

Université de Montréal

**Étude sur le rôle des déséquilibres génomiques dans le  
Syndrome d'Impatiences Musculaires de l'Éveil**

par

Simon L. Girard

Département de Physiologie

Faculté de Médecine

Mémoire présenté à la Faculté de Médecine, département de Physiologie  
en vue de l'obtention du grade de Maître ès Sciences  
en Sciences Neurologiques

22 Juillet 2010

© Simon L. Girard, 2010  
Université de Montréal  
Faculté des études supérieures et postdoctorales

Ce mémoire intitulé :

Étude sur le rôle des déséquilibres génomiques dans le Syndrome d'Impatiences Musculaires de  
l'Éveil

Présenté par :  
Simon L. Girard

a été évalué par un jury composé des personnes suivantes :

Edward Bradley, président-rapporteur  
Guy Rouleau, directeur de recherche  
Frédéric Charron, membre du jury

## Résumé

Le Syndrome d'Impatiences Musculaires de l'Éveil (SIME) est une maladie neurologique caractérisée par un besoin urgent de bouger les jambes. C'est également l'une des causes les plus fréquentes d'insomnie. C'est une maladie très répandue, avec une prévalence de presque 15 % dans la population générale. Les maladies multifactorielles comme le SIME sont souvent le résultat de l'évolution d'une composante génétique et d'une composante environnementale. Dans le cadre du SIME, les études d'association génomique ont permis l'identification de 4 variants à effet modéré ou faible. Cependant, ces quatre variants n'expliquent qu'une faible partie de la composante génétique de la maladie, ce qui confirme que plusieurs nouveaux variants sont encore à identifier. Le rôle des déséquilibres génomiques (Copy Number Variations ou CNVs) dans le mécanisme génétique du SIME est à ce jour inconnu. Cependant, les CNVs se sont récemment positionnés comme une source d'intérêt majeur de variation génétique potentiellement responsable des phénotypes. En collaboration avec une équipe de Munich, nous avons réalisé deux études CNVs à échelle génomique (biopuces à SNP et hybridation génomique comparée (CGH)) sur des patients SIME d'ascendance germanique. À l'aide d'une étude cas-contrôle, nous avons pu identifier des régions avec une occurrence de CNVs différentes pour les patients SIME, comparés à différents groupes contrôles. L'une de ces régions est particulièrement intéressante, car elle est concordante à la fois avec des précédentes études familiales ainsi qu'avec les récentes études d'associations génomiques.

**Mots-clés :** Syndrome d'Impatiences Musculaires de l'Éveil, Génétique Humaine, Neurogénétique, Aberrations chromosomiques, Études d'associations génomiques, PTPRD

## Abstract

Restless Legs syndrome (RLS) is a neurological disorder characterized by the urge to move one's limbs. It is also one of the most frequent causes of insomnia. The prevalence of RLS is estimated to be around 15% in the general population. Complex disorders like RLS are often the result of the evolution of genetic and environmental components. For RLS, recent Genome Wide Association Study (GWAS) have identified four variants with mild to moderate effects. However, those four variants explain only a small part of the disease heritability and thus, we expect that many new variants are still to be found. The impact of Copy-Number Variation (CNV) in the genetic mechanism of RLS is still unknown. However, many studies have recently position the CNVs as a significant source of genetic variation potentially responsible of phenotypes. In collaboration with a team from Munich, we conducted two genome-wide CNVs studies (Genome Wide SNP chips and Comparative Genomic Hybridization (CGH)) on RLS patients from Germany. Using cases-controls studies, we identified regions with a different occurrence of CNVs for RLS patients, compared to different groups of controls. One of these regions is particularly interesting, as it has already been identified by both linkage and association studies.

**Keywords :** Restless Legs Syndrome, Human Genetic, Neurogenetic, Copy Number Variation (CNV), Genome Wide Association Study (GWAS), PTPRD

## Table des matières

Introduction .....	10
1.1 L'étiologie du SIME.....	3
1.1.1 Les causes du SIME symptomatique .....	3
1.1.2 Les mécanismes métaboliques .....	4
1.2 Épidémiologie .....	7
1.2.1 Prévalence.....	8
1.2.2 Comorbidité .....	8
2. L'étiologie génétique du SIME.....	9
2.1 Introduction .....	9
2.2 Études de liaison génétique.....	11
2.2.1 Loci identifiés par des études de liaisons pour le SIME .....	12
2.2.2 Perspectives .....	16
2.3 Les études d'associations.....	17
2.3.1 Les associations positives faites par EAG pour le SIME .....	18
2.4 Perspectives génétiques .....	21
2.4.1 Le séquençage de nouvelle génération (SNG) .....	23
2.5 Les études de déséquilibre génomique .....	25
2.5.1 La nature des CNVs .....	25
2.5.2 Les méthodes expérimentales pour détecter les CNVs .....	27
2.5.3 L'effet des CNVs sur les phénotypes.....	29
3. But de l'étude .....	31
4. Méthodologie .....	33
4.1 Recrutement des patients.....	33
4.2 Détection de CNVs à l'aide de bio-puces à SNPs .....	34
4.2.1 Génotypage sur les bio-puces à SNPs .....	34
4.2.2 Détection des CNVs.....	34
4.2.3 Étude cas-contrôles des données SNP-chip.....	35
4.2.4 Annotations des régions CNVs.....	36
4.3 Détection de CNVs à l'aide de la plateforme CGH .....	36
4.3.1 Hybridation des ADN.....	36
4.3.2 Traitement des images.....	37
4.3.3 Algorithme de détection des CNVs .....	37
4.3.4 Étude cas-contrôles des données CGH .....	38
5. Résultats .....	39
5.1 Étude préliminaire sur les individus HapMap .....	39
5.2 Détection de CNVs par SNP-chip.....	41
5.3 Détection des CNVs par CGH .....	44
6. Discussion.....	45
7. Perspectives .....	51

8. Conclusion.....	53
9. Références.....	54

## Liste des figures

<b>FIGURE 1</b> ILLUSTRATION DES DIFFÉRENTS TYPES DE DÉSÉQUILIBRES GÉNOMIQUES (CNVS).....	1
<b>FIGURE 2</b> DIAGRAMME DE L'INTERACTION DES DIFFÉRENTS MODULES DU LOGICIEL BIRDSUITE <sup>1</sup> (BROAD INSTITUTE).....	1
<b>FIGURE 3.</b> OCCURRENCE MOYENNE DES CNVS PAR INDIVIDUS, CLASSÉE PAR CHROMOSOME..	1
<b>FIGURE 4</b> PROCESSUS DE SÉLECTION DES CNVS SPÉCIFIQUES AU SIME.....	1
<b>FIGURE 5</b> RÉSULTATS DE L'ÉTUDE DE CNVS À ÉCHELLE GÉNOMIQUE. L'AXE DES X PRÉSENTE.....	1
<b>FIGURE 6.</b> CHEVAUCHEMENT DES ÉTUDES DE SNP-CHIPS, CGH ET ASSOCIATIONS AU NIVEAU DE LA RÉGION NON-CODANTE DU GÈNE PTPRD.....	1

## Liste des tableaux

<b>TABLEAU 1 :</b> TABLEAU RÉCAPITULATION DES ASSOCIATIONS GÉNÉTIQUES IDENTIFIÉES PAR ÉTUDE DE LIAISONS POUR LE SIME.....	13
<b>TABLEAU 2</b> TABLEAU RÉCAPITULATIF DES LOCI IDENTIFIÉS PAR ÉTUDES D'ASSOCIATION.....	19
<b>TABLEAU 3</b> COMPARAISON ENTRE DIVERSES ÉTUDES DE CNVS À ÉCHELLE GÉNOMIQUE.....	26
<b>TABLEAU 4</b> RÉSUMÉ DÉTAILLÉ DES RÉGIONS DE CNVS OBSERVÉES LORS DE L'ÉTUDE À ÉCHELLE GÉNOMIQUE.....	44
<b>TABLEAU 5</b> VALIDATION DES RÉGIONS EN UTILISANT UNE TECHNOLOGIE CGH.....	45
<b>TABLEAU 6</b> TABLEAU RÉCAPITULATIF DES ÉTUDES GÉNÉTIQUES AYANT DÉCOUVERT DES ASSOCIATIONS SIGNIFICATIVES AVEC LA RÉGION CHROMOSOMIQUE 9P.....	46

**Liste des sigles et abréviations**

ADN = Acide DésoxyriboNucléique  
ARN = Acide RiboNucléique  
BMP = Bone Morphogenic Proteins  
BTBD9 = BTB/POZ domain containing 9  
CFTR = Cystic Fibrosis transmembrane conductance regulator  
CGH = Hybridation par Génomique Comparée  
cM = CentiMorgans  
CNVs = Déséquilibres Génomiques (Copy Number Variations)  
DMD = Dystrophin  
DL = Déséquilibre de Liaison  
DLRS = Dérivée du ratio logarithmique  
EAG = Étude d'association génomique  
ELG = Étude de liaison génétique  
FAM – Fréquence de l'Allèle Mineur  
FDOPA = fluoro-(18F)-dihydroxyphénylalanine  
HMM = Modèle caché de Markov  
HOXA9 = Homeobox A9  
Kb = Kilobase  
LBX1 = Ladybird Homeobox 1  
LBXCOR1 = Ladybird Homeobox 1 corepressor  
LOD-score = Logarithme des probabilités  
LCS = Liquide Cérébro-Spinal  
MAPK = Mitogen-Activated Protein Kinase  
MAP2K5 = MAP Kinase Kinase 5  
Mbp = Mega paire de base  
MEIS1 = MEis Homeobox 1  
MPJ = Mouvement périodique des jambes  
MPJE = MPJ à l'éveil  
MPJS = MPJ durant le sommeil  
PBX1 = pre-B-cell leukemia homeobox 1  
PCR = Réaction en chaîne de polymérase  
PTP = Protéine Tyrosine Phosphatase  
PTPRD = Protéine Tyrosine Phosphatase Receptor D  
SIME = Syndrome d'Impatiences Musculaires de l'Éveil  
SJSR = Syndrome des Jambes sans Repos  
SNC = Système Nerveux Central  
SN = Substantia Nigra  
SNG = Séquençage de Nouvelle Génération  
SNP = Polymorphisme à simple nucléotide  
SNP-chip = Biopuce à SNP  
TEP = tomographie par émissions de positrons  
TEMP = tomographie d'émission monophotonique  
TH = Tyrosine Hydroxylase  
WTCCC = Welcome Trust Case Control Consortium

*[Cette symphonie est dédiée à Amélie, mais aussi un peu à son papa, qui comprend très bien toutes les étapes d'interprétation]*

## Remerciements

Je veux tout d'abord remercier Guy Rouleau, mon directeur de recherche. Sans un mentor, il est toujours difficile pour un étudiant de s'orienter et de progresser. Lorsque l'on a la chance d'avoir un mentor qui partage la même passion que soi-même, de grandes choses sont possibles.

Je veux également remercier Juliane Winkelmann et Barbara Schormair, de l'équipe de Munich. Je n'aurai jamais pu réaliser ce projet, si ce n'était de l'ouverture et de l'aide de Barbara et Juliane. De l'équipe de Munich, je tiens également à remercier Franziska Knauf, Sebastian Eck, Jelena Golic, Thomas Meitinger, Derek Spieler et Maria Kaffe.

Évidemment, un gros merci à tous les gens avec qui j'ai travaillé dans le laboratoire de Guy Rouleau. Une mention spéciale pour Patrick Dion, mais également pour Jean-Baptiste, Judith, Cynthia, Véronique, Nancy, Hussein, Julie, Paule et Ron de l'équipe génétique. Évidemment, merci à Lan, Anne et Hélène de l'équipe RLS. Une mention *sans inhibition* pour l'équipe du labo Cossette (Maude, Caro et Caro, Pamela et Maxime).

Je tiens à terminer en remerciant les personnes les plus importantes pour moi. Ma fiancée Amélie (et sa famille!), ma famille Martine, Bruno et Marco, ma complice de toujours Jessie, et mes amies Sarah, Laurence, Isabelle, Naomi. Merci pour le support à Rach et Tchaikov.

Et à tous ceux que j'oublie sur papier, ma reconnaissance envers vous n'est pas moindre!

## Introduction

Le Syndrome d'Impatiences Musculaires de l'Éveil (SIME) est un trouble neurologique fréquemment observé dans l'ensemble de la population mondiale. Toutefois, les études épidémiologiques demeurent encore partagées en ce qui concerne la prévalence de la maladie puisqu'il n'y a pas de consensus définitif sur un choix de critères diagnostiques. Ainsi, il est difficile de déterminer une valeur pour celle-ci au niveau de la population mondiale. On situe généralement cette valeur entre 7 et 15 %<sup>2,3</sup>, le tiers de cette valeur correspondant aux cas sévères requérant de la médication<sup>2</sup>. Il a également été démontré que cette fréquence augmente de façon significative dans la population plus âgée (> 50 ans)<sup>4</sup>. La question d'une association spécifique avec un sexe demeure toutefois nébuleuse, certaines études proposant un ratio de l'ordre de 2 femmes pour 1 homme<sup>2,5</sup> alors que d'autres études ne rapportent aucune association significative avec le genre<sup>3,6</sup>.

Il existe actuellement une polémique sur la désignation de la maladie, particulièrement au niveau de la langue française. Tout d'abord observée en 1672 par le médecin anglais Thomas Willis, elle fut désignée comme « Restless Legs ». Par la suite, en 1861, Wittmaack fait une description détaillée du syndrome, cette fois sous la nomination d'« Anxietas Tibiarum ». Toujours selon lui, l'Anxietas Tibiarum serait une manifestation hystérique plutôt que neurologique. Pratiquement un siècle plus tard, en 1945, le neurologue suédois Karl Axel Ekbom fait la première description clinique et neurologique du syndrome qui sera dorénavant connu sous l'appellation « Restless Legs Syndrome ». Quelques fois, en l'honneur des deux médecins qui furent à l'origine de l'identification du syndrome, on le désigne comme le syndrome de Wittmaack-Ekbom. La désignation française « Syndrome d'Impatiences Musculaires de l'Éveil » a été établie dans les années 1980 par les spécialistes du sommeil. Au cours des dernières années, le SIME a été de plus en plus popularisé au niveau de la population générale et est maintenant plus connu sous la traduction libre « Syndrome des Jambes Sans Repos (SJSR) ». Les spécialistes reconnaissent que cette dernière appellation est plus proche de la réalité et que la désignation SIME peut porter à confusion sur la nature des symptômes. Ainsi, les deux appellations (SIME et SJSR) sont acceptées et couramment utilisées. À des fins pratiques, pour le reste de cet ouvrage, l'abréviation SIME sera utilisée pour faire référence au syndrome.

---

Le diagnostic du SIME se fait selon des critères bien précis<sup>7,8</sup> :

- Un vif besoin de bouger le membre (il est plus fréquent de ressentir les symptômes du SIME au niveau des jambes, mais on peut également les retrouver au niveau des bras), souvent accompagné de sensations douloureuses. Les symptômes peuvent varier entre des sensations légères (souvent décrites comme « fourmis dans les jambes ») à des sensations beaucoup plus douloureuses (dites « sensations de brûlure »).
- Les sensations douloureuses sont partiellement ou complètement soulagées par des activités de mouvement, comme la marche ou des étirements.
- Les inconforts sont exclusivement présents, ou aggravés, lors des périodes de repos ou d'inactivité.
- Les inconforts sont exclusivement présents, ou aggravés durant la soirée ou pendant la nuit.

De plus, il existe également des critères de supports qui peuvent aider dans l'établissement d'un diagnostic :

- Présence d'autres cas de SIME dans la famille
- Réponse positive à un traitement à la dopamine
- Présence de mouvements périodiques de la jambe (MPJ) à l'éveil (MPJE) ou au sommeil (MPJS) tel qu'établi par un actigraphe ou autre appareil de l'activité de la jambe.

On estime que la plupart des gens souffrants de SIME n'ont pas de symptômes qui affectent significativement leur qualité de vie et ainsi, n'ont pas besoin de traitement thérapeutique. Or, une portion significative des patients souffrants de SIME (symptômes présents au moins deux fois par semaine) voit sa qualité de vie largement diminuée par la présence de la maladie. Les symptômes du SIME sont parmi les causes les plus répandues d'insomnie. En outre, les patients souffrants de SIME seraient également plus à risque de développer des troubles d'anxiété ou de dépression<sup>9</sup>.

## 1.1 L'étiologie du SIME

Il existe deux formes principales du SIME, la forme idiopathique (que l'on appelle également SIME génétique ou primaire) et la forme symptomatique (aussi nommé SIME secondaire). On qualifie de SIME idiopathique tout patient qui présente les symptômes caractéristiques sans cependant présenter de causes sous-jacentes qui pourraient être responsables de l'apparition de ces symptômes. En revanche, lorsque l'on peut lier les symptômes du SIME à une origine bien déterminée, ce dernier sera qualifié de symptomatique. Trois causes principales du SIME symptomatique sont connues : l'anémie (ainsi que les dons de sang), la grossesse et l'insuffisance rénale.

### 1.1.1 Les causes du SIME symptomatique

#### 1.1.1.1 L'anémie et les dons de sang

Plusieurs études ont démontré qu'un taux de fer dans le sérum inférieur à 50 ng par ml pourrait être une des causes potentielles du SIME<sup>10,11</sup>. L'insuffisance en fer est l'une des causes les plus fréquentes du SIME symptomatique, comptant pour presque 20 % de tous les cas recensés<sup>12</sup>. Parallèlement, il a également été possible d'observer une corrélation entre les patients SIME qui donnent régulièrement du sang et un taux de fer plus faible que la moyenne<sup>13</sup>. Cette dernière étude a été réalisée sur un nombre restreint de patients (n=8) et bien qu'une association soit déjà connue entre le SIME et l'anémie, l'association entre le SIME et les dons de sang est toutefois loin d'être établie de façon définitive. Il faut donc être prudent lors du diagnostic d'un patient car il n'est pas définitivement acquis qu'un patient SIME donneur de sang soit atteint par la forme secondaire de la maladie. Ce point aura une grande importance lors des analyses génétiques, traitées plus loin. Quant à la cause métabolique de l'implication du fer, elle sera traitée dans la sous-section suivante.

#### 1.1.1.2 La grossesse

La grossesse est un autre facteur de risque important lorsqu'il s'agit de la forme symptomatique du SIME. En effet, on estime qu'une femme enceinte est deux à trois fois plus à risque de développer le syndrome<sup>14</sup>. Dans la majorité des cas, les symptômes seront de très légers à modérés. Il a été démontré que les symptômes sont plus sévères lors du troisième trimestre de la grossesse et que, dans la majorité des cas, on observe une disparition complète du SIME après la naissance du bébé<sup>15</sup>. On ignore encore la raison pour laquelle les femmes ont une plus forte

tendance à développer les symptômes du SIME lors d'une grossesse. Au cours d'études biochimiques tentant de faire un lien entre l'anémie de grossesse et le SIME, une association a été observée entre la vitamine B9 et le SIME de grossesse<sup>16</sup>. Jusqu'à maintenant, les fondements de cette association ne sont pas encore compris, mais une hypothèse intéressante est que des suppléments de vitamine B9 pourraient traiter non seulement les femmes enceintes souffrant de SIME, mais également les patients souffrants de SIME idiopathique.

### ***1.1.1.3 L'insuffisance rénale***

Plusieurs études épidémiologiques ont pu déterminer que la fréquence du SIME est significativement plus élevée chez les patients recevant des traitements de dialyse (généralement pour traiter l'insuffisance rénale, aussi appelée urémie) que chez la population normale<sup>17</sup>. À l'image des symptômes causés par la grossesse, les symptômes chez les patients urémiques recevant des traitements de dialyse sont encore mal expliqués. Il a été observé que les symptômes du SIME ne sont en aucun cas améliorés après un traitement de dialyse. Par contre, ceux-ci disparaissent complètement après une transplantation rénale<sup>18</sup>. Un fait très intéressant est qu'une corrélation a été établie entre les fréquences et le nombre de sessions de dialyse et la sévérité du SIME<sup>19</sup>. Plusieurs études ont tenté de dresser des profils pour les patients recevant des traitements de dialyse souffrant de SIME, mais aucune différence n'a pu être établie lorsque comparé à une population urémique standard<sup>20,21</sup>.

### **1.1.2 Les mécanismes métaboliques**

Jusqu'à présent, aucun mécanisme n'a été établi de façon non équivoque comme étant le facteur causatif du SIME. Ainsi, plusieurs doutes planent encore au sujet des mécanismes métaboliques responsables de la pathogenèse du SIME. Si certains indices établissent de forts liens entre la maladie et certains mécanismes, le système métabolique à la base du SIME est encore, pour la plus large portion, inconnu. Cependant, deux hypothèses, soit celle de l'implication du transport du fer et celle de l'implication de la dopamine sont très sérieusement étudiées. Comme ces deux hypothèses ne sont pas mutuellement exclusives, il est envisageable que les deux soient impliquées dans le mécanisme du SIME.

#### ***1.1.2.1 L'hypothèse de l'implication du fer***

L'origine de l'hypothèse de l'implication du métabolisme du fer dans le SIME provient du fait que pour certains patients, un soulagement complet et persistant est observé lorsqu'ils sont traités avec des suppléments de fer<sup>11</sup>. De plus, les trois causes majeures du SIME secondaire, soit la

grossesse, l'anémie et l'insuffisance rénale, sont toutes des troubles qui compromettent le niveau de fer.

Les premières études qui ont établi un lien métabolique entre le SIME et le niveau de fer ont montré une corrélation entre la sévérité de la maladie et le niveau de ferritine du sérum (système périphérique)<sup>11,22</sup>. La ferritine est une protéine qui emmagasine le fer de manière contrôlée, permettant ainsi d'éviter les déficits ou les surplus de fer dans l'organisme. D'ailleurs, une faible mesure du taux de la ferritine dans le sérum sanguin est l'élément diagnostique principal de l'anémie<sup>23</sup>; cette mesure est généralement la plus répandue lorsqu'il s'agit de mesurer le niveau de stockage du fer. Récemment, une étude à double insu a démontré que lorsque l'on traite des patients SIME présentant un faible taux de ferritine du sérum avec un surplus de fer, les symptômes de la maladie s'en trouvent réduits<sup>24</sup>.

Afin d'affiner l'hypothèse de l'implication du fer, plusieurs équipes de recherche se sont concentrées sur les niveaux de ferritine du liquide cérébro-spinal (LCS). En effet, comme une majorité des patients SIME répond très bien à un traitement dopaminergique<sup>25</sup> il est rationnel de se pencher sur le rôle du système nerveux central (SNC) et d'y étudier le mécanisme du fer. Ainsi, deux études ont analysé le mécanisme du fer dans le LCS chez des patients SIME<sup>26,27</sup>. Dans ces deux études, une baisse du niveau de ferritine ainsi qu'une augmentation du niveau de la transferrine (protéine globulaire qui transporte le fer de l'intestin vers les réserves hépatiques) dans le LCS ont été observées chez les patients SIME. Paradoxalement, aucune différence n'a été notée pour le niveau de ferritine du sérum, ce qui est contraire avec des observations faites par un autre groupe<sup>24</sup>. On pourrait donc conclure que les patients SIME ont à la fois une carence en fer au niveau du cerveau, mais aussi une réduction de l'activité de transfert du fer du sérum vers le LCS. De plus, une troisième étude a pu établir un lien entre la ferritine du LCS et le cycle circadien chez les patients avec une apparition précoce des symptômes de la maladie<sup>28</sup>. Enfin, une étude récente a permis d'identifier une différence substantielle du taux de ferritine mitochondriale au niveau de la substantia nigra (SN) du ganglion de la base<sup>29</sup>, entre des patients SIME et des sujets contrôles. Cette étude est d'une grande importance pour l'avancement de la recherche sur le SIME, non seulement parce qu'il s'agit d'une région du cerveau contenant un nombre important de neurones dopaminergiques, mais aussi parce que la SN est également impliquée dans un autre trouble du mouvement notoire, la maladie de Parkinson. Les liens connus entre le SIME et la maladie de Parkinson seront discutés dans une section subséquente.

### *1.1.2.2 L'hypothèse de l'implication du système dopaminergique*

L'hypothèse impliquant le système dopaminergique dans la pathophysiologie du SIME a été avancée suite à l'observation du soulagement des symptômes de la maladie lors de l'administration d'agents thérapeutiques agissant sur le profil dopaminergique<sup>25</sup>. À la suite de ces observations, plusieurs études ont été mises en place pour tenter d'identifier le rôle de la dopamine dans le mécanisme du SIME.

La tomographie par émissions de positrons (TEP) en utilisant le fluoro-(18F)-dihydroxyphénylalanine (FDOPA, le précurseur métabolique de la dopamine) comme traceur, fut utilisée afin d'établir le profil différentiel de la dopamine dans le cerveau d'un patient SIME. Un profil différent fut observé dans le cas des présynapses négrostriatale<sup>30</sup> par rapport à des individus contrôles. Une baisse de la consommation de FDOPA fut également observée au niveau du noyau caudé et du putamen. Une autre étude, cette fois en utilisant la tomographie d'émission monophotonique (TEMP), a permis de distinguer un profil anormal au niveau de la liaison des récepteurs dopaminergiques D2 dans le striatum, lorsque comparé avec un groupe de patients contrôles<sup>31</sup>. Ces deux études supportent l'hypothèse d'une dysfonction centrale du système dopaminergique. Cependant, il est très important de prendre en considération le fait que ces études ont été réalisées sur des échantillons très petits. De ce fait, il est difficile de lier les faits observés avec le SIME.

Étant donné la puissance limitée et le manque de reproductibilité de l'imagerie médicale, les groupes de recherches se sont tournés vers des approches plus directes, notamment en procédant à des analyses du LCS. En comparant les mesures dopaminergiques du LCS à deux moments précis de la journée (10:00 AM et 10:00 PM), il a été possible d'observer une plus grande variation des dynamiques dopaminergiques circadiennes chez les patients SIME par comparaison à un groupe contrôle d'individus en santé<sup>32</sup>.

Plus récemment, des études ont été réalisées à partir de cerveaux prélevés sur des patients SIME décédés. L'analyse de ces cerveaux a notamment permis d'obtenir un profil dopaminergique plus complet que les analyses tomographiques par positron ou monophotoniques. Le profil d'expression de protéines impliquées dans le système de la dopamine a été établi à partir de huit cerveaux de patients SIME<sup>33</sup> et une hausse significative de la tyrosine hydroxylase (TH) a ainsi été observée. La TH est une protéine clé du système dopaminergique car elle est responsable de la conversion de la tyrosine en dopamine<sup>34</sup>. Des résultats ont également été observés dans des

modèles animaux et cellulaires de déficience en fer. Une importante baisse du niveau de récepteur dopaminergique dans la moelle épinière a été observée chez des souris avec une carence en fer<sup>35</sup>. Ces observations fournissaient pour la première fois une indication claire de l'implication de la dopamine au niveau de la pathogenèse SIME. Ces résultats sont également en accord avec l'hypothèse de l'implication du fer. Ainsi, une insuffisance de fer pourrait être à l'origine d'une dysfonction dopaminergique dans la SN. Afin de clarifier le rôle du métabolisme du fer dans le système dopaminergique, une autre équipe s'est penchée sur des rats avec une déficience en fer. L'équipe a observé une régulation fonctionnelle positive des récepteurs adénosique A2 au niveau des synapses *pre* et *post* synaptiques du striatum<sup>36</sup>. Il est connu que l'adénosine est un puissant modulateur du système dopaminergique<sup>37</sup>, ce qui fortifie davantage la co-implication du métabolisme du fer ainsi que du système dopaminergique dans la pathogenèse du SIME.

La dopamine étant également un neurotransmetteur clé impliqué dans la formation de la mémoire motrice, une équipe a récemment testé les effets du SIME sur cette dernière composante. En utilisant une technique appelée la simulation associative couplée, il a été possible de démontrer que les patients SIME ont une déficience au niveau de la potentialisation à long terme de la mémoire motrice, ce qui implique également un affaiblissement de la plasticité sensori-motrice associative<sup>38</sup>. Cependant, un traitement dopaminergique rétablit partiellement la plasticité associative, ce qui constitue un indice additionnel de l'implication du système dopaminergique dans la pathophysiologie du SIME.

## 1.2 Épidémiologie

Il faut être très avisé lorsque l'on consulte les études épidémiologiques portant sur le SIME. En effet, dû à la très grande prévalence de ce syndrome dans la population générale (7 à 15 %), il est très facile de recruter de grandes cohortes de patients. Cependant, les symptômes du SIME n'étant pas toujours très inconfortables, il peut être facile de mal interpréter certains facteurs et d'établir des diagnostics erronés. La présence de ces faux positifs et faux négatifs doit être minimisée au maximum, afin de permettre aux études génétiques de posséder une puissance suffisante pour détecter les variations rares (voir chapitre suivant). Ainsi, lorsque l'on consulte les experts de la maladie, un consensus est que pour qu'une association entre le SIME et une autre condition biologique soit acceptée, il faut qu'elle soit répliquée au moins une fois, voir deux fois, dans des populations indépendantes.

### 1.2.1 Prévalence

Les premières études épidémiologiques portant sur le SIME ont été effectuées presque exclusivement dans les populations de descendance européenne, ce qui a eu pour effet de répandre l'idée que le SIME est présent à forte prévalence uniquement chez les populations caucasiennes. Ce n'est que pendant la dernière décennie que des études épidémiologiques chez d'autres populations ont été publiées. Plusieurs études ont rapporté des prévalences plutôt contradictoires pour l'Asie. L'exemple le plus flagrant est la Corée du Sud. En 2005, une étude a rapporté une prévalence du SIME de 12.1 %<sup>39</sup>. Deux années plus tard, en 2007, un autre groupe a publié une nouvelle étude rapportant une prévalence de 3.9 %<sup>40</sup>. Finalement, en 2009, une nouvelle étude fut réalisée, cette fois présentant une prévalence de 0.9 %<sup>41</sup>. Ces données contradictoires démontrent la grande nécessité de standardiser le diagnostic du SIME à grande échelle. D'autres études similaires, réalisées en Inde<sup>42</sup> et au Japon<sup>43,44</sup>, présentent également des données contradictoires. Par contre, un fait observable dans toutes ces études est la prévalence significativement plus grande chez les femmes que chez les hommes.

### 1.2.2 Comorbidité

Certaines associations sont globalement acceptées, notamment les trois exemples présentés dans la section 1.1.1 (anémie, grossesse et insuffisance rénale). Une autre de ces associations qui n'est plus remise en question est celle avec les mouvements périodiques des jambes durant le sommeil (MPJS)<sup>45</sup>. Cependant, bien que l'association soit incontestable, il est très difficile de quantifier les liens entre le SIME et les MPJS, si bien qu'il est pratiquement impossible d'établir une valeur témoin en ce qui concerne la prévalence des MPJS dans le SIME. Des valeurs aussi élevées que 80 % ont été utilisées, mais les preuves sont insuffisantes pour les confirmer<sup>46</sup>. Cependant, il apparaît correct de dire qu'une forte majorité des patients SIME souffre également de MPJS<sup>46,47</sup>. Cependant, il existe une portion significative des patients SIME qui ne présente aucun trouble de MPJS, et vice-versa. Une autre interrogation intéressante est de déterminer si le SIME et les MPJS résultent d'un mécanisme biologique commun ou très rapproché. Cependant, comme il a été démontré que les traitements généralement utilisés pour le SIME sont très peu efficaces pour les MPJS, les hypothèses penchent plus en faveur de deux mécanismes différents<sup>48</sup>, potentiellement liés. Ce facteur accroît grandement la complexité des études génétiques et l'établissement d'associations génétiques. Un de ces exemples est le gène *BTBD9*<sup>49</sup> qui sera traité dans une section suivante.

Une autre association qui est de plus en plus acceptée est celle avec la maladie de Parkinson<sup>50-52</sup>. Bien que la majorité des études associant le SIME et la maladie de Parkinson ait été faites avant les récentes études portant sur le système dopaminergique, elles prennent aujourd'hui encore plus de sens à la suite de l'implication de la SN dans le SIME<sup>29,33</sup>, implication déjà connue pour la maladie de Parkinson. Dans les dernières années, plusieurs autres troubles neurologiques ont également été associés avec le SIME, notamment le trouble du déficit de l'attention<sup>53-55</sup>, les troubles d'anxiétés et dépressions<sup>9,56</sup>, la sclérose en plaques<sup>57-59</sup> et divers types de neuropathies<sup>60-62</sup>. En outre, de solides associations ont également été identifiées avec le diabète de type II<sup>63-65</sup> et avec diverses maladies cardiovasculaires, dont l'hypertension<sup>66-68</sup>. Au niveau des études génétiques, il est crucial de bien déterminer les liens sous-jacents entre le SIME et les autres phénotypes afin d'être en mesure de déterminer la nature idiopathique ou non du syndrome.

## 2. L'étiologie génétique du SIME

### 2.1 Introduction

Le SIME est un trouble neurologique avec une forte composante héréditaire. En effet, on rapporte qu'une grande partie des cas de SIME idiopathique seraient des cas héréditaires (au moins un membre de la famille rapproché est également atteint)<sup>2</sup>. Cette affirmation a été depuis validée dans plusieurs couples de jumeaux d'origine canadienne-française. En comparant les prévalences pour les paires de jumeaux dizygotes et monozygotes (10 et 50 %, respectivement), il a été possible d'établir le facteur d'héritabilité du SIME à presque 50 %<sup>69</sup>. Il n'a pas encore été possible de prouver un facteur similaire au niveau des cas de SIME symptomatiques, mais des résultats préliminaires (non publiés) laissent présager une héritabilité significative.

Outre les études sur les paires de jumeaux, il est également possible de confirmer le grand facteur héréditaire du SIME en examinant le nombre de grandes familles atteintes du SIME rapporté. Un nombre important de familles avec un grand nombre d'atteints (plus de dix personnes atteintes) ont été rapportées, notamment dans les populations d'origine canadienne-française<sup>70-72</sup>, germanique (et sud-tyrolienne)<sup>73</sup>, italienne<sup>74</sup>, et nord-américaine<sup>75</sup>. Notre équipe est à l'origine de plusieurs publications portant sur la génétique du SIME et possède une collection de près de 300 familles affectées par le SIME. Si beaucoup de ces familles sont trop petites pour nous permettre une conclusion génétique, certaines familles possèdent un nombre impressionnant de membres

atteints et ne laissent aucun doute quant à l'implication d'un facteur génétique au niveau de la famille.

En présence d'une maladie rare à faible prévalence, le premier réflexe du généticien sera d'assumer une transmission selon les lois de Mendel (un gène = une maladie) selon un mode dominant (lorsqu'une seule copie de la mutation est suffisante pour développer le phénotype) ou récessif (une copie de la mutation doit être présente au niveau de chaque haplotype). Dans le cas d'un mode de transmission dominant, les individus atteints peuvent être hétérozygotes (une seule copie de l'allèle) ou homozygotes (deux copies de l'allèle) alors que pour une transmission récessive, les individus atteints sont forcément homozygotes. Si plusieurs maladies rares suivent les lois de Mendel, on connaît également plusieurs maladies rares qui peuvent être causées par plusieurs dysfonctions géniques et ainsi ne sont pas forcément en accord avec les lois de Mendel. La sclérose latérale amyotrophique est l'une de ces maladies, avec des mutations causatives identifiées sur neuf gènes différents<sup>76</sup>. Le portrait se complexifie grandement lorsque l'on tente d'élucider la composante génétique des maladies à forte prévalence. Plusieurs hypothèses doivent alors être considérées : un mode de transmission mendélien avec pénétrance incomplète, une composition hétérozygote complexe ou un composé de facteurs de susceptibilités menant à un risque accru de développer le phénotype.

La fréquence élevée du SIME laisse penser que le mécanisme génétique responsable des symptômes de la maladie serait plutôt un mécanisme polygénique, comme le démontrent plusieurs études sur les maladies à prévalence élevée<sup>77-79</sup>. Dans le cas des maladies complexes, il est plus difficile d'identifier des facteurs génétiques en raison de la complexité de la transmission héréditaire qui ne suit pas forcément les lois mendéliennes établies pour les maladies monogéniques. Ainsi, il est donc important de réaliser différents types d'analyses, à la fois au niveau familial, mais également au niveau populationnel, afin de maximiser le pouvoir de détection des variants à effet variable.

Jusqu'à présent, deux approches d'analyses génétiques ont été majoritairement utilisées : les études de liaisons génétiques, et les études d'associations. Dans la section suivante, les deux techniques seront brièvement discutées. Par la suite, un recensement des loci identifiés sera présenté, ainsi qu'une discussion sur la pertinence des travaux précédents sur l'étude ici présentée.

## 2.2 Études de liaison génétique

Les études de liaison génétique (ELG) sont les outils de bases pour démystifier les bases génétiques d'un phénotype partagé par une même famille<sup>80</sup>. Les ELG sont des analyses génétiques basées sur des rapports de recombinaisons génétiques et leur ségrégation dans la famille.

La recombinaison génétique ayant lieu naturellement lors d'une méiose permet de fragmenter le génome en blocs de recombinaisons de tailles variables. Au niveau familial, la distribution de ces blocs de recombinaisons sera différente pour chaque individu et permettra ainsi d'établir des fréquences de recombinaisons à l'échelle du génome. Dans le cas d'une famille, il est possible d'établir l'occurrence génétique d'un phénotype particulier (une maladie ou un trait quantitatif), en établissant une probabilité de coségrégation entre le phénotype et les blocs de recombinaisons. Si une liaison positive est établie, il est alors possible de combiner les limites des blocs de recombinaisons de chaque individu afin de déterminer l'étendue génétique du locus ainsi identifié. L'échelle utilisée pour l'étendue des loci est le centimorgan (cM). 1 cM représente ainsi une probabilité de 1 % que les deux allèles observés soient séparés par une recombinaison.

Afin de déterminer la présence significative de liaisons génétiques entre un phénotype et une région génétique, les généticiens utilisent un score appelé LOD-score (Logarithme des probabilités). Le LOD-score compare la liaison observée avec la probabilité que celle-ci soit un événement aléatoire et permet donc d'obtenir une mesure de la certitude de l'observation sur l'échelle logarithmique. Le LOD-score se calcule selon la formule suivante :

$$LOD = \log_{10} \frac{(1 - \theta)^{NR} \times \theta^R}{0.5^{(NR+R)}}$$

Où R et NR représentent respectivement le nombre de recombinant et de non-recombinant parmi la progéniture et  $\theta$  représente le taux de recombinaison. Il est généralement accepté qu'un LOD-score au-dessus de 3 est un indice suggestif pour une liaison génétique (un LOD-score au-dessus de 5 est généralement un indice concluant pour une liaison génétique)<sup>81</sup>.

Il existe deux modes principaux pour les études de liaisons : le mode paramétrique et le mode non paramétrique. Le premier mode assume que certaines informations sont connues sur le mode de transmission héréditaire du trait observé. Ainsi, il est possible de diriger les résultats selon la

transmission (par exemple : Autosomale récessive, Autosomale Dominante, sexuelle récessive) en ajustant des paramètres comme le taux de phénotopies ou la fréquence attendue de la mutation. Le mode non paramétrique, à l'opposé, n'assume aucun mode de transmission. De ce fait, il ne souffre pas d'un biais si le mode de transmission assumé diffère du mode de transmission réel. Par contre, le mode non paramétrique n'est pas aussi sensible que le mode paramétrique et est plus à risque de ne pas détecter les liaisons plus faibles.

La méthode la plus couramment utilisée pour mettre au point une étude de liaison est de procéder à une étude d'exploration à l'échelle du génome entier en utilisant des marqueurs connus et distribués sur tous les chromosomes; ces marqueurs peuvent être des microsatellites ou des polymorphismes à nucléotide simple (SNP)<sup>80</sup>. Il est important que la densité des marqueurs choisis soit suffisante afin de ne laisser aucune région chromosomique non examinée et il faut également tenir compte de la nature des marqueurs. Ainsi, vu le caractère hétérogène et diversifié des microsatellites, ils seront requis dans un nombre plus petit que celui des marqueurs SNP qui eux, sont à caractère binaire. Par contre, comme la fréquence des marqueurs SNP est beaucoup plus grande, ils permettent une meilleure résolution du profil de recombinaison et ainsi, les marqueurs microsatellites sont de moins en moins utilisés pour les études à large échelle génomique.

Les phases subséquentes sont des phases d'affinement. On utilise des combinaisons de marqueurs ayant des résolutions de plus en plus fines, en se basant sur les résultats positifs préalablement obtenus lors de la phase d'exploration. Lorsque le bloc de recombinaison est clairement défini (locus), on tente par d'autres approches (séquençages, fluorescences, FISH, etc....) d'identifier le variant causal à l'intérieur du locus identifié.

### 2.2.1 Loci identifiés par des études de liaisons pour le SIME

Au niveau du SIME, plusieurs études de liaisons ont été réalisées sur des familles de taille moyenne (3 à 5 personnes atteintes du SIME) à très grande (plus de 10 personnes atteintes du SIME). Jusqu'à présent, plusieurs loci ont été identifiés dans plusieurs populations, mais aucune étude n'a pu faire le lien direct entre un locus et un variant causatif de la maladie (mutation, délétion, etc...). Dans la section suivante, une revue exhaustive des loci identifiés par études de liaisons pour le SIME sera réalisée afin de mieux comprendre les acquis qui serviront de base à l'établissement du présent projet. **(Voir Tableau 1).**

Loci	Position Chromosomique	Mode	Transmission	Lod-Score Maximal	Population	Réplication	Refs
RLS1	12q22-23	Para.	Récessive	3.59	Canadienne-Française	-Canadienne-Française -Allemande	[70, 82-83]
RLS2	14q13-21	Para	Dominant	3.23	Italienne	-Canadienne-Française	[73,84]
RLS3	9p24-23	Non-Para	-	3.91	Américaine	Allemande	[75, 85-87]
RLS3*	9p21.3-21.1	Para	Dominant	3.60	Allemande	Canadienne-Française*	[88]
RLS4	2q	Non-Para	-	5.5	Italienne (Tyrol du Sud)	-	[90]
RLS5	20p13	Para	Pseudo-Dominant	3.86	Canadienne-Française	-	[72]
RLS6	19p13	Para	Dominant	2.61**	Italienne	-	[74]
RLS7	16p12.1	Para	Pseudo-Dominant	3.18	Canadienne-Française	-	[71]

Tableau 1 : Tableau récapitulatif des associations génétiques identifiées par étude de liaisons pour le SIME. Aucune de ces investigations n'a mené à l'identification d'un variant causatif. Deux modes d'analyses ont été utilisés : Para pour paramétrique et Non-Para pour non-paramétrique. \* Données non publiées \*\*Cette valeur n'est pas significative selon les conventions génétiques

### 2.2.1.1 RLS1 (chromosome 12)

Le premier locus identifié pour le SIME (RLS1) est sur le chromosome 12q (position chromosomique (Mbp) : 92,7 – 102,7)<sup>70</sup>. L'étude a été réalisée sur une famille d'origine canadienne-française en utilisant un modèle de transmission autosomal récessif et une haute fréquence de l'allèle prédisposant à la maladie. Les études de liaisons ont donné un LOD-score de 3.59 dans une région du chromosome 12 qui contient plusieurs gènes candidats intéressants. Cependant, aucun variant n'a pu être identifié lors des analyses de ségrégations des génomes de la famille d'origine. Ce locus a été répliqué dans la population d'origine, à nouveau par le même groupe<sup>82</sup>, et également dans des familles bavaroises<sup>83</sup>.

Une étude d'association locale a été effectuée au niveau du locus RLS1<sup>160</sup>. Plusieurs SNPs ont été considérés lors de la phase initiale et la phase de réplication. Les auteurs ont identifiés plusieurs associations positives entre des SNPs et le SIME. Tous les signaux sont situés à l'intérieur du gène

*NOS1* qui encode une synthétase de l'acide nitrique. Bien que ce gène soit situé à l'intérieur du locus RLS1, la méthodologie utilisée par les auteurs de l'étude pour la sélection des SNPs ainsi que pour l'établissement de la phase de réplification laisse matière à réflexion. En effet, les corrections pour essais multiples ne sont pas cohérentes entre les différentes phases. De plus, le séquençage du gène n'a pas résulté en une identification concluante de variation génétique. L'étude, bien qu'intéressante, devra être répliquée plus solidement avec que le gène *NOS1* puisse être sérieusement considéré dans la pathogénèse du SIME.

#### **2.2.1.2 RLS2 (Chromosome 14)**

Le deuxième locus du SIME (RLS2) a été identifié dans une famille italienne, qui présentait à la fois des symptômes de SIME et de MPJS<sup>73</sup>, sur le chromosome 14 (position chromosomique (Mbp) : 33,5 – 46,2). L'analyse a été réalisée en utilisant un modèle autosomal dominant et un modèle de dominance approprié pour les maladies monogéniques. Le LOD-score maximal rapporté est de 3.23. Les résultats ont ensuite été confirmés en utilisant des études de liaisons non paramétriques. Cependant, ce locus n'a jamais été lié à aucun variant causatif, mais a depuis été répliqué dans la population canadienne-française<sup>84</sup>.

#### **2.2.1.3 RLS3 (Chromosome 9)**

Le troisième locus du SIME (RLS3) sur le chromosome 9 (position chromosomique (Mbp) : 0,5 – 10,5) est celui qui a généré le plus de controverse. En effet, suite à la publication par Chen et coll. de la première identification de ce locus dans une famille américaine<sup>75</sup>, la méthodologie utilisée par le groupe a été largement critiquée<sup>85</sup>, en partie à cause de l'utilisation d'une analyse non paramétrique testée avec les critères de significances d'un mode paramétrique. En effet, le LOD-score maximal reporté par Chen et coll. est de 3.91. Or, selon une échelle de correspondance entre les liaisons paramétriques et non paramétriques, un LOD-score d'au moins 4.08 serait nécessaire pour établir une liaison suggestive<sup>86</sup>. Cependant, ce locus a été répliqué dans une famille allemande, dans une étude qui a identifié une liaison suggestive à quelques centimorgans de la liaison originale<sup>87</sup>. Subséquemment, une nouvelle répllication indépendante a été réalisée, à nouveau dans une famille allemande. Cette fois, le locus identifié se trouve à plus de dix centimorgans du locus original, ce qui pousse l'auteur à suggérer que ce locus serait peut-être complètement distinct du RLS3 (il est donc noté RLS3\*)<sup>88</sup>. De plus, une étude non publiée montre que ce locus serait également lié à quelques petites familles dans la population canadienne-française.

En dehors des analyses de liaisons, RLS3 est le seul locus à avoir été partiellement répliqué en utilisant une autre méthode d'analyse génétique. En effet, un gène (*PTPRD*) à l'intérieur du locus RLS3 original a également été impliqué dans le SIME par une étude d'association<sup>89</sup>, dont nous traiterons dans une section suivante.

#### **2.2.1.4 RLS4 (chromosome 2)**

Le quatrième locus (RLS4) identifié par étude de liaisons se trouve sur le chromosome 2q (position chromosomique (Mbp) : 192,2 – 213,2). Il a été identifié dans une famille originaire du Tyrol du Sud. La liaison a tout d'abord été identifiée dans une étude à échelle génomique en utilisant un mode non paramétrique, et a ensuite été affinée, toujours en utilisant le mode non paramétrique<sup>90</sup>. Cette famille sud-tyrolienne est encore aujourd'hui la famille où l'association entre le SIME et un trait génétique est la plus forte avec un LOD-score de 5.5. Cependant, aucun variant génétique n'a été identifié, et ce locus n'a jamais été répliqué.

#### **2.2.1.5 RLS5 (chromosome 20)**

Le cinquième locus du SIME (RLS5) est également le deuxième locus identifié sur le chromosome 20 (position chromosomique (Mbp) : 0 – 5,0) dans la population canadienne-française<sup>72</sup>. Il a été identifié en utilisant un mode paramétrique de pseudodominance autosomale (pénétrance de 70 %) produisant un LOD-score maximal de 3.86. À ce jour, aucune répllication n'a été effectuée, et aucun variant d'intérêt n'a été identifié.

#### **2.2.1.6 RLS6 (chromosome 19)**

Le sixième locus du SIME (RLS 6) a été suggéré dans une famille italienne, sur le chromosome 19 (position chromosomique (Mbp) : 16,6 – 18,2)<sup>74</sup>. À nouveau, le mode utilisé était un mode paramétrique autosomal dominant avec un LOD-score maximal de 2.61. Aucune répllication n'a été reportée, et aucun variant génétique n'a été identifié.

#### **2.2.1.7 RLS7 (chromosome 16)**

Le septième locus du SIME (RLS7) est le troisième locus identifié sur le chromosome 16 (position chromosomique (Mbp) : 22,5 – 25,5) dans la population canadienne-française<sup>71</sup>. Un mode paramétrique de pseudodominance a été utilisé pour calculer le LOD-score maximal de 3.18. À nouveau, aucune répllication n'a été reportée, et aucun variant n'a été positivement lié.

### 2.2.3 Perspectives

En parcourant la liste des loci identifiés par études de liaisons pour le SIME, une première constatation peut être établie : il n'existe pas de standard pour les études de liaisons sur des maladies complexes. Dans le cas présent, lorsque l'on regarde les modes utilisés, deux études ont utilisé le mode non paramétrique (RLS2 et RLS4) tandis que les autres ont utilisé le mode paramétrique (RLS1 : autosomal récessif, RLS3 et RLS6 : autosomal dominant, RLS5 et RLS7 : autosomal pseudodominant). Lorsque l'on compare les paramètres utilisés par chaque étude (phénotypes, fréquences alléliques, etc....), il s'avère qu'aucune étude n'a utilisé les mêmes paramètres. Il est donc difficile de confirmer la justesse des résultats de ces études, surtout qu'à l'exception de RLS4, tous les résultats sont marginalement significatifs, voire pas du tout significatifs (RLS6).

La complexité génétique n'est pas la seule explication du succès limité des études de liaisons pour le SIME. La difficulté d'établir un diagnostic clair influence grandement les résultats. En effet, comme les symptômes du SIME peuvent parfois être relativement légers, il peut être difficile de faire la distinction entre une personne souffrant réellement du SIME et une personne non atteinte. L'imprécision relative au diagnostic influence grandement les résultats finaux. En effet, à l'insu des études populationnelles, un cas de mauvais diagnostic peut influencer radicalement le LOD-score et, le cas échéant, faire la différence entre une forte liaison génétique et une liaison génétique absente. Afin de pallier à cette difficulté, il serait intéressant de refaire les analyses déjà réalisées, mais cette fois, en utilisant un trait quantitatif<sup>91</sup> facilement mesurable (par exemple, la fréquence des symptômes, ou l'indice du nombre de MPJS). En tentant de lier des traits génétiques à des symptômes précis plutôt qu'à un ensemble de symptômes (maladie), il serait possiblement plus facile de déterminer les origines moléculaires de chacun des symptômes et d'ensuite, retracer les mécanismes de la maladie à l'aide de chacune de ses composantes.

Il serait également conseillé de revoir le mode utilisé pour la transmission du SIME. En effet, il a longtemps été assumé que le SIME suivait un mode de transmission autosomal-dominant, mais suite aux récentes découvertes dans le domaine des études d'association, cette assumption n'est plus juste. En effet, dans le cadre d'une maladie complexe, il est à prévoir qu'un grand nombre d'acteurs géniques interviendront dans le processus de la maladie. À cela, il faut également ajouter tous les facteurs environnementaux (c.-à-d. non génétiques) qui peuvent influencer chacun des mécanismes impliqués (fer, dopamine). Il serait donc intéressant de réanalyser les familles

déjà rapportées, mais avec des paramètres plus proches de la réalité du SIME (pénétrance réduite, phénotypes élevés). Cependant, au final, ces analyses seraient encore biaisées par des paramètres qui ne sont peut-être pas justifiés. Le mode le plus approprié pour le SIME est sans doute le mode non paramétrique. L'ennui est que ce mode n'utilise que les membres affectés de la famille, et ainsi, le pouvoir de détection en est diminué.

### 2.3 Les études d'associations

D'une manière très simplifiée, les études d'associations représentent l'équivalent des études de liaisons mais les études d'association sont basées sur une population de patients plutôt que sur des familles. Comme il n'est pas possible d'observer la ségrégation d'un trait génétique parmi les membres affectés, on utilise plutôt la fréquence allélique pour tenter d'établir s'il y a une différence entre un groupe de patients et un groupe de contrôles. Le grand essor des études d'associations, particulièrement celui des études d'associations à l'échelle génomique (EAG) est très récent. En effet, avant la complétion du « projet génome humain » en 2003, les connaissances des variations génétiques chez l'humain étaient trop limitées pour établir une carte génétique des SNPs à l'échelle génomique. Aujourd'hui, plus de 400 EAG ont été réalisées et près de 12,000 associations positives ont été réalisées (certains gènes sont associés à plusieurs maladies)<sup>92</sup>.

En ce qui concerne les EAG, même si le principe est similaire, on ne parle plus de bloc de recombinaison comme pour les études de liaisons. À l'échelle populationnelle, on parle plutôt de déséquilibre de liaison (DL). Si des combinaisons d'allèles se produisent plus souvent que si ces combinaisons étaient dictées par le hasard, on parle de déséquilibre de liaison. Lorsqu'on a commencé à s'intéresser plus sérieusement aux variations génétiques au niveau populationnel, on a réalisé que le génome était fragmenté en bloc de DL de tailles variables. Un bloc de DL est un segment chromosomique continu qui est déterminé en comparant les haplotypes de plusieurs membres d'une même population, comme le fait le projet HapMap<sup>93</sup>. Avec une connaissance précise des blocs de DL, il est possible de déterminer une liste de SNPs « étiquette » qui permet de réduire le nombre de SNP à étudier d'environ dix millions à moins d'un million. Plusieurs compagnies ont mis en place des plateformes qui permettent aujourd'hui de génotyper un demi-million de SNP en une seule expérimentation.

À l'instar des études de liaisons, une association positive entre un marqueur SNP et un trait phénotypique n'est pas nécessairement concluante de l'identification d'un variant génétique causatif. Il n'est pas exclu que le SNP ainsi identifié soit le variant causal recherché mais, dans la plupart des cas, il faut envisager que le variant causal soit en haut DL avec le marqueur SNP mais qu'il devra être identifié par séquençage ou autre technique.

La plus grande limitation des études d'associations est sans doute le grand nombre d'individus non-apparentés requis pour atteindre un niveau de détection acceptable. De ce fait, il est très difficile de réaliser une EAG pour des maladies à faible prévalence. En ce qui concerne les maladies complexes à haute prévalence, elles deviennent d'excellentes candidates pour les études d'associations.

### 2.3.1 Les associations positives faites par EAG pour le SIME

Bien que la prévalence du SIME soit très élevée, le nombre d'équipes de recherches traitant de la génétique du SIME est très limité, d'où le nombre relativement peu élevé d'EAG qui ont été conduites dans le domaine. À ce jour, deux études originales ont été effectuées sur des patients du SIME. La première a été conduite par une équipe de Munich en utilisant 401 cas familiaux du SIME<sup>94</sup> de provenance allemande et autrichienne. Les résultats ont par la suite été répliqués dans 251 patients SIME canadiens-français. Lors de cette étude, quatre gènes ont été liés de façon positive au SIME, Meis homeobox 1 (*MEIS1*), BTB-POZ domain containing 9 (*BTBD9*) et le complexe formé des deux gènes mitogen-activated protein kinase kinase 5 (*MAP2K5*) et Lbxcor1 homolog (*LBXCOR1*), sur les chromosomes 2, 6 et 15, respectivement.

Une deuxième étude réalisée auprès de 306 patients SIME d'origine islandaise a également identifié le signal sur le chromosome 6 (*BTBD9*)<sup>49</sup>. Tous les patients présentaient également des symptômes de MPJS pathologiques. Les résultats ont pu être répliqués dans une population des États-Unis. Cependant, aucun signal significatif n'a été identifié pour les deux autres loci identifiés lors de la première étude, soit le chromosome 2 (*MEIS1*) et chromosome 15 (*MAP2K5/LBXCOR1*).

Une troisième étude, réalisée par le groupe à l'origine de la première étude, a pu identifier un quatrième locus lié au SIME en testant les signaux suggestifs de leur première étude chez une cohorte plus nombreuse de patients SIME (n = 628)<sup>89</sup>. La position chromosomique de ce locus coïncide avec un gène : protein tyrosine phosphatase, receptor type, D (*PTPRD*). Pour la première fois, le locus identifié sur le chromosome 9 (*PTPRD*) se situe dans une même proximité qu'une

région (RLS3) précédemment identifiée par études de liaisons chez une famille<sup>75</sup>. Le **tableau 2** présente un récapitulatif des loci identifiés par études d'association.

	Chromosome	Position (Mbp)	FAM* Patients	FAM* Contrôles	Risque Relatif	Valeur P	Population**
<b>MEIS1</b>	2	66,6	36,7%	24,1%	1.74	9.08x10 <sup>-23</sup>	Allemande
<b>BTBD9</b>	6	38,5	16,5%	23,8%	1.67	9.44x10 <sup>-13</sup>	Allemande + Islandaise
<b>MAP2K5/ LBXCOR1</b>	15	65,8 – 65,9	23,9%	31,1%	1.51	6.09x10 <sup>-17</sup>	Allemande
<b>PTPRD</b>	9	8,8 – 9,2	21.6 %	16,4%	1,4	5,81x10 <sup>-9</sup>	Allemande

**Tableau 2** Tableau récapitulatif des loci identifiés par études d'association

\* FAM = Fréquence de l'allèle mineur

\*\* Tous les loci identifiés dans la population allemande ont été répliqués dans la population canadienne-française

### 2.3.1.1 MEIS1 (chromosome 2)

*MEIS1* est un gène faisant partie de la famille des homéogènes ou gènes à « Homéobox », un ensemble de gènes connu pour jouer un rôle important dans le développement de l'embryon et la néoplasie. *MEIS1* encode une protéine faisant partie de la famille TALE (Extension en boucle de Trois Acides aminés). On connaît depuis quelques années déjà l'implication de *MEIS1* dans le mécanisme de la leucémie myéloïde par le biais de son interaction avec pre-B-cell leukemia homeobox 1 (*PBX1*)<sup>95</sup>. Tout récemment, il a été démontré que dans un système sain, l'expression de *MEIS1* est modulée par homeobox A9 (*HOXA9*)<sup>96</sup> afin d'influencer l'hématopoïèse<sup>96</sup>. Hormis son rôle confirmé dans la leucémie, *MEIS1* est également un acteur important du développement embryonnaire du système nerveux. En effet, il a récemment été démontré que *MEIS1* ainsi que *PBX1*, un autre gène de la même famille, influencent l'expression de « sex determining region Y-box 3 (*SOX3*) », un gène important pour la régulation du développement du cerveau embryonnaire<sup>97</sup>.

Notre équipe a récemment mené une étude spécifique au SIME pour évaluer le rôle de *MEIS1* dans la maladie. En prenant pour base les signaux observés lors de l'étude d'associations originale<sup>94</sup>, il a été possible d'identifier d'autres variations également liées au SIME. Par la suite, en examinant l'expression du gène dans des lymphoblastes et des cerveaux de patients SIME

(provenant de la banque de cerveau de la Restless Legs Foundation), il a été possible d'identifier une baisse marquée de l'expression de *MEIS1* pour les patients porteurs de l'haplotype à risque<sup>98</sup>. Cette observation, ajoutée aux résultats initiaux, solidifie l'hypothèse d'un rôle du gène *MEIS1* dans le SIME.

### 2.3.1.2 *BTBD9* (chromosome 6)

Outre le fait qu'il contient un domaine POZ (domaine d'interaction protéine-protéine à doigts de zinc), on dispose de très peu d'informations sur les fonctions et implications de *BTBD9*. De par ses structures et domaines, il est possible d'inférer que *BTBD9* est un gène potentiellement impliqué dans la régulation de l'expression génique par le biais de la conformation de la chromatine<sup>99,100</sup>.

Si les signaux obtenus lors des deux EAG pour *BTBD9* sont convaincants, quelques doutes subsistent quant à l'association entre *BTBD9* et le SIME. En effet, pour l'étude sur la population islandaise, il a été observé que le signal de *BTBD9* était encore plus important pour les patients souffrants de MPJS que pour les patients souffrants de SIME<sup>49</sup>. Comme il existe une forte association entre les deux maladies, il est fort possible que le signal identifié par la première étude soit fortement produit par des patients SIME souffrant également de MPJS. De plus, pour renforcer cette hypothèse, une association entre *BTBD9* et des patients atteints du Syndrome de Gilles de la Tourette (SGT) a également été rapportée, avec une association plus robuste que pour le SIME<sup>101</sup>. Les études sur le sujet ne sont pas assez nombreuses pour pouvoir conclure à une association entre les MPJS et le SGT mais étant donné la présence de tics moteurs dans le SGT, il serait fort logique de conclure à une prévalence plus élevée des MPJS dans le SGT. De ce fait, il est très difficile de conclure à l'implication réelle de *BTBD9* dans le SIME et ainsi, ce gène est très peu étudié dans le cadre de la maladie.

### 2.3.1.3 *MAP2K5/LBXCOR1* (chromosome 15)

Le troisième site d'association est un signal qui pourrait correspondre à deux gènes, soit *MAP2K5* et *LBXCOR1* qui sont situés à environ 20 kb de distances. *MAP2K5* est une protéine kinase-kinase activée par une activité mitogène double. Elle fait partie de la famille des MAP kinases-kinases (MAP2K), composante du mécanisme plus général des MAP kinases (MAPK). Les MAP2K sont des régulateurs des MAPK, kinases impliquées dans l'action de la majorité des oncogènes extracellulaires<sup>102</sup>. Plus spécifiquement, *MAP2K5* interagit et active directement *MAPK7/ERK5*<sup>103</sup>, jouant un rôle important dans la cascade de signalisation des facteurs de croissance stimulant la

prolifération cellulaire et la différenciation des cellules musculaires. Cette cascade serait impliquée dans plusieurs mécanismes cancéreux<sup>104</sup>.

LBXCOR1 est un corépresseur du facteur de transcription *LBX1*. Chez la souris, ce facteur de transcription est un régulateur de la migration cellulaire requise pour l'acquisition de l'identité dorsale des muscles supérieurs des membres<sup>105</sup>. LBXCOR1 est également impliqué dans la régulation négative de la voie de signalisation de plusieurs gènes du développement, par exemple « Bone Morphogenetic Protein (BMP)<sup>106</sup>, voie très importante pour l'organogénèse du système nerveux central ainsi que le développement postnatal des os<sup>107,108</sup>.

#### 2.3.1.4 *PTPRD* (chromosome 9)

*PTPRD* est un récepteur de type « tyrosine phosphatase delta », membre de la famille des protéines tyrosine-phosphatases (PTP). Cette famille de protéines de signalisation est connue pour ses diverses activités de régulation, notamment dans la croissance cellulaire, la différenciation, le cycle mitotique et la transformation oncogénique<sup>109</sup>. De plus, des études sur des gènes similaires chez le poulet et la drosophile suggèrent que les PTP pourraient également jouer un rôle important dans la croissance des neurites et la régulation du processus d'orientation des axones<sup>110</sup>.

Son rôle dans les oncogènes en tant que suppresseur de tumeur a récemment été l'objet de plusieurs études<sup>111,112</sup>. De plus, il a été démontré que des délétions du gène *PTPRD* sont une cause fréquente de glioblastome multiformes, le type le plus commun et le plus agressif de tumeurs du cerveau<sup>113</sup>. Des délétions du même type ont également été observées dans les carcinomes des cellules squameuses de la tête et du cou, ainsi que dans les cancers du poumon.

## 2.4 Perspectives génétiques

Depuis la découverte du premier locus lié à une famille<sup>70</sup> jusqu'à l'identification des quatre gènes candidats par les études d'associations<sup>49,89,94</sup>, le domaine de la génétique du SIME a été en constante évolution. Cependant, depuis pratiquement deux ans, aucune avancée majeure n'a été réalisée. Quelques études de répliques ont été effectuées, mais aucune percée génétique n'a apporté d'informations nouvelles. Il semble qu'aussi bien au niveau des études de liaisons que des études d'associations, une limite ait été atteinte.

Le pouvoir de détection des études d'associations a toujours été étroitement lié avec la clarté des phénotypes. Or, dans le cas du SIME, comme il n'existe aucun marqueur objectif de l'intensité des symptômes (c.-à-d. marqueurs biochimiques ou anatomiques), la seule option restante est un entretien direct ou téléphonique avec les patients. Lorsqu'on lie la variabilité de la tolérance à la douleur avec les sensations imprécises causées par les symptômes du SIME, il est très difficile d'établir un seuil objectif pour la quantification des phénotypes. Ainsi, comme les études de liaisons génétiques sont très sensibles aux erreurs de phénotypage en raison d'un échantillon beaucoup plus faible, il n'est pas facile d'obtenir un réel signal de liaison. De plus, la probabilité est forte pour que plusieurs locus précédemment identifiés soient des faux positifs dus à des mésinterprétations phénotypiques.

Les études d'associations sont beaucoup moins sensibles au niveau de la clarté du diagnostic, mais sont en revanche directement dépendantes de la taille de l'échantillon. Les premières études d'associations génomiques publiées entre 2007 et 2008 étaient généralement réalisées sur moins d'un millier de patients et autant de contrôles, comme c'est le cas pour le SIME. Cependant, les généticiens ont réalisé la limite des études se basant sur des grandes fréquences alléliques (fréquence de l'allèle mineur (FAM) > 5 %). En effet, si les premières études ont permis l'identification des variations génétiques à effet modéré, on a vite constaté que le pouvoir d'identification des variations génétiques à faible effet était très limité. Afin d'augmenter la détection de ces derniers, il a fallu augmenter exponentiellement la taille des échantillons, passant de quelques milliers à quelques dizaines de milliers et aujourd'hui à plus de cent mille échantillons pour une seule étude. Pour de telles études, il est facile de se baser sur des traits quantitatifs présents chez tous les individus (par exemple : la taille des humains<sup>114</sup>) ou sur des maladies fréquemment étudiées comme le cancer. Cependant, pour des cas comme le SIME où le nombre de groupes recrutant des patients est très limité, il est beaucoup plus difficile d'identifier les variations génétiques à faible risque relatif.

Dans le cas du SIME, la démarche qui permet d'identifier des variations à effet faible n'a pas encore été menée. En effet, les efforts requis pour collecter un nombre aussi important de patients ne justifient peut-être pas l'apport de l'identification de telles variations. D'autant plus que certains généticiens mettent aujourd'hui en doute les calculs actuels du facteur de l'hérédité des maladies complexes, et par le fait même, le calcul des effets génétiques des variations déjà identifiées<sup>115</sup>. Dans le cas du SIME, la question n'est pas de savoir si l'identification des variations

à faible effet est pertinente, mais plutôt de savoir si les nouvelles technologies génomiques permettraient une identification plus facile et moins coûteuse de ces variations. De plus, il est également possible de penser que ces nouvelles technologies pourraient également révéler une partie du portrait du mécanisme génétique de la maladie qui est jusqu'à présent resté inconnu. Ces technologies se divisent en deux domaines bien distincts : le séquençage de nouvelle génération et les études de déséquilibres génomiques.

#### 2.4.1 Le séquençage de nouvelle génération (SNG)

Le principe du séquençage de l'ADN est un principe qui remonte à 1977 lorsque Sanger et son équipe ont mis au point la technique du séquençage didéoxynucléotidique, qui a ensuite été renommée séquençage « Sanger »<sup>116</sup>. Cette méthode a été le standard universel pour le séquençage durant trente ans. L'apogée de cette technique a évidemment été le projet « Génome » qui avait pour but de séquencer le génome humain en entier. Ce projet, qui s'est étendu sur plus de quinze ans, est estimé à un coût global de 3 milliards de dollars américains<sup>159</sup>. Les avancées de ce projet ont été gigantesques avec un accès universel pour tous les scientifiques à la première version complète du génome humain<sup>117</sup>. Plusieurs projets ont été engendrés par ce projet original, chacun d'eux ayant pour but de mieux comprendre les variations génétiques présentes chez l'humain. Peu à peu, ces projets ont permis de mieux comprendre la fonction et la distribution de ces variations génétiques. Parallèlement, plusieurs percées technologiques ont permis la mise au point de nouvelles technologies de séquençage, qui permettent de séquencer des génomes complets dans une période de temps très courte. Ce séquençage de nouvelle génération a également permis de réduire radicalement le coût de séquençage d'un génome individuel. De plusieurs milliards en 2003, le coût est passé à 500 000 \$ en 2007 et en 2009, le même exercice pouvait être réalisé pour 50 000 \$<sup>118</sup>. On estime que le séquençage de nouvelle génération fera son entrée dans la médecine génétique et dans le diagnostic de masse lorsque le coût d'un séquençage individuel sera de l'ordre de 1 000 dollars. Bien que nous soyons encore loin de cette barre psychologique, il est permis de croire que cet objectif sera atteint au cours des cinq prochaines années.

À l'insu des trente dernières années où une seule technologie de séquençage (Sanger) était prédominante, il existe aujourd'hui plusieurs techniques de SNG. On peut les grouper en deux groupes : les techniques requérant une amplification soit par PCR d'émulsion ou par amplification d'état solide, et les techniques qui n'en requièrent pas. La technologie permettant de procéder à

un SNG sans amplification n'est pas encore tout à fait au point et ne permet pas encore d'obtenir une couverture satisfaisante du génome. Le SNG avec amplification est une technique plus mature qui est très largement utilisée. Bien que trois systèmes se partagent le marché (Illumina Solexa, ABI SOLiD et Roche 454), le procédé est similaire pour toutes les technologies<sup>118</sup> :

- Préparation de la bibliothèque : l'ADN génomique est aléatoirement fragmenté et transformé en millions de petits fragments. Puis, chaque fragment, après avoir été lié avec un modèle spécifique, est attaché à une plateforme ou à un support solide.
- Séquençage et imagerie : des millions de réactions de séquençage ont lieu en parallèle. Un système d'imagerie très puissant est utilisé afin de pouvoir superviser chacune des analyses.
- Alignement et assemblage : Toutes les images capturées par le système sont analysées de manière à produire les séquences correspondant aux fragments de la bibliothèque. Par la suite, à l'aide d'algorithmes spécifiques, on peut aligner ces fragments sur un génome connu, ou produire l'assemblage d'un génome de novo.

Avec l'avènement du SNG, de nouvelles méthodes ont été développées pour permettre des applications plus spécifiques. En effet, bien que le coût de séquençage d'un génome entier ait dramatiquement baissé, il arrive bien souvent qu'il soit trop élevé pour la majorité des études. Ainsi, deux nouvelles applications du SNG peuvent être utilisées pour réduire les coûts :

- Séquençage de l'exome entier : La proportion de l'ADN qui sera directement transcrit pour servir à la production de protéine est minime (moins de 3%). Il est donc possible de capturer toutes les parties codantes du génome et de les séquencer, réduisant de beaucoup la couverture nécessaire pour un séquençage.
- Capture de régions : À l'instar du séquençage de l'exome entier, la capture de régions permet de se concentrer sur une région précise du génome (par exemple : une région identifiée lors d'une étude de liaison) et de diminuer grandement la région à séquencer.

En ce qui concerne le SIME, aucune étude utilisant le SNG n'a été rapportée. Cependant, il y a fort à parier que ce n'est qu'une question de temps avant que certaines équipes se tournent vers le SNG pour tenter d'élucider une nouvelle partie du mécanisme génétique de la maladie.

## 2.5 Les études de déséquilibre génomique

Les déséquilibres génomiques (Copy Number Variation, CNVs) sont des variations génétiques qui ont longtemps été négligées. En effet, les CNVs sont très difficiles à valider expérimentalement, car ils échappent généralement aux méthodes de séquençage (incluant le SNG). Ainsi, pour les détecter, il faut procéder à des études indépendantes qui, jusqu'à tout récemment, étaient très onéreuses. Le domaine a vraiment pris son essor avec l'avènement des biopuces génomiques, l'une

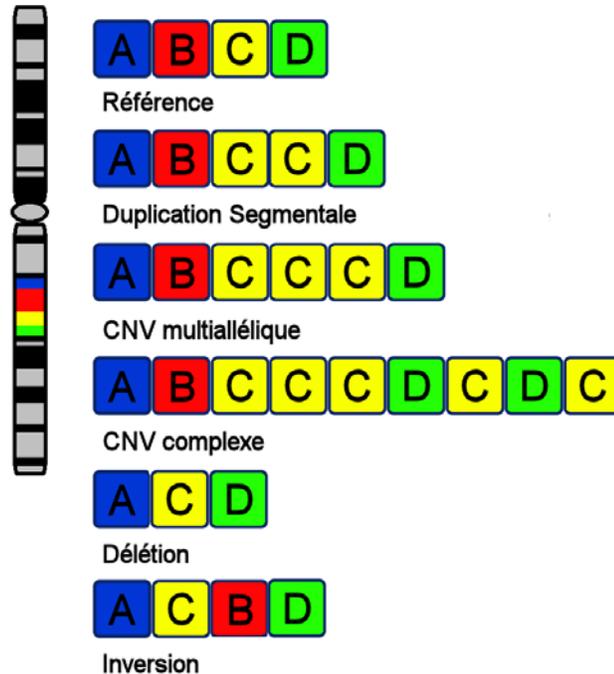


Figure 1 Illustration des différents types de déséquilibres génomiques (CNVs). Tous les types incluent un changement direct du nombre de copies, à l'exception de l'inversion qui ne présente aucune différence. De ce fait, les inversions sont les réarrangements les plus difficiles à identifier car aucune différence du nombre de copie n'est observée lors des analyses.

des méthodes utilisées lors de la présente étude. Comme le rôle des CNVs dans le mécanisme du SIME est le sujet principal de cet ouvrage, la théorie entourant ce type de variation sera traitée plus exhaustivement dans les sections suivantes.

### 2.5.1 La nature des CNVs

Un CNV est un segment d'ADN dans lequel des différences du nombre de copies ont été observées lorsque l'on compare un sujet avec au moins deux autres génomes de référence. Pour être appelée CNV, une variation génétique doit être d'au moins 1 Kb et peut s'étendre jusqu'à plusieurs mégabases (Mbp)<sup>119</sup>. Les organismes diploïdes comme l'être humain possèdent deux copies de chaque chromosome. Ainsi, le nombre copie normal pour chaque élément génétique est de deux (une copie pour chaque chromosome). Lorsque ce nombre varie, on peut inférer la présence d'un CNV. Il existe plusieurs types de CNVs (**figure 1**).

À l'instar des autres variations génétiques, les CNVs peuvent être transmis héréditairement par les parents ou bien causés par des mutations sporadiques de novo. Le mécanisme de formation des CNVs n'est pas encore bien connu. Il a récemment été proposé que ce type de variation génétique puisse être causé par la recombinaison allélique non homologue durant la méiose<sup>120</sup>. Un autre mécanisme qui pourrait être responsable de la formation des CNVs est la jonction des extrémités non homologues. Cette hypothèse a récemment été émise lors de l'étude des jonctions de plusieurs CNVs responsables de la maladie de Pelizaeus-Merzbacher<sup>121</sup>.

La complétion du projet « Génome » a révélé que les CNVs, jusqu'alors très peu connus, pourraient être beaucoup plus fréquents qu'on ne le croyait jusqu'alors et malgré ce projet gigantesque il est encore difficile d'estimer l'occurrence définitive des CNVs. La dernière mise à jour de la base de données Database of Genomic Variants enregistre près de 8,500 loci de CNVs<sup>122</sup>. Plusieurs études ont tenté d'apporter des réponses sur la distribution des CNVs chez l'humain. Cependant, ces études se révèlent très différentes les unes des autres et brillent par leurs contrastes plutôt que par leurs similitudes. On réalise rapidement que selon les méthodes expérimentales et analytiques employées, les résultats varient énormément. **Le tableau 3** présente quelques études des CNVs à échelle génomique.

Tableau 3 Comparaison entre diverses études de CNVs à échelle génomique

Étude	Méthode	Échantillons	CNV / individu	Taille CNV moyenne	CNVr	Couverture Génomique
Redon et al. <sup>123</sup>	CGH	270	70	341 Kb	913	
Redon et al. <sup>123</sup>	Biopuce-SNP	270	24	206 KB	980	
Redon et al. <sup>123</sup>	Combiné	270	-	-	1447	12 %
Perry et al. <sup>124</sup>	CGH	30	> 50	10-50 Kb	1153	-
Kidd et Al <sup>125</sup>	Clonage Fosmid	8	> 180	5-10 Kb	1491	> 25 %

La disparité entre les études est marquante, surtout lorsque l'on compare le nombre moyen de CNVs par individu. Il semble que ce paramètre soit très fortement lié au choix de la méthodologie expérimentale utilisée. Selon Redon et coll.<sup>123</sup>, deux des méthodes couramment utilisées (CGH et biopuce-SNP) seraient complémentaires, plutôt que similaires. De plus, il apparaît que la taille des CNVs est directement proportionnelle avec leur observation dans différentes études; plus le CNV est grand, plus le facteur de corrélation est élevé. D'ailleurs, Kidd et coll.<sup>125</sup> dans une autre étude, observe que plus de 50 % des CNVs qu'il observe sont de nouveaux éléments par rapport à ce qui est déjà répertorié dans la littérature. Il faut donc être très prudent lorsqu'on réalise une étude sur l'occurrence des CNVs car l'absence d'un CNV lorsque l'on n'utilise qu'une seule méthode peut ne pas être conclusive sur l'absence de ce même CNV chez l'individu testé.

### 2.5.2 Les méthodes expérimentales pour détecter les CNVs

Il existe plusieurs méthodes disponibles pour détecter les déséquilibres génomiques. Si chacune des méthodes possède des avantages et des inconvénients, certaines méthodes sont largement plus répandues que d'autres. Nous ferons ici une courte description de certaines méthodes pertinentes dans le cadre du présent projet.

#### 2.5.2.1 Les biopuces à SNP

L'utilisation de biopuces à SNP (SNP-chip) pour la détection de CNVs est considérée comme une méthode indirecte. En effet, l'utilité première des SNP-chips est le génotypage à grande échelle des SNPs pour les études d'associations génomiques. Ces puces sont généralement composées de plusieurs centaines de milliers de sondes qui permettent de définir le génotype bien précis d'un SNP. Lors des études de CNVs, on utilise ces mêmes sondes, mais au lieu du génotype, on compare les niveaux d'intensité. On procède à de nombreuses normalisations et il est alors possible d'établir un seuil d'intensité normal pour chaque sonde. Par la suite, en se référant aux valeurs individuelles, il est possible de déterminer des régions génomiques où l'intensité diverge du seuil moyen. Selon certains critères, il est alors possible de déterminer la nature et l'étendue du CNV. Plus récemment, les derniers modèles de SNP-chip incluent un nombre important de sondes spécifiques à la détection de CNV.

Cette méthode est très fréquemment utilisée, en grande partie parce que les SNP-chips ont été très utilisés pour les études d'associations génomiques. Ainsi, on est en mesure de récupérer les données produites lors d'études précédentes et de les interpréter sous un nouvel angle, avec des coûts minimes. Par contre, comme la SNP-chip n'est pas une plateforme spécifique à l'étude des

CNVs, elle s'avère très souvent confuse. En effet, les différents niveaux de normalisation introduisent un biais significatif et augmentent de beaucoup le nombre de faux positifs<sup>1,133,135</sup>. Il existe un nombre important d'algorithmes d'analyse et il est très important de considérer les particularités de chacun avant de procéder à l'analyse.

#### ***2.5.2.2 L'hybridation par génomique comparée***

L'hybridation par génomique comparée (Comparative Genomic Hybridization, CGH) est une méthode cytogénétique qui est spécifique pour l'analyse du nombre de copies dans un échantillon d'ADN. Le principe du CGH est plutôt simple. On procède à l'hybridation d'un échantillon d'ADN du sujet avec un échantillon d'un ADN contrôle (ou avec un mélange de plusieurs ADN contrôles). On utilise par la suite une micropuce contenant plusieurs milliers de sondes où se fixeront les ADN hybridés. En utilisant la fluorescence, il est alors possible de détecter les différences de nombre de copies et ainsi, d'inférer la présence de CNVs.

Le point fort de cette méthode est sa spécificité pour l'analyse de CNVs. Ainsi, contrairement à la SNP-chip, il n'y a pas nécessité d'une normalisation aussi précise. De ce fait, la méthode est à la fois plus sensible et plus fiable (moins de faux positifs). Par contre, il est très difficile de réemployer les résultats de la CGH pour d'autres analyses et ainsi, la méthode est beaucoup plus onéreuse, surtout lorsque l'on considère les nombreuses applications de la SNP-chip.

#### ***2.5.2.3 La détection de CNVs par PCR quantitative***

La détection de CNVs par PCR quantitative est probablement la méthode la plus fiable actuellement existante. En effet, elle permet de vérifier directement le niveau du nombre de copies pour une sonde avec un nombre minimal de normalisations. C'est également une excellente méthode pour définir très précisément les limites du CNV car elle permet d'ajuster progressivement les limites chromosomiques en fonction des résultats obtenus.

Cependant, cette méthode est limitée à des usages de validations. En effet, il ne serait pas envisageable de procéder à une détection exploratrice à l'aide d'un PCR quantitative car les coûts financiers et temporels seraient beaucoup trop importants. Cette méthode n'est donc d'aucune utilité lorsqu'il s'agit de sonder une vaste région du génome, mais devient très intéressante lorsqu'il s'agit de valider certains CNVs observés dans des étapes antérieures.

### 2.5.3 L'effet des CNVs sur les phénotypes

À l'instar des autres types de variations génétiques, les effets que peuvent avoir les CNVs sur les phénotypes sont très variés. Cependant, comme l'occurrence des CNVs semble beaucoup plus élevée que ce que l'on pensait lors de la complétion du projet Génome en 2003<sup>123</sup>, il devient évident que la majorité des CNVs observés dans un génome ont un effet bénin ou nul. Il est donc nécessaire de procéder à plusieurs validations avant de conclure qu'un CNV affecte directement un gène pour causer un phénotype.

Il existe plusieurs exemples de CNVs qui sont fortement associés avec des maladies. L'autisme<sup>126</sup>, le déficit d'attention<sup>127</sup> et la schizophrénie<sup>128</sup> sont de très bons exemples de maladies pouvant être partiellement ou complètement liés à une variation du nombre de copies. Récemment, une étude a pu établir un lien entre des CNVs précédemment observés chez des patients présentant des retards développementaux, et l'obésité morbide<sup>129</sup>. Cependant, l'occurrence de CNVs que l'on peut lier directement à des phénotypes est généralement très faible, et ainsi, responsable d'une minime partie des maladies génétiques.

On peut ainsi s'interroger sur l'effet génétique pouvant être causé par la présence générale d'un CNV, qu'il soit lié à un phénotype ou non. Afin de répondre à cette question, plusieurs études ont tenté de vérifier l'effet de la variation du nombre de copies sur l'expression génique. Une première équipe a tenté d'évaluer l'impact respectif des CNVs et des SNPs sur le niveau d'expression des gènes<sup>130</sup>. Ils ont pu conclure que chaque type de variation pouvait avoir un impact relatif sur l'expression génique, mais ils n'ont pas pu établir une corrélation entre SNPs et CNVs provenant de la même région génomique. Une deuxième étude a permis d'établir un facteur direct entre les CNVs et l'expression génique chez la souris. L'étude a mis en évidence l'impact des CNVs sur les gènes superposés, mais également sur les gènes flanquants. Dans la majorité des cas, l'expression des gènes corrélait également avec le nombre de copies<sup>148</sup>.

Un important consortium de chercheurs (Welcome Trust Case Control Consortium, WTCCC) a tenté de répondre à cette question<sup>131</sup>. Pour ce faire, ils ont utilisé des cohortes de 2000 patients pour six différentes maladies (trouble bipolaire, cancer du sein, cardiopathie coronarienne, hypertension, arthrite rhumatoïde et diabète (type I et II)) et un groupe commun de 3000 contrôles. Pour toutes les maladies, ils ont généré une plateforme d'expérimentation CGH ciblant 3432 CNVs bien définis. En utilisant une étude cas-contrôle, ils ont pu comparer l'occurrence de chacun de ces CNVs pour les six maladies. De leur étude, ils ont pu tirer deux conclusions

majeures. La première est qu'il faut porter un grand soin aux choix de plateformes ainsi qu'aux méthodes de normalisation, les analyses pour l'identification des CNVs étant très sensibles. Leur deuxième conclusion est que l'identification des CNVs est très vulnérable vis-à-vis des artéfacts expérimentaux. Ainsi, plusieurs CNVs peuvent en réalité être des faux positifs. Bien que cette étude du WTCCC a identifié plusieurs CNVs liées avec des maladies, ils ont également observé que ces CNVs correspondent avec des signaux observés en EAG. De ce fait, leur conclusion principale est que les CNVs qui peuvent être identifiés par les plateformes existantes ont peu de chance de contribuer significativement à la génétique des maladies communes s'ils sont situés dans des régions précédemment non-liées aux maladies.

S'il est démontré que les CNVs doivent être considérés comme une source de variation génétique très importante, il faut cependant garder en tête que l'impact réel des CNVs est encore méconnu et c'est par le biais d'études phénotypiques, que nous pourrions mieux comprendre l'effet réel de ce type de variation.

### 3. But de l'étude

L'avènement des études d'associations génomiques a permis une grande avancée au niveau de la génétique des maladies multifactorielles comme le SIME. Ainsi quatre loci ont été identifiés (*MEIS1*, *BTBD9*, *LBXCOR1/MAP2K5*, *PTPRD*)<sup>49,89,94</sup> et l'un de ces locus (*MEIS1*)<sup>98</sup> a été validé en utilisant des approches de biologie fonctionnelle. Cependant, aucun de ces loci ne semble être impliqué dans les ségrégations familiales, à l'exception du locus sur le chromosome 9 (*PTPRD*)<sup>75,87</sup>.

Ainsi, la majorité de la composante héréditaire n'étant pas expliquée par les loci identifiés, il est nécessaire de se mettre à la recherche de nouveaux gènes impliqués dans la pathogenèse de la maladie. Comme les CNVs se sont récemment positionnés comme une potentielle source de variation génétique, **notre postulat de départ est qu'il sera possible d'identifier des CNVs causatifs impliqués dans le mécanisme génétique du SIME**. Afin d'identifier ces variations et de les impliquer dans la maladie, notre étude sera composée de deux objectifs principaux :

- **Objectif #1 : Identification des CNVs en utilisant des données déjà existantes (SNP-chip).** Cette phase aura pour base des données déjà générées par le biais de puces de génotypages<sup>89,94</sup>. Ces données seront analysées avec des algorithmes spécifiques à la détection de CNVs. Une étude cas-contrôle sera établie afin de déterminer l'occurrence spécifique au SIME pour les CNVs.
- **Objectif #2 : Validation de la phase 1 en utilisant une méthode expérimentale indépendante (CGH).** Cette phase a pour but de valider les loci identifiés lors de la première phase. En utilisant une étude cas-contrôle, il sera possible de déterminer définitivement l'implication du locus dans la maladie.

L'identification de loci avec une occurrence différente au niveau des CNVs aurait un impact majeur au niveau du SIME. En effet, cela permettrait l'implication éventuelle de nouveaux gènes avec la maladie. En affinant le mécanisme génétique du SIME, il serait envisageable d'identifier de nouvelles cibles thérapeutiques afin de perfectionner le soulagement des symptômes. Cependant, les impacts de l'identification de CNVs causatifs ne seraient pas

importants seulement pour le domaine de la recherche sur le SIME. En effet, il s'agirait d'un élément supplémentaire quant à l'implication des CNVs dans les phénotypes et c'est tout le domaine de la génétique humaine qui serait bénéficiaire.

Récemment, une étude du WTCCC a conclu qu'il est peu probable que des CNVs communs dans des régions précédemment inconnues puissent contribuer à la génétique des maladies complexes<sup>131</sup>. Nous pensons que notre postulat de base est toujours valable, même à la lumière des observations du WTCCC. La présence de CNVs dans des régions précédemment identifiées par études de liaisons ou par études d'associations permettrait potentiellement l'identification de variations causales pour le SIME. De plus, nous pensons qu'il n'est pas exclu d'identifier des CNVs dans des régions précédemment non associées au SIME. En effet, en consultant la littérature sur le sujet, on observe que beaucoup de CNVs communs qui ont été liés à un phénotype le sont dans le cadre de maladies neurologiques ou psychiatriques, notamment la schizophrénie<sup>128</sup>, l'autisme<sup>126</sup> et le déficit d'attention<sup>127</sup>. L'étude du WTCCC se concentre sur six maladies différentes, mais une seule de ces maladies est considérée comme neurologique ou psychiatrique, soit le trouble bipolaire. Ainsi, comme le SIME est un trouble neurologique, nous pensons qu'il sera possible d'identifier des CNVs à occurrence spécifique à ce syndrome.

## 4. Méthodologie

### 4.1 Recrutement des patients

#### 4.1.1 Patients SIME

Les patients SIME (n=1012, âge moyen 60.2 +/- 12.35, 26 % hommes) ont été recrutés dans des cliniques spécialisées sur le SIME à travers l'Allemagne et l'Autriche. Tous les diagnostics ont été réalisés à la suite d'un entretien individuel. Tous les patients sélectionnés pour l'étude sont de descendance européenne et répondent aux quatre critères standards déterminés par le regroupement international de recherche sur le SIME<sup>7</sup>. Deux types de diagnostics ont été utilisés : SIME familial (au moins un membre de la famille atteint) ou sporadique (aucun autre membre de la famille atteint). Les patients avec un statut familial inconnu ou incertain ont été ajoutés au groupe sporadique afin de préserver le caractère homogène du groupe familial. Tous les patients ont été examinés pour des diagnostics de SIME secondaire (urémie, dialyse ou anémie) et les cas douteux ont été exclus.

#### 4.1.2 Contrôles POPGEN (contrôles externes)

Les contrôles ont été fournis par l'organisme POPGEN, qui recrute des individus provenant du nord de l'Allemagne, pour servir à des fins de contrôles populationnels. Tous les individus POPGEN ont dû remplir un questionnaire exhaustif sur leur santé et leur condition physique. Aucune information spécifique au SIME n'était disponible. Toutes ces informations sont disponibles sur demande pour les scientifiques utilisant les services de POPGEN.

#### 4.1.3 Contrôles internes

Des patients atteints du cancer du sein ont été recrutés par un consortium de cliniciens allemands, après le premier diagnostic. Tous les patients ont été testés négativement pour des mutations sur les gènes connus pouvant affecter le cancer du sein (*BRCA1*, *BRCA2*). Aucune information spécifique au SIME n'était disponible.

#### 4.1.4 Population test

44 individus non apparentés du projet HapMap ont été utilisés à des fins de développement de méthodes<sup>132</sup>. Ces individus provenant de quatre origines différentes (caucasienne, africaine,

chinoise et japonaise) ont été utilisés pour une phase préliminaire de notre étude. Ainsi, les informations phénotypiques n'ont pas été utilisées.

## 4.2 Détection de CNVs à l'aide de bio-puces à SNPs

### 4.2.1 Génotypage sur les bio-puces à SNPs

1,012 patients SIME et 1,164 contrôles (cancer du sein = 688 et POPGEN = 476) ont été génotypés en utilisant la plateforme Affymetrix © GenomeWide Chip 5.0. Cette puce contient 500 658 sondes pour détecter des SNPs. En plus, sur la même puce, on trouve également 420 000 sondes non polymorphiques qui sont utilisées pour la détection de CNVs. Deux algorithmes ont été utilisés pour générer une mesure de contrôle de qualité appelé fréquence d'appel (call rate) : brlmm-p (une extension de brlmm, l'algorithme développé par Affymetrix © pour l'identification des SNPs) et Birdsuite<sup>1</sup>. Tous les échantillons avec une fréquence d'appel inférieure à la valeur standard 98 % ont été exclus de l'étude.

### 4.2.2 Détection des CNVs

Afin d'optimiser la détection des CNVs pour l'étude, trois algorithmes ont été testés dans le cadre d'une étude préliminaire afin de déterminer la solution la plus optimale pour les étapes subséquentes. Les paramètres utilisés lors de l'étude préliminaire ont ensuite été utilisés lors de la phase principale de l'étude.

#### 4.2.2.1 Modèle caché de Markov

Le modèle caché de Markov (Hidden Markov Model, HMM) est un algorithme très souvent utilisé en bio-informatique. Nous avons utilisé l'implantation du HMM dans un logiciel libre appelé PennCNV<sup>133</sup>. La particularité de ce programme est qu'il ne considère pas seulement l'intensité du signal, mais également le ratio allélique et d'autres facteurs. De plus, ce logiciel introduit un score de confiance bayésien qui permet d'utiliser d'autres membres de la famille à des fins de validations des CNVs. Cependant, comme dans la présente étude, un seul membre de chaque famille est disponible, il ne nous est pas possible d'utiliser ce score de confiance. De plus, comme PennCNV ne supporte pas nativement la plateforme Affymetrix ©, nous avons utilisé le module complémentaire PennCNV-Affy<sup>134</sup> afin de convertir les fichiers bruts Affymetrix © en un format compréhensible. PennCNV étant très restrictif de par nature, nous avons appliqué un filtre de qualité minimal : un minimum de cinq sondes et une longueur minimale de 20 Kb, déterminé selon les spécifications de la puce 5.0 d'Affymetrix. Un filtre au niveau de la taille a été utilisé car

la distribution des sondes n'est pas constante et certaines régions génomiques contiennent beaucoup plus de sondes que d'autres régions.

#### *4.2.2.2 Segmentation circulaire binaire*

L'algorithme de segmentation circulaire binaire (circular binary segmentation, CBS) a d'abord été développé pour la technologie de CGH et a été adapté pour les données de SNP-chip<sup>135</sup>. Le principe est de diviser le génome en petites parties. Chaque partie est ensuite testée pour un changement du nombre de copies par permutation. Nativement, l'algorithme ne supporte pas le format Affymetrix © mais le modèle complémentaire PennCNV-Affy<sup>134</sup> a été adapté afin qu'il puisse supporter l'algorithme CBS. Un filtre de qualité de cinq SNPs et de 20 Kb a été appliqué.

#### *4.2.2.3 Analyse intégrative*

L'algorithme Birdsuite<sup>1</sup>, développé par le Broad Institute, propose une philosophie différente des deux algorithmes. Il est composé de plusieurs modules analytiques et d'un module intégratif qui permet d'intégrer les résultats de chaque module analytique. Le premier module, Canary, se charge de sonder les sondes non polymorphiques afin de détecter les CNVs les plus communs. Le deuxième module, Birdseed, permet d'utiliser l'intensité des sondes de SNPs afin de procéder à la détection de CNVs. Le troisième module analytique, Birdseye analyse indépendamment les données des sondes afin de détecter de nouveaux CNVs dans des régions inconnues jusqu'à présent. Finalement, c'est le module intégratif Fawkes qui se charge de compiler les résultats des trois modules analytiques et de générer une détection de CNVs fiable (**Figure 2**). De plus, cet algorithme supporte nativement les puces génomiques 5.0 et 6.0 d'Affymetrix ©. Comme le seuil de détection de BirdSuite est plus faible que pour les autres algorithmes utilisés, un filtre de qualité de 10 SNPs et de 20 Kb a été utilisé.

#### *4.2.3 Étude cas-contrôles des données SNP-chip*

Des fichiers cumulatifs du résultat de l'algorithme de détection de CNVs ont été produits par chacun des ensembles de données. Suite à une occurrence plus fréquente des CNVs chez les patients SIME, une normalisation restrictive d'un facteur 1.15 a été appliquée sur tous les patients SIME. Cette dernière mesure réduit le nombre de faux positifs, mais augmente également les probabilités de manquer des faux négatifs.

Afin de déterminer les degrés de différence entre les données, un test Fisher a été utilisé sur des fenêtres de 10 Kb qui couvrent la totalité du génome. Après une correction d'essais multiples, le seuil de significativité a été placé à une valeur P de  $1 \times 10^{-7}$ . Des valeurs P

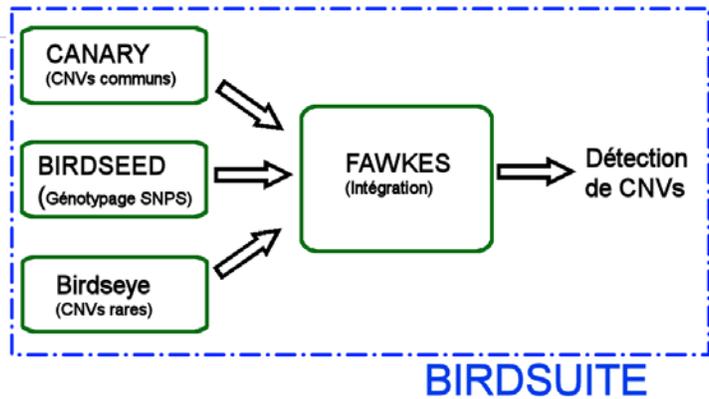


Figure 2 Diagramme de l'interaction des différents modules du logiciel BirdSuite<sup>1</sup> (Broad Institute)

ont été déterminées à l'échelle génomique avec des intervalles de 10 Kb. L'ensemble de l'étude cas-contrôles a été effectué par un algorithme « maison » jumelé avec un module du logiciel R afin de pouvoir réaliser un graphique pour chaque chromosome.

#### 4.2.4 Annotations des régions CNVs

Une fois les régions CNVs identifiées par l'étude cas-contrôle, une annotation détaillée a été réalisée en utilisant comme outil de base le fureteur génomique de l'Université de Californie à Santa Cruz (UCSC Genome Browser)<sup>136</sup>. Afin de s'assurer une couverture optimale, les régions CNVs identifiées ont été agrandies de 100 Kb en 5' ainsi qu'en 3'. Tous les gènes connus ont été examinés<sup>137</sup>, avec une attention particulière portée aux gènes surexprimés dans le cerveau ou connus pour être associés avec d'autres maladies neurologiques. Par la suite, la conservation des séquences, chez les mammifères<sup>138</sup> d'une part et chez tous les vertébrés<sup>139</sup> d'autre part, a été vérifiée. Finalement, l'existence de CNVs déjà connus dans ces régions a été vérifiée à l'aide de la base de données des variations génomiques<sup>122</sup>.

### 4.3 Détection de CNVs à l'aide de la plateforme CGH

#### 4.3.1 Hybridation des ADN

Nous avons choisi 125 patients et 125 contrôles parmi les résultats de la première phase (SNP-chip) afin de valider les résultats. À partir d'ADN sanguin, nous avons procédé à l'hybridation sur des puces SurePrint G3 Human CNV 400K de Agilent ©. Chacune des puces comprend deux plages identiques de plus de 412 000 sondes permettant de procéder à l'hybridation de deux ADN différents. Chaque hybridation a été faite avec un mélange d'ADN de références (cinq échantillons d'ADN, mâle ou femelle selon le sexe de l'individu testé) produit par Proméga ©. Toutes les étapes de l'hybridation (à l'exception de la fixation sur la puce et de la numérisation de l'image)

ont été faites par un système de manutention automatisé des liquides Bravo © fourni par Agilent©. La numérisation a été faite par un numériseur DNA Microarray Scanner C de Agilent ©.

#### 4.3.2 Traitement des images

Les fichiers d'images produites par le numériseur ont été traités par le logiciel Feature Extraction Software (Agilent ©). Dans un premier temps, un alignement automatique de grilles était réalisé par le logiciel. Dans certains cas où l'alignement n'était pas satisfaisant, la grille a été réalignée manuellement selon le format des puces utilisées. Une fois les images transformées en fichier d'intensités, nous avons calculé la valeur de la propagation de la dérivée du ratio logarithmique (Derivated Log Ration spread, DLRS). Cette valeur est une mesure de la qualité de l'hybridation sur la puce. Nous avons utilisé un filtre où nous avons exclu toutes les images avec une valeur DLRS supérieure à 0.3, qui est la valeur standard définie par l'équipe technique d'Agilent ©. D'autres valeurs de contrôle de qualité ont été utilisées afin d'estimer le seuil entre l'intensité du signal réel et des signaux parasites. Nous avons également appliqué un filtre restrictif au niveau du bruit de fond de chaque détecteur (rouge et vert). Ainsi, tous les échantillons avec une valeur de bruit de fond pour l'un des deux signaux supérieure à 35 ont été retirés de l'analyse.

#### 4.3.3 Algorithme de détection des CNVs

Afin de procéder à la détection de CNV dans les individus, nous avons utilisé le logiciel Genomic Workbench (Agilent ©). Les fichiers d'intensités sont tout d'abord centralisés avec un seuil de 6.0 et une taille de boîte de 10. De plus, un filtre est ajouté pour les réplicons intrapuce. L'algorithme d'analyse utilisé est le modèle caché de Markov (HMM) à sept états avec un seuil de 0.5. Le HMM est un processus stochastique qui mesure la probabilité d'observer un signal aberrant lié à une sonde en tenant compte de la présence possible d'un biais dans le signal. Les sondes sont ainsi séquentiellement marquées sur les chromosomes selon leur probabilité de présenter une aberration chromosomique. Ainsi, le résultat d'une analyse de HMM est une série d'états où chaque état représente la façon dont les sondes de chaque individu ont été identifiées<sup>140</sup>. La version du HMM adaptée pour les données CGH permet d'optimiser les données en utilisant un algorithme d'apprentissage Baum-Welch.

L'avantage de l'algorithme de HMM par rapport aux autres méthodes d'analyses CGH est qu'il permet de diminuer le bruit de fond présent lors d'analyses à grande échelle, permettant ainsi d'identifier des régions d'aberrations chromosomiques. De plus, le HMM permet d'établir un cadre de probabilités afin de conférer une mesure de confiance à chaque observation. En plus de

cela, nous avons choisi d'utiliser le HMM car c'est une méthode comparable à celle utilisée lors de l'analyse des SNP-chips qui permettra une comparaison directe des résultats provenant des deux méthodes.

#### **4.3.4 Étude cas-contrôles des données CGH**

Les résultats de la détection des CNVs ont été regroupés en deux fichiers cumulatifs : l'un pour les patients SIME et l'autre pour les contrôles. À l'instar des données SNP-chip, nous avons utilisé un test de différence Fisher afin de déterminer si l'un des deux ensembles différait significativement de l'autre pour une position génomique précise. Comme la taille des CNVs détectés par CGH était fortement différente de la taille des CNVs détectés par SNP-chip, nous avons adapté l'algorithme utilisé lors de la première phase. Des fenêtres de 500 Kb ont été établies sur l'ensemble du génome. Pour chaque fenêtre, une valeur P a été calculée. Le même seuil de significativité établi lors de l'analyse des SNP-chips a été utilisé.

## 5. Résultats

### 5.1 Étude préliminaire sur les individus HapMap

Afin de limiter le nombre de corrections pour essais multiples nécessaire pour la présente étude, la méthodologie a tout d'abord été développée sur un ensemble de données indépendantes de l'ensemble principal (patients SIME). Afin que la méthodologie soit adéquate pour les patients SIME, nous avons utilisé des données réalisées sur le support Affymetrix © GenomeWide SNP chip 5.0. Les données, provenant directement du site web d'Affymetrix ©, consistent en dix trios constitués de 30 individus du projet HapMap<sup>132</sup>. Le but de l'étude préliminaire étant de déterminer l'algorithme le plus approprié pour l'étude, la nature et la position des CNVs n'ont pas été prises en considération. Ainsi, la relation entre les individus HapMap n'a pas été utilisée et chaque individu a été considéré comme indépendant des autres. En effet, au final, seules les données des contrôles de qualité et de détection moyenne de CNVs ont été utilisées pour établir la méthodologie.

En se basant sur un article comparant les différents algorithmes de détection de CNVs sur les plateformes de biopuces à SNPs<sup>141</sup>, deux algorithmes différents correspondant le mieux à nos besoins ont été sélectionnés; le modèle caché de Markov (PennCNV)<sup>133</sup> et un algorithme intégratif (Birdsuite)<sup>1</sup>. En plus de ces deux algorithmes, un algorithme de segmentation circulaire binaire (DNAcopy)<sup>135</sup>, recommandé par des communications orales de spécialistes du domaine, a également été utilisé.

Afin d'évaluer adéquatement les capacités de chaque algorithme, 44 fichiers Affymetrix bruts provenant des 30 individus HapMap (certains individus étant présents plus d'une fois) ont pu être analysés. La détection de CNV a été réalisée nativement par Birdsuite et par le biais d'un module complémentaire<sup>134</sup> pour PennCNV et DNAcopy. Par la suite, les différents CNVs détectés pour chaque individu selon les filtres des différents algorithmes (voir méthodologie) ont été comptabilisés. À l'aide de ces résultats, un nombre d'occurrences moyen de CNVs pour chaque chromosome a pu être produit. **La figure 3** présente ces résultats.

À la suite de discussions avec plusieurs spécialistes, il a été décidé d'orienter le choix vers l'algorithme intégratif (Birdsuite). Plusieurs raisons ont motivé ce choix :

- Le nombre de CNVs par individu est beaucoup plus grand avec Birdsuite qu'avec les deux autres algorithmes.
- L'algorithme intégratif permet de différencier les phases de détection de CNVs communs et de CNVs rares.
- Birdsuite a été spécialement conçu pour la plateforme Affymetrix © alors que les deux autres algorithmes la supportent par un module complémentaire.

L'un des points faibles de Birdsuite est qu'étant donné le nombre très élevé de CNVs détectés, il est fort probable qu'une proportion importante de cet ensemble soit des faux positifs. Ces résultats faux positifs seraient vraisemblablement liés à un bruit de fond trop élevé ou à des conditions expérimentales parfois sub-optimales. Cependant, dans le cas de l'étude actuelle, l'analyse cas-contrôles nous permet de contrôler ces deux paramètres car il est attendu que les biais expérimentaux soient similaires dans le groupe des patients et le groupe des contrôles. Ainsi,

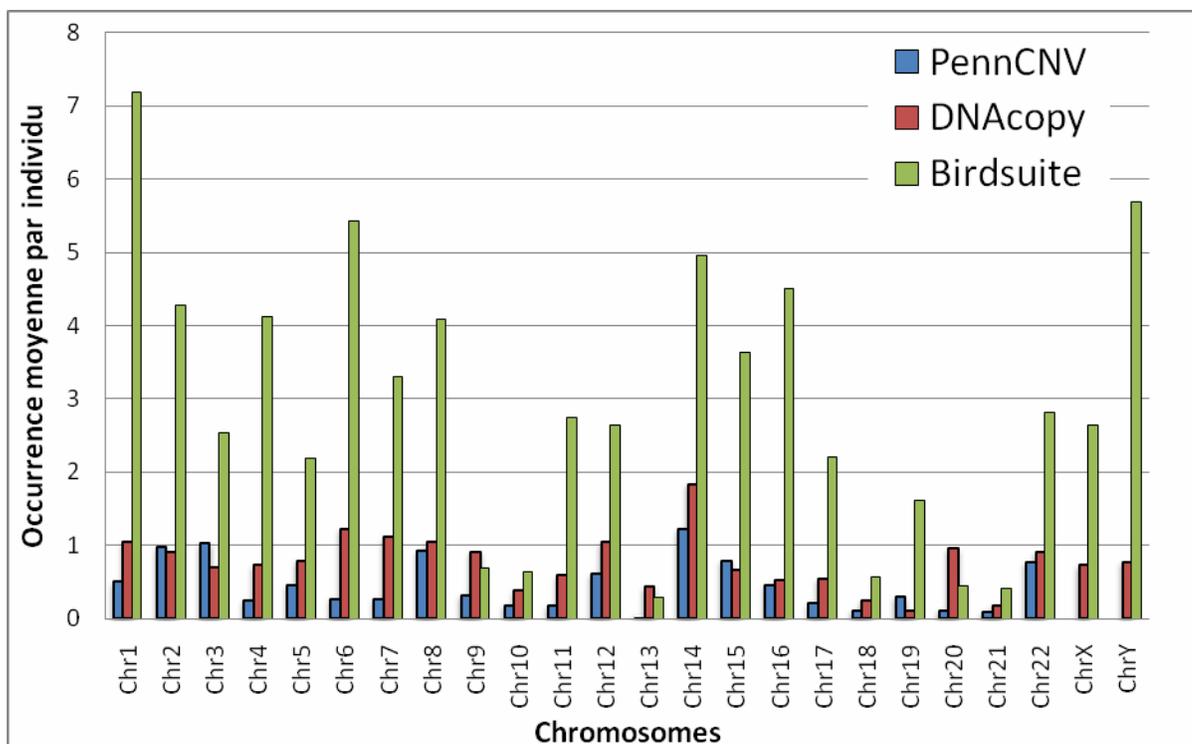


Figure 3. Occurrence moyenne des CNVs par individus, classée par chromosome. L'axe des Y présente l'occurrence moyenne pour chaque individu. Les CNVs identifiés par PennCNV sont présentés en bleu, les CNVs identifiés par DNACopy sont présentés en rouge et les CNVs identifiés par Birdsuite sont présentés en vert. Il est à noter que PennCNV n'a pas été en mesure de détecter de CNVs sur les chromosomes sexuels.

après la comparaison, il ne resterait que les CNVs spécifiques à l'un des groupes. Il est vrai qu'il reste des possibilités d'obtenir des résultats faux positifs, mais cette méthode représente un bon compromis entre la minimisation des faux positifs et des faux négatifs. De plus, même en utilisant un algorithme très restrictif, les données de SNP-chip ne sont pas assez fiables pour pouvoir tirer des conclusions en n'utilisant qu'une seule méthode expérimentale. De ce fait, comme dans tous les cas une validation expérimentale sera nécessaire, il est optimal de choisir un algorithme produisant un nombre plus élevé de détections.

## 5.2 Détection de CNVs par SNP-chip

La première phase de l'étude consistait en une détection de CNVs sur une plateforme SNP-chip (Genomewide Chip 5.0 d'Affymetrix ©). Cette plateforme a été choisie, en très grande partie, parce que les données avaient subséquemment été générées par une autre étude<sup>89,94</sup>; cette dernière portait également sur le SIME et avait pour but d'effectuer une étude d'association génomique sur des patients SIME, mais n'avait pas tenté d'évaluer la présence des CNVs.

Les fichiers bruts de plus de mille patients (n= 1,120) étaient disponibles. Tous ces patients ont été recrutés par des médecins allemands et autrichiens et sont tous de descendance européenne, une grande majorité provenant du sud de l'Allemagne (Bavière). Cependant, le phénotype du SIME étant d'une nature très variable, certains critères ont été appliqués afin d'homogénéiser l'ensemble. Après avoir retiré tous les échantillons pouvant provenir d'un SIME secondaire (dialyse, anémie, grossesse), nous avons créé le groupe d'analyse en y incluant seulement les patients avec une histoire familiale de SIME positive. Pour cela, les patients devaient pouvoir rapporter la présence d'au moins un membre de leur famille souffrant également de SIME, le diagnostic étant confirmé par entretien téléphonique ou individuel. Le groupe primaire d'analyse (**groupe patients**) était constitué de 493 patients souffrants de SIME avec histoire familiale positive. Tous les autres patients SIME avec une histoire familiale négative ou inconnue furent placés dans un groupe de validation (**groupe validation**).

L'une des lacunes de la première phase de l'étude est l'absence d'un groupe contrôle traité dans les mêmes conditions expérimentales que les patients SIME. Certaines populations contrôles étaient disponibles sous d'autres supports (500k, GenomeWideSNP chip 6.0) ou d'autres plateformes (Illumina ©), mais aucune dans des conditions similaires à celle des patients SIME. Afin de pallier à ce manque, un partenariat a été établi avec l'organisme POPGEN qui a pu nous

fournir près de 500 contrôles provenant d'une population allemande (**groupe contrôles externes**). Dans le cadre de l'accord avec l'organisme, nous avons pu obtenir les fichiers bruts pour chaque patient, le sexe ainsi que quelques informations phénotypiques relatives aux troubles du sommeil. Cependant, comme l'ensemble de données contrôle a été traité dans un lieu et une période de temps différent, il est envisageable que cela introduise un biais expérimental non désirable. Afin d'estimer ce biais et de vérifier son impact sur l'étude, une population contrôle interne a également été définie (**groupe contrôle interne**). Cette population est composée de deux sous-

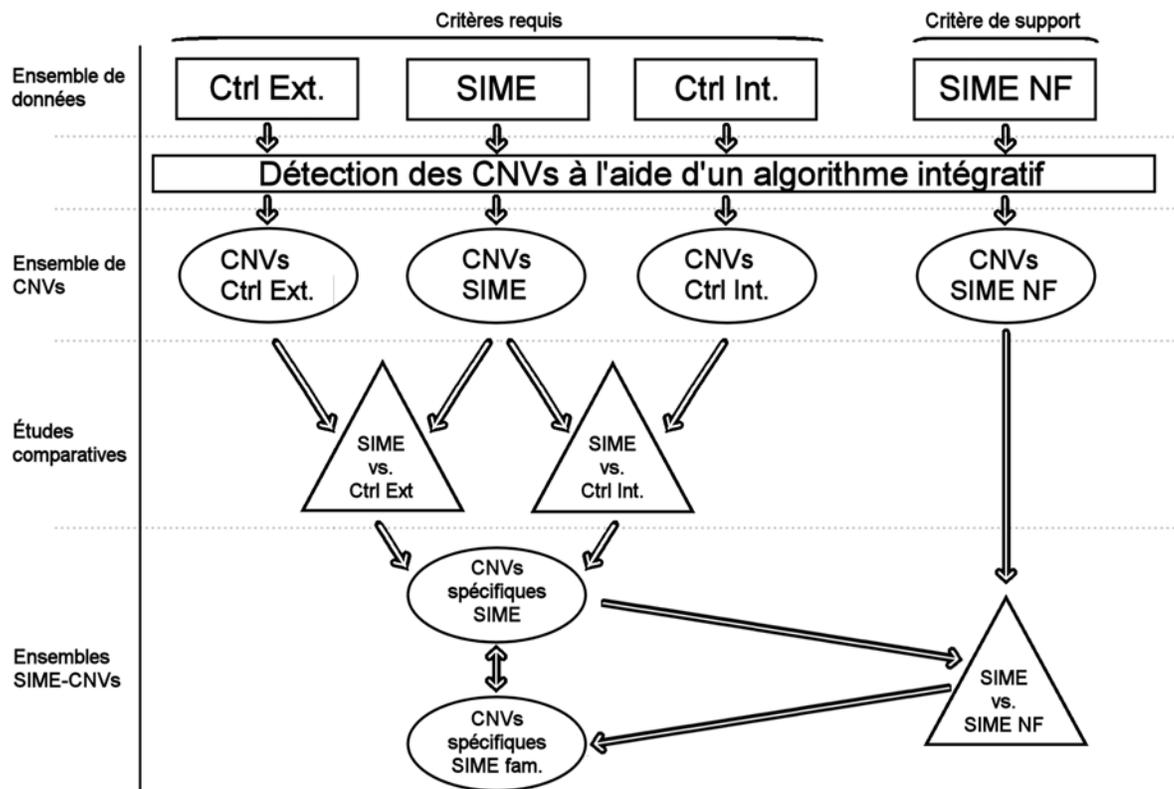


Figure 4 Processus de sélection des CNVs spécifiques au SIME. Les données sont tout d'abord regroupées en ensemble de phénotype similaire. Par la suite, une détection des CNVs est réalisée pour chaque ensemble (en utilisant l'algorithme Birdsuite). Puis, en premier lieu, une étude comparative est réalisée entre l'ensemble SIME et les ensembles contrôles. Une occurrence spécifique Sine Qua non doit être observée pour chaque ensemble contrôle. Finalement, une comparaison est faite avec l'ensemble de support. (Ext = Externes, Int = Internes, NF = Non-familiaux, Fam = familiaux)

populations principales: l'une contenant des patients atteints d'un cancer du sein et l'autre contenant des patients ayant souffert de septicémie. La collection des échantillons de ces deux populations a été réalisée dans les mêmes conditions expérimentales (lieu, équipement et personnel technique) et dans la même période de temps que les échantillons SIME. En mixant les

deux populations avec une proportion 1:1, il est attendu que les effets spécifiques à un trait phénotypique seront largement diminués.

Ainsi pour déterminer les CNVs à occurrence plus élevée ou plus faible, il est nécessaire que l'effet soit significatif lorsque comparé aux deux contrôles (interne et externe). Quant au groupe composé des patients SIME non-familiaux, il est utilisé comme groupe de validation. Comme il est composé de plusieurs patients avec un statut inconnu, il est implicite que certains de ces patients possèdent une histoire familiale, mais qu'il n'est pas possible de le confirmer. Ainsi, lorsque l'on compare les patients SIME avec le groupe de validation, il est attendu que l'effet sera atténué

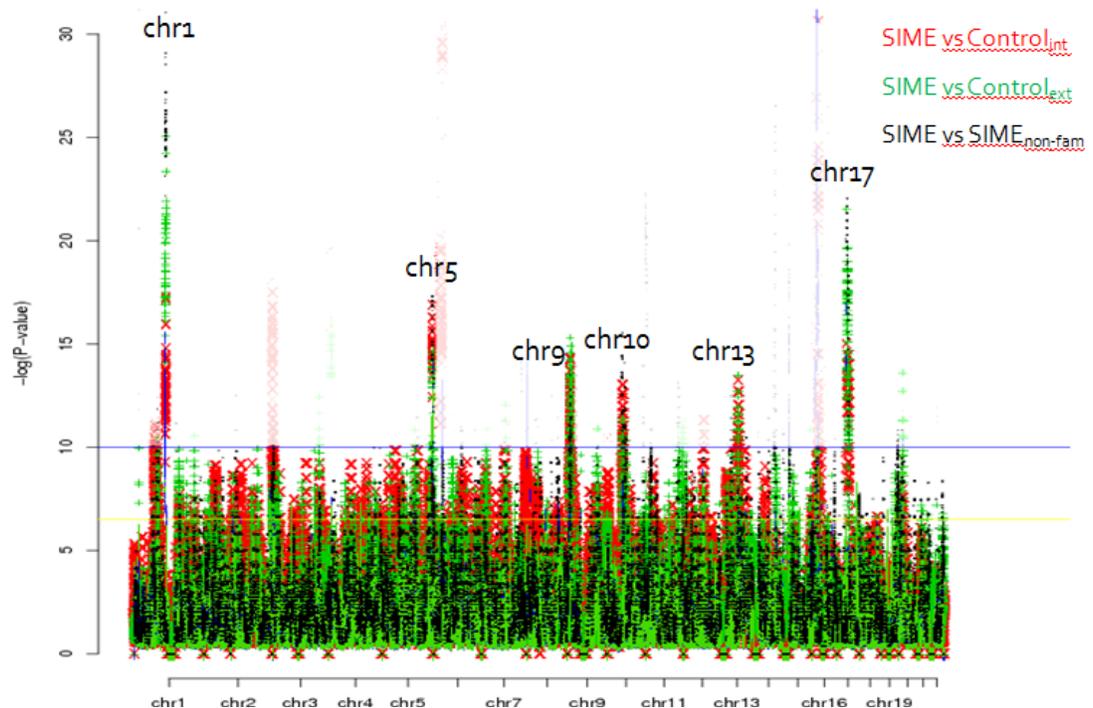


Figure 5 Résultats de l'étude de CNVs à échelle génomique. L'axe des X présente la position chromosomique (regroupée par chromosome) et l'axe des Y présente le log négatif de la valeur P. Au total, six régions ont été retenues. Les régions ne répondant pas à tous les critères de sélection ont été ombragées afin de faciliter la visualisation.

mais devrait tout de même être observable. La comparaison avec le groupe de validation est utilisée comme preuve de support, mais n'est pas un critère *sine qua non*. La **figure 4** présente un schéma démonstratif des étapes de l'étude comparative.

Tous les CNVs répondant aux critères de qualité (voir section méthodologie) ont été catalogués dans leurs ensembles respectifs. À ce stade, la nature des nombres de copies (délétion ou duplication) n'est pas prise en considération. Pour la première phase de l'étude, l'accent est mis

sur la détermination de régions de CNVs pouvant avoir une occurrence différente pour le SIME. Une fois ces régions identifiées, il sera plus facile, au cours de la deuxième phase de l'étude, de vérifier quelle nature de CNV peut avoir le plus d'impact sur le phénotype. Les études comparatives ont été réalisées en effectuant des tests Fisher à des intervalles de 10Kb sur tout le génome en utilisant les deux types d'hypothèses Fisher. Les comparaisons ont été effectuées comme décrites dans le paragraphe précédent. À l'aide d'une correction pour tests multiples Bonferonni, un seuil de significativité a été établi.

Lors du regroupement des études comparatives, 6 régions distinctes répondaient aux critères de sélection et ont été retenues (**figure 5**). Le **tableau 4** présente une description détaillée des résultats.

**Tableau 4** Résumé détaillé des régions de CNVs observées lors de l'étude à échelle génomique

Région	Dimension (Mb)	Coordonnées (Mega paire de bases)	Nombre de gène	Associations précédentes	Nombre de CNVs connus
<b>1p13-12</b>	2,5	110,2 – 112,7	39	Cancers	64
<b>5q35.3</b>	0.98	177,02 – 178,0	13	-	41
<b>9p24.1-23</b>	3.5	8,35 – 11,85	1	SIME, Cancer	131
<b>10q21.1</b>	1.25	54,75 – 56,0	1	Surdit�	35
<b>13q21.31</b>	1.18	61,04 – 62,22	1	-	3
<b>17q21.2-31</b>	5.2	36,8 – 42,0	> 100	Cancer du sein, Ob�sitt�, Alzheimer, Narcolepsie, Hypertension	136

### 5.3 D tection des CNVs par CGH

Comme l'utilisation des SNP-chips pour la d tection des CNVs est une m thode indirecte et qu'elle est tr s souvent critiqu e, il nous est apparu essentiel de valider les r sultats de la premi re phase de l' tude en utilisant une m thode ind pendante. Nous avons choisi la m thode d'hybridation par g nomique compar e (CGH), car c'est une m thode sp cifique   la d tection d'aberrations chromosomiques. De plus, comme l'hybridation se fait par rapport   un mix d'ADN contr les, il n'est pas n cessaire de proc der   une normalisation pour l'ensemble des individus,  

l'opposé de la technologie des SNP-chips. De ce fait, chaque échantillon peut-être analysé séparément, ce qui permet d'avoir des résultats plus précis.

Pour la deuxième phase d'analyse, nous avons dû développer la méthodologie pour une étude de CGH à large échelle. En partenariat avec la compagnie Agilent, nous avons installé une plateforme de manipulation automatisée (système BRAVO) afin de minimiser la variabilité inter-échantillons. De plus, dû à la nature onéreuse des analyses CGH, seulement un sous-ensemble (n=125) de l'échantillon initial a été analysé. Un nombre équivalent de contrôles a été utilisé afin de pouvoir comparer les résultats CGH aux résultats de la première phase de l'étude.

Nous avons à nouveau opté pour une stratégie d'analyse à l'échelle génomique afin d'être en mesure de comparer les deux technologies. Afin de simplifier le processus analytique, nous n'avons considéré que les régions identifiées lors de la première phase de l'étude. **Le tableau 5** présente les résultats obtenus lors de l'analyse avec un algorithme de modèle caché de Markov (HMM).

**Tableau 5 Validation des régions en utilisant une technologie CGH**

Région	Dimension (Mb)	Valeur P*	Type CNV	Nombre de copies
1p13-12	-	0.06	-	-
5q35.3	2	0.0004	-	-
9p24.1-23	0.5	$6.38 \cdot 10^{-7}$	Deletion	1
10q21.1	-	0.3	-	-
13q21.31	1.3	$5.17 \cdot 10^{-7}$	Duplication	3
17q21.2-31	-	0.009	-	--

\*Seuil corrigé pour essais multiples : Valeur P <  $1 \times 10^{-5}$

## 6. Discussion

Des six régions précédemment identifiées par l'étude des déséquilibres génomiques par SNP-chips (voir Tableau 5), deux régions ont également été observées dans les études CGH. La région

sur le chromosome 9p est toutefois celle qui retient plus particulièrement l'attention de la communauté du SIME. En effet, un facteur qui accroît davantage l'intérêt de cette région est qu'elle est également la seule région identifiée à la fois par des études de liaisons et d'associations génétiques. Identifiée pour la première fois dans une famille américaine<sup>75</sup> (mais abondamment critiquée dans la littérature<sup>85</sup>), une association de SIME à cette région a été répliquée dans une famille allemande<sup>87</sup>. Si la méthodologie utilisée dans le rapport initial peut paraître douteuse, le fait que le même locus ait été identifié dans une étude indépendante consolide les résultats pour cette région. Une autre étude a également rapporté une liaison génétique sur le chromosome 9p, mais cette fois séparée de 10 cM du locus original<sup>88</sup>. Puis, plus récemment, la réplique d'une étude d'associations génomiques a permis l'identification de deux variants situés dans la région intronique du gène *PTPRD*, également situé sur le chromosome 9p. **Le tableau 6** présente une récapitulation des études portant sur la région 9p.

**Tableau 6** Tableau récapitulatif des études génétiques ayant découvert des associations significatives avec la région chromosomique 9p

Étude	Type d'étude	Chromosome	Locus
<b>Chen et coll.</b> <sup>75</sup>	Liaison	9p22-24	RLS3
<b>Liebetanz et coll.</b> <sup>87</sup>	Liaison	9p22-24	RLS3
<b>Lohmann-Hedrich et coll.</b> <sup>88</sup>	Liaison	9p21.3-9p21.1	RLS3*
<b>Schormair et coll.</b> <sup>89</sup>	Association	9p23-9p24.1	RLS3
<b>Étude en cours</b>	CNVs	9p23-9p24.1	RLS3

Lorsque l'on regroupe toutes les études portant sur une association entre le SIME et la région chromosomique 9p, les délimitations cumulatives font en sorte qu'un seul gène est présent dans la région : *PTPRD*. Ce gène encode un récepteur de protéine de type tyrosine phosphatase. La phosphorylation des tyrosines est généralement effectuée en rapport égal par les protéines tyrosines kinases (PTKs) et les protéines de type tyrosine phosphatase (PTPs). Dans le cas de *PTPRD*, cette protéine est une PTP transmembranaire classique<sup>142</sup>. Si on connaît 38 PTPs classiques<sup>143</sup>, on répertorie aujourd'hui au moins 107 protéines avec une activité PTP<sup>142</sup>. En se basant sur la séquence d'acide aminé, il est possible de classer toutes les PTPs en quatre familles, chacune ayant une classe de substrat spécifique. *PTPRD* fait partie de la classe I des PTPs

à base de cystéine, classe à activité strictement tyrosine spécifique<sup>143</sup>. Si les PTPs sont connues pour avoir un grand rôle à jouer dans les activités du système immunitaire<sup>144</sup>, on leur connaît de nombreux autres rôles dans divers systèmes biologiques. Pour *PTPRD*, il serait en particulier impliqué dans plusieurs processus cancéreux<sup>113</sup>. Cependant, un rôle dans les maladies neurologiques comme le SIME n'est pas exclu, et même probable. En effet, plusieurs autres PTPs ont été associées avec plusieurs maladies neurologiques, notamment la sclérose en plaques (*PTPRC*)<sup>145</sup>, l'autisme (*PTPN9*)<sup>146</sup> et l'épilepsie progressive myoclonique (*EPM2A*)<sup>147</sup>. De plus, il existe différents modèles de souris knock-out pour le gène *PTPRD*. Un de ces modèles, une souris double mutante pour les récepteurs de protéines de type tyrosine phosphatase  $\delta$  et  $\sigma$ , meure instantanément après une césarienne<sup>161</sup>. La souris, paralysée, montre également des signes de dysgénésie musculaire sévère ainsi qu'une lourde perte des motoneurons de la moelle épinière. Le second modèle, une souris knock-out pour le gène protéine tyrosine phosphatase delta, résulte en un animal avec des problèmes d'apprentissages<sup>162</sup>. De plus, la souris meure rapidement du à des problèmes d'apport alimentaire. Ces deux modèles suggèrent un rôle important pour le gène *PTPRD* dans le système nerveux et renforce l'hypothèse de l'implication du gène dans le SIME.

Lorsque nous comparons les signaux obtenus lors de l'identification de CNVs en utilisant les SNP-chips avec les signaux obtenus lors de l'identification de CNVs par CGH, nous observons un chevauchement presque parfait au niveau d'une partie du gène *PTPRD*. De plus, cette région de chevauchement est exactement la même région que celle identifiée lors de l'étude d'association de Schormair et coll.<sup>89</sup> (**voir figure 6**). La région consensus ainsi observée ne couvre aucun exon codant. En effet, le gène *PTPRD* peut posséder jusqu'à 45 exons, mais les 10 premiers exons sont toujours non-codants. De ce fait, il est fort probable que ces exons non-codants soient impliqués dans le repliement de l'ARN ainsi que dans la régulation de l'expression du gène. Une hypothèse plausible est que les effets phénotypiques des CNVs pourraient être totalement différents de ceux d'un polymorphisme de nucléotide. En effet, il a été montré que les CNVs ont un effet global sur l'expression des gènes<sup>130,148</sup>. Des CNVs localisés dans des régions non-codantes (promoteurs, 3' UTR) ont déjà été associés à des maladies différentes. Un exemple de ce genre de CNVs est la duplication trouvée dans la proximité directe du gène *peripheral myelin protein 22 (PMP22)* qui cause le syndrome de Charcot-Marie-Tooth<sup>149</sup>. Un autre exemple est une délétion trouvée sur le chromosome 1 qui est associée avec les neuroblastomes. Cette délétion ne contient aucun gène connu, mais un transcrit à l'intérieur serait vraisemblablement un nouveau membre de la famille

de gène neuroblastoma breakpoint family, member 1 (*NBPF1*), précédemment liée aux neuroblastomes<sup>150</sup>. De plus, la région du chromosome 9 contient un grand nombre de CNVs précédemment identifiés<sup>122</sup>. Ce nombre peut en partie s'expliquer par le fait qu'un seul gène se trouve dans cette région de 3.5 Mb et que de ce fait, l'implication fonctionnelle de cette région chromosomique serait vraisemblablement limitée. Par contre, cela pourrait aussi appuyer notre hypothèse concernant l'implication des CNVs dans le SIME. En effet, avec une prévalence autour de 10 % dans la population générale, il est fort probable que si des CNVs sont liés au SIME, ils aient déjà été observés dans des études populationnelles portant sur l'occurrence des CNVs au niveau génomique.

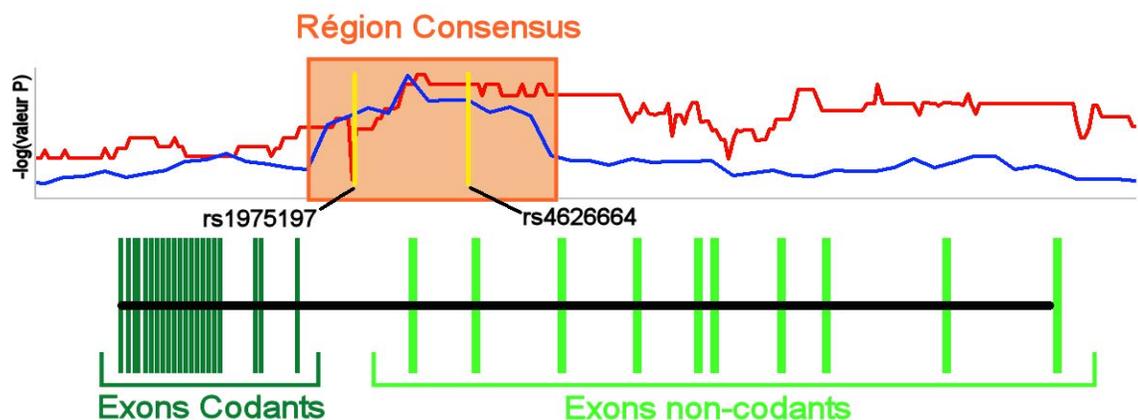


Figure 6. Chevauchement des études de SNP-chips, CGH et associations au niveau de la région non-codante du gène PTPRD. Le signal en rouge représente le test Fisher réalisé lors de l'analyse des SNP-chips. Le signal en bleu représente le test Fisher réalisé lors de l'analyse CGH. Les deux barres jaunes correspondent à la position des SNPs identifiés lors de l'étude d'association.

Des cinq autres régions identifiées lors de la détection de CNVs par SNP-chips, une seule autre a été répliquée dans la deuxième phase de l'étude, celle sur le chromosome 13. Cependant, une analyse détaillée de la région révèle qu'elle ne contient aucun gène connu. On peut y noter la présence de certains marqueurs de séquences exprimés connus, mais aucun ne révèle d'homologie significative avec des gènes connus<sup>151</sup>. De ce fait, il devient difficile de lier cette région au SIME, car nous ne disposons d'aucun indice biologique qui nous permette de supporter les liens entre la région et le phénotype. La région sur le chromosome 5, bien que non-significative dans la deuxième phase de l'étude, montre tout de même une tendance (Valeur P =  $4 \times 10^{-4}$ ). Cette région contient 13 gènes différents, dont quatre ont été associés à différents

phénotypes, selon la base de données OMIM<sup>152</sup>. Cependant, aucune de ces associations ne semble correspondre au phénotype du SIME.

Lorsque l'on compare les résultats de la présente étude avec ceux du consortium WTCCC publiés récemment<sup>131</sup>, nos observations corroborent, jusqu'à un certain point, les conclusions du WTCCC. Leur étude traitant de l'occurrence spécifique de certains CNVs pour six maladies différentes (trouble bipolaire, cancer du sein, cardiopathie coronarienne, hypertension, arthrite rhumatoïde et diabète de type I et II), ils ont conclu qu'il est peu probable que les CNVs communs identifiés par les méthodes actuelles contribue significativement à la composante génétique des maladies complexes<sup>131</sup>. En effet, selon leurs résultats, ces régions auraient déjà été identifiées par les EAG conduites auparavant. Pour la présente étude, une des deux régions de CNVs spécifique au SIME coïncide avec des données précédemment identifiées par EAG, celle du gène *PTPRD* sur le chromosome 9<sup>89</sup>. En collaboration avec Schormair et coll., nous avons revisité les données de l'EAG pour la région du chromosome 13 et avons pu conclure à l'absence complète d'un signal d'association. De plus, étant donné l'absence de points d'intérêts fonctionnellement intéressants au niveau du chromosome 13 rend très ardu une potentielle confirmation de l'implication de cette région au niveau du SIME.

Les résultats obtenus nous permettent de suggérer qu'une variation du nombre de copies au niveau de la région non-codante du gène *PTPRD* pourrait agir comme facteur de susceptibilité pour le SIME. En utilisant deux études cas-contrôles indépendantes, il a été possible de valider le signal pour chacune des études. De plus, ce signal correspond exactement à la région identifiée lors d'une étude d'association<sup>89</sup>. De ce fait, il est justifié de se demander si le signal observé par Schormair et coll.<sup>89</sup> lors de leur étude d'association ne serait pas un marqueur de l'effet causatif des CNVs. Cependant, comme l'effet observé au niveau des CNVs est plus important qu'au niveau des SNPs, il serait aussi pertinent d'envisager que les signaux observés par les SNPs et par les CNVs soient additifs. Cette hypothèse serait supportée par une étude réalisée par Stranger et coll. sur l'effet respectif des SNPs et des CNVs sur l'expression des gènes. L'étude du groupe montre que les SNPs et les CNVs ont tous deux un effet sur l'expression génique. Cependant, lorsque l'on tente d'établir un lien entre les CNVs et les SNPs provenant d'une même région chromosomique, le chevauchement est très faible<sup>130</sup>. Ainsi, selon eux, pour comprendre l'effet de variations génétiques sur un phénotype, il serait essentiel de consulter à la fois les SNPs et les CNVs.

Par contre, il n'est pas acquis qu'une hausse de l'expression génique se traduise au niveau protéique, et encore moins au niveau phénotypique. Abondant dans ce sens, une étude sur des cellules cancéreuses a démontré qu'au niveau transcriptomique, on observe des corrélations entre les différences d'expression géniques et les régions à haute prévalence en CNVs dans moins de 4% des fragments transcrits à l'intérieur de ces régions<sup>153</sup>. Cependant, une nouvelle méthode d'analyse bio-informatique permettrait peut-être d'établir une meilleure corrélation<sup>154</sup>. Cette méthode se base sur un ensemble de gènes physiquement proches plutôt que sur des gènes individuels. En analysant deux ensembles indépendants de données sur le cancer du sein, le groupe à l'origine de la méthode a pu constater une très bonne corrélation entre le dosage des gènes et leur niveau d'expression<sup>154</sup>.

Bien que nous ne disposions d'aucun élément pouvant directement lier la présence de CNVs dans le chromosome 9p avec le phénotype du SIME, la présence redondante de la région dans plusieurs études génétiques suggère que la présence de CNVs dans cette région serait un facteur de susceptibilité pour la maladie. Il existe d'autres exemples de CNVs communs qui agiraient comme facteur de susceptibilité pour une maladie, notamment pour les neuroblastomes. En utilisant une méthode similaire à celle utilisée pour notre étude, Diskin et coll. ont défini une région à haute prévalence en CNV qui augmenterait les prédispositions de développer un neuroblastome d'un facteur très significatif (risque relatif rapproché  $> 3$ )<sup>150</sup>.

## 7. Perspectives

Une étape importante à venir pour la présente étude est la validation de la région du chromosome 9 dans une population indépendante. En effet, jusqu'à présent, tous les échantillons testés ont été collectés dans la population autrichienne-allemande. Or, à l'image de toutes les autres variations génétiques, les CNVs possèdent une forte composante populationnelle. En effet, en observant les trois populations HapMap (africaine, asiatique et caucasienne), une étude a démontré une différence significative au niveau de la fréquence de certains CNVs lorsque comparés entre les trois populations<sup>155</sup>. Cette étude va même jusqu'à suggérer qu'une partie de la variation phénotypique interpopulationnelle pourrait être expliquée par une occurrence différente de certains CNVs. Des différences du même ordre ont également été observées dans une comparaison entre la population caucasienne et la population asiatique<sup>156</sup> ainsi que dans le cadre d'une comparaison entre la population caucasienne et la population du Moyen-Orient<sup>157</sup>. Dans le cadre de notre étude, il est donc important de valider l'implication de la région identifiée dans des cohortes de patients SIME provenant d'autres origines ethniques. Dans le cas présent, la population canadienne-française serait le choix idéal. En effet, toutes les variations génétiques identifiées par études d'associations génomiques dans la population autrichienne-allemande ont été répliquées dans la population canadienne-française<sup>89,94</sup>. Ainsi, il est donc attendu que si les changements du nombre de copies du gène *PTPRD* sont bien des facteurs de susceptibilité au SIME, ces variations seraient également présentes dans la population canadienne-française.

L'utilisation de la population canadienne-française à des fins de réplication aurait également un autre avantage majeur. En effet, la banque d'échantillons d'ADN de patients SIME du Dr. Rouleau, l'une des plus importantes au niveau mondial, est également la seule qui possède un important nombre de lignées cellulaires lymphoblastoïdes prélevées sur des patients SIME. De ce fait, il serait donc possible de vérifier si un facteur de différence du nombre de copies serait observable en termes de différence du niveau d'expression sur des lymphoblastes. Une approche similaire a été utilisée par notre groupe pour valider l'effet du gène *MEIS1*, probablement impliqué le mécanisme génétique du SIME<sup>98</sup>. Il serait également intéressant de vérifier si la présence d'un changement du nombre de copies du gène *PTPRD* pourrait se refléter sur le niveau d'expression des autres gènes importants pour le SIME (*MEIS1*, *BTBD9* et *LBXCOR1*). Une association positive

pourrait permettre d'établir des interactions protéiques entre les différents gènes impliqués dans la pathogenèse du SIME.

Un autre avantage de l'utilisation des lymphoblastes de patients SIME canadienne-française serait de pouvoir déterminer l'impact relatif des variations SNPs et CNVs sur l'expression du gène PTPRD. En effet, il serait possible d'identifier des changements d'expressions reliés à l'allèle mineur identifié lors de l'EAG et de les comparer avec les changements d'expression potentiellement liés avec l'occurrence des CNVs dans la région du chromosome 9. Cette analyse aurait un impact majeur non seulement au niveau du mécanisme génétique du SIME, mais également au niveau de la pathogenèse génétique de plusieurs maladies génétiquement complexes. Il serait également intéressant de pouvoir compter sur un modèle animal pour le SIME. Cependant, étant donné le caractère imprécis et variable du phénotype, l'établissement d'un tel modèle n'a pas encore été réalisé. Cependant, les avantages d'un tel modèle (une souris par exemple) seraient énormes. Il serait entre autre possible d'observer le niveau d'expression des gènes discutés précédemment. Notre groupe a réalisé des études similaires chez *c. elegans*<sup>98</sup> mais l'éloignement entre l'homme et le ver complique beaucoup les études génétiques.

Une autre étude intéressante serait de procéder à une détection d'aberrations chromosomiques au niveau de la famille allemande précédemment liée à la région du chromosome 9 par une étude de liaisons génétiques<sup>87</sup>. L'ADN de cette famille nous est accessible par le biais de nos collaborateurs allemands, avec qui cette étude a été réalisée. L'identification de CNVs à ségrégation familiale serait une avancée majeure pour le SIME car il s'agirait de la première identification de variation génétique à la fois liée avec la forme familiale et sporadique de la maladie.

Finalement, pour les autres familles précédemment liées au SIME par des études de liaisons génétiques, mais pour lesquelles aucune variation génétique n'a été identifiée, nous avons comme projet d'analyser ces familles en utilisant la nouvelle technologie de séquenceur à haut débit. Les possibilités offertes par cette percée technologique semblent très prometteuses. Une étude récente a pu identifier des mutations dans une région chromosomique (40 Mb) précédemment liée à la rétinopathie vitreuse<sup>158</sup>. En utilisant une approche similaire, nous pensons qu'il serait possible d'élucider la cause génétique du SIME dans plusieurs familles et d'impliquer

de nouveaux gènes dans la pathogenèse de la maladie. L'identification de ces gènes pourrait permettre d'identifier des interactions moléculaires ayant un fort impact sur la maladie qui deviendrait ainsi une importante cible pour de nouvelles approches thérapeutiques.

## 8. Conclusion

Nous avons réalisé la première étude génomique ayant pour but d'identifier l'impact des aberrations chromosomiques sur le Syndrome des Impatiences Musculaires de l'Éveil (SIME). En utilisant une première méthode de détection des CNVs par le biais des puces de génotypes, nous avons pu identifier six régions avec une occurrence plus élevée au niveau du SIME. De ces six régions, deux ont été validées par une deuxième méthode indépendante utilisant la technologie de l'hybridation par génomique comparée (CGH). Parmi ces deux régions se trouve une région sur le chromosome 9 qui a précédemment été identifiée par d'autres études, aussi bien par associations génétiques que par liaisons génétiques. Le jumelage de tous ces indices amassés jusqu'à maintenant permet de conclure que la présence des CNVs dans les exons non-codants du gène *PTPRD* pourrait agir comme facteur de susceptibilité pour le développement du SIME. Le SIME étant une maladie génétique complexe, il n'est pas surprenant qu'elle puisse être causée par différents types de facteurs génétiques. En plus de partiellement lever le voile sur la pathogenèse du Syndrome des Impatiences Musculaires de l'Éveil, cette étude permet aussi d'apporter un nouvel indice quant à l'implication possible des aberrations chromosomiques comme causes potentielles de plusieurs phénotypes.

## 9. Références

1. Korn, J.M., *et al.* Integrated genotype calling and association analysis of SNPs, common copy number polymorphisms and rare CNVs. *Nat Genet* **40**, 1253-1260 (2008).
2. Allen, R.P., *et al.* Restless legs syndrome prevalence and impact: REST general population study. *Arch Intern Med* **165**, 1286-1292 (2005).
3. Winkelman, J.W., Finn, L. & Young, T. Prevalence and correlates of restless legs syndrome symptoms in the Wisconsin Sleep Cohort. *Sleep Med* **7**, 545-552 (2006).
4. Hornyak, M. & Trenkwalder, C. Restless legs syndrome and periodic limb movement disorder in the elderly. *J Psychosom Res* **56**, 543-548 (2004).
5. Tison, F., *et al.* Epidemiology of restless legs syndrome in French adults: a nationwide survey: the INSTANT Study. *Neurology* **65**, 239-246 (2005).
6. Phillips, B., Hening, W., Britz, P. & Mannino, D. Prevalence and correlates of restless legs syndrome: results from the 2005 National Sleep Foundation Poll. *Chest* **129**, 76-80 (2006).
7. Allen, R.P., *et al.* Restless legs syndrome: diagnostic criteria, special considerations, and epidemiology. A report from the restless legs syndrome diagnosis and epidemiology workshop at the National Institutes of Health. *Sleep Med* **4**, 101-119 (2003).
8. Trenkwalder, C., Paulus, W. & Walters, A.S. The restless legs syndrome. *Lancet Neurol* **4**, 465-475 (2005).
9. Winkelmann, J., *et al.* "Anxietas tibiaram". Depression and anxiety disorders in patients with restless legs syndrome. *J Neurol* **252**, 67-71 (2005).
10. Aul, E.A., Davis, B.J. & Rodnitzky, R.L. The importance of formal serum iron studies in the assessment of restless legs syndrome. *Neurology* **51**, 912 (1998).
11. Sun, E.R., Chen, C.A., Ho, G., Earley, C.J. & Allen, R.P. Iron and the restless legs syndrome. *Sleep* **21**, 371-377 (1998).
12. Restless legs syndrome: detection and management in primary care. National Heart, Lung, and Blood Institute Working Group on Restless Legs Syndrome. *Am Fam Physician* **62**, 108-114 (2000).
13. Silber, M.H. & Richardson, J.W. Multiple blood donations associated with iron deficiency in patients with restless legs syndrome. *Mayo Clin Proc* **78**, 52-54 (2003).
14. Manconi, M., *et al.* Pregnancy as a risk factor for restless legs syndrome. *Sleep Med* **5**, 305-308 (2004).
15. Manconi, M., *et al.* Restless legs syndrome and pregnancy. *Neurology* **63**, 1065-1069 (2004).
16. Lee, K.A., Zaffke, M.E. & Baratte-Beebe, K. Restless legs syndrome and sleep disturbance during pregnancy: the role of folate and iron. *J Womens Health Gen Based Med* **10**, 335-341 (2001).
17. Kavanagh, D., Siddiqui, S. & Geddes, C.C. Restless legs syndrome in patients on dialysis. *Am J Kidney Dis* **43**, 763-771 (2004).
18. Richert A, Osuna E, Sateia M & B, R. A survey of sleep and restless legs in haemodialysis, peritoneal dialysis and renal transplant population. *Sleep* **21**(1998).

19. Huiqi, Q., Shan, L. & Mingcai, Q. Restless legs syndrome (RLS) in uremic patients is related to the frequency of hemodialysis sessions. *Nephron* **86**, 540 (2000).
20. Collado-Seidel, V., *et al.* Clinical and biochemical findings in uremic patients with and without restless legs syndrome. *Am J Kidney Dis* **31**, 324-328 (1998).
21. Takaki, J., *et al.* Clinical and psychological aspects of restless legs syndrome in uremic patients on hemodialysis. *Am J Kidney Dis* **41**, 833-839 (2003).
22. O'Keefe, S.T., Gavin, K. & Lavan, J.N. Iron status and restless legs syndrome in the elderly. *Age Ageing* **23**, 200-203 (1994).
23. Guyatt, G.H., *et al.* Diagnosis of iron-deficiency anemia in the elderly. *Am J Med* **88**, 205-209 (1990).
24. Wang, J., O'Reilly, B., Venkataraman, R., Mysliwiec, V. & Mysliwiec, A. Efficacy of oral iron in patients with restless legs syndrome and a low-normal ferritin: A randomized, double-blind, placebo-controlled study. *Sleep Med* **10**, 973-975 (2009).
25. Montplaisir, J., Lorrain, D. & Godbout, R. Restless legs syndrome and periodic leg movements in sleep: the primary role of dopaminergic mechanism. *Eur Neurol* **31**, 41-43 (1991).
26. Earley, C.J., *et al.* Abnormalities in CSF concentrations of ferritin and transferrin in restless legs syndrome. *Neurology* **54**, 1698-1700 (2000).
27. Mizuno, S., Mihara, T., Miyaoka, T., Inagaki, T. & Horiguchi, J. CSF iron, ferritin and transferrin levels in restless legs syndrome. *J Sleep Res* **14**, 43-47 (2005).
28. Earley, C.J., Connor, J.R., Beard, J.L., Clardy, S.L. & Allen, R.P. Ferritin levels in the cerebrospinal fluid and restless legs syndrome: effects of different clinical phenotypes. *Sleep* **28**, 1069-1075 (2005).
29. Snyder, A.M., *et al.* Mitochondrial ferritin in the substantia nigra in restless legs syndrome. *J Neuropathol Exp Neurol* **68**, 1193-1199 (2009).
30. Ruottinen, H.M., *et al.* An FDOPA PET study in patients with periodic limb movement disorder and restless legs syndrome. *Neurology* **54**, 502-504 (2000).
31. Michaud, M., Soucy, J.P., Chabli, A., Lavigne, G. & Montplaisir, J. SPECT imaging of striatal pre- and postsynaptic dopaminergic status in restless legs syndrome with periodic leg movements in sleep. *J Neurol* **249**, 164-170 (2002).
32. Earley, C.J., Hyland, K. & Allen, R.P. Circadian changes in CSF dopaminergic measures in restless legs syndrome. *Sleep Med* **7**, 263-268 (2006).
33. Connor, J.R., *et al.* Altered dopaminergic profile in the putamen and substantia nigra in restless leg syndrome. *Brain* **132**, 2403-2412 (2009).
34. Fernstrom, J.D. & Fernstrom, M.H. Tyrosine, phenylalanine, and catecholamine synthesis and function in the brain. *J Nutr* **137**, 1539S-1547S; discussion 1548S (2007).
35. Zhao, H., *et al.* Spinal cord dopamine receptor expression and function in mice with 6-OHDA lesion of the A11 nucleus and dietary iron deprivation. *J Neurosci Res* **85**, 1065-1076 (2007).
36. Quiroz, C., *et al.* Up-regulation of striatal adenosine A(2A) receptors with iron deficiency in rats Effects on locomotion and cortico-striatal neurotransmission. *Exp Neurol*.
37. Schiffmann, S.N., Fisone, G., Moresco, R., Cunha, R.A. & Ferre, S. Adenosine A2A receptors and basal ganglia physiology. *Prog Neurobiol* **83**, 277-292 (2007).
38. Rizzo, V., *et al.* Dopamine agonists restore cortical plasticity in patients with idiopathic restless legs syndrome. *Mov Disord* **24**, 710-715 (2009).

39. Kim, J., *et al.* Prevalence of restless legs syndrome and associated factors in the Korean adult population: the Korean Health and Genome Study. *Psychiatry Clin Neurosci* **59**, 350-353 (2005).
40. Cho, Y.W., *et al.* Epidemiology of restless legs syndrome in Korean adults. *Sleep* **31**, 219-223 (2008).
41. Cho, S.J., *et al.* Restless legs syndrome in a community sample of Korean adults: prevalence, impact on quality of life, and association with DSM-IV psychiatric disorders. *Sleep* **32**, 1069-1076 (2009).
42. Rangarajan, S., Rangarajan, S. & D'Souza, G.A. Restless legs syndrome in an Indian urban population. *Sleep Med* **9**, 88-93 (2007).
43. Nomura, T., Inoue, Y., Kusumi, M., Uemura, Y. & Nakashima, K. Prevalence of restless legs syndrome in a rural community in Japan. *Mov Disord* **23**, 2363-2369 (2008).
44. Mizuno, S., Miyaoka, T., Inagaki, T. & Horiguchi, J. Prevalence of restless legs syndrome in non-institutionalized Japanese elderly. *Psychiatry Clin Neurosci* **59**, 461-465 (2005).
45. Karatas, M. Restless legs syndrome and periodic limb movements during sleep: diagnosis and treatment. *Neurologist* **13**, 294-301 (2007).
46. Vetrugno, R., Provini, F. & Montagna, P. Restless legs syndrome and periodic limb movements. *Rev Neurol Dis* **3**, 61-70 (2006).
47. Wetter, T.C. & Pollmacher, T. Restless legs and periodic leg movements in sleep syndromes. *J Neurol* **244**, S37-45 (1997).
48. Leissner, L. & Sandelin, M. Periodic limb movements in sleep: to treat or not to treat? *Sleep Med* **3 Suppl**, S27-30 (2002).
49. Stefansson, H., *et al.* A genetic risk factor for periodic limb movements in sleep. *N Engl J Med* **357**, 639-647 (2007).
50. Krishnan, P.R., Bhatia, M. & Behari, M. Restless legs syndrome in Parkinson's disease: a case-controlled study. *Mov Disord* **18**, 181-185 (2003).
51. Ondo, W.G., *et al.* Daytime sleepiness and other sleep disorders in Parkinson's disease. *Neurology* **57**, 1392-1396 (2001).
52. Tan, E.K., Lum, S.Y. & Wong, M.C. Restless legs syndrome in Parkinson's disease. *J Neurol Sci* **196**, 33-36 (2002).
53. Chervin, R.D., *et al.* Associations between symptoms of inattention, hyperactivity, restless legs, and periodic leg movements. *Sleep* **25**, 213-218 (2002).
54. Philipsen, A., Hornyak, M. & Riemann, D. Sleep and sleep disorders in adults with attention deficit/hyperactivity disorder. *Sleep Med Rev* **10**, 399-405 (2006).
55. Wagner, M.L., Walters, A.S. & Fisher, B.C. Symptoms of attention-deficit/hyperactivity disorder in adults with restless legs syndrome. *Sleep* **27**, 1499-1504 (2004).
56. Sevim, S., *et al.* Correlation of anxiety and depression symptoms in patients with restless legs syndrome: a population based survey. *J Neurol Neurosurg Psychiatry* **75**, 226-230 (2004).
57. Auger, C., Montplaisir, J. & Duquette, P. Increased frequency of restless legs syndrome in a French-Canadian population with multiple sclerosis. *Neurology* **65**, 1652-1653 (2005).
58. Gomez-Choco, M.J., *et al.* Prevalence of restless legs syndrome and REM sleep behavior disorder in multiple sclerosis. *Mult Scler* **13**, 805-808 (2007).

59. Manconi, M., *et al.* High prevalence of restless legs syndrome in multiple sclerosis. *Eur J Neurol* **14**, 534-539 (2007).
60. Gemignani, F., *et al.* Restless legs syndrome and polyneuropathy. *Mov Disord* **21**, 1254-1257 (2006).
61. Gemignani, F., Brindani, F., Vitetta, F., Marbini, A. & Calzetti, S. Restless legs syndrome in diabetic neuropathy: a frequent manifestation of small fiber neuropathy. *J Peripher Nerv Syst* **12**, 50-53 (2007).
62. Nineb, A., *et al.* Restless legs syndrome is frequently overlooked in patients being evaluated for polyneuropathies. *Eur J Neurol* **14**, 788-792 (2007).
63. Lopes, L.A., *et al.* Restless legs syndrome and quality of sleep in type 2 diabetes. *Diabetes Care* **28**, 2633-2636 (2005).
64. Merlino, G., *et al.* Association of restless legs syndrome in type 2 diabetes: a case-control study. *Sleep* **30**, 866-871 (2007).
65. Skomro, R.P., Ludwig, S., Salamon, E. & Kryger, M.H. Sleep complaints and restless legs syndrome in adult type 2 diabetics. *Sleep Med* **2**, 417-422 (2001).
66. Minai, O.A., Malik, N., Foldvary, N., Bair, N. & Golish, J.A. Prevalence and characteristics of restless legs syndrome in patients with pulmonary hypertension. *J Heart Lung Transplant* **27**, 335-340 (2008).
67. Schlesinger, I., Erikh, I., Avizohar, O., Sprecher, E. & Yarnitsky, D. Cardiovascular risk factors in restless legs syndrome. *Mov Disord* **24**, 1587-1592 (2009).
68. Winkelman, J.W., Shahar, E., Sharief, I. & Gottlieb, D.J. Association of restless legs syndrome and cardiovascular disease in the Sleep Heart Health Study. *Neurology* **70**, 35-42 (2008).
69. Xiong, L., Jang, K., Montplaisir, J., Levchenko, A., Thibodeau, P., Gaspar, C., Turecki, G., Rouleau, G.A., Canadian restless legs syndrome twin study. *Neurology* **68**, 1631-1633 (2007).
70. Desautels, A., Turecki, G., Montplaisir, J., Sequeira, A., Verner, A., Identification of a major susceptibility locus for restless legs syndrome on chromosome 12q. *Am J Hum Genet* **69**, 1266-1270 (2001).
71. Levchenko, A., Montplaisir, J., Asselin, G., Provost, S., Girard, S.L., Xiong, L., Lemyre, E., St-Onge, J., Thibodeau, P., Desautels, A., Turecki, G., Gaspar, C., Dube, M.P., Rouleau, G.A., Autosomal-dominant locus for restless legs syndrome in French-Canadians on chromosome 16p12.1. *Mov Disord* (2009) **24(1)**, 40-50
72. Levchenko, A., Provost, S., Montplaisir, J., Xiong, L., St-Onge, J., Thibodeau, P., Riviere, J.B., Desautels, A., Turecki, G., Dube, M.P., Rouleau, G.A., A novel autosomal dominant restless legs syndrome locus maps to chromosome 20p13. *Neurology* **67**, 900-901 (2006).
73. Bonati, M.T., *et al.* Autosomal dominant restless legs syndrome maps on chromosome 14q. *Brain* **126**, 1485-1492 (2003).
74. Kemlink, D., *et al.* Suggestive evidence for linkage for restless legs syndrome on chromosome 19p13. *Neurogenetics* **9**, 75-82 (2008).
75. Chen, S., *et al.* Genomewide linkage scan identifies a novel susceptibility locus for restless legs syndrome on chromosome 9p. *Am J Hum Genet* **74**, 876-885 (2004).
76. Dion, P.A., Daoud, H. & Rouleau, G.A. Genetics of motor neuron disorders: new insights into pathogenic mechanisms. *Nat Rev Genet* **10**, 769-782 (2009).
77. Grant, S.F. & Hakonarson, H. Genome-wide association studies in type 1 diabetes. *Curr Diab Rep* **9**, 157-163 (2009).

78. Austin, E.D., Loyd, J.E. & Phillips, J.A., 3rd. Genetics of pulmonary arterial hypertension. *Semin Respir Crit Care Med* **30**, 386-398 (2009).
79. Hardy, J., Lewis, P., Revesz, T., Lees, A. & Paisan-Ruiz, C. The genetics of Parkinson's syndromes: a critical review. *Curr Opin Genet Dev* **19**, 254-265 (2009).
80. Tom Strachan, A.P.R. *Human Molecular Genetics*, (Wiley-Liss, 1997).
81. AJF, G., JH, M., DT, S., RC, L. & WM, G. *An Introduction to Genetic Analysis*, (WH Freeman and Company, 1993).
82. Desautels, A., *et al.* Restless legs syndrome: confirmation of linkage to chromosome 12q, genetic heterogeneity, and evidence of complexity. *Arch Neurol* **62**, 591-596 (2005).
83. Winkelmann, J., *et al.* Evidence for further genetic locus heterogeneity and confirmation of RLS-1 in restless legs syndrome. *Mov Disord* **21**, 28-33 (2006).
84. Levchenko, A., *et al.* The 14q restless legs syndrome locus in the French Canadian population. *Ann Neurol* **55**, 887-891 (2004).
85. Ray, A. & Weeks, D.E. No convincing evidence of linkage for restless legs syndrome on chromosome 9p. *Am J Hum Genet* **76**, 705-707; author reply 707-710 (2005).
86. Lander, E. & Kruglyak, L. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nat Genet* **11**, 241-247 (1995).
87. Liebetanz, K.M., *et al.* RLS3: fine-mapping of an autosomal dominant locus in a family with intrafamilial heterogeneity. *Neurology* **67**, 320-321 (2006).
88. Lohmann-Hedrich, K., *et al.* Evidence for linkage of restless legs syndrome to chromosome 9p: are there two distinct loci? *Neurology* **70**, 686-694 (2008).
89. Schormair, B., *et al.* PTPRD (protein tyrosine phosphatase receptor type delta) is associated with restless legs syndrome. *Nat Genet* **40**, 946-948 (2008).
90. Pichler, I., *et al.* Linkage analysis identifies a novel locus for restless legs syndrome on chromosome 2q in a South Tyrolean population isolate. *Am J Hum Genet* **79**, 716-723 (2006).
91. Plomin, R., Haworth, C.M. & Davis, O.S. Common disorders are quantitative traits. *Nat Rev Genet* **10**, 872-878 (2009).
92. Becker, K.G., Barnes, K.C., Bright, T.J. & Wang, S.A. The genetic association database. *Nat Genet* **36**, 431-432 (2004).
93. A haplotype map of the human genome. *Nature* **437**, 1299-1320 (2005).
94. Winkelmann, J., *et al.* Genome-wide association study of restless legs syndrome identifies common variants in three genomic regions. *Nat Genet* **39**, 1000-1006 (2007).
95. Moskow, J.J., Bullrich, F., Huebner, K., Daar, I.O. & Buchberg, A.M. Meis1, a PBX1-related homeobox gene involved in myeloid leukemia in BXH-2 mice. *Mol Cell Biol* **15**, 5434-5443 (1995).
96. Hu, Y.L., Fong, S., Ferrell, C., Largman, C. & Shen, W.F. HOXA9 modulates its oncogenic partner Meis1 to influence normal hematopoiesis. *Mol Cell Biol* **29**, 5181-5192 (2009).
97. Mojsin, M. & Stevanovic, M. PBX1 and MEIS1 up-regulate SOX3 gene expression by direct interaction with a consensus binding site within the basal promoter region. *Biochem J* **425**, 107-116.
98. Xiong, L., *et al.* MEIS1 intronic risk haplotype associated with restless legs syndrome affects its mRNA and protein expression levels. *Hum Mol Genet* **18**, 1065-1074 (2009).

99. Ahmad, K.F., *et al.* Mechanism of SMRT corepressor recruitment by the BCL6 BTB domain. *Mol Cell* **12**, 1551-1564 (2003).
100. Bardwell, V.J. & Treisman, R. The POZ domain: a conserved protein-protein interaction motif. *Genes Dev* **8**, 1664-1677 (1994).
101. Riviere, J.B., *et al.* Association of intronic variants of the BTBD9 gene with Tourette syndrome. *Arch Neurol* **66**, 1267-1272 (2009).
102. Chang, L. & Karin, M. Mammalian MAP kinase signalling cascades. *Nature* **410**, 37-40 (2001).
103. Zhou, G., Bao, Z.Q. & Dixon, J.E. Components of a new human protein kinase signal transduction pathway. *J Biol Chem* **270**, 12665-12669 (1995).
104. Mehta, P.B., *et al.* MEK5 overexpression is associated with metastatic prostate cancer, and stimulates proliferation, MMP-9 expression and invasion. *Oncogene* **22**, 1381-1389 (2003).
105. Yu, M., *et al.* A developmentally regulated inducer of EMT, LBX1, contributes to breast cancer progression. *Genes Dev* **23**, 1737-1742 (2009).
106. Arndt, S., Poser, I., Moser, M. & Bosserhoff, A.K. Fussel-15, a novel Ski/Sno homolog protein, antagonizes BMP signaling. *Mol Cell Neurosci* **34**, 603-611 (2007).
107. Sampath, T.K., Muthukumar, N. & Reddi, A.H. Isolation of osteogenin, an extracellular matrix-associated, bone-inductive protein, by heparin affinity chromatography. *Proc Natl Acad Sci U S A* **84**, 7109-7113 (1987).
108. Chen, D., Zhao, M. & Mundy, G.R. Bone morphogenetic proteins. *Growth Factors* **22**, 233-241 (2004).
109. Pulido, R., Krueger, N.X., Serra-Pages, C., Saito, H. & Streuli, M. Molecular characterization of the human transmembrane protein-tyrosine phosphatase delta. Evidence for tissue-specific expression of alternative human transmembrane protein-tyrosine phosphatase delta isoforms. *J Biol Chem* **270**, 6722-6728 (1995).
110. Wang, J. & Bixby, J.L. Receptor tyrosine phosphatase-delta is a homophilic, neurite-promoting cell adhesion molecular for CNS neurons. *Mol Cell Neurosci* **14**, 370-384 (1999).
111. Chan, T.A. & Heguy, A. The protein tyrosine phosphatase receptor D, a broadly inactivated tumor suppressor regulating STAT function. *Cell Cycle* **8**, 3063-3064 (2009).
112. Kohno, T., *et al.* A catalog of genes homozygously deleted in human lung cancer and the candidacy of PTPRD as a tumor suppressor gene. *Genes Chromosomes Cancer*.
113. Veeriah, S., *et al.* The tyrosine phosphatase PTPRD is a tumor suppressor that is frequently inactivated and mutated in glioblastoma and other human cancers. *Proc Natl Acad Sci U S A* **106**, 9435-9440 (2009).
114. McEvoy, B.P. & Visscher, P.M. Genetics of human height. *Econ Hum Biol* **7**, 294-306 (2009).
115. Manolio, T.A., *et al.* Finding the missing heritability of complex diseases. *Nature* **461**, 747-753 (2009).
116. Sanger, F., Nicklen, S. & Coulson, A.R. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* **74**, 5463-5467 (1977).
117. Finishing the euchromatic sequence of the human genome. *Nature* **431**, 931-945 (2004).
118. Metzker, M.L. Sequencing technologies - the next generation. *Nat Rev Genet* **11**, 31-46.

119. Cook, E.H., Jr. & Scherer, S.W. Copy-number variations associated with neuropsychiatric conditions. *Nature* **455**, 919-923 (2008).
120. Kim, P.M., *et al.* Analysis of copy number variants and segmental duplications in the human genome: Evidence for a change in the process of formation in recent evolutionary history. *Genome Res* **18**, 1865-1874 (2008).
121. Lee, J.A., Carvalho, C.M. & Lupski, J.R. A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders. *Cell* **131**, 1235-1247 (2007).
122. Iafrate, A.J., *et al.* Detection of large-scale variation in the human genome. *Nat Genet* **36**, 949-951 (2004).
123. Redon, R., *et al.* Global variation in copy number in the human genome. *Nature* **444**, 444-454 (2006).
124. Perry, G.H., *et al.* The fine-scale and complex architecture of human copy-number variation. *Am J Hum Genet* **82**, 685-695 (2008).
125. Kidd, J.M., *et al.* Mapping and sequencing of structural variation from eight human genomes. *Nature* **453**, 56-64 (2008).
126. Glessner, J.T., *et al.* Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. *Nature* **459**, 569-573 (2009).
127. Elia, J., *et al.* Rare structural variants found in attention-deficit hyperactivity disorder are preferentially associated with neurodevelopmental genes. *Mol Psychiatry* (2009).
128. Vrijenhoek, T., *et al.* Recurrent CNVs disrupt three candidate genes in schizophrenia patients. *Am J Hum Genet* **83**, 504-510 (2008).
129. Bochukova, E.G., *et al.* Large, rare chromosomal deletions associated with severe early-onset obesity. *Nature* (2009).
130. Stranger, B.E., *et al.* Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science* **315**, 848-853 (2007).
131. Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls. *Nature* **464**, 713-720 (2010).
132. Thorisson, G.A., Smith, A.V., Krishnan, L. & Stein, L.D. The International HapMap Project Web site. *Genome Res* **15**, 1592-1593 (2005).
133. Wang, K., *et al.* PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Res* **17**, 1665-1674 (2007).
134. Bioinformatics, O. PennCNV-Affy Protocol for CNV detection in Affymetrix SNP arrays. Vol. 2009 (2009).
135. Venkatraman, E.S. & Olshen, A.B. A faster circular binary segmentation algorithm for the analysis of array CGH data. *Bioinformatics* **23**, 657-663 (2007).
136. Kent, W.J., *et al.* The human genome browser at UCSC. *Genome Res* **12**, 996-1006 (2002).
137. Hsu, F., *et al.* The UCSC Known Genes. *Bioinformatics* **22**, 1036-1046 (2006).
138. Pollard, K.S., Hubisz, M.J., Rosenbloom, K.R. & Siepel, A. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res* **20**, 110-121 (2010).
139. Siepel, A., *et al.* Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res* **15**, 1034-1050 (2005).
140. (C), A. CGH Interactive Analysis *Agilent Genomic Workbench 5.0*(2009).
141. Winchester, L., Yau, C. & Ragoussis, J. Comparing CNV detection methods for SNP arrays. *Brief Funct Genomic Proteomic* **8**, 353-366 (2009).

142. Alonso, A., *et al.* Protein tyrosine phosphatases in the human genome. *Cell* **117**, 699-711 (2004).
143. Andersen, J.N., *et al.* A genomic perspective on protein tyrosine phosphatases: gene structure, pseudogenes, and genetic disease linkage. *FASEB J* **18**, 8-30 (2004).
144. Mustelin, T., Vang, T. & Bottini, N. Protein tyrosine phosphatases and the immune response. *Nat Rev Immunol* **5**, 43-57 (2005).
145. Jacobsen, M., *et al.* A point mutation in PTPRC is associated with the development of multiple sclerosis. *Nat Genet* **26**, 495-499 (2000).
146. Smith, M., *et al.* Analysis of a 1-megabase deletion in 15q22-q23 in an autistic patient: identification of candidate genes for autism and of homologous DNA segments in 15q22-q23 and 15q11-q13. *Am J Med Genet* **96**, 765-770 (2000).
147. Minassian, B.A., *et al.* Mutations in a gene encoding a novel protein tyrosine phosphatase cause progressive myoclonus epilepsy. *Nat Genet* **20**, 171-174 (1998).
148. Orozco, L.D., *et al.* Copy number variation influences gene expression and metabolic traits in mice. *Hum Mol Genet* (2009).
149. Weterman, M.A., *et al.* Copy number variation upstream of PMP22 in Charcot-Marie-Tooth disease. *Eur J Hum Genet* (2009).
150. Diskin, S.J., *et al.* Copy number variation at 1q21.1 associated with neuroblastoma. *Nature* **459**, 987-991 (2009).
151. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. Basic local alignment search tool. *J Mol Biol* **215**, 403-410 (1990).
152. Hamosh, A., *et al.* Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res* **30**, 52-55 (2002).
153. Platzer, P., *et al.* Silence of chromosomal amplifications in colon cancer. *Cancer Res* **62**, 1134-1138 (2002).
154. Menezes, R.X., Boetzer, M., Sieswerda, M., van Ommen, G.J. & Boer, J.M. Integrated analysis of DNA copy number and gene expression microarray data using gene sets. *BMC Bioinformatics* **10**, 203 (2009).
155. Armengol, L., *et al.* Identification of copy number variants defining genomic differences among major human groups. *PLoS One* **4**, e7230 (2009).
156. Li, J., *et al.* Whole genome distribution and ethnic differentiation of copy number variation in Caucasian and Asian populations. *PLoS One* **4**, e7958 (2009).
157. Yim, S.H., *et al.* Copy number variations in East-Asian population and their evolutionary and functional implications. *Hum Mol Genet* **19**, 1001-1008.
158. Nikopoulos, K., *et al.* Next-generation sequencing of a 40 Mb linkage interval reveals TSPAN12 mutations in patients with familial exudative vitreoretinopathy. *Am J Hum Genet* **86**, 240-247.
159. National Human Genome Research Institute The Human Genome Project Completion : Frequently Asked Questions, <http://www.genome.gov/11006943>, consulté le 05/07/2010
160. Winkelmann J, Lichtner P, Schormair B, Uhr M, Hauk S, Stiasny-Kolster K, Trekwaldner C, *et al.* Variants in the neuronal nitric oxide synthase (nNOS, NOS1) gene are associated with restless legs syndrome. *Mov Disorders* **23**, 350-358
161. Uetani N, Chagnon MJ, Kennedy TE, Iwakura Y, Tremblay ML, Mammalian motoneuron axon targeting requires receptor protein tyrosine phosphatases sigma and delta, *J Neurosci* **31**, 5872-80

162. Uetani N, Kato K, Ogura H, Mizuna K, Kawano K, Mikoshiba K, Yakura H, Asano M, Iwakura Y, Impaired learning with enhanced hippocampal long-term potentiation in PTPdelta-deficient mice, *EMBO J* **15**, 2775-85