène un certain nombre de défis, dont celui de l'interprétation des variants non codants. L'étude de l'architecture tridimensionnelle du génome, en apportant une information sur l'ensemble des interactions régulant la transcription, pourrait être une aide précieuse dans ce domaine.

## V. Annexes

### A. Détail de la méthode de tri des CNV

Un premier filtre a été effectué via Cartagenia Bench Lab CNV afin d'extraire l'ensemble des patients pour lesquels au moins un CNV était étiqueté dans les catégories « Variant de signification inconnue » ou « Probablement bénin », ou dans une catégorie synonyme. Ces catégories synonymes correspondent à des anciens classements ou erreurs de saisie de données de la cohorte et tendent à être remplacées. La liste de ces catégories est la suivante : attente parents, Benign?, en cours, Inférieur à 1M DPN, Inférieur à 200 kb, Inférieur à 400 kb, probablement bénin, Unknown?, Variant de signification inconnue, Sans étiquette.

On vérifie également que le type d'aberration a bien été renseigné pour ces CNVs, en réalisant un tri ne conservant que ces types d'aberration : Homozygous deletion, Deletion, Duplication, Multiplication.

Un filtre permet également d'éliminer les patients uniquement porteurs de CNV hérités d'un des parents, dont la pathogénicité est plus difficile à établir. Ces CNVs sont triés en ne conservant que ceux identifiés dans ce type d'hérédité : Non spécifié, Attente parents, De novo, Etude parentale incomplète, Non confirmé, Unknown.

A ce stade, les données de 1175 patients, porteurs de 4026 CNV au total sont récupérées.

Sur cet ensemble de patients, on identifie les patients porteurs d'au moins un CNV étiqueté dans les catégories Pathogène, Pathogenic, ou Probablement pathogène : 153 patients sont concernés, porteurs de 764 CNV. Ces 764 CNV représentent à la fois les CNV pathogènes ou probablement pathogènes de ces patients, mais également leurs CNV étiquetés dans toutes les autres catégories. En faisant l'hypothèse d'une origine unique des symptômes présentés par ces patients, on peut donc considérer que les variants de signification inconnue et probablement bénins dont ils sont porteurs ne sont définitivement pas pathogènes. Ces 764 CNVs sont donc retirés des 4026 CNVs obtenus précédemment, via l'utilisation d'une feuille de calcul Excel : on obtient alors 1784 CNVs. Cette réduction du nombre de CNVs s'explique par la présence de nombreux CNVs bénins très récurrents.

Si ces 1784 CNVs sont bien portés par les patients à analyser, ils correspondent à la fois aux variants de signification inconnue et étiquetés probablement bénins, mais également aux variants définitivement classés comme bénins. Afin d'éliminer ces derniers de l'analyse, un tri est réalisé via une feuille Excel, ne conservant que les CNV étiquetés dans ces catégories :

attente parents, Benign?, en cours, Inférieur à 1M DPN, Inférieur à 200 kb, Inférieur à 400 kb, probablement bénin, Unknown?, Variant de signification inconnue, Sans étiquette.

Les CNV portés par le chromosome Y sont supprimés, les données de Hi-C actuellement disponibles ne permettant pas d'interpréter la présence de TADs sur ce chromosome. Les doublons de patients sont également retirés.

Au final, 1045 CNV dont 649 duplications et 396 délétions ont été retenus pour une analyse utilisant les données d'Hi-C disponibles.

## B. Légende de la segmentation de la chromatine ChromHMM(90) visualisée sur UCSC Genome Browser

L'algorithme ChromHMM permet d'annoter l'ensemble du génome à partir de données de ChiP-Seq, notamment issues du projet ENCODE. Les différentes régions annotées sont représentés par des bandes de couleurs différentes sur les visualisateurs de données génomiques.

Rouge vif - Promoteur actif

Rouge clair - Promoteur faible

Pourpre – Promoteur inactif

Orange - Enhancer fort

Jaune - Enhancer faible

Bleu - Insulateur

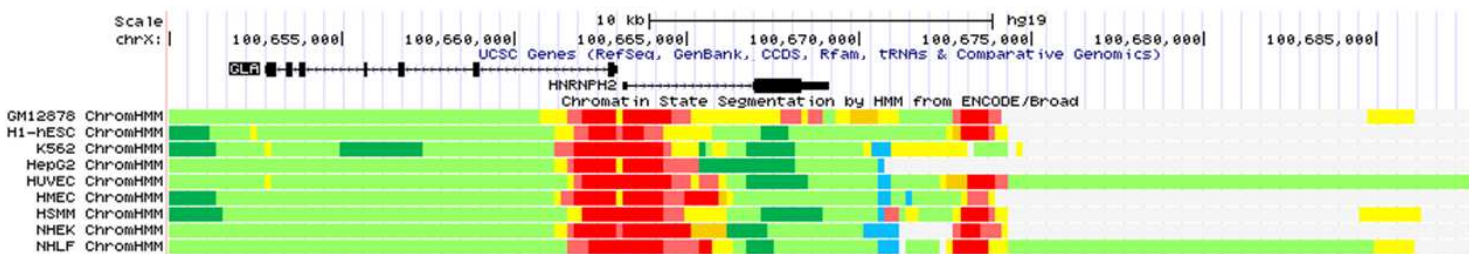
Vert foncé - Transition transcriptionnelle ou élongation

Vert clair - Transcription faible

Gris - Réprimé par Polycomb

Gris clair- Hétérochromatine ou zone répétée

Exemple : segmentation de la chromatine au locus *GLA/HNRNPH2*



C. Liste de 44 gènes impliqués dans la déficience intellectuelle.

<i>ANKRD11</i>	<i>FOXP1</i>	<i>MED13L</i>	<i>SLC2A1</i>
<i>ARID1B</i>	<i>GATAD2B</i>	<i>NAA10</i>	<i>SLC6A8</i>
<i>ATRX</i>	<i>GRIA3</i>	<i>PQBP1</i>	<i>SLC9A6</i>
<i>CASK</i>	<i>GRIN1</i>	<i>PTCHD1</i>	<i>SMARCA2</i>
<i>CDKL5</i>	<i>GRIN2A</i>	<i>RAI1</i>	<i>SMC1A</i>
<i>CTNNB1</i>	<i>GRIN2B</i>	<i>SATB2</i>	<i>STXBP1</i>
<i>CUL4B</i>	<i>IL1RAPL1</i>	<i>SCN2A</i>	<i>SYNGAP1</i>
<i>DLG3</i>	<i>IQSEC2</i>	<i>SCN8A</i>	<i>TBR1</i>
<i>DYRK1A</i>	<i>KDM5C</i>	<i>SETBP1</i>	<i>TCF4</i>
<i>EP300</i>	<i>KMT2A</i>	<i>SHANK3</i>	<i>UPF3B</i>
<i>FOXP1</i>	<i>MECP2</i>	<i>SLC16A2</i>	<i>WDR45</i>

## VI. Références

1. Flemming W. Zellsubstanz, Kern und Zellteilung. Verlag von FCW Vogel. 1882;
2. Rabl C. über Zellteilung. *Morphol Jahrb.* 1885;10:210-330.
3. Bonev B, Cavalli G. Organization and function of the 3D genome. *Nat Rev Genet.* 14 oct 2016;17(11):661-78.
4. Luger K, Mäder AW, Richmond RK, Sargent DF, Richmond TJ. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature.* 18 sept 1997;389(6648):251-60.
5. Bednar J, Horowitz RA, Grigoryev SA, Carruthers LM, Hansen JC, Koster AJ, et al. Nucleosomes, linker DNA, and linker histone form a unique structural motif that directs the higher-order folding and compaction of chromatin. *Proc Natl Acad Sci.* 24 nov 1998;95(24):14173-8.
6. Robinson PJJ, Fairall L, Huynh VAT, Rhodes D. EM measurements define the dimensions of the « 30-nm » chromatin fiber: Evidence for a compact, interdigitated structure. *Proc Natl Acad Sci.* 25 avr 2006;103(17):6506-11.
7. Schalch T, Duda S, Sargent DF, Richmond TJ. X-ray structure of a tetranucleosome and its implications for the chromatin fibre. *Nature.* 7 juill 2005;436(7047):138-41.
8. Song F, Chen P, Sun D, Wang M, Dong L, Liang D, et al. Cryo-EM Study of the Chromatin Fiber Reveals a Double Helix Twisted by Tetranucleosomal Units. *Science.* 25 avr 2014;344(6182):376-80.
9. Belmont AS, Bruce K. Visualization of G1 chromosomes: a folded, twisted, supercoiled chromonema model of interphase chromatid structure. *J Cell Biol.* oct 1994;127(2):287-302.
10. Belmont AS, Braunfeld MB, Sedat JW, Agard DA. Large-scale chromatin structural domains within mitotic and interphase chromosomes in vivo and in vitro. *Chromosoma.* août 1989;98(2):129-43.
11. McDowall AW, Smith JM, Dubochet J. Cryo-electron microscopy of vitrified chromosomes in situ. *EMBO J.* juin 1986;5(6):1395-402.
12. Nishino Y, Eltsov M, Joti Y, Ito K, Takata H, Takahashi Y, et al. Human mitotic chromosomes consist predominantly of irregularly folded nucleosome fibres without a 30-nm chromatin structure. *EMBO J.* 4 avr 2012;31(7):1644-53.
13. Ahmed K, Dehghani H, Rugg-Gunn P, Fussner E, Rossant J, Bazett-Jones DP. Global chromatin architecture reflects pluripotency and lineage commitment in the early mouse embryo. *PloS One.* 7 mai 2010;5(5):e10531.

14. Ou HD, Phan S, Deerinck TJ, Thor A, Ellisman MH, O'Shea CC. ChromEMT: Visualizing 3D chromatin structure and compaction in interphase and mitotic cells. *Science*. 28 juill 2017;357(6349):eaag0025.
15. Birman S, Rusu M, Wriggers W. Using Sculptor and Situs for simultaneous assembly of atomic components into low-resolution shapes. *J Struct Biol*. mars 2011;173(3):428-35.
16. Stack SM, Brown DB, Dewey WC. Visualization of interphase chromosomes. *J Cell Sci*. août 1977;26:281-99.
17. Zorn C, Cremer C, Cremer T, Zimmer J. Unscheduled DNA synthesis after partial UV irradiation of the cell nucleus. Distribution in interphase and metaphase. *Exp Cell Res*. nov 1979;124(1):111-9.
18. Lichter P, Cremer T, Borden J, Manuelidis L, Ward DC. Delineation of individual human chromosomes in metaphase and interphase cells by in situ suppression hybridization using recombinant DNA libraries. *Hum Genet*. nov 1988;80(3):224-34.
19. Cremer T, Cremer C. Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat Rev Genet*. avr 2001;2(4):292-301.
20. Bolzer A, Kreth G, Solovei I, Koehler D, Saracoglu K, Fauth C, et al. Three-Dimensional Maps of All Chromosomes in Human Male Fibroblast Nuclei and Prometaphase Rosettes. Misteli T, éditeur. *PLoS Biol*. 26 avr 2005;3(5):e157.
21. Guelen L, Pagie L, Brassat E, Meuleman W, Faza MB, Talhout W, et al. Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature*. 12 juin 2008;453(7197):948-51.
22. van Steensel B, Belmont AS. Lamina-Associated Domains: Links with Chromosome Architecture, Heterochromatin, and Gene Repression. *Cell*. mai 2017;169(5):780-91.
23. Németh A, Conesa A, Santoyo-Lopez J, Medina I, Montaner D, Péterfia B, et al. Initial Genomics of the Human Nucleolus. Akhtar A, éditeur. *PLoS Genet*. 26 mars 2010;6(3):e1000889.
24. van Koningsbruggen S, Gierlinski M, Schofield P, Martin D, Barton GJ, Ariyurek Y, et al. High-Resolution Whole-Genome Sequencing Reveals That Specific Chromatin Domains from Most Human Chromosomes Associate with Nucleoli. *Mol Biol Cell*. 1 nov 2010;21(21):3735-48.
25. Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragozy T, Telling A, et al. Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. *Science*. 9 oct 2009;326(5950):289-93.
26. Rao SSP, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, et al. A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. *Cell*. déc 2014;159(7):1665-80.

27. Münkler C, Langowski J. Chromosome structure predicted by a polymer model. *Phys Rev E*. 1 mai 1998;57(5):5888-96.
28. Grosberg AY, Nechaev SK, Shakhnovich EI. The role of topological constraints in the kinetics of collapse of macromolecules. *J Phys*. 1988;49(12):2095-100.
29. Mirny LA. The fractal globule as a model of chromatin architecture in the cell. *Chromosome Res*. janv 2011;19(1):37-51.
30. Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*. 11 avr 2012;485(7398):376-80.
31. Dixon JR, Gorkin DU, Ren B. Chromatin Domains: The Unit of Chromosome Organization. *Mol Cell*. juin 2016;62(5):668-80.
32. Nagano T, Lubling Y, Stevens TJ, Schoenfelder S, Yaffe E, Dean W, et al. Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature*. 25 sept 2013;502(7469):59-64.
33. Stevens TJ, Lando D, Basu S, Atkinson LP, Cao Y, Lee SF, et al. 3D structures of individual mammalian genomes studied by single-cell Hi-C. *Nature*. 13 mars 2017;544(7648):59-64.
34. Ulianov SV, Tachibana-Konwalski K, Razin SV. Single-cell Hi-C bridges microscopy and genome-wide sequencing approaches to study 3D chromatin organization. *BioEssays*. 9 août 2017;1700104.
35. Nagano T, Lubling Y, Várnai C, Dudley C, Leung W, Baran Y, et al. Cell-cycle dynamics of chromosomal organization at single-cell resolution. *Nature*. 5 juill 2017;547(7661):61-7.
36. Chiariello AM, Annunziatella C, Bianco S, Esposito A, Nicodemi M. Polymer physics of chromosome large-scale 3D organisation. *Sci Rep [Internet]*. sept 2016 [cité 17 août 2017];6(1). Disponible sur: <http://www.nature.com/articles/srep29775>
37. Splinter E. CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus. *Genes Dev*. 1 sept 2006;20(17):2349-54.
38. Ong C-T, Corces VG. CTCF: an architectural protein bridging genome topology and function. *Nat Rev Genet*. 11 mars 2014;15(4):234-46.
39. Remeseiro S, Losada A. Cohesin, a chromatin engagement ring. *Curr Opin Cell Biol*. févr 2013;25(1):63-71.
40. Sanborn AL, Rao SSP, Huang S-C, Durand NC, Huntley MH, Jewett AI, et al. Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc Natl Acad Sci*. 24 nov 2015;112(47):E6456-65.



41. Fudenberg G, Imakaev M, Lu C, Goloborodko A, Abdennur N, Mirny LA. Formation of Chromosomal Domains by Loop Extrusion. *Cell Rep.* mai 2016;15(9):2038-49.
42. Dolgin E. DNA's secret weapon against knots and tangles. *Nature.* 19 avr 2017;544(7650):284-6.
43. Nora EP, Lajoie BR, Schulz EG, Giorgetti L, Okamoto I, Servant N, et al. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature.* 11 avr 2012;485(7398):381-5.
44. Jiang Y, Loh Y-HE, Rajarajan P, Hirayama T, Liao W, Kassim BS, et al. The methyltransferase SETDB1 regulates a large neuron-specific topological chromatin domain. *Nat Genet* [Internet]. 3 juill 2017 [cité 9 juill 2017]; Disponible sur: <http://www.nature.com/doifinder/10.1038/ng.3906>
45. Andrey G, Montavon T, Mascrez B, Gonzalez F, Noordermeer D, Leleu M, et al. A Switch Between Topological Domains Underlies HoxD Genes Collinearity in Mouse Limbs. *Science.* 7 juin 2013;340(6137):1234167-1234167.
46. Lonfat N, Duboule D. Structure, function and evolution of topologically associating domains (TADs) at *HOX* loci. *FEBS Lett.* 7 oct 2015;589(20PartA):2869-76.
47. Woodfine K, Fiegler H, Beare DM, Collins JE, McCann OT, Young BD, et al. Replication timing of the human genome. *Hum Mol Genet.* 15 janv 2004;13(2):191-202.
48. Taylor JH. The mode of chromosome duplication in *Crepis capillaris*. *Exp Cell Res.* oct 1958;15(2):350-7.
49. Pope BD, Ryba T, Dileep V, Yue F, Wu W, Denas O, et al. Topologically associating domains are stable units of replication-timing regulation. *Nature.* 19 nov 2014;515(7527):402-5.
50. Cullen K, Kladde M, Seyfred M. Interaction between transcription regulatory regions of prolactin chromatin. *Science.* 9 juill 1993;261(5118):203-6.
51. Dekker J. Capturing Chromosome Conformation. *Science.* 15 févr 2002;295(5558):1306-11.
52. Hagège H, Klous P, Braem C, Splinter E, Dekker J, Cathala G, et al. Quantitative analysis of chromosome conformation capture assays (3C-qPCR). *Nat Protoc.* juill 2007;2(7):1722-33.
53. Stadhouders R, Kolovos P, Brouwer R, Zuin J, van den Heuvel A, Kockx C, et al. Multiplexed chromosome conformation capture sequencing for rapid genome-scale high-resolution detection of long-range chromatin interactions. *Nat Protoc.* 14 févr 2013;8(3):509-24.
54. Simonis M, Kooren J, de Laat W. An evaluation of 3C-based methods to capture DNA interactions. *Nat Methods.* nov 2007;4(11):895-901.

55. Wright JB, Brown SJ, Cole MD. Upregulation of c-MYC in cis through a Large Chromatin Loop Linked to a Cancer Risk-Associated Single-Nucleotide Polymorphism in Colorectal Cancer Cells. *Mol Cell Biol.* 15 mars 2010;30(6):1411-20.
56. Amano T, Sagai T, Tanabe H, Mizushina Y, Nakazawa H, Shiroishi T. Chromosomal Dynamics at the Shh Locus: Limb Bud-Specific Differential Regulation of Competence and Active Transcription. *Dev Cell.* janv 2009;16(1):47-57.
57. Gheldof N, Smith EM, Tabuchi TM, Koch CM, Dunham I, Stamatoyannopoulos JA, et al. Cell-type-specific long-range looping interactions identify distant regulatory elements of the CFTR gene. *Nucleic Acids Res.* juill 2010;38(13):4325-36.
58. Dekker J, Marti-Renom MA, Mirny LA. Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data. *Nat Rev Genet.* 9 mai 2013;14(6):390-403.
59. Krijger PHL, de Laat W. Regulation of disease-associated gene expression in the 3D genome. *Nat Rev Mol Cell Biol.* 9 nov 2016;17(12):771-82.
60. Simonis M, Klous P, Splinter E, Moshkin Y, Willemsen R, de Wit E, et al. Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nat Genet.* nov 2006;38(11):1348-54.
61. Giresi PG, Kim J, McDaniell RM, Iyer VR, Lieb JD. FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. *Genome Res.* 1 juin 2007;17(6):877-85.
62. Imakaev M, Fudenberg G, McCord RP, Naumova N, Goloborodko A, Lajoie BR, et al. Iterative correction of Hi-C data reveals hallmarks of chromosome organization. *Nat Methods.* 2 sept 2012;9(10):999-1003.
63. Dostie J, Richmond TA, Arnaout RA, Selzer RR, Lee WL, Honan TA, et al. Chromosome Conformation Capture Carbon Copy (5C): A massively parallel solution for mapping interactions between genomic elements. *Genome Res.* 1 oct 2006;16(10):1299-309.
64. Sanyal A, Lajoie BR, Jain G, Dekker J. The long-range interaction landscape of gene promoters. *Nature.* 5 sept 2012;489(7414):109-13.
65. Sati S, Cavalli G. Chromosome conformation capture technologies and their impact in understanding genome function. *Chromosoma.* févr 2017;126(1):33-44.
66. Lajoie BR, Dekker J, Kaplan N. The Hitchhiker's guide to Hi-C analysis: Practical guidelines. *Methods.* janv 2015;72:65-75.
67. Dali R, Blanchette M. A critical assessment of topologically associating domain prediction tools. *Nucleic Acids Res.* 7 avr 2017;45(6):2994-3005.

68. Forcato M, Nicoletti C, Pal K, Livi CM, Ferrari F, Bicciato S. Comparison of computational methods for Hi-C data analysis. *Nat Methods*. 12 juin 2017;14(7):679-85.
69. Fullwood MJ, Liu MH, Pan YF, Liu J, Xu H, Mohamed YB, et al. An oestrogen-receptor- $\alpha$ -bound human chromatin interactome. *Nature*. 5 nov 2009;462(7269):58-64.
70. Lupiáñez DG, Spielmann M, Mundlos S. Breaking TADs: How Alterations of Chromatin Domains Result in Disease. *Trends Genet*. avr 2016;32(4):225-37.
71. Giorgio E, Robyr D, Spielmann M, Ferrero E, Di Gregorio E, Imperiale D, et al. A large genomic deletion leads to enhancer adoption by the lamin B1 gene: a second path to autosomal dominant adult-onset demyelinating leukodystrophy (ADLD). *Hum Mol Genet*. 1 juin 2015;24(11):3143-54.
72. Gordon CT, Attanasio C, Bhatia S, Benko S, Ansari M, Tan TY, et al. Identification of Novel Craniofacial Regulatory Domains Located far Upstream of *SOX9* and Disrupted in Pierre Robin Sequence. *Hum Mutat*. août 2014;35(8):1011-20.
73. Benko S, Gordon CT, Mallet D, Sreenivasan R, Thauvin-Robinet C, Brendehaug A, et al. Disruption of a long distance regulatory region upstream of *SOX9* in isolated disorders of sex development. *J Med Genet*. 1 déc 2011;48(12):825-30.
74. Kurth I, Klopocki E, Stricker S, van Oosterwijk J, Vanek S, Altmann J, et al. Duplications of noncoding elements 5' of *SOX9* are associated with brachydactyly-anonychia. *Nat Genet*. août 2009;41(8):862-3.
75. Franke M, Ibrahim DM, Andrey G, Schwarzer W, Heinrich V, Schöpflin R, et al. Formation of new chromatin domains determines pathogenicity of genomic duplications. *Nature*. 5 oct 2016;538(7624):265-9.
76. Redin C, Brand H, Collins RL, Kammin T, Mitchell E, Hodge JC, et al. The genomic landscape of balanced cytogenetic abnormalities associated with human congenital anomalies. *Nat Genet*. 14 nov 2016;49(1):36-45.
77. Hu J, Verzi MP, Robinson AS, Tang PL-F, Hua LL, Xu S-M, et al. Endothelin signaling activates *Mef2c* expression in the neural crest through a *MEF2C*-dependent positive-feedback transcriptional pathway. *Development*. 15 août 2015;142(16):2775-80.
78. Zweier M, Gregor A, Zweier C, Engels H, Sticht H, Wohlleber E, et al. Mutations in *MEF2C* from the 5q14.3q15 microdeletion syndrome region are a frequent cause of severe mental retardation and diminish *MECP2* and *CDKL5* expression. *Hum Mutat*. 6 avr 2010;31(6):722-33.
79. Gordon CT, Tessier A, Demir Z, Goldenberg A, Oufadem M, Voisin N, et al. The association of severe encephalopathy and question mark ear is highly suggestive of loss of *MEF2C* function. *Clin Genet* [Internet]. mai 2017 [cité 11 juill 2017]; Disponible sur: <http://doi.wiley.com/10.1111/cge.13046>

80. Le Meur N, Holder-Espinasse M, Jaillard S, Goldenberg A, Joriot S, Amati-Bonneau P, et al. MEF2C haploinsufficiency caused by either microdeletion of the 5q14.3 region or mutation is responsible for severe mental retardation with stereotypic movements, epilepsy and/or cerebral malformations. *J Med Genet.* 1 janv 2010;47(1):22-9.
81. Vergult S, Van Binsbergen E, Sante T, Nowak S, Vanakker O, Claes K, et al. Mate pair sequencing for the detection of chromosomal aberrations in patients with intellectual disability and congenital malformations. *Eur J Hum Genet.* mai 2014;22(5):652-9.
82. Floris C, Rassa S, Boccone L, Gasperini D, Cao A, Crisponi L. Two patients with balanced translocations and autistic disorder: CSMD3 as a candidate gene for autism found in their common 8q23 breakpoint area. *Eur J Hum Genet.* juin 2008;16(6):696-704.
83. Saitsu H, Igarashi N, Kato M, Okada I, Kosho T, Shimokawa O, et al. De novo 5q14.3 translocation 121.5-kb upstream of MEF2C in a patient with severe intellectual disability and early-onset epileptic encephalopathy. *Am J Med Genet A.* nov 2011;155(11):2879-84.
84. Döcker D, Schubach M, Menzel M, Munz M, Spaich C, Biskup S, et al. Further delineation of the SATB2 phenotype. *Eur J Hum Genet.* août 2014;22(8):1034-9.
85. Brunetti-Pierrri N, Paciorkowski AR, Ciccone R, Mina ED, Bonaglia MC, Borgatti R, et al. Duplications of FOXP1 in 14q12 are associated with developmental epilepsy, mental retardation, and severe speech impairment. *Eur J Hum Genet.* janv 2011;19(1):102-7.
86. Rauch A, Wieczorek D, Graf E, Wieland T, Ende S, Schwarzmayr T, et al. Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: an exome sequencing study. *The Lancet.* nov 2012;380(9854):1674-82.
87. Zepeda-Mendoza CJ, Ibn-Salem J, Kammin T, Harris DJ, Rita D, Gripp KW, et al. Computational Prediction of Position Effects of Apparently Balanced Human Chromosomal Rearrangements. *Am J Hum Genet* [Internet]. juill 2017 [cité 26 juill 2017]; Disponible sur: <http://linkinghub.elsevier.com/retrieve/pii/S000292971730246X>
88. Higgins AW, Alkuraya FS, Bosco AF, Brown KK, Bruns GAP, Donovan DJ, et al. Characterization of Apparently Balanced Chromosomal Rearrangements from the Developmental Genome Anatomy Project. *Am J Hum Genet.* mars 2008;82(3):712-22.
89. Visel A, Minovitsky S, Dubchak I, Pennacchio LA. VISTA Enhancer Browser--a database of tissue-specific human enhancers. *Nucleic Acids Res.* 3 janv 2007;35(Database):D88-92.
90. Ernst J, Kellis M. ChromHMM: automating chromatin-state discovery and characterization. *Nat Methods.* 28 févr 2012;9(3):215-6.
91. Köhler S, Doelken SC, Mungall CJ, Bauer S, Firth HV, Bailleul-Forestier I, et al. The Human Phenotype Ontology project: linking molecular biology and disease through phenotype data. *Nucleic Acids Res.* janv 2014;42(D1):D966-74.

92. Ibn-Salem J, Köhler S, Love MI, Chung H-R, Huang N, Hurles ME, et al. Deletions of chromosomal regulatory boundaries are associated with congenital disease. *Genome Biol* [Internet]. sept 2014 [cité 1 août 2017];15(9). Disponible sur: <http://genomebiology.biomedcentral.com/articles/10.1186/s13059-014-0423-1>
93. Ceol CJ, Houvras Y, Jane-Valbuena J, Bilodeau S, Orlando DA, Battisti V, et al. The histone methyltransferase SETDB1 is recurrently amplified in melanoma and accelerates its onset. *Nature*. 24 mars 2011;471(7339):513-7.
94. Cukier HN, Lee JM, Ma D, Young JI, Mayo V, Butler BL, et al. The Expanding Role of MBD Genes in Autism: Identification of a *MECP2* Duplication and Novel Alterations in *MBD5*, *MBD6*, and *SETDB1*: ASD patients with novel variants in MBD genes. *Autism Res*. déc 2012;5(6):385-97.
95. Xu Q, Goldstein J, Wang P, Gadi IK, Labreche H, Rehder C, et al. Chromosomal microarray analysis in clinical evaluation of neurodevelopmental disorders-reporting a novel deletion of SETDB1 and illustration of counseling challenge. *Pediatr Res*. sept 2016;80(3):371-81.
96. Chen WV, Maniatis T. Clustered protocadherins. *Development*. 15 août 2013;140(16):3297-302.
97. McGowan PO, Suderman M, Sasaki A, Huang TCT, Hallett M, Meaney MJ, et al. Broad Epigenetic Signature of Maternal Care in the Brain of Adult Rats. Sirigu A, éditeur. *PLoS ONE*. 28 févr 2011;6(2):e14739.
98. Yokota S, Hirayama T, Hirano K, Kaneko R, Toyoda S, Kawamura Y, et al. Identification of the Cluster Control Region for the Protocadherin- $\beta$  Genes Located beyond the Protocadherin- $\gamma$  Cluster. *J Biol Chem*. 9 sept 2011;286(36):31885-95.
99. MacDonald JR, Ziman R, Yuen RKC, Feuk L, Scherer SW. The Database of Genomic Variants: a curated collection of structural variation in the human genome. *Nucleic Acids Res*. janv 2014;42(Database issue):D986-992.
100. Kaminsky EB, Kaul V, Paschall J, Church DM, Bunke B, Kunig D, et al. An evidence-based approach to establish the functional and clinical significance of copy number variants in intellectual and developmental disabilities: *Genet Med*. sept 2011;13(9):777-84.
101. Firth HV, Richards SM, Bevan AP, Clayton S, Corpas M, Rajan D, et al. DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. *Am J Hum Genet*. avr 2009;84(4):524-33.
102. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, et al. The Human Genome Browser at UCSC. *Genome Res*. 16 mai 2002;12(6):996-1006.
103. Yates A, Akanni W, Amode MR, Barrell D, Billis K, Carvalho-Silva D, et al. Ensembl 2016. *Nucleic Acids Res*. 4 janv 2016;44(D1):D710-716.

104. Online Mendelian Inheritance in Man, OMIM®. McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, MD), 2017. World Wide Web URL: <https://omim.org/>.
105. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 15 mars 2010;26(6):841-2.
106. Afgan E, Baker D, van den Beek M, Blankenberg D, Bouvier D, Čech M, et al. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic Acids Res*. 8 juill 2016;44(W1):W3-10.
107. Wang Y, Zhang B, Zhang L, An L, Xu J, Li D, et al. The 3D Genome Browser: a web-based browser for visualizing 3D genome organization and long-range chromatin interactions. 27 févr 2017 [cité 7 août 2017]; Disponible sur: <http://biorxiv.org/lookup/doi/10.1101/112268>
108. Lupiáñez DG, Kraft K, Heinrich V, Krawitz P, Brancati F, Klopocki E, et al. Disruptions of Topological Chromatin Domains Cause Pathogenic Rewiring of Gene-Enhancer Interactions. *Cell*. mai 2015;161(5):1012-25.
109. Flöttmann R, Wagner J, Kobus K, Curry CJ, Savarirayan R, Nishimura G, et al. Microdeletions on 6p22.3 are associated with mesomelic dysplasia Savarirayan type. *J Med Genet*. juill 2015;52(7):476-83.
110. Di Gregorio E, Riberi E, Belligni EF, Biamino E, Spielmann M, Ala U, et al. Copy number variants analysis in a cohort of isolated and syndromic developmental delay/intellectual disability reveals novel genomic disorders, position effects and candidate disease genes. *Clin Genet* [Internet]. 25 juill 2017 [cité 17 août 2017]; Disponible sur: <http://doi.wiley.com/10.1111/cge.13009>
111. Andersson R, Gebhard C, Miguel-Escalada I, Hoof I, Bornholdt J, Boyd M, et al. An atlas of active enhancers across human cell types and tissues. *Nature*. 26 mars 2014;507(7493):455-61.
112. Wright KL, Ting JP-Y. Epigenetic regulation of MHC-II and CIITA genes. *Trends Immunol*. sept 2006;27(9):405-12.
113. Pattenden SG. Interferon-gamma-induced chromatin remodeling at the CIITA locus is BRG1 dependent. *EMBO J*. 15 avr 2002;21(8):1978-86.
114. Ni Z, Abou El Hassan M, Xu Z, Yu T, Bremner R. The chromatin-remodeling enzyme BRG1 coordinates CIITA induction through many interdependent distal enhancers. *Nat Immunol*. juill 2008;9(7):785-93.
115. Abou El Hassan M, Bremner R. A rapid simple approach to quantify chromosome conformation capture. *Nucleic Acids Res*. 9 janv 2009;37(5):e35-e35.
116. Briscoe J, Théron PP. The mechanisms of Hedgehog signalling and its roles in development and disease. *Nat Rev Mol Cell Biol*. 30 mai 2013;14(7):418-31.

117. Gao B, Guo J, She C, Shu A, Yang M, Tan Z, et al. Mutations in *IHH*, encoding Indian hedgehog, cause brachydactyly type A-1. *Nat Genet.* août 2001;28(4):386-8.
118. Hellemans J, Coucke PJ, Giedion A, De Paepe A, Kramer P, Beemer F, et al. Homozygous mutations in *IHH* cause acrocapitofemoral dysplasia, an autosomal recessive disorder with cone-shaped epiphyses in hands and hips. *Am J Hum Genet.* avr 2003;72(4):1040-6.
119. Klopocki E, Lohan S, Brancati F, Koll R, Brehm A, Seemann P, et al. Copy-Number Variations Involving the *IHH* Locus Are Associated with Syndactyly and Craniosynostosis. *Am J Hum Genet.* janv 2011;88(1):70-5.
120. Barroso E, Berges-Soria J, Benito-Sanz S, Rivera-Pedroza CI, Ballesta-Martínez MJ, López-González V, et al. Identification of the fourth duplication of upstream *IHH* regulatory elements, in a family with craniosynostosis Philadelphia type, helps to define the phenotypic characterization of these regulatory elements. *Am J Med Genet A.* avr 2015;167(4):902-6.
121. Umehara F, Tate G, Itoh K, Yamaguchi N, Douchi T, Mitsuya T, et al. A Novel Mutation of desert hedgehog in a Patient with 46,XY Partial Gonadal Dysgenesis Accompanied by Minifascicular Neuropathy. *Am J Hum Genet.* nov 2000;67(5):1302-5.
122. Belloni E, Muenke M, Roessler E, Traverse G, Siegel-Bartelt J, Frumkin A, et al. Identification of Sonic hedgehog as a candidate gene responsible for holoprosencephaly. *Nat Genet.* nov 1996;14(3):353-6.
123. Kang S, Graham JM, Olney AH, Biesecker LG. *GLI3* frameshift mutations cause autosomal dominant Pallister-Hall syndrome. *Nat Genet.* mars 1997;15(3):266-8.
124. Pettigrew AL, Greenberg F, Caskey CT, Ledbetter DH. Greig syndrome associated with an interstitial deletion of 7p: confirmation of the localization of Greig syndrome to 7p13. *Hum Genet.* août 1991;87(4):452-6.
125. Al-Qattan M. A novel frameshift mutation of the *GLI3* gene in a family with broad thumbs with/without big toes, postaxial polydactyly and variable syndactyly of the hands/feet. *Clin Genet.* nov 2012;82(5):502-4.
126. Schmidt C, McGonnell IM, Allen S, Otto A, Patel K. *Wnt6* controls amniote neural crest induction through the non-canonical signaling pathway. *Dev Dyn.* sept 2007;236(9):2502-11.
127. Geetha-Loganathan P, Nimmagadda S, Christ B, Huang R, Scaal M. Ectodermal *Wnt6* is an early negative regulator of limb chondrogenesis in the chicken embryo. *BMC Dev Biol.* 2010;10(1):32.
128. Fabes J, Anderson P, Brennan C, Bolsover S. Regeneration-enhancing effects of EphA4 blocking peptide following corticospinal tract injury in adult rat spinal cord: EphA4 blocking peptide evokes corticospinal tract recovery. *Eur J Neurosci.* 26 oct 2007;26(9):2496-505.

129. Helmbacher F, Schneider-Maunoury S, Topilko P, Tiret L, Charnay P. Targeting of the EphA4 tyrosine kinase receptor affects dorsal/ventral pathfinding of limb motor axons. *Dev Camb Engl.* août 2000;127(15):3313-24.
130. Van Hoecke A, Schoonaert L, Lemmens R, Timmers M, Staats KA, Laird AS, et al. EPHA4 is a disease modifier of amyotrophic lateral sclerosis in animal models and in humans. *Nat Med.* 26 août 2012;18(9):1418-22.
131. Monsoro-Burq AH. PAX transcription factors in neural crest development. *Semin Cell Dev Biol.* août 2015;44:87-96.
132. Tassabehji M, Read AP, Newton VE, Harris R, Balling R, Gruss P, et al. Waardenburg's syndrome patients have mutations in the human homologue of the Pax-3 paired box gene. *Nature.* 13 févr 1992;355(6361):635-6.
133. Hoth CF, Milunsky A, Lipsky N, Sheffer R, Clarren SK, Baldwin CT. Mutations in the paired domain of the human PAX3 gene cause Klein-Waardenburg syndrome (WS-III) as well as Waardenburg syndrome type I (WS-I). *Am J Hum Genet.* mars 1993;52(3):455-62.
134. Dundar M, Gordon TM, Ozyazgan I, Oguzkaya F, Ozkul Y, Cooke A, et al. A novel acropectoral syndrome maps to chromosome 7q36. *J Med Genet.* mai 2001;38(5):304-9.
135. Yuksel-Apak M, Bögershausen N, Pawlik B, Li Y, Apak S, Uyguner O, et al. A large duplication involving the IHH locus mimics acrocallosal syndrome. *Eur J Hum Genet.* juin 2012;20(6):639-44.
136. Russell MR, Penikis A, Oldridge DA, Alvarez-Dominguez JR, McDaniel L, Diamond M, et al. CASC15-S Is a Tumor Suppressor lncRNA at the 6p22 Neuroblastoma Susceptibility Locus. *Cancer Res.* 1 août 2015;75(15):3155-66.
137. Lessard L, Liu M, Marzese DM, Wang H, Chong K, Kawas N, et al. The CASC15 Long Intergenic Noncoding RNA Locus Is Involved in Melanoma Progression and Phenotype Switching. *J Invest Dermatol.* oct 2015;135(10):2464-74.
138. Dy P, Penzo-Méndez A, Wang H, Pedraza CE, Macklin WB, Lefebvre V. The three SoxC proteins--Sox4, Sox11 and Sox12--exhibit overlapping expression patterns and molecular properties. *Nucleic Acids Res.* mai 2008;36(9):3101-17.
139. Hempel A, Pagnamenta AT, Blyth M, Mansour S, McConnell V, Kou I, et al. Deletions and de novo mutations of *SOX11* are associated with a neurodevelopmental disorder with features of Coffin-Siris syndrome. *J Med Genet.* mars 2016;53(3):152-62.
140. Penzo-Méndez AI. Critical roles for SoxC transcription factors in development and cancer. *Int J Biochem Cell Biol.* mars 2010;42(3):425-8.
141. Cheung M, Abu-Elmagd M, Clevers H, Scotting PJ. Roles of Sox4 in central nervous system development. *Brain Res Mol Brain Res.* 23 juin 2000;79(1-2):180-91.



142. Shim S, Kwan KY, Li M, Lefebvre V, Sestan N. Cis-regulatory control of corticospinal system development and evolution. *Nature*. 7 juin 2012;486(7401):74-9.
143. Thein DC, Thalhammer JM, Hartwig AC, Crenshaw EB, Lefebvre V, Wegner M, et al. The closely related transcription factors Sox4 and Sox11 function as survival factors during spinal cord development. *J Neurochem*. oct 2010;115(1):131-41.
144. Huang J, Arsenault M, Kann M, Lopez-Mendez C, Saleh M, Wadowska D, et al. The transcription factor Sry-related HMG box-4 (SOX4) is required for normal renal development in vivo. *Dev Dyn Off Publ Am Assoc Anat*. juin 2013;242(6):790-9.
145. Pelletier J, Bruening W, Li FP, Haber DA, Glaser T, Housman DE. WT1 mutations contribute to abnormal genital system development and hereditary Wilms' tumour. *Nature*. oct 1991;353(6343):431-4.
146. Schumacher V, Schäfer K, Wühl E, Altrogge H, Bonzel K-E, Guschmann M, et al. Spectrum of early onset nephrotic syndrome associated with WT1 missense mutations. *Kidney Int*. juin 1998;53(6):1594-600.
147. Barbaux S, Niaudet P, Gubler M-C, Grünfeld J-P, Jaubert F, Kuttent F, et al. Donor splice-site mutations in WT1 are responsible for Frasier syndrome. *Nat Genet*. déc 1997;17(4):467-70.
148. Cardoso L, De Souza K, De O. Reis A, Andrade R, Britto A, De Lima M, et al. WT1, WTX and CTNNB1 mutation analysis in 43 patients with sporadic Wilms' tumor. *Oncol Rep [Internet]*. 19 oct 2012 [cité 20 juill 2017]; Disponible sur: <http://www.spandidos-publications.com/10.3892/or.2012.2096>
149. Kreidberg JA, Sariola H, Loring JM, Maeda M, Pelletier J, Housman D, et al. WT-1 is required for early kidney development. *Cell*. 27 août 1993;74(4):679-91.
150. Kato K, Bhattaram P, Penzo-Méndez A, Gadi A, Lefebvre V. SOXC Transcription Factors Induce Cartilage Growth Plate Formation in Mouse Embryos by Promoting Noncanonical WNT Signaling. *J Bone Miner Res Off J Am Soc Bone Miner Res*. sept 2015;30(9):1560-71.
151. Seelan RS, Mukhopadhyay P, Warner DR, Webb CL, Pisano M, Greene RM. Epigenetic regulation of Sox4 during palate development. *Epigenomics*. avr 2013;5(2):131-46.
152. Penzo-Méndez A, Dy P, Pallavi B, Lefebvre V. Generation of mice harboring a Sox4 conditional null allele. *Genes N Y N* 2000. déc 2007;45(12):776-80.
153. Wang C, Zhao H, Lu J, Yin J, Zang L, Song N, et al. Clinicopathological significance of SOX4 expression in primary gallbladder carcinoma. *Diagn Pathol*. 2012;7:41.
154. Wang W, Zhang J, Zhan X, Lin T, Yang M, Hu J, et al. SOX4 is associated with poor prognosis in cholangiocarcinoma. *Biochem Biophys Res Commun*. 26 sept 2014;452(3):614-21.

155. Wang L, Zhang J, Yang X, Chang YWY, Qi M, Zhou Z, et al. SOX4 is associated with poor prognosis in prostate cancer and promotes epithelial-mesenchymal transition in vitro. *Prostate Cancer Prostatic Dis.* déc 2013;16(4):301-7.
156. Wang D, Hao T, Pan Y, Qian X, Zhou D. Increased expression of SOX4 is a biomarker for malignant status and poor prognosis in patients with non-small cell lung cancer. *Mol Cell Biochem.* avr 2015;402(1-2):75-82.
157. Pramoonjago P, Baras AS, Moskaluk CA. Knockdown of Sox4 expression by RNAi induces apoptosis in ACC3 cells. *Oncogene.* 14 sept 2006;25(41):5626-39.
158. Chen J, Ju HL, Yuan XY, Wang TJ, Lai BQ. SOX4 is a potential prognostic factor in human cancers: a systematic review and meta-analysis. *Clin Transl Oncol.* janv 2016;18(1):65-72.
159. Savarirayan R, Cormier-Daire V, Curry CJ, Nashelsky MB, Rappaport V, Rimoin DL, et al. New mesomelic dysplasia with absent fibulae and triangular tibiae. *Am J Med Genet.* 4 sept 2000;94(1):59-63.
160. Ladinsky HT, Elizalde A, Schickler R, Dees PB, Crenshaw ML, Sleasman JW. Hypereosinophilic syndrome and hemimelia in a patient with chromosome 6p22.3 deletion. *Pediatr Allergy Immunol.* août 2014;25(5):500-3.
161. Cotney J, Leng J, Yin J, Reilly SK, DeMare LE, Emera D, et al. The Evolution of Lineage-Specific Regulatory Activities in the Human Embryonic Limb. *Cell.* juill 2013;154(1):185-96.
162. Tokuzawa Y, Yagi K, Yamashita Y, Nakachi Y, Nikaido I, Bono H, et al. Id4, a New Candidate Gene for Senile Osteoporosis, Acts as a Molecular Switch Promoting Osteoblast Differentiation. *Barsh GS, éditeur. PLoS Genet.* 8 juill 2010;6(7):e1001019.
163. Dauwerse JG, de Vries BBA, Wouters CH, Bakker E, Rappold G, Mortier GR, et al. A t(4;6)(q12;p23) translocation disrupts a membrane-associated O-acetyl transferase gene (MBOAT1) in a patient with a novel brachydactyly-syndactyly syndrome. *Eur J Hum Genet EJHG.* juill 2007;15(7):743-51.
164. Arragain S, Handelman SK, Forouhar F, Wei F-Y, Tomizawa K, Hunt JF, et al. Identification of eukaryotic and prokaryotic methylthiotransferase for biosynthesis of 2-methylthio-N<sup>6</sup>-threonylcarbamoyladenine in tRNA. *J Biol Chem.* 10 sept 2010;285(37):28425-33.
165. Cooper CS, Nicholson AG, Foster C, Dodson A, Edwards S, Fletcher A, et al. Nuclear overexpression of the E2F3 transcription factor in human lung cancer. *Lung Cancer.* nov 2006;54(2):155-62.
166. Olsson AY, Feber A, Edwards S, Te Poele R, Giddings I, Merson S, et al. Role of E2F3 expression in modulating cellular proliferation rate in human bladder and prostate cancer cells. *Oncogene.* 15 févr 2007;26(7):1028-37.

167. Tordai A, Wang J, Andre F, Liedtke C, Yan K, Sotiriou C, et al. Evaluation of biological pathways involved in chemotherapy response in breast cancer. *Breast Cancer Res.* 2008;10(2):R37.
168. An Q, Wang Y, An R, Li Y, Yao T, Zhai B, et al. Association of E2F3 expression with clinicopathological features of Wilms' tumors. *J Pediatr Surg.* nov 2013;48(11):2187-93.
169. Ernst J, Kheradpour P, Mikkelsen TS, Shores N, Ward LD, Epstein CB, et al. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature.* 5 mai 2011;473(7345):43-9.
170. Trimouille A, Barouk-Simonet E, Charron S, Bouron J, Bernhard J-C, Lacombe D, et al. Deletion of the transcription factor SOX4 is implicated in syndromic neuroblastoma. *Clin Genet.* 16 mars 2017;
171. Pomper JK, Wilhelm H, Tayebati SK, Asmus F, Schüle R, Sievert K-D, et al. A novel clinical syndrome revealing a deficiency of the muscarinic M3 receptor. *Neurology.* 1 févr 2011;76(5):451-5.
172. Weber S, Thiele H, Mir S, Toliat MR, Sozeri B, Reutter H, et al. Muscarinic Acetylcholine Receptor M3 Mutation Causes Urinary Bladder Disease and a Prune-Belly-like Syndrome. *Am J Hum Genet.* 11 nov 2011;89(5):668-74.
173. Matsui M, Motomura D, Karasawa H, Fujikawa T, Jiang J, Komiya Y, et al. Multiple functional defects in peripheral autonomic organs in mice lacking muscarinic acetylcholine receptor gene for the M3 subtype. *Proc Natl Acad Sci U S A.* 15 août 2000;97(17):9579-84.
174. Sch?nichen A, Geyer M. Fifteen formins for an actin filament: A molecular view on the regulation of human formins. *Biochim Biophys Acta BBA - Mol Cell Res.* févr 2010;1803(2):152-63.
175. Leader B, Lim H, Carabatsos MJ, Harrington A, Ecsedy J, Pellman D, et al. Formin-2, polyploidy, hypofertility and positioning of the meiotic spindle in mouse oocytes. *Nat Cell Biol.* déc 2002;4(12):921-8.
176. Leader B, Leder P. Formin-2, a novel formin homology protein of the cappuccino subfamily, is highly expressed in the developing and adult central nervous system. *Mech Dev.* mai 2000;93(1-2):221-31.
177. Peleg S, Sananbenesi F, Zovoilis A, Burkhardt S, Bahari-Javan S, Agis-Balboa RC, et al. Altered Histone Acetylation Is Associated with Age-Dependent Memory Impairment in Mice. *Science.* 7 mai 2010;328(5979):753-6.
178. Law R, Dixon-Salazar T, Jerber J, Cai N, Abbasi AA, Zaki MS, et al. Biallelic truncating mutations in FMN2, encoding the actin-regulatory protein Formin 2, cause nonsyndromic autosomal-recessive intellectual disability. *Am J Hum Genet.* 4 déc 2014;95(6):721-8.

179. Perrone MD, Rocca MS, Bruno I, Faletra F, Pecile V, Gasparini P. De novo 911 Kb interstitial deletion on chromosome 1q43 in a boy with mental retardation and short stature. *Eur J Med Genet.* févr 2012;55(2):117-9.
180. Almuqbil M, Hamdan FF, Mathonnet G, Rosenblatt B, Srour M. De novo deletion of FMN2 in a girl with mild non-syndromic intellectual disability. *Eur J Med Genet.* déc 2013;56(12):686-8.
181. Charfi C, Voisin V, Levros L-C, Edouard E, Rassart E. Gene profiling of Graffi murine leukemia virus-induced lymphoid leukemias: identification of leukemia markers and Fmn2 as a potential oncogene. *Blood.* 10 févr 2011;117(6):1899-910.
182. Bruinsma CF, Savelberg SMC, Kool MJ, Jolfaei MA, Van Woerden GM, Baarends WM, et al. An essential role for UBE2A/HR6A in learning and memory and mGLUR-dependent long-term depression. *Hum Mol Genet.* 1 janv 2016;25(1):1-8.
183. Nascimento RMP, Otto PA, de Brouwer APM, Vianna-Morgante AM. UBE2A, Which Encodes a Ubiquitin-Conjugating Enzyme, Is Mutated in a Novel X-Linked Mental Retardation Syndrome. *Am J Hum Genet.* sept 2006;79(3):549-55.
184. Budny B, Badura-Stronka M, Materna-Kiryluk A, Tzschach A, Raynaud M, Latos-Bielenska A, et al. Novel missense mutations in the ubiquitination-related gene UBE2A cause a recognizable X-linked mental retardation syndrome. *Clin Genet.* juin 2010;77(6):541-51.
185. de Leeuw N, Bulk S, Green A, Jaeckle-Santos L, Baker LA, Zinn AR, et al. UBE2A deficiency syndrome: Mild to severe intellectual disability accompanied by seizures, absent speech, urogenital, and skin anomalies in male patients. *Am J Med Genet A.* déc 2010;152A(12):3084-90.
186. Vandewalle J, Bauters M, Van Esch H, Belet S, Verbeeck J, Fieremans N, et al. The mitochondrial solute carrier SLC25A5 at Xq24 is a novel candidate gene for non-syndromic intellectual disability. *Hum Genet.* oct 2013;132(10):1177-85.
187. Peng C, Huang TH, Jeung EB, Donaldson CJ, Vale WW, Leung PC. Expression of the type II activin receptor gene in the human placenta. *Endocrinology.* déc 1993;133(6):3046-9.
188. Matzuk MM, Kumar TR, Bradley A. Different phenotypes for mice deficient in either activins or activin receptor type II. *Nature.* mars 1995;374(6520):356-60.
189. Guernsey DL, Matsuoka M, Jiang H, Evans S, Macgillivray C, Nightingale M, et al. Mutations in origin recognition complex gene ORC4 cause Meier-Gorlin syndrome. *Nat Genet.* 27 févr 2011;43(4):360-4.
190. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature.* 17 août 2016;536(7616):285-91.

191. Jaillard S, Dubourg C, Gerard-Blanluet M, Delahaye A, Pasquier L, Dupont C, et al. 2q23.1 microdeletion identified by array comparative genomic hybridisation: an emerging phenotype with Angelman-like features? *J Med Genet.* 1 déc 2009;46(12):847-55.
192. Mullegama SV, Elsea SH. Clinical and Molecular Aspects of MBD5-Associated Neurodevelopmental Disorder (MAND). *Eur J Hum Genet.* août 2016;24(9):1235-43.
193. Talkowski ME, Mullegama SV, Rosenfeld JA, van Bon BWM, Shen Y, Repnikova EA, et al. Assessment of 2q23.1 Microdeletion Syndrome Implicates MBD5 as a Single Causal Locus of Intellectual Disability, Epilepsy, and Autism Spectrum Disorder. *Am J Hum Genet.* oct 2011;89(4):551-63.
194. Mullegama SV, Rosenfeld JA, Orellana C, van Bon BWM, Halbach S, Repnikova EA, et al. Reciprocal deletion and duplication at 2q23.1 indicates a role for MBD5 in autism spectrum disorder. *Eur J Hum Genet.* janv 2014;22(1):57-63.
195. Bonnet C, Ali Khan A, Bresso E, Vigouroux C, Béri M, Lejczak S, et al. Extended spectrum of MBD5 mutations in neurodevelopmental disorders. *Eur J Hum Genet.* déc 2013;21(12):1457-61.
196. Payne F, Colnaghi R, Rocha N, Seth A, Harris J, Carpenter G, et al. Hypomorphism in human NSMCE2 linked to primordial dwarfism and insulin resistance. *J Clin Invest.* 2 sept 2014;124(9):4028-38.
197. Clemen CS, Tangavelou K, Strucksberg K-H, Just S, Gaertner L, Regus-Leidig H, et al. Strumpellin is a novel valosin-containing protein binding partner linking hereditary spastic paraplegia to protein aggregation diseases. *Brain.* oct 2010;133(10):2920-41.
198. Derivery E, Helfer E, Henriot V, Gautreau A. Actin Polymerization Controls the Organization of WASH Domains at the Surface of Endosomes. Keen J, éditeur. *PLoS ONE.* 21 juin 2012;7(6):e39774.
199. Valdmanis PN, Meijer IA, Reynolds A, Lei A, MacLeod P, Schlesinger D, et al. Mutations in the KIAA0196 Gene at the SPG8 Locus Cause Hereditary Spastic Paraplegia. *Am J Hum Genet.* janv 2007;80(1):152-61.
200. Jahic A, Khundadze M, Jaenisch N, Schüle R, Klimpe S, Klebe S, et al. The spectrum of KIAA0196 variants, and characterization of a murine knockout: implications for the mutational mechanism in hereditary spastic paraplegia type SPG8. *Orphanet J Rare Dis* [Internet]. déc 2015 [cité 8 août 2017];10(1). Disponible sur: <http://www.ijrd.com/content/10/1/147>
201. Elliott AM, Simard LR, Coghlan G, Chudley AE, Chodirker BN, Greenberg CR, et al. A novel mutation in *KIAA0196*: identification of a gene involved in Ritscher-Schinzel/3C syndrome in a First Nations cohort. *J Med Genet.* déc 2013;50(12):819-22.

202. Goldenberg A, Riccardi F, Tessier A, Pfundt R, Busa T, Cacciagli P, et al. Clinical and molecular findings in 39 patients with KBG syndrome caused by deletion or mutation of *ANKRD11*. *Am J Med Genet A*. nov 2016;170(11):2847-59.
203. Isidor B, Pichon O, Redon R, Day-Salvatore D, Hamel A, Siwicka KA, et al. Mesomelia-Synostoses Syndrome Results from Deletion of *SULF1* and *SLCO5A1* Genes at 8q13. *Am J Hum Genet*. juill 2010;87(1):95-100.
204. Ea V, Court F, Forné T. Quantitative Analysis of Intra-chromosomal Contacts: The 3C-qPCR Method. In: Haggarty P, Harrison K, éditeurs. *Population Epigenetics* [Internet]. New York, NY: Springer New York; 2015 [cité 18 août 2017]. p. 75-88. Disponible sur: [http://link.springer.com/10.1007/7651\\_2015\\_269](http://link.springer.com/10.1007/7651_2015_269)
205. Gregor A, Oti M, Kouwenhoven EN, Hoyer J, Sticht H, Ekici AB, et al. De Novo Mutations in the Genome Organizer CTCF Cause Intellectual Disability. *Am J Hum Genet*. juill 2013;93(1):124-31.

## Summary

The study of tridimensional organization of chromatin by Hi-C technology, derived from Chromosome Conformation Capture (3C), revealed chromatin domains named Topologically Associated Domains (TADs). Alterations of TADs can lead to several developmental defects.

The interpretation of array Comparative Genomic Hybridization (array-CGH) used in routine for diagnosis of developmental defects can be difficult, with many Copy Number Variations (CNVs) of unknown significance. However, the effect of these CNVs on TADs has only recently been taken into account for this interpretation.

During this work, 735 CNVs detected by array-CGH and initially considered of unknown significance have been reanalyzed, to determine if they can alter the TADs organization. Furthermore, a 3C protocol has been developed.

The new interpretation has been performed using public data of Hi-C, epigenetic marks of the ENCODE project, and enhancers databases.

This study found a possible alteration of TADs on 7 patients. The yield of this analyze (7/735), is similar to another study, and highlights that the disruption of TADs is probably an uncommon pathophysiological mechanism.

Furthermore, the technical limits of array-CGH and public databases annotations are a barrier to the interpretation of the CNVs effect on TADs.

If the study of TADs for diagnostic purpose is possible nowadays, especially on well-characterized locus, it is still a complex study, with the need of confirmation with *in vitro* techniques, such as 3C derived technologies.

## Résumé

L'étude de l'architecture tridimensionnelle de la chromatine par la technologie de Hi-C, dérivée de la Chromosome Conformation Capture (3C), a mis en lumière l'existence de domaines chromatiniens dénommés Topologically Associated Domains (TADs), dont l'altération peut être à l'origine d'anomalies du développement.

La technique d'Analyse Chromosomique sur Puce à ADN (ACPA), utilisée en routine dans le diagnostic des anomalies du développement, se heurte à des difficultés d'interprétation, de nombreuses variations du nombre de copies (CNVs) ainsi détectées restant de signification inconnue. Cependant, l'effet de ces CNVs sur les TADs n'est que récemment pris en compte dans cette interprétation.

Dans le cadre de cette thèse, 735 CNVs détectés par ACPA et initialement considérés comme de signification inconnue, ont été ré-analysés afin de déterminer si ceux-ci peuvent altérer l'organisation des TADs. Par ailleurs, la mise au point d'un protocole de 3C a été réalisée.

Une nouvelle interprétation a ainsi été effectuée en utilisant différents jeux de données publiques d'expériences de Hi-C et de marques épigénétiques issues du projet ENCODE, ainsi que de différentes bases de données d'enhancers.

Cette étude a mis en évidence une possible altération de TADs chez 7 patients. Le rendement de la ré-analyse de notre cohorte (7/735), concordant avec une étude similaire, montre que la perturbation de TADs par un CNV est probablement un mécanisme physiopathologique rare.

De plus, les limites techniques de l'ACPA, ainsi que celles des annotations disponibles dans les bases de données publiques, sont autant d'obstacles à l'interprétation de l'effet des CNVs sur les TADs.

Si l'étude des TADs dans le cadre du diagnostic apparaît donc aujourd'hui comme possible, notamment pour les locus dont l'architecture est désormais caractérisée, celle-ci reste complexe, et nécessite une confirmation par des techniques *in vitro* telles que celles dérivées de la 3C.

Discipline : Génétique Médicale

Mots clés : Topologically Associated Domains, 3D Genome, CNV, ACPA, Chromosome Conformation Capture