

# **Distribution et transmission de variants génétiques impliqués dans la marche atopique et dans l'asthme dans la population canadienne**

**Mémoire**

**Mathieu Simard**

**Maîtrise en sciences cliniques et biomédicales de l'Université Laval offerte en  
extension à l'Université du Québec à Chicoutimi  
Maître ès. Sciences (M.Sc.)**

Québec, Canada

© Mathieu Simard, 2020

# **Distribution et transmission de variants génétiques impliqués dans la marche atopique et dans l'asthme dans la population canadienne**

**Mémoire**

**Mathieu Simard**

Sous la direction de :

Catherine Laprise, directrice de recherche  
Simon Girard, co-directeur de recherche

## RÉSUMÉ

La dermatite atopique (DA) est caractérisée par une barrière cutanée endommagée qui laisse pénétrer les allergènes ce qui peut mener à une sensibilisation accrue aux allergènes chez les individus ayant une prédisposition génétique et à une augmentation du risque de développer les autres phénotypes de la marche atopique, dont les allergies alimentaires, la rhinite allergique et l'asthme. Ce projet avait pour but de développer un indice de risque polygénique (de l'anglais *polygenic risk score* [PRS]) visant à identifier les nouveau-nés qui sont plus à risque de développer la DA dans la population canadienne afin de prévenir cette pathologie ainsi que sa progression dans la marche atopique. La cohorte régionale d'asthme du Saguenay–Lac-Saint-Jean et la cohorte de naissances CHILd qui incluent respectivement des données génomiques pour 1 200 individus canadiens français et 5 300 individus des provinces canadiennes anglophones (Vancouver, Edmonton, Winnipeg et Toronto) incluant diverses ethnies ont été employées pour les analyses. Deux scénarios ont été évalués : un avec les meilleurs loci de la couverture pangénomique (*Genome-wide association study* [GWAS] et données imputées) réalisée dans la population canadienne et un second avec les meilleures associations issues de la littérature scientifique. Les PRS des patients ont été calculés à partir des rapports des chances et du nombre d'allèles à risque pour chacun des loci sélectionnés. Un PRS avec une aire sous la courbe de 85% expliquant donc 36% de la variance dans la DA a été établi avec 25 loci à l'aide d'un modèle incluant l'âge, le sexe ainsi que l'ethnie parentale comme covariables. Un modèle avec uniquement les 6 variants les plus fortement associés (avec un p-value  $< 1 \times 10^{-40}$ ) pourrait également être utilisé car il préserve une bonne discrimination (AUC de 85%), ce qui renforce la possibilité de concevoir un outil diagnostique, un PRS pour la DA, qui permettra d'identifier les nouveau-nés (<3 mois) à risque de développer la DA dans l'optique de prévenir son développement et de surcroît la progression de la marche atopique.

## ABSTRACT

Atopic dermatitis (AD) is characterized by a damaged skin barrier that allows allergens to penetrate the body, leading to a sensitization and a higher risk to develop food allergy, asthma and/or allergic rhinitis, all characterizing the atopic march. This study aims to prevent these diseases by building a polygenic risk score (PRS) to identify newborns with a higher risk of developing AD in the Canadian population. The Saguenay–Lac-Saint-Jean (SLSJ) asthma cohort and the CHILD birth cohort, which includes respectively imputed GWAS data from 1,200 French-Canadian individuals and 5,300 individuals from Vancouver, Edmonton, Winnipeg and Toronto, have been used for analyses. The pooling of these two cohorts have increased the power of analysis and have considered the ethnic variability of the Canadian population. Loci associated with AD in the literature were tested with a general regression model and best associations will be included in the PRS. Patients PRS' were calculated according to the odds ratio and the number of risk allele(s) for selected loci. A ROC curve will be built to determine the cut-off threshold of the PRS, and analyses have been run on the risk categories and random samples, tested with and without potential covariates, to assess its sensitivity and specificity. A PRS having an area under the curve (AUC) of 85% and explaining 36% of AD variance was built with loci identified from GWAS results and with the contribution of covariates. The PRS developed is aimed to allow to discriminate newborns (<3 months old) that are at risk of developing AD in the perspective

# TABLE DES MATIÈRES

RÉSUMÉ .....	ii
ABSTRACT .....	iii
TABLE DES MATIÈRES.....	iv
LISTE DES FIGURES .....	vi
LISTE DES TABLEAUX.....	vi
LISTE DES ABBRÉVIATIONS .....	viii
REMERCIEMENTS .....	xi
AVANT-PROPOS .....	xii
INTRODUCTION .....	1
CHAPITRE 1 – GÉNOMES ET ÉTUDES D’ASSOCIATION PANGÉNOMIQUES .....	3
1.1 Composition et structure du génome humain.....	3
1.2 Motifs génomiques d’intérêts en analyse génétique .....	6
1.3 Études d’association pangénomiques.....	8
CHAPITRE 2 – ÉPIDÉMIOLOGIE ET GÉNÉTIQUE DES MALADIES ATOPIQUES ET DE L’ASTHME .....	10
2.1 Asthme et maladies atopiques : définitions .....	10
2.2 Développement de la marche atopique.....	12
2.3 Génétique de la dermatite atopique .....	14
2.4 Génétique de l’asthme .....	21
2.5 Prévalence de la dermatite atopique et de l’asthme .....	25
2.5.1 Dermite atopique .....	25
2.5.2 Asthme .....	26
CHAPITRE 3 – ÉTUDE DE COHORTES ET COHORTES .....	28
3.1 Études génétiques dans les cohortes .....	28
3.2 Histoire démogénétique de la population québécoise.....	28
3.3 Cohorte d’asthme du Saguenay–Lac-Saint-Jean.....	32
3.4 Histoire démogénétique de la population canadienne anglaise .....	33
3.4 Canadian Healthy Infant Longitudinal Development Study – Cohorte CHILD .....	34
CHAPITRE 4 – SCORE DE RISQUE POLYGÉNIQUE .....	36
4.1 Génétique à l’ère post-GWAS .....	36
4.2 Fonction d’efficacité du récepteur .....	39
CHAPITRE 5 – DÉVELOPPEMENT D’UN SCORE DE RISQUE POLYGÉNIQUE POUR LA DERMITE ATOPIQUE DANS LA POPULATION CANADIENNE .....	43

5.2 Abstract .....	45
5.3 Method and results .....	46
5.4 Conclusion.....	50
CHAPITRE 6 – AUTRES CONTRIBUTIONS .....	51
6.1 Identité par descendants .....	51
6.2 Intérêt et apports des segments IBD dans l’asthme.....	54
6.3 Étude des haplotypes de l’asthme dans la population du Saguenay–Lac-Saint-Jean par identité par descendants .....	55
DISCUSSION .....	65
CONCLUSION .....	69
Annexe 1.....	71
Annexe 2.....	75
Impact of paternal age at conception on human health .....	76
Résumé .....	77
Abstract .....	77
Introduction.....	78
Testicular morphological changes.....	79
Mutations in spermatogonial cells .....	80
<i>De novo</i> mutations .....	81
Selfish sperm selection .....	82
Telomere length .....	83
A link between molecular biology mechanisms and evolutionary biology .....	84
Chromosomal aneuploidies.....	85
Paternal age effect disorders .....	86
Autism Spectrum Disorders (ASD).....	86
Schizophrenia.....	87
Childhood Cancers.....	87
Epigenetics.....	88
BIBLIOGRAPHIE.....	90

## LISTE DES FIGURES

Figure 1 Progression des manifestations de la marche atopique dès l'enfance [39].....	12
Figure 2 Composantes génétiques et environnementales de la dermatite atopique [66].....	19
Figure 3 Prévalence mondiale de la dermatite atopique [54] .....	26
Figure 4 Prévalence mondiale de l'asthme [121].....	27
Figure 5 Courbe de Kaplan-Meier du risque cumulatif (%) en fonction de l'âge [150] .....	37
Figure 6 Structure d'une courbe ROC [162] .....	40
Figure 7 Chromosomes homologues avec segments IBD partagés [187] .....	52
Figure 8 Schématisation de la distribution des IBD en fonction de divers scénarios démographiques [188]....	53

## LISTE DES TABLEAUX

Tableau 1 Classification des ARN non codants du génome humain [5].....	3
Tableau 2 Architecture du génome humain [7].....	4
Tableau 3 Proportion de l'héritabilité expliquée par des loci associés par GWAS [31] .....	9
Tableau 4 Gènes associés à la dermatite atopique dans la littérature scientifique [54] .....	16
Tableau 5 Variantes génétiques des interleukines identifiées dans la dermatite atopique .....	19
Tableau 6 miARN associés à la dermatite atopique [76].....	20
Tableau 7 Quelques gènes connus ayant été associés à l'asthme [87].....	22
Tableau 8 Apport des différentes vagues migratrices sur la composition du génome québécois contemporain [129].....	29
Tableau 9 Quelques traits mendéliens à prévalence et pénétrance plus élevée dans la population canadienne-française [136] .....	31
Tableau 10 Description des différents types de famille composant la cohorte d'asthme SLSJ.....	32
Tableau 11 Nombre d'individus affectés par famille et par phénotypes dans la cohorte <i>d'asthme du SLSJ</i> [137] .....	33
Tableau 12 Origine des immigrants venus s'établir au Canada au 20 <sup>e</sup> siècle [145].....	34
Tableau 13 Caractéristiques des parents participants dans la cohorte CHILD [147] .....	35
Tableau 14 Les 25 meilleures associations pangénomiques dans la littérature.....	72
Tableau 15 Les 25 meilleures associations pangénomiques dans les cohortes canadiennes combinées .....	73
Tableau 16 Caractéristiques phénotypiques des participants de deux cohortes canadiennes .....	74





## LISTE DES ABBRÉVIATIONS

ADN : acide désoxyribonucléique

ARN : acide ribonucléique

ARNm : ARN messenger

ARNr : ARN ribosomique

ASD : autism spectrum disorder (trouble du spectre de l'autisme)

AUC : area under the curve (aire sous la courbe)

BRE : B recognition element (élément de reconnaissance B)

CC : case-control (cas-témoin)

CHILD : Canadian Healthy Infant Longitudinal Development study

CIUSS : centre intégré universitaire en santé et services sociaux

CNS : central nervous system (système nerveux central)

crasiARN : ARN lié au séquence centromérique

DA : dermite atopique

DL : déséquilibre de liaison

FN : proportion de faux négatifs

FP : proportion de faux positifs

GINA : global initiative for asthma

GSC : germ stem cells (cellules souches germinales)

GWAS : genome-wide association analysis

IBD : identité par descendants

IBS : identité par état

IgE : immunoglobuline E

Inr : élément initiateur

LINEs : long interspersed nuclear elements

lncARN : long ARN non-codant

LTR : long tandem repeat

MFQLS : family-based association analysis

miARN : micro-ARN

msAD : moderate to severe atopic dermatitis

ncARN : ARN non-codant

NRI : net reclassification index

NTA : nombre de patients réellement atteints

NT : nombre total de patients

NTN : nombre de patients réellement non atteints

OR : odds ratio

PARC : predicting asthma risk in children

piARN : ARN interagissant avec Piwi

PRS : Polygenic Risk Score

ROC : receiver operating characteristic

SCZ : schizophrenia

SINEs : short interspersed nuclear elements

SLSJ : Saguenay–Lac-Saint-Jean

snARN : petit ARN nucléolaire

SNP : single-nucleotide polymorphism

SRP : score de risque polygénique

TDT : family trio

Tel-sARN : petit ARN télomérique

TSS : site d'initiation de la transcription

VN : proportion de vrais négatifs

VP : proportion de faux négatifs

WHO : world health organization

*“We now have the possibility of achieving all  
we ever hoped for from medicine”*  
-David Sainsbury

## REMERCIEMENTS

Suite à ces deux années de maîtrise, j'ai eu l'énorme, et surtout l'incroyable chance de rencontrer et de travailler avec des personnes exceptionnelles. Parmi ces personnes, je tiens à remercier mes codirecteurs de maîtrise Catherine Laprise et Simon Girard pour votre dévotion et votre temps précieux que vous m'avez accordé tout au long de mon parcours. Merci de m'avoir appris à persévérer et à ne pas abandonner à la première embûche. Vos encouragements et conseils m'ont permis de grandir et de m'épanouir sur le plan personnel.

Je tiens également à remercier l'équipe du laboratoire dans laquelle j'ai évolué. Premièrement, un gargantuesque merci à Anne-Marie Madore pour toutes les heures que j'ai passé dans ton bureau à te demander des conseils et de l'aide qui m'ont été plus qu'utiles. Un gros merci également à Jolyane Meloche pour les cours enseignés. Merci à Joanie Bouchard pour tous les moments passés ensemble à explorer la bioinformatique. Merci également à tous les autres membres de ce laboratoire : Anne-Marie Boucher-Lafleur, Marie-Ève Lavoie, Pierre-Alexandre Gagnon ainsi que Mbarka Bchtenia.

Un énorme merci à mon amie Sabrina Bourget-Gaudreault qui a su me comprendre, me soutenir et m'encourager tout au long de ses deux années de maîtrise. Ton soutien et ton support externe nous ont permis de traverser et de cheminer dans ce long périple qu'est la maîtrise.

Ces remerciements ne pourraient s'achever sans remercier Claudia Moreau. Merci d'avoir littéralement passé milles heures avec moi à essayer de m'enseigner la bioinformatique « online » et également pour tous tes conseils, ta patience et ta disponibilité malgré la distance. Sans toi, je ne crois pas que ce mémoire aurait pu prendre forme.

# AVANT-PROPOS

Ce mémoire de maîtrise est sous-divisé en six chapitres qui font état des recherches auxquelles j'ai contribué lors de ces deux dernières années sur la génétique de la marche atopique. Il se concentre sur ce qui est reconnu comme le premier phénotype de cette marche atopique, la dermatite atopique (DA), au dernier statut clinique de celle-ci, l'asthme. Le chapitre 1 aborde l'architecture du génome humain ainsi que les études d'association pangénomiques. Le chapitre 2 poursuit sur les différentes définitions cliniques de l'asthme et des maladies atopiques et traite de leur épidémiologie génétique. Le chapitre 3 trace un survol des études génétiques effectuées dans les cohortes et décrit les deux cohortes utilisées pour les recherches de ce mémoire. Le chapitre 4 se concentre sur les statistiques génétiques, leur utilité clinique en portant une attention particulière aux indices de risque polygéniques (PRS). Par la suite, l'article de recherche : développement d'un score de risque polygénique pour la dermatite atopique dans la population canadienne (*polygenic risk score for atopic dermatitis in the Canadian population*) est présenté dans le cinquième chapitre. Le chapitre 6 ouvre sur une autre contribution qui ont a été effectuée en lien avec l'étude génétique utilisant des cohortes familiales. Spécifiquement, un article traitant des identités par descendants dans l'asthme dans la population du Saguenay–Lac-Saint-Jean. Les dernières parties de ce mémoire comprennent la discussion générale abordant les forces et les limites de notre étude et la conclusion trace les principales découvertes en plus d'ouvrir sur certaines perspectives. De plus, deux annexes sont également greffées à la fin. La première fournit du matériel supplémentaire en lien avec l'article scientifique du chapitre 5. De son côté, l'annexe 2 présente une revue de la littérature scientifique sur l'impact de l'âge paternel sur la santé humaine (publiée dans la revue *Clinical Chemistry* en janvier 2019).

# INTRODUCTION

Les coûts associés aux différents soins de santé et traitements exercent une pression sur les sociétés contemporaines. Avec une explosion démographique et un vieillissement rapide de sa population, le Canada n'échappe pas aux problèmes de santé globaux. Dans les dernières décennies, l'essor de la génétique médicale a permis de mieux caractériser et identifier les phénotypes d'une multitude de pathologies. Cette connaissance de la génétique individuelle va permettre d'adapter les traitements médicaux en fonction des caractéristiques génomiques pour viser la meilleure intervention possible [1]. Cependant, pour certaines maladies dites complexes, l'analyse génétique peut s'avérer être compliquée car plusieurs autres facteurs viennent influencer le cours de ces maladies soit, par exemple, l'impact de l'environnement sur l'expression des gènes, les interactions gène-gène ou encore gène-environnement. À ce jour, avec plus de 2500 études d'association pangénomiques effectuées sur diverses populations, plus de dix milles associations génotype-phénotype ont été effectuées [2]. Ces nombres peuvent sembler incommensurables mais des questions demeurent toujours sans réponse. De nombreux outils statistiques ont été développés pour permettre aux médecins d'interpréter et de mieux employer ces outils dans le système de santé. Une de ces maladies dites complexe est la dermite atopique. Cette maladie atopique est caractérisée par une rougeur et une démangeaison cutanée pouvant mener à des complications irréversibles affectant la qualité de vie des personnes qui en sont atteintes. Ultimement, une grande proportion des jeunes enfants qui en souffrent vont progresser dans une autre maladie nommée marche atopique. Par une simple affection cutanée les enfants vont donc développer des allergies alimentaires, de l'asthme ainsi qu'une rhinite allergique venant donc aggraver le phénotype initial. Pour contrer ce problème et, potentiellement, contribuer à aider certains patients, les objectifs visés par cette étude sont donc de :

- 1) Développer un outil de prédiction pour la dermatite atopique et les manifestations de la marche atopique basé sur la génétique des individus de la population canadienne.
- 2) Identifier les enfants à risque de développer la dermatite atopique et, de manière subséquente, la marche atopique dans la population canadienne.

Également, dans une perspective à long terme, permettre aux cliniciens de mettre en œuvre cet outil prédictif dans le but d'améliorer la qualité de vie des patients atopiques en leur prodiguant les soins médicaux appropriés.

Pour mettre en œuvre ces objectifs, deux cohortes distinctes de la population canadienne ont été employées : la cohorte d'asthme du Saguenay–Lac-Saint-Jean et la cohorte *Canadian Healthy Infant Longitudinal Development* [CHILD] [3, 4]. De plus, les hypothèses de recherche sont les suivantes :

Hypothèse nulle ( $H_0$ ) : Il n'y a pas d'incidence tant qu'au choix des marqueurs génétiques utilisés (littérature ou données pangénomiques) quant à l'identification des enfants qui sont le plus à risque de développer la dermatite atopique.

Hypothèse alternative ( $H_1$ ) : Les marqueurs génétiques issus des données imputées pangénomiques permettent de mieux identifier les enfants à risque de développer la dermatite atopique car ils prennent en compte la structure génétique de la population.

# CHAPITRE 1 – GÉNOMES ET ÉTUDES D'ASSOCIATION PANGÉNOMIQUES

## 1.1 Composition et structure du génome humain

Le génome humain diploïde est composé de 22 paires de chromosomes dits non sexuels nommés autosomes ainsi que d'une paire de chromosomes sexuels (gonosomes). Ces 46 chromosomes représentant 3,2 milliards de paires de bases (répartis sous forme de bases nucléiques adénines [A], guanines [G], cytosines [C] et thymines [T]) et environ 30 000 gènes, codant pour autant de protéines, sont retrouvés tout au long de cette séquence nucléotidique (Human Genome Project, 2018). D'autres éléments génomiques sont retrouvés avec un nombre de copies plus ou moins important tout au long du génome. De ces éléments, diverses séquences régulatrices (éléments régulateurs) sont réparties à travers le génome humain et certaines sont regroupées dans la grande classe des acides ribonucléiques non codants [ncARN]. Les sous-classes retrouvées chez la famille des ncARN sont présentées dans le tableau 1.

**Tableau 1** Classification des ARN non codants du génome humain [tiré de 5]

Sous-classes d'ARN non codants (ncARN)	Fonctions et caractéristiques génomiques	Taille (Nucléotides [nt])
Longs ARN non codants (lncARN)	Régulation épigénétique et génétique	> 200 nt
Petits ARN interférents (siARN)	Régulation génique, transposons et défense contre les pathogènes viraux	~ 21-22 nt
Micro-ARN (miARN)	Associé à la protéine Argonaute et agit sur la régulation post-traductionnelle des gènes	~ 22 nt
ARN interagissant avec Piwi (piARN)	Régulation des transposons et de la chromatine dans les cellules germinales	~ 26-30 nt
ARN associé aux promoteurs	Rôle potentiel dans la régulation de l'expression génique	~ 200-500 nt
Petits ARN nucléaires (snARN)	Guides pour la méthylation à l'ARNr et la pseudouridylation	~ 60-45 nt
Petits ARN télomériques (tel-sARN)	Maintenance et stabilisation des télomères	~ 24 nt
ARN associé au centrosome (crasiARN)	Guide dans les processus locaux de modification de la chromatine	~ 34-42 nt



Les éléments mobiles dispersés (transposons, rétrotransposons à séquence terminale longue répétée, longs éléments nucléaires intercalés et les petits éléments nucléaires intercalés) comptent pour environ 45% de l'acide désoxyribonucléique [ADN] humain. Finalement, des éléments mobiles répétés en tandem (minisatellites et microsatellites) (~6%) et des régions intergènes (25%) sont les derniers éléments qui composent le génome humain (Tableau 1). En plus de cette séquence de 3,2 milliards de paires de bases, une séquence circulaire additionnelle d'ADN mitochondrial de 16 569 paires de bases est retrouvée à l'intérieur des mitochondries [6].

**Tableau 2** Architecture du génome humain [tiré de 7]

Éléments génomiques	Longueur (pb)	Nombre de copies	Fraction du génome
Gènes codants pour des protéines	0,5 – 2 200 kpb	~ 21 000	~ 40% <sup>†</sup> (2% <sup>††</sup> )
Gènes codants pour de longs ARN non-codants (lncARN)	0,2 – 50 kpb	~ 10 000	~ 15%
ARN ribosomal (ARNr)	43 kpb	~ 300	0,4%
Petits ARN nucléaire U2 (ARNsn U2)	6,1 kpb	~ 20	< 0,0001%
Séquences simples (microsatellites, minisatellites)	1 – 500 pb	Variable	~ 6%
Transposons	2 – 3 kpb	300 000	3%
Rétrotransposons à LTR	6 – 11 kpb	440 000	8%
Longs éléments nucléaires intercalés (LINEs)	6 – 8 kpb	860 000	21%
Petits éléments nucléaires intercalés (SINEs)	100 – 400 pb	1 600 000	13%
Pseudogènes processés	Variable	~ 12 500	~ 0,4%
Régions intergéniques	Variable	Inconnu	~ 25%

<sup>†</sup>En incluant les introns et les exons

<sup>††</sup>En incluant seulement les exons

La structure typique d'un gène eucaryote comprend une région amplificatrice, une région promotrice ainsi qu'une région transcriptionnelle. La région amplificatrice d'un gène comprend plusieurs amplificateurs (de tailles variables pouvant aller de 50 pb à 1 500 pb) qui ont pour fonction d'activer (ou bien de réprimer) le promoteur et la transcription du gène dans un type cellulaire particulier ou encore, à un moment précis, lors du développement où d'autres produits protéiques sont requis par les cellules par la liaison de protéines nommées facteurs de transcription [8].

En amont du site d'initiation de la transcription [**TSS**], à environ 35 paires de bases, se trouve le promoteur du gène, élément critique à la régulation de l'expression génique chez les eucaryotes [9]. Les promoteurs sont reconnus par les facteurs de transcription grâce à des motifs nucléotidiques distinctifs tels les boîtes TATA (ou également boîtes de Goldberg-Hogness) et les régions riches en CpG. D'autres motifs nucléotidiques existent également dans certains promoteurs eucaryotes [**Inr**, **BRE**] [10]. En aval du promoteur, une courte séquence d'acides nucléiques nommée région 5' non traduite) est retrouvée et y joue un rôle dans la régulation de l'expression génique en ajoutant un second contrôle dans la transcription d'ARN messenger [**ARNm**] [11]. La région 5' non traduite peut s'avérer être un handicap pour le profil mutationnel dû à leur séquence relativement longue dans les gènes eucaryotes augmentant considérablement, dans le un tiers des cas, un allongement du produit protéique sur la séquence N-terminale [12]. La section codante (ou unité transcriptionnelle) est composée de séquences introniques et exoniques en alternance. Après épissage de l'unité transcriptionnelle, les introns sont retirés et les exons sont regroupés pour former les ARNm retrouvés dans le cytoplasme. Les gènes humains sont ceux qui possèdent le plus d'introns ainsi que les introns contenant le plus de nucléotides avec une moyenne de 3 413 nucléotides par intron (quoique la longueur la plus commune soit de 75 à 150 nucléotides dans la plupart des gènes) [13]. En regroupant les 23 paires de chromosomes, il est estimé que la longueur intronique totale est de

1 123 657 235 paires de bases, avec le chromosome 7 contenant à lui seul 20 537 introns, pour seulement 39 841 315 paires de bases d'exons dans le génome diploïde complet [14]. Ce déséquilibre introns-exons semble être bénéfique pour la cellule malgré le coût bioénergétique élevé à leur stabilité. En effet, les introns sont essentiels à la régulation positive de l'expression des gènes, dans le contrôle de la qualité de la dégradation des ARNm non-sens et d'augmenter l'efficacité de la sélection naturelle en augmentant les possibilités de recombinaisons [15].

## **1.2 Motifs génomiques d'intérêts en analyse génétique**

De nombreux marqueurs génomiques sont utilisés en génétique des populations pour mettre en évidence certaines distinctions génétiques interindividuelles et des structures populationnelles. En estimant qu'un génome individuel comporte environ 4,1-5,0 millions de paires de bases qui diffèrent de celui de référence établie par le Projet Génome Humain (Human Genome Project, 2018), il est primordial d'être en mesure d'identifier ces différences plus en détail pour établir des associations et des liens avec des pathologies et des structures de populations [16]. Les microsatellites (séquences simples répétées) sont de courtes séquences nucléotidiques de 1 à 6 paires de bases répétées à plusieurs reprises et dispersées dans tout le génome et représentant environ 3% de ce dernier et hérités selon une transmission mendélienne [17, 18]. Plusieurs motifs de microsatellites semblent être plus récurrents dans le génome humain. Les unités monomériques répétées d'adénine sont les plus fréquentes tandis que les unités dimériques AC et AT sont les plus communes. Pour les répétitions de plus longue séquence, les trimères AAT et AAG sont les plus fréquents, les tétramères AAAT, AAAG, AAAC et AAGG sont les plus communs à l'échelle du génome tandis que les pentamères AAAAT et AAAAC sont retrouvés en majorité [17]. L'analyse de ces séquences simples répétées permet donc de faire ressortir certaines régions génomiques. En allongeant la séquence nucléotidique des microsatellites, on retrouve les répétitions en tandem en nombre variable (**VNTR**)

qui sont classés comme microsatellites et minisatellites [19, 20]. Ces séquences de 6 à plus de 50 nucléotides sont réparties à travers le génome mais spécialement plus rassemblées dans les régions télomériques des chromosomes. Leur grande nature polymorphique les rend utiles pour des analyses d'association et l'étude des traits monogéniques [21]. Une autre classe de marqueur génomique qui a gagné une certaine notoriété depuis les années 1980 est celle des polymorphismes d'un seul nucléotide (**SNP** – *Single Nucleotide Polymorphism*). Avec l'avènement de l'étude des traits complexes dans les années 1990, les SNP se sont montrés plus efficaces pour l'étude des maladies multifactorielles que les minisatellites et microsatellites qui sont plus propices à subir un taux de mutation plus élevé en raison de leur taille [22]. Les éléments simples répétés étaient utilisés pour identifier de potentielles régions du génome qui pouvait avoir un lien avec l'état de la pathologie mais les informations qu'ils révélaient ne renseignaient pas sur le phénotype à proprement parlé, un changement de marqueur s'avérait donc nécessaire [22]. À l'opposé, les SNP peuvent se retrouver dans les sections codantes et donc engendrer des conséquences phénotypiques qui seront, dans quelques cas, visibles. En date de 2015, 84,7 millions de SNP ont été répertoriés dans le génome humain et représentent 99,9% des différences entre les génomes soulignant leur importance ainsi que leur contribution pour de futures études génétiques et génomiques au côté d'autres variations structurales touchant plusieurs bases à la fois [16]. Un autre élément utilisé en génomique est le génome mitochondrial (mitogénome), une séquence d'ADN exclusivement transmise par la mère [23]. Le mitogénome ne subit pas de recombinaisons contrairement au génome et rend donc la tâche plus facile pour retracer des liens entre des individus [24]. Plus de 300 variants ont été recensés à ce jour dans le mitogénome soulevant donc la contribution de cette courte séquence de 16,5 kpb [25].

### 1.3 Études d'association pangénomiques

Historiquement, la première étude d'association pangénomique (**GWAS** – *Genome-wide Association Study*) remonte au début des années 2000 avec la publication d'un article scientifique par une équipe japonaise, dirigée par le généticien Yusuke Nakamura, portant sur l'infarctus du myocarde [26]. Cette première tentative d'étude à l'échelle globale du génome a marqué un grand pas en génétique moderne. La structure d'une GWAS est basée sur une approche expérimentale visant à identifier des associations entre un trait et des variants génétiques. Pour pouvoir effectuer une étude GWAS typique une population d'intérêt doit être recrutée. Sous des bases phénotypiques, un sous-groupe d'individus étant atteint d'une maladie sera classé comme patients (cas) et un second sous-groupe, composé d'individus sains, sera classé comme témoins. Ces deux sous-groupes d'individus seront alors échantillonnés pour obtenir une fraction de leur ADN dans l'optique d'un génotypage, c'est-à-dire, l'identification moléculaire des variations interindividuelles, appelées SNP devant posséder une fréquence allélique se situant aux alentours de 5% dans la population, entre les individus atteints et non atteints de la pathologie étudiée. À partir des informations récoltées lors du génotypage des individus, il est alors possible de cartographier les variations et, à l'aide d'outils statistiques, déterminer quel(s) SNP se démarquent plus particulièrement dans la maladie étudiée [27]. Pour qu'un SNP soit associé à une maladie, à l'échelle populationnelle, il doit être en déséquilibre de liaison avec le locus de susceptibilité, c'est-à-dire une forte association non-aléatoire entre les allèles du locus d'intérêt et le locus de susceptibilité (dans le cas d'une étude GWAS, on parle de tagSNP) [28]. Les études GWAS ont permis de mieux documenter plus d'une centaine de pathologies allant des maladies cardiovasculaires aux différents types de cancer et également de parfaire les connaissances aux sujets de gènes suspectés d'être impliqués dans certaines pathologies. En effet, seulement 9,5% des 348 premières études GWAS ont identifié des SNPs qui étaient localisés dans

les exons ou bien dans les régions non traduites 5' et 3' tandis qu'environ 45% des SNPs identifiés étaient retrouvés dans les introns et un autre 45% dans des régions associées à aucun gène [28]. Toutefois, malgré le fait que de nombreuses variations génétiques soient identifiées, les GWAS sont toujours peu prédictifs de l'héritabilité des traits complexes et l'expliquent seulement en partie (Tableau 3). Cela peut être expliqué par le fait que la sélection naturelle tend à faire disparaître les variants génétiques qui ont une trop grande influence sur la maladie [28]. D'autres auteurs mentionnent également le fait que les GWAS sont peu puissants face à l'identification de variants rares et ayant un effet moins prononcé sur le phénotype [29]. La contribution des études GWAS en génétique humaine a donc permis de développer des outils forts utiles comme les scores de risque polygénique (*Polygenic Risk Score* [**PRS**]), l'évaluation du déséquilibre de liaison, la randomisation mendélienne ou encore l'estimation de l'héritabilité pour des traits simples comme complexes [30].

**Tableau 3** Proportion de l'héritabilité expliquée par des loci associés par GWAS [adapté de 31]

Pathologies	Nombre de loci associés	%Héritabilité expliqué
Diabète de type 2	76	~ 10%
Cancer du sein	67	~ 14%
Arthrite rhumatoïde	48	~ 51%
Troubles bipolaires	56	~ 2%
Schizophrénie	108	~ 3-7%

## CHAPITRE 2 – ÉPIDÉMIOLOGIE ET GÉNÉTIQUE DES MALADIES ATOPIQUES ET DE L'ASTHME

### 2.1 Asthme et maladies atopiques : définitions

L'asthme et les maladies atopiques (eczéma, rhinite allergique et l'asthme atopique) sont des traits complexes dont l'apparition et le développement sont gouvernés par une multitude de facteurs environnementaux et génétiques ce qui rend la caractérisation et l'établissement d'un phénotype clinique difficile car, dans la plupart des cas, ces maladies présentent un degré complexe de sous-divisions phénotypiques (par exemple, faible, modéré ou sévère) classés selon des marqueurs génétiques et moléculaires. Les présentes définitions se basent sur les différentes définitions retrouvées dans la littérature scientifique et médicale.

La définition de l'asthme donnée par le *Global Initiative for Asthma (GINA)* est : « une maladie hétérogène, habituellement caractérisée par une inflammation chronique des voies respiratoires. Elle se manifeste par un historique de symptômes respiratoires tel un sifflement, un souffle court, un serrement à la poitrine et une toux qui varient en intensité sur une période de temps en relation avec une limitation chronique du débit d'air expiré » [32]. La définition peut être élargie pour prendre en compte les caractéristiques cellulaires de l'asthme en y ajoutant à la définition précédente : « où différents types cellulaires et éléments cellulaires y jouent un rôle en particulier, les mastocytes, les éosinophiles, les lymphocytes T, les macrophages, les neutrophiles ainsi que les cellules épithéliales » [33].

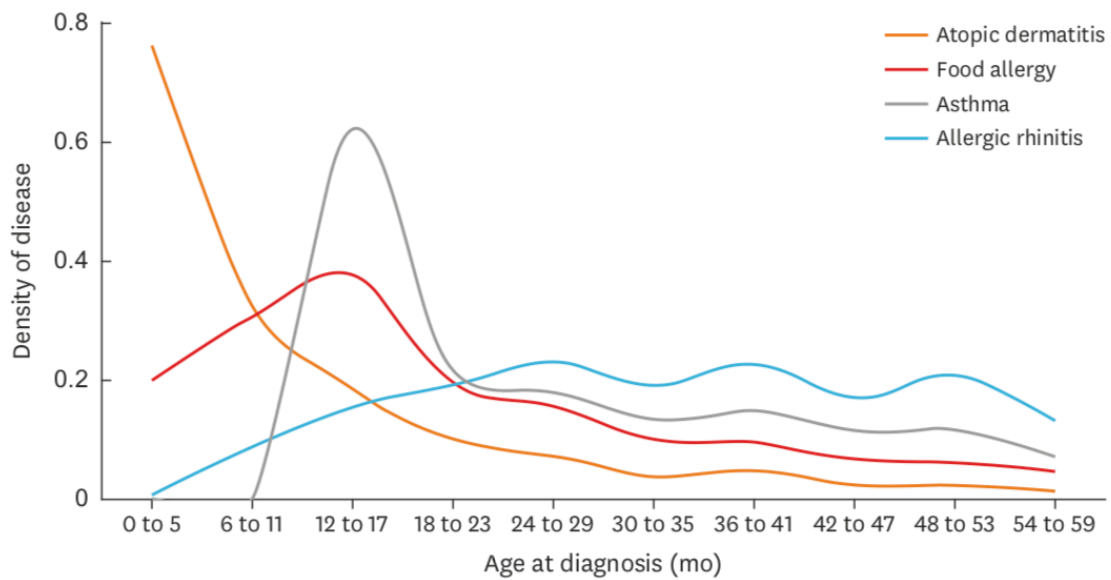
De leur côté, les termes « atopique » et « atopie » réfèrent, selon la *European Academy of Allergy and Clinical Immunology*, à : « une tendance individuelle et/ou familiale à développer, habituellement lors de l'enfance ou de l'adolescence, une sensibilité accrue accompagnée d'une production d'immunoglobuline E (**IgE**) prolongée en réponse à des allergènes, habituellement des protéines » [34]. En somme, les maladies atopiques sont caractérisées par une réponse immunitaire dirigée par les IgE. L'asthme atopique, la rhinite allergique ainsi que l'eczéma font partie de cette famille de maladie à prédispositions individuelles/familiales. Toutefois, le terme « eczéma » englobe un ensemble de maladies atopiques dont la dermatite atopique, l'eczéma atopique, l'eczéma juvénile, la dermatite atopiforme et la dermatite flexurale ce qui rend la définition de cette atopie confuse et complexe en raison de leur pathophysiologie toujours mal comprise [35]. Pour des raisons de compréhension et de rigueur, le terme privilégié pour l'eczéma est la dermatite atopique. La définition retenue pour le terme dermatite est : « manifestation clinique d'un prurit (démangeaison cutanée), d'un érythème (rougeur de la peau), de la présence de vésicules, d'une lichénification chronique (épaississement de la peau en plaques de coloration variant du rose au brun), d'une hyperkératose et de fissures cutanées » [36]. À l'heure actuelle, l'outil diagnostique clinique le plus utilisé par les dermatologues pour diagnostiquer la dermatite atopique et ses alias est le *UK Diagnostic Criteria for Atopic Dermatitis* (avec une spécificité de 97% et une sensibilité de 80%) qui est reconnu pour être valide dans les cas de diagnostics posés en milieux hospitaliers et dans les études basées sur des populations hétérogènes à l'échelle globale là où les autres outils diagnostiques ne se sont pas prouvés rigoureux. Pour obtenir un diagnostic de dermatite atopique, le patient doit présenter une irritation cutanée provoquant une démangeaison durant les 12 derniers mois ainsi que trois des cinq conditions suivantes : 1) des rides de peau (plis de coudes et de genoux, entre autres) présentant des symptômes eczémateux; 2) une histoire personnelle d'asthme ou de rhinite allergique; 3) une peau à tendance sèche dans la dernière année; 4) manifestation de symptômes eczémateux avant



l'âge de 2 ans et 5) une dermatite avec des lésions au niveau des grands plis flexuraux ou affectant la tête, le visage, le cou, les mamelons ou les membres [37].

## 2.2 Développement de la marche atopique

Le terme « marche atopique » réfère à la progression des symptômes de la dermatite atopique chez l'enfant en bas âge vers d'autres manifestations allergiques tels les allergies alimentaires, la rhinite allergique et l'asthme allergique [38]. Cette progression est initiée par la manifestation de symptômes de dermatite atopique dès la naissance suivie de près par l'apparition d'allergies alimentaires lors de l'introduction de nouveaux aliments après la période d'allaitement. Dès l'âge d'un an, l'incidence de l'asthme augmente pour atteindre son pic vers l'âge de 2 ans et s'accompagne, de manière générale, de l'apparition des symptômes de la rhinite allergique comme l'illustre la Figure 1.



**Figure 1** Progression des manifestations de la marche atopique dès l'enfance [tiré de 39]

Il est admis que seule une faible proportion des patients atteints de dermatite atopique vont progresser vers toutes les manifestations atopiques. Une étude génétique populationnelle britannique estime

qu'environ 7% des enfants présentant des manifestations de la dermatite atopique vont progresser dans la séquence complète de la marche atopique [40]. Toutefois, une proportion beaucoup plus significative d'enfants atteints (40%) vont développer des allergies alimentaires [38]. Les enfants étant allergiques aux arachides avaient 31% plus de risque de développer de l'asthme et 35% plus susceptibles de souffrir de rhinite allergique [41].

Majoritairement en raison d'une perte de fonction dans le gène de la filaggrine (*FLG*) qui résulte en une barrière cutanée endommagée chez les enfants atteints de dermatite atopique, les allergènes sont plus susceptibles de pénétrer le derme et d'induire une réponse spécifique de type IgE menant à la production d'antigènes spécifiques provoquant une sensibilisation avancée [42]. Comme les allergènes ne pénètrent pas les voies traditionnelles (tractus digestif) mais par la peau lésée, ils vont causer un élargissement des mastocytes intestinaux ainsi qu'une prédisposition à une anaphylaxie médiée aux IgE [42]. Le lien entre la dermatite atopique et les maladies inflammatoires des voies respiratoires semble reposer sur des loci communs et liés (*overlapped loci*) à ces différentes affections ainsi que sur l'hypothèse que les gènes impliqués sont de nature immunitaire et commune [38]. Récemment, une méta-analyse (analyse qui regroupe des données indépendantes de diverses études afin d'en tirer une conclusion) basée sur douze populations a mis en relation les variants rs9357733 (*EFHC1*) et rs993226 (*TMTC2-SLC6A15*) comme étant liés au phénotype commun asthme-dermatite atopique [43]. La réponse immunitaire et inflammatoire provoquée par la pénétration des allergènes dans la barrière cutanée lésée provoque une production accélérée de cytokines de type 2 (IL-4, IL-13, IL-25, IL-33 et de TSLP) menant ultimement à une circulation systémique de ces cytokines [38, 44].

Dans une optique de prévention globale de la dermatite atopique chez les nouveau-nés les équipes scientifiques et médicales ont évalué l'effet de l'application d'une crème émolliente sur la peau endommagée pour éviter la pénétration par les allergènes environnementaux et cela résulterait en une diminution de la prévalence de la dermatite atopique de 32% à 50% [45]. Des thérapies médicamenteuses visant à contourner la biosynthèse altérée de la profilaggrine (*FLG*) en venant corriger les mutations non-sens, en appliquant topiquement des métabolites entrant dans la biosynthèse de la FLG ou en stimulant la synthèse de cette protéine essentielle aux kératinocytes pourrait constituer des solutions à plus long terme pour les patients ne bénéficiant pas des effets protecteurs des crèmes émollientes [45, 46].

Dans le contexte de ce mémoire, sur la base de la structure des deux cohortes utilisées [cohorte de naissance (CHILD) et cohorte familiale d'asthme allergique (cohorte familiale d'asthme du Saguenay–Lac-Saint-Jean)], la phase initiale de la marche atopique soit la dermatite atopique de même que la dernière phase de cette marche, i.e. l'asthme seront étudiées. Pour cette raison, les sections suivantes ne traiteront que de l'état des données scientifiques sur ces deux phénotypes de la marche atopique. Comme la dermatite atopique [**DA**] est la première manifestation clinique de la marche atopique, son identification en bas âge (<3 mois) est un facteur clef dans le processus de prévention d'où l'intérêt de développer un score de risque polygénique [**SRP**] pour cette maladie.

### **2.3 Génétique de la dermatite atopique**

La dermatite atopique a une importante composante génétique et suit un mode de transmission non-Mendélien. Une étude populationnelle effectuée sur des jumeaux, au Danemark, a identifié un taux de concordance de 72% pour les jumeaux monozygotes ainsi que 23% chez les jumeaux dizygotes suggérant ainsi l'importance de la composante génétique [47]. Une empreinte quant à la

transmission parentale serait également reliée à la dermatite atopique. En effet, le statut atopique de l'enfant serait corrélé plus fortement avec celui de la mère que celui du père. Plusieurs hypothèses ont été mises en évidence pour tenter d'expliquer le plus grand impact maternel tel les gènes soumis à empreinte parental (*genomic imprinting*), l'environnement *in utero* ou encore la transmission mitochondriale [48]. Plusieurs gènes candidats ont été mis en évidence quant à leur rôle potentiel dans cette pathologie [35, 48]. Un de ces gènes, ayant été répliqué dans plus de 20 études indépendantes, est le gène de la profilaggrine situé dans le complexe de différenciation épidermique (1q21) et qui explique jusqu'à 10% de l'héritabilité de la dermatite atopique [49]. Les deux mutations causant une perte de fonction (*loss-of-function mutation*) R501X et 2282del4 expliquent respectivement 18% et 48% de la dermatite atopique modérée à sévère dans la population caucasienne [50]. Sa perte de fonction entraîne la formation incomplète de la couche de kératinocytes morts entourant les cellules du derme et ayant pour conséquence de laisser pénétrer les allergènes, les bactéries et les polluants plus facilement à travers la barrière cutanée [51, 52]. Plus de 40 mutations différentes résultant en une perte de fonction ont été répertoriées pour le gène *FLG* dans les populations caucasiennes et asiatiques ainsi que 289 variantes chez les afro-américains [53]. Bien que l'on estime que le risque de développer une dermatite atopique modérée à sévère soit évalué à 50% (héritabilité) dans la population européenne, plusieurs autres gènes ont été associés à la dermatite atopique. Le Tableau 4 résume les principales associations génétiques documentées pour la dermatite atopique.

**Tableau 4** Gènes associés à la dermatite atopique dans une revue de la littérature scientifique [adapté de 54]

Loci	Gènes	Fonctions géniques
1q21.2	<i>CIART</i>	Répression transcriptionnelle circadienne
1q21.3	<i>IL6R</i>	Sous-unité du récepteur IL-6, différenciation des lymphocytes B
1q21.3	<i>FLG</i>	Biosynthèse de la profilaggrine
2p13.2	<i>CD207</i>	Lectine de type C spécifique au mannose, présentation d'antigènes
2p16.1-p15	<i>PUS10</i>	Modification post-traductionnelle d'ARN structuraux et apoptose
2p25.1	<i>LINC00299</i>	Long ARN non-codant
	<i>IL1RL1</i>	Sous-unité du récepteur IL-33, impliqué dans la fonction des T <sub>H</sub>
2q12.1	<i>IL18R1</i>	Sous-unité du récepteur IL-18, régulation de l'inflammation via NF- $\kappa$ B
	<i>IL18RAP</i>	Sous-unité accessoire du récepteur IL-18
2q24.3	<i>XIRP2</i>	Protection contre la dépolymérisation des filaments d'actine
3p21.1	<i>RFT1</i>	Transport d'oligosaccharides, N-glycosylation des protéines
3p22.3	<i>CCR4</i>	Transport des leucocytes, récepteur aux chémokines CC
3q13.2	<i>CCDC80</i>	Adhésion cellulaire et assemblage de la matrice cellulaire
4q27	<i>IL2</i>	Différenciation, prolifération et activation des cellules immunitaires (lymphocytes B et T, macrophages et cellules NK)
	<i>IL21</i>	
5p13.2	<i>IL7R</i>	Sous-unité des récepteurs IL-7 et TSLP, promotion de la réponse T <sub>H</sub> 2
5q22.1	<i>TSLP</i>	Promotion de la réponse T <sub>H</sub> 2, peptides antimicrobiens oraux et cutanés
5q31.1	<i>IL4</i>	Prolifération des lymphocytes B et communication isotypique des IgE
5q31.1	<i>IL13</i>	Fonctions des T <sub>H</sub> 2
6p21.32	<i>HLA-DRB1</i>	Traitement des antigènes, présentation des cellules nucléées pour CMH 2
6p21.33	<i>HLA-B</i>	Traitement des antigènes, présentation des cellules nucléées pour CMH 1
7p22.2	<i>CARD11</i>	Activation du récepteur TCR des lymphocytes T et activation du NF- $\kappa$ B
8q21.13	<i>ZBTB10</i>	Régulation de la transcription
9p21.3	<i>DMRTA1</i>	Régulation de la transcription
10p15.1	<i>IL15RA</i>	Sous-unité du récepteur IL-15, stimulation et prolifération de lymphocytes
	<i>IL2RA</i>	Sous-unité du récepteur IL-2, stimulation et prolifération de lymphocytes
10q21.2	<i>ZNF365</i>	Cytokinèse mitotique et maintenance de la stabilité du génome
11p13-p12	<i>PRR5L</i>	Régulation de la phosphorylation de la protéine kinase C
11p15.4	<i>NLRP10</i>	Régulation de l'immunité innée, libération de cytokines pro-inflammatoire
11q13.5	<i>LRRC32</i>	Régulation des lymphocytes T régulateurs et activation de TGF $\beta$
11q24.3	<i>ETS1</i>	Facteur de transcription, contrôle des cytokines
14q13.2	<i>PPP2R30</i>	Régulation des protéines phosphatases et mort cellulaire induite par activation des lymphocytes B
16p13.13	<i>CLEC16A</i>	Lectine de type C, régulation de l'autophagie et mort mitochondriale
17q21.2	<i>STAT3</i>	Activation et médiation de la réponse cellulaire aux interleukines
17q21.32-q21.33	<i>ZNF652</i>	Régulation de la transcription

19p13.2	<i>ACTL9</i>	Fonctions cytosquelettiques
20q13.2	<i>CYP24A1</i>	Métabolisme de la vitamine D
20q13.33	<i>TNFRSF6B</i>	Protection contre l'apoptose

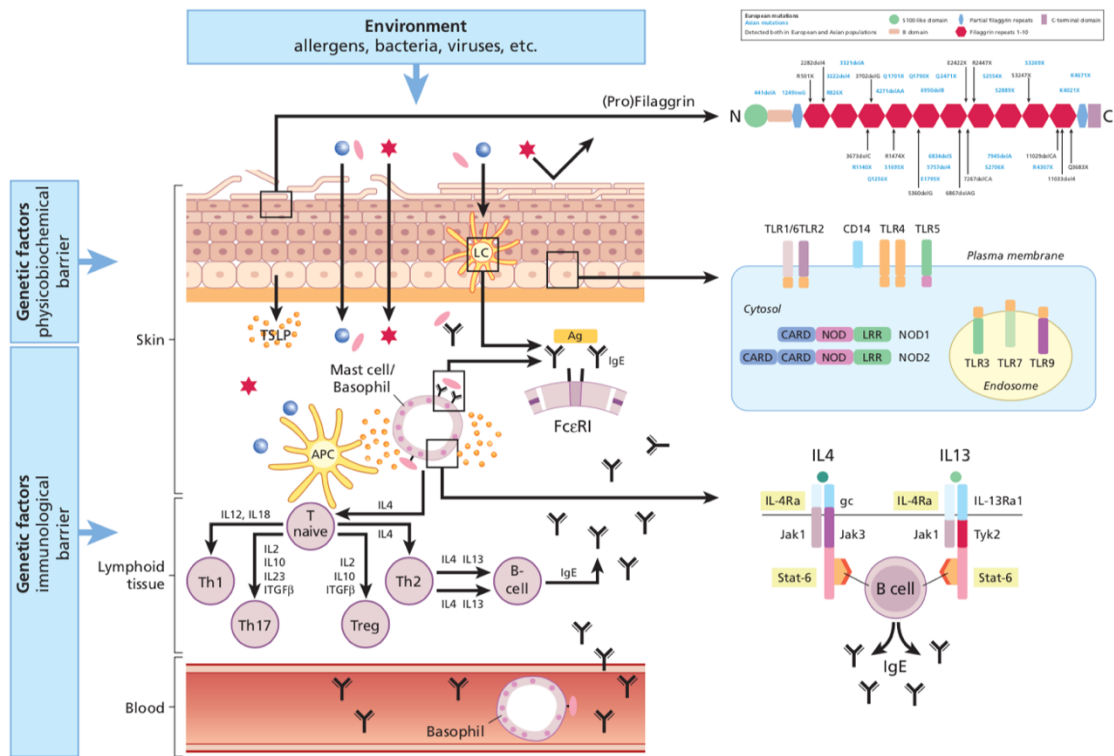
Les gènes associés à la dermatite atopique semblent jouer un rôle dans la composition de la barrière cutanée ainsi que dans la modulation de la réponse immunitaire étant généralement grandement affectés par les influences environnementales (allergènes, virus, polluants).

L'inhibiteur de la sérine protéase de type Kazal-5 (*SPINK5*), est un autre gène montrant une forte association dans la dermatite atopique. Retrouvé dans l'agglomération des inhibiteurs de la sérine peptidase sur le chromosome 5 codant pour la protéine LEKT1 (inhibiteur de sérine peptidase de type Kazal-5) qui joue un rôle dans la différenciation des kératinocytes et de la formation de l'épithélium [55]. Une perte de fonction chez *SPINK5* mène à une activité protéolytique plus élevée résultant en une inhibition du gène *FLG* et une augmentation de l'expression de la lymphopoïétine stromale thymique (*TSLP*) causant une réponse pro-inflammatoire [50]. Une mutation particulière, la Glu420Lys (1258G/A), causant une perte de fonction chez *LEKT1*, serait un facteur de risque accru pour la dermatite atopique [56]. Les gènes codant pour les récepteurs de type Toll (*TLR1* à *TLR11*) jouent un rôle dans la reconnaissance d'organismes pathogènes et dans la défense de l'organisme face à ces pathogènes en induisant une réponse inflammatoire aux cytokines et aux interférons de type 1 [57]. Sur le récepteur de la membrane plasmique TLR2, le variant non-sens Arg753Gln a été associée à un phénotype sévère de dermatite atopique en raison d'une élévation du taux d'IgE sanguin et d'une susceptibilité accrue à des infections causées par *Staphylococcus aureus* [58, 59]. Le gène *TLR4* sur le chromosome 9q32-q33, s'est montré deux fois moins exprimé chez des nouveau-nés aux Pays-Bas résultant en une diminution des taux d'interleukine 10, une cytokine impliquée dans une réduction de la réponse immunitaire [60]. Une méta-analyse basée sur neuf populations caucasiennes a mis en évidence les variants rs5743708 (*TLR2*) et rs4986790 (*TLR4*) comme

facteurs de risque mettant une plus grande emphase sur le rôle potentiel de cette famille de récepteurs dans la dermatite atopique [61]. Deux SNPs des récepteurs TLR7 et TLR8, situés sur la membrane plasmique des endosomes, qui ont pour fonction la reconnaissance d'ARN bactériens ont été associés, dans une étude danoise, à une susceptibilité accrue au développement de la dermatite atopique [62]. La variante C-127T du promoteur de l'homologue évolutif TLR9 des récepteurs TLR7 et TLR8 qui est spécifique aux motifs CpG de l'ADN bactérien a été mise en évidence dans un sous-type de dermatite atopique dans une étude allemande [63].

Les variantes Arg130Gln (*IL13*) et -589C/T (*IL4*) des récepteurs IL4R de la membrane plasmique des mastocytes et des basophiles qui sont activés par les interleukines 4 et 13 sont associés à risque plus élevé de développer la dermatite atopique et d'autres manifestations atopiques dès l'âge de 2 ans dans la population canadienne [64]. En date de l'année 2019, 40 interleukines différentes ont été identifiées dans la littérature scientifique comme étant associées à la dermatite atopique [65]. Comme plusieurs variants ont été identifiées comme modulateurs de l'activité immunitaire et inflammatoire dans les maladies atopiques, le Tableau 5 résume les différentes percées scientifiques qui ont été faites en lien avec le rôle potentiel des différentes interleukines dans l'atopie.

La Figure 2 représente une synthèse des différentes composantes génétiques et environnementales connues et documentées comme ayant une influence sur le développement de la DA [66].



**Figure 2** Composantes génétiques et environnementales de la dermatite atopique [tiré de 66]

**Tableau 5** Variantes génétiques des interleukines identifiées dans la dermatite atopique

Locus	Gènes	Variants	Population	Références
5q31-q33	<i>IL12B</i>	1188A/C	Japonaise	[67]
4q27	<i>IL2-IL21</i>	-	Japonaise	[68]
5q31	<i>IL4</i>	-590C/T Gln551Arg	Japonaise, allemande, canadienne États-unienne, japonaise, britannique	[64, 67] [67]
16p12	<i>IL4RA</i>	-3223C/T	Allemande, japonaise	[67]
5q31	<i>IL5</i>	-703C/T	Japonaise	[67]
1q21.3	<i>IL6R</i>	Asp358Ala	Britannique, allemande	[69]
2q12	<i>IL1RL1-IL8R1</i>	rs13015714	Japonaise	[70]
5q31.1	<i>IL9*</i>	-4091G/A	Coréenne (Corée du Sud)	[71]
Xq28/Yq12	<i>IL9R*</i>	rs3093467		
1q32.1	<i>IL10</i>	-819A/G -592A/C Arg130Gln	Coréenne (Corée du Sud) Canadienne, japonaise, allemande	[72] [64, 67]
5q31	<i>IL13</i>	-1111C/T	Néerlandaise	[67]
12q24.31	<i>IL31</i>	rs7977932	Taiwanaise	[73]

\*Interaction gène-gène IL9-IL9R des chromosomes 5 et XY



D'autres mécanismes de régulation du génome ont été proposés quant à leur rôle potentiel dans le développement de la dermatite atopique. De courtes séquences d'ARN nucléaires endogènes nommées micro-ARN [**miARN**]. Les miARN sont composés d'environ 22 nucléotides et sont impliqués dans la régulation génique via le complexe de répression induit par les miARN (RISC). Ce dernier vient inhiber l'activité des ARNm bloquant donc la transcription génique [74]. Un seul miARN peut bloquer plusieurs ARNm inhibant donc rapidement l'expression d'un gène et pouvant amener d'importantes conséquences physiologiques. Des micro-ARN (miARN) tel miARN-155, un miARN impliqué dans la régulation des réponses immunitaires innées et acquises qui présente une expression plus élevée dans les lésions cutanées des patients souffrant de dermatite atopique résultant en une augmentation de l'inflammation par une réponse stimulée des cellules T<sub>H</sub> [75]. D'autres miARN présents dans les échantillons sanguins de patients atteints de dermatite atopique ont été mis en relation avec la dermatite atopique et sont présentés dans le Tableau 6.

**Tableau 6** miARN associés à la dermatite atopique [adapté de 76]

<b>miARN à expression élevée</b>		<b>miARN à expression faible</b>	
let-7i	miARN-106b	miARN-122a	miARN-335
miARN-17-5p	miARN-146a	miARN-133a	miARN-483
miARN-20a	miARN-155	miARN-133b	miARN-515-5p
miARN-21	miARN-193a	miARN-215	miARN-519d
miARN-24	miARN-199a	miARN-326	
miARN-27	miARN-222		
miARN-29a			

Récemment, des études ont commencé à investiguer le rôle que pourraient avoir les longs ARN non codants [**lncARN**] dans la pathophysiologie de la dermatite atopique. Ces lncARN sont des ARN comprenant au moins 200 nucléotides et sont sous-classés en 3 catégories : 1) les petits lncARN avec 200 à 950 nucléotides, 2) les moyens lncARN qui comptent entre 950 à 4 800 nucléotides et 3) les grands lncARN de séquence d'un peu plus de 4 800 nucléotides. Les lncARN, comme les miARN, sont impliqués dans la régulation de l'expression des gènes [77]. Dans le cas de la dermatite atopique, quelques lncARN ont été identifiés par différentes équipes de recherche à l'échelle mondiale. Les lncARN humanlincRNA0016+, uc008thl.1, uc029qxr.1 et AK077345 possèdent une activité élevée contribuant à augmenter les réponses inflammatoires et immunitaires tandis que uc029ycn.1, ENSMUST00000164311 et ENSMUST00000149791 sont documentés comme étant moins exprimés chez les patients atopiques comparativement au niveau mesuré chez les témoins sans atopie. [78]. De plus, l'ARN non codant LINC00299 a été associé à la dermatite atopique dans quelques études GWAS et semble influencer sur la fonction cutanée endommagée chez les eczémateux [45, 79, 80].

## **2.4 Génétique de l'asthme**

Encore aujourd'hui, la composante génétique de l'asthme demeure le sujet de nombreuses études. L'héritabilité estimée par des études sur des jumeaux se retrouve entre 50% et 90%, mettant donc à l'évidence une composante génétique importante et toujours à documenter [81, 82]. Avec les nombreuses études GWAS effectuées depuis une trentaine d'années, de nombreuses variations génétiques ont été identifiées dans différentes populations à travers le monde [83-86]. Le Tableau 7 expose les différents gènes ayant été associés à l'asthme.

**Tableau 7** Gènes connus ayant été associés à l'asthme [adapté de 87]

Loci	Gènes	Fonctions géniques	Variants
1p13.3	<i>GSTM1</i>	Stress oxydatif et détoxification	+/nulle
1q21.3	<i>FLG</i>	Intégrité de la barrière épithéliale	Arg510X, 2282del4
1q31-32	<i>IL10</i>	Immunorégulation	-1082A/G, -571C/A
2q33	<i>CTLA4</i>	Inhibition de la réponse des lymphocytes T	-318C/T, 49A/G
5q31	<i>IL13</i>	Fonctions des T <sub>H</sub> 2	-1112C/T, Arg130Gln
5q31.1	<i>IL14</i>	Différentiation des T <sub>H</sub> 2 et induction des IgE	-589C/T, +33C/T
5q31.1	<i>CD14</i>	Immunité innée	-1721G/A, -260C/T
5q32	<i>SPINK5</i>	Inhibiteur de la protéase à sérine épithéliale	Glu420Lys
5q31-32	<i>ADRB2</i>	Relaxation des muscles lisses bronchiaux	Arg16Gly, Gln27Glu
5q33.2	<i>HAVCR1</i>	Régulation de la réponse des lymphocytes T	5583_5397del
5q35	<i>LTC4S</i>	Biosynthèse des leucotriènes cystéinés	-444A/C
6p21.3	<i>LTA</i>	Inflammation	Ncol (intron 1)
6p21.3	<i>TNF</i>	Inflammation	-308G/A, -857C/T
6p21	<i>HLA-DRB1</i>	Présentation d'antigènes	SNPs multiples
6p21	<i>HLA-DQB1</i>	Présentation d'antigènes	SNPs multiples
6p21	<i>HLA-DPB1</i>	Présentation d'antigènes	SNPs multiples
7p14.3	<i>GPRA</i>	Régulation de la croissance cellulaire	Haplotypes
8p22	<i>NAT2</i>	Détoxification des carcinogènes	SNPs faiblement Ac
11q13	<i>FCER1B</i>	Récepteur de haute affinité aux IgE	Ile181Leu, Gly237Glu
11q12.3-q13.1	<i>CC16</i>	Protéine anti-inflammatoire de l'épithélium	38A/G
11q13	<i>GSTP1</i>	Stress oxydatif et détoxification	Ile105Val
11q22.2-q22.3	<i>IL18</i>	Induction de l'interféron $\gamma$ et du TNF	-656T/G, -137G/C
12q13	<i>STAT6</i>	Signalisation de IL4 et IL13	2964G/A, (GT) <sub>n</sub> exon 1
12q24.2-q24.31	<i>NOS1</i>	Biosynthèse de l'oxyde nitrique	3391C/T, 5266C/T
14q11.2	<i>CMA1</i>	Sérine protéase des mastocytes	BstX1, -1903G1A
16p12.1-p12.2	<i>IL4R</i>	Chaîne $\alpha$ des récepteurs IL13 et IL4	Ile50Val, Glu551Arg
17q21.1-q21.2	<i>CCL11</i>	Chimioattracteur d'éosinophiles	Ala23Thr, -1328G/A
17q11.2-q12	<i>CCL5</i>	Chimioattracteur de monocytes et d'éosinophiles	-403A/G, -28C/G
17q23.3	<i>ACE</i>	Inactivation des médiateurs inflammatoires	Indel
19p13.3	<i>TBXA2R</i>	Contraction des muscles lisses, inflammation	924T/C, 795T/C
19q13.2	<i>TGFB1</i>	Immunorégulation et prolifération cellulaire	-509C/T
20p13	<i>ADAM33</i>	Interactions cellules-cellules et cellule-matrice	SNPs multiples
22q11.23	<i>GSTT1</i>	Stress oxydatif et détoxification	A/nulle

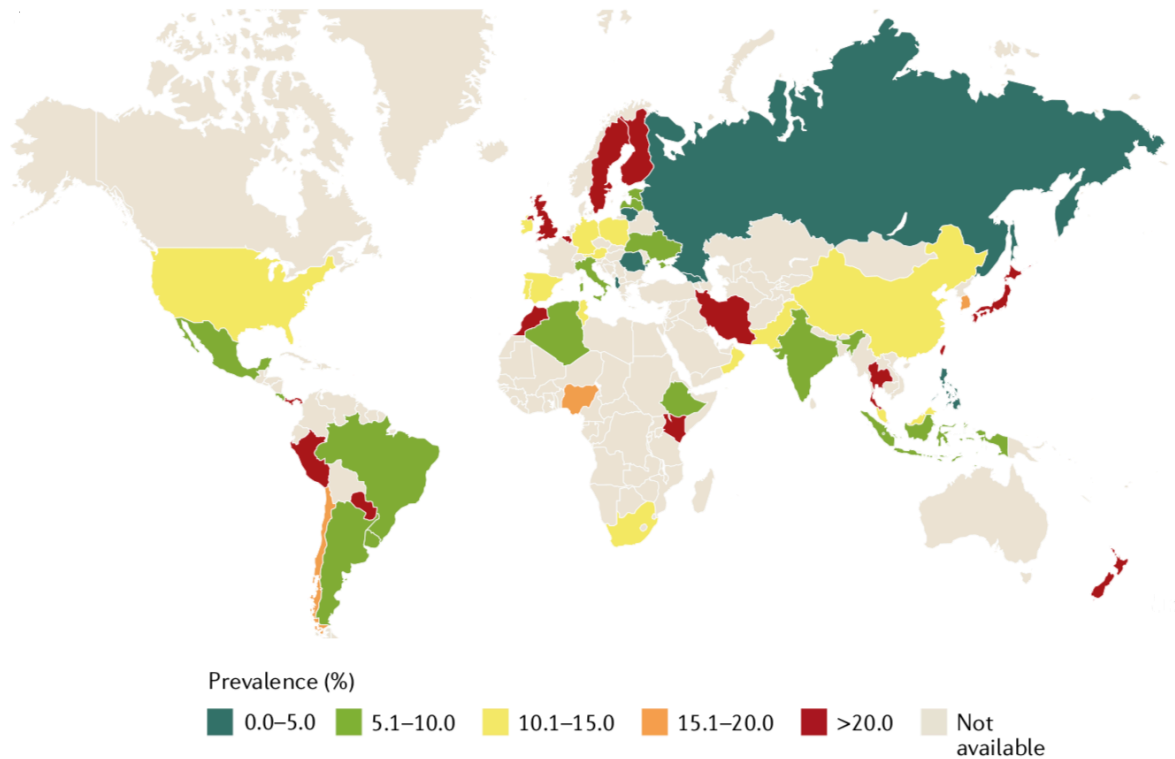
Les nombreux gènes candidats répertoriés dans le génome humain ayant une influence sur le développement de l'asthme allergique peuvent être sous-classés en trois catégories : les gènes influençant la composante immune, ceux influençant les voies inflammatoires et les gènes affectant le remodelage des voies respiratoires (*ADAM33*, *IL13*, *GSDMB*, *VEGF*, *PLAUR*) [88]. Dans la première sous-classe de gènes, trois régions génomiques particulières ont été associées plus significativement à l'inflammation des voies respiratoires soit 5q23-33, 5p15 et 12q14-24.2 [87-89]. La région étendue 5q23-33 comprend le gène *IL-9* qui a été identifié comme facteur de risque à une hypersensibilité bronchique chez l'humain [90]. Un lien entre le gène de l'interleukine 4 (*IL-4*) a été établi dans la population japonaise comme étant un autre facteur de risque dans le développement de l'asthme infantile [91]. D'autres gènes liés aux interleukines ont aussi été associés avec un risque accru de développer de l'asthme dans diverses études dont les gènes *IL-3*, *IL-5* et *IL-13*, tous impliqués dans les réponses immunitaires et inflammatoires [92, 93]. Ces interleukines étant les responsables de la production d'immunoglobuline E (IgE) par les cellules des voies respiratoires infiltrées par des lymphocytes T<sub>H</sub>2 et caractérisant la réponse inflammatoire observée [94]. La région chromosomique 5q32 contient le gène hautement polymorphique *ADRB2* codant pour l'adrénorécepteur  $\beta_2$ . Deux polymorphismes (Arg16Gly et Gln27Glu) semblent causer respectivement une bronchodilatation retardée ainsi qu'un phénotype d'asthme sévère tandis que le second polymorphisme semble être associé à l'asthme chez l'enfant ainsi qu'à une concentration élevée d'IgE sanguins [95]. Le complexe majeur d'histocompatibilité sur le chromosome 6p21 constitue un regroupement de gènes impliqués dans la présentation d'antigènes à la surface cellulaire et représente une des régions les plus associées à l'asthme [96]. En effet, plusieurs études ont mis en relation de nombreux polymorphismes susceptibles de jouer un rôle dans le développement et l'apparition de l'asthme comme les variantes HLA-DRB1\*11, HLA-DQB1\*0301, HLA-DQA1\*0505 et HLA-G\*964 [96-100]. De plus, étant donné leur grande association avec

l'asthme, les loci du locus HLA peuvent être couplés avec un endotype d'asthme particulier tel le gène *HLA-DRB1* avec l'asthme allergique, *HLA-DQB1* avec l'asthme occupationnel et *HLA-DPB1* avec l'asthme sensible à l'aspirine apportant donc des spécifications quant au type traitement nécessaire pour les patients [101]. Une autre région fortement associée est retrouvée sur le chromosome 11 (11q13-11q22.3) qui regroupe plusieurs loci associés à l'asthme. Ainsi, le gène *FCERB1* (rs2583476) codant pour des récepteurs à haute affinité pour les IgE a été associé à un risque accru de souffrir d'asthme chez les hommes pakistanais tandis qu'une diminution de l'expression du gène *CC16* semblerait être associée avec une fonction pulmonaire plus faible chez les espagnols [102, 103]. Plus récemment, la région chromosomique candidate 17q21 représente le locus le plus fréquemment associé à l'asthme dans les études GWAS. Cette région comprend les gènes candidats *ORMDL3*, *GSDMA* et *GSDMB* qui sont les gènes les plus associés et répliqués dans l'asthme infantile [104]. De façon intéressante, *ORMDL3* constitue une des plus fortes associations ayant été effectuée avec cette région du génome. Plusieurs chercheurs ont établi un lien entre l'expression d'*ORMDL3* dans les cellules épithéliales pulmonaires et un risque accru d'asthme [105-107]. Cette association s'est même avérée vraie dans trois populations ethniquement différentes [108]. Une possible interaction entre les gènes *ORMDL3* (rs8076131, rs12603332 et rs3744246) et *GSDMB* (rs7216389) impliquant les quatre SNP pouvant donc accroître de manière plus significative la dysfonction pulmonaire chez les asthmatiques [109]. Finalement, il importe de mentionner le gène *ADAM33* sur le chromosome 20p13 dont plusieurs variants ont été associés à l'asthme. De manière plus spécifique, le polymorphisme V4 a été associé à un risque plus élevé d'asthme chez la population caucasienne tandis que les polymorphismes Q1 et V4 sont associés à un risque élevé d'asthme chez les adultes [110].

## **2.5 Prévalence de la dermatite atopique et de l'asthme**

### **2.5.1 Dermite atopique**

D'un point de vue pharmacoéconomique, on estime que la dermatite atopique coûte, à elle seule, 1,4 milliard de dollars canadiens aux systèmes de santé provinciaux et territoriaux du Canada [111]. Comme les manifestations cliniques et phénotypiques de la dermatite atopique sont variables et que les définitions utilisées pour en définir son statut dans les différentes études le sont également, il est difficile d'obtenir des données précises quant à sa prévalence (Figure 3). Au Canada, il est estimé que la prévalence de la dermatite atopique se situe aux alentours de 10% (8,5% chez les 6-7 ans et 9,4% chez les 13-14 ans) [112]. Aux États-Unis, une étude populationnelle (basée sur un questionnaire demandant si l'enfant a été diagnostiqué par un médecin ou non) comprenant 102 353 enfants âgés de 0 à 17 ans a pris en considération l'âge, l'ethnie et le genre pour identifier une prévalence à l'échelle nationale. L'étude a conclu que la prévalence de la DA aux États-Unis était de 10,7% mais avec des variations entre les différents États allant de 8,7% jusqu'à 18,1% soulignant donc la variabilité accrue de cette maladie ainsi que ses manifestations phénotypiques variées [113].

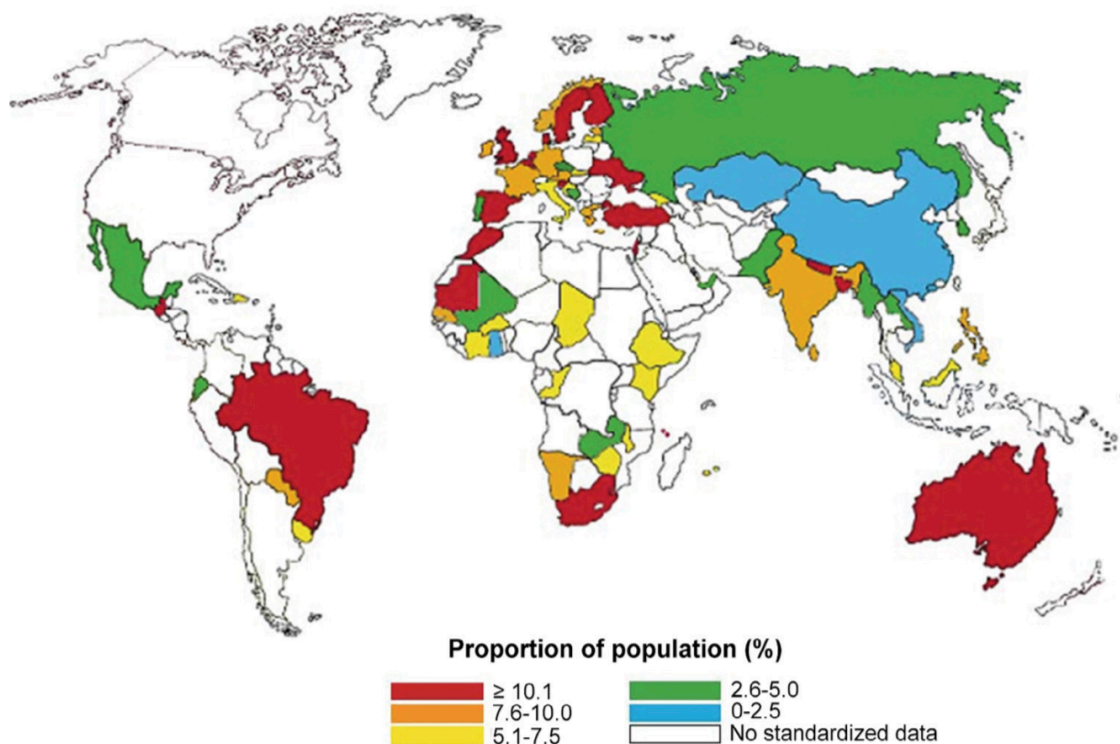


**Figure 3** Prévalence mondiale de la dermatite atopique [tiré de 54]

### 2.5.2 Asthme

L'asthme, en tant que maladie affectant la qualité de vie des patients qui en souffrent, représente un grand fardeau pour l'économie canadienne en termes de coûts directs et indirects sur les systèmes de santé des provinces et territoires avec, en 2012, un coût estimé à 2,1 milliards de dollars. L'estimation de son fardeau économique est revue à la hausse pour 2030 avec un coût approximatif sur les systèmes de santé de 4,2 milliards de dollars canadiens [114]. La prévalence de l'asthme est une donnée épidémiologique variable suivant les différentes études dans divers pays puisqu'elles ne sont pas soumises aux mêmes définitions phénotypiques et aux mêmes conditions environnementales (différentes saisons, pollution atmosphérique, zones industrielles ou bien la présence de diverses espèces végétales). En prenant les résultats de différentes études mondiales sur l'asthme chez les enfants âgés de 13-14 ans, on recense une prévalence de 15,3% en Australie,

11,5% en Angleterre, 9,5% en Finlande, 7,7% au Chili, 6,9% au Canada, 5,0% au Japon, 4,1% en Italie, 3,5% en Égypte et 1,8% en Chine [115]. Au Québec, en 2013-2014, selon le Ministère de la Santé et des Services sociaux, la prévalence de l'asthme chez les 12 ans et plus est estimée à 8,7% dans la population (9,7% pour les filles et 7,7% pour les garçons) [116]. La prévalence de l'asthme augmente depuis 1950 dans les pays développés et semble être corrélée avec l'industrialisation et l'urbanisation dans les pays en développement. Par exemple, au Canada, dans une période d'uniquement 3 ans (1980-1983), une augmentation de 2,7% en prévalence a été répertoriée [117-119]. Une autre découverte épidémiologique intéressante met en relation une prévalence plus élevée dans les pays anglophones (Australie, Nouvelle-Zélande, Angleterre, États-Unis et Irlande) qui n'a pas encore atteint de plateau comme celui situé entre 8-12% qui a été atteint dans les pays en voie de développement [118, 120].



**Figure 4** Prévalence mondiale de l'asthme [tiré de 121]



## CHAPITRE 3 – ÉTUDE DE COHORTES ET COHORTES

### 3.1 Études génétiques dans les cohortes

Les cohortes familiales puisqu'elles possèdent un environnement plus similaire sont de précieux outils en génétique des populations [122]. Outre leur coût plus important que les études cas-témoins, les études de cohortes sont plus complètes au niveau des renseignements de l'état de santé des patients car un suivi est assuré après quelques années (*follow-up*) ce qui ajoute des informations quant à la progression de certaines maladies [122]. De plus, elles permettent de mieux évaluer les interactions gènes-environnement possibles [28]. Par exemple, lorsque les facteurs de risque ont été identifiés et que les chercheurs désirent les associer à des biomarqueurs qui se développent plus tard dans le développement de l'individu (comme l'exposition à des polluants dans le milieu de travail ou encore l'effet de la cigarette sur la santé) les études de cohortes peuvent être pertinentes pour mieux tenir compte de l'évolution de la symptomatologie due aux différentes expositions environnementales. Malgré leurs points positifs, les études de cohorte comportent quelques biais notamment des biais de sélection, de mémoire, de surveillance (certaines maladies ont tendance à échapper aux examens de routine qui ne sont pas approfondis), d'exposition ou encore des biais de diagnostic [123].

### 3.2 Histoire démogénétique de la population québécoise

Le peuplement européen du Québec au XVII<sup>e</sup> siècle par les immigrants venus des régions du nord de la France, majoritairement de Normandie, de Bretagne et de Poitou-Charentes, a comme point de départ la ville de Québec avec l'arrivée, en 1608, de Samuel de Champlain et ses compatriotes [124]. En quelques décennies, une expansion rapide s'est fait le long des berges du fleuve Saint-Laurent de Québec jusqu'à Montréal. Suite à la conquête britannique de 1760, la province de

Québec tombe sous emprise anglaise et l'immigration française cesse donc drastiquement [125]. L'immigration française se voit remplacée par l'arrivée de migrants des îles et colonies britanniques ainsi que d'Allemagne [125, 126]. Suite à de nombreuses années où l'immigration européenne est demeurée relativement faible, ce n'est qu'au début des années 1800, qu'un fort taux de natalité et de nuptialité frappent de plein fouet la province de Québec. D'après le recensement de cette époque, la population totale est d'environ 210 000 habitants. La colonisation des régions intérieures du territoire québécois : le Saguenay–Lac-Saint-Jean, la Côte Nord, Charlevoix, l'Outaouais et la Mauricie est alors amorcée à partir de petits groupes de migrants sillonnant les terres hors de la vallée du Saint-Laurent [127, 128]. Les différentes vagues d'immigration ayant eu lieu au Québec ont façonné le patrimoine génétique de la population canadienne-française contemporaine (Tableau 8). La vague migratrice qui a le plus laissé sa trace sur la génétique québécoise est celle qui s'est déroulée à la naissance de la colonie française en Amérique du Nord, soit celle de 1608 à 1679.

**Tableau 8** Apport des différentes vagues migratrices sur la composition du génome québécois contemporain [tiré de 129]

<b>Vagues migratrices</b>	<b>Nombre d'immigrants</b>	<b>% dans l'ensemble des immigrants</b>	<b>Importance génétique</b>
1608-1679	3 380	6,9	68%
1680-1699	1 290	2,6	14%
1700-1729	1 480	3,0	6%
1730-1759	4 000	8,1	8%
1760-1799	4 000	8,1	2%
1800-1899	10 000	20,3	1%
1900-1949	25 000	50,9	1%
Total	49 150	100,0	100%

Avec la croissance rapide de la population, le dispersement géographique des fiefs s'accroît et de nombreuses paroisses sont ouvertes sur le territoire québécois. En 1675, la région de Charlevoix est ouverte à la colonisation. Malgré sa modeste croissance démographique et un maigre 11 767 habitants, il est estimé qu'environ 15% des couples recensés à cette époque dans Charlevoix ont immigré dans la région actuelle du Saguenay–Lac-Saint-Jean ce qui représente environ 13 800

Charlevoisiens (entre 1852 et 1911) [130-132]. En raison de son éloignement géographique et accentué par la rudesse de l'hiver, les mouvements migratoires demeurent faibles dans les premières années de la colonisation de la région. Cette faible vague migratrice en provenance de la région de Charlevoix pourrait laisser croire que la population régionale présente un taux de consanguinité plus élevé que dans les régions ayant reçu des vagues d'immigration plus diversifiées, laissant place à ce que l'on appelle le paradigme Saguenéen. Toutefois, en calculant les coefficients de consanguinité et de parenté moyens des Saguenéens qui étaient respectivement de  $17,7 \times 10^{-4}$  (avec  $6,5 \times 10^{-4}$  chez le groupe contrôle) et  $9,8 \times 10^{-4}$  ( $1,8 \times 10^{-4}$  chez les contrôles), on remarque que la population régionale n'est pas plus homogène que dans les autres régions du Québec [131]. Le fait que certaines maladies monogéniques soient plus présentes dans la région serait dû au triple effet fondateur régional (augmentant ainsi l'homogénéité de la population de part de la faiblesse de son pool génétique initial) se reflétant ainsi par une perte notable de diversité allélique à certains loci. Le premier effet fondateur correspond à celui des premiers migrants du nord de la France établis en Nouvelle-France dans les débuts de la colonie. Dès 1675, les 599 fondateurs français ayant immigrés dans Charlevoix sur sept générations sont à l'origine de la deuxième perte de diversité allélique avec une population ayant passée de 599 habitants à plus de 18 000 en seulement 180 ans. C'est lorsque que certains de ces migrants se sont dirigés vers le Saguenay, s'y sont établis et ont commencé à former des familles nombreuses en peu de temps que le troisième effet fondateur a pris place [133]. Plusieurs chercheurs ont mis en évidence les conséquences de ces effets fondateurs sur la population régionale. Citons, par exemple, que seulement trois mutations seraient responsables de 94% des cas de fibrose kystique dans la région ou encore qu'une seule mutation homozygote serait la cause de 80% des cas de tyrosinémie héréditaire de type I [133-135]. La bonne connaissance de la structure génétique et migratoire de la population québécoise a permis d'identifier 22 maladies mendéliennes avec une prévalence accrue au Québec [131]. Ces allèles délétères connus dans la

population du Québec peuvent donc être utilisés pour évaluer certains paramètres en génétique des populations. C'est ainsi que la mutation dans le gène *FAH* (tyrosinémie héréditaire de type I) peut être employée pour évaluer les effets démographiques sur la diversité allélique interrégionale et que ceux des gènes de la  $\beta$ -thalassémie (*HBB*) et de la maladie de Tay-Sachs (*HEXA*) pour identifier des migrants ayant amené certaines maladies non communes dans une région particulière [131]. Quelques-uns des traits mendéliens les plus fréquents ainsi que leurs allèles pathogènes à pénétrance élevée sont présentés dans le Tableau 9.

**Tableau 9** Quelques traits mendéliens à prévalence et pénétrance plus élevée dans la population canadienne-française [tiré de 136]

Trait mendélien	Transmission	Prévalence <sup>a</sup>	Gènes	Variants
Ataxie spastique de type Charlevoix-Saguenay	Récessive	1/1952	<i>SACS</i>	c.6594delT c.5254C/T c.2436delG
Neuropathie sensitivomotrice avec ou sans agénésie du corps calleux	Récessive	1/2117	<i>SLC12A6</i>	c.1584-1585delCTinsG
Syndrome de Leigh de type canadien-français	Récessive	1/2178	<i>LRPPRC</i>	c.1119C/T
Tyrosinémie de type I	Récessive	1/1846	<i>FAH</i>	IVS12+5G/A
Rachitisme vitamino-dépendant de type I	Récessive	1/2916	<i>CYP27B1</i>	c.958delG
Dystrophie musculaire oculo-pharyngée	Dominante	>1/7500	<i>PABPN1</i>	Répétition (GCN) <sub>n</sub>
Cystinose	Récessive	1/6237	<i>CTNS</i>	W138X, I133F
Dystrophie myotonique	Dominante	1/530	<i>Myotonine</i>	Répétition (GCN) <sub>n</sub>
Hypercholestérolémie familiale	Dominante	1/122	<i>LDL-R</i>	del>15kb W66G
Fibrose kystique	Récessive	1/902	<i>CFTR</i>	delF508, L206W
Hémochromatose héréditaire	Récessive	1/71	<i>HFE</i>	C282Y, H63D

<sup>a</sup> Dans la population du Saguenay–Lac-Saint-Jean [136]

### 3.3 Cohorte d'asthme du Saguenay–Lac-Saint-Jean

La cohorte d'asthme du Saguenay–Lac-Saint-Jean (abrégé SLSJ) est composée de 1 394 individus identifiés via des probants asthmatiques et répartis dans 271 familles indépendantes réparties sur le territoire de la région situé au nord-est du Québec (Tableau 10). La population du SLSJ est caractérisée par une succession de trois effets fondateurs récents combinés avec une isolation géographique ce qui contribue à la rendre attrayante pour des études en génétique des populations [131, 133]. La cohorte d'asthme du SLSJ comprend une collection de plus de 75 caractéristiques phénotypiques variées (âge, statut respiratoire, sexe, mesures de biomarqueurs, etc.) collectées depuis 20 ans [137]. La composition par phénotypes des individus affectés de la cohorte est présentée dans le Tableau 11. Depuis sa mise en place en 1997, la cohorte a identifié 22 SNP liés aux phénotypes d'asthme, d'atopie et de rhinite, 11 autres SNP liés aux différents types cellulaires des leucocytes ainsi que 59 loci pangénomiques [137]. Deux loci se sont également démarqué soit le locus 5q31.1 associé à l'atopie ainsi que le locus 17q21.2 avec 7 SNP associés à l'asthme. Globalement, la cohorte d'asthme du SLSJ a grandement contribué à la caractérisation génétique d'une partie de la population régionale par l'entremise de plusieurs études scientifiques d'envergure [3, 138-143].

**Tableau 10** Description des différents types de famille composant la cohorte d'asthme SLSJ

<b>Nombre de familles nucléaires</b>	484
Nombre de familles avec plus d'un enfant	151
<b>Nombre de familles indépendantes</b>	254
Nombre de familles avec plus de deux générations	131
Nombres de trios (parents et probants)	65

**Tableau 11** Nombre d'individus affectés par famille et par phénotypes dans la cohorte d'asthme du SLSJ [tiré de 137]

Phenotype	Number of affected individuals per family												
	1	2	3	4	5	6	7	8	9	10	11	0	
AHR	102	48	19	7	3	1							74
Asthma	58	105	55	24	4	2	2	1	1				2
Allergy	49	70	65	35	12	3	3	1	2	1	1		12
Allergy with asthma	104	82	34	7	1	4							22
Rhinitis	90	56	34	9	4	2							59
Rhinitis with asthma	96	46	10	4	1								97
Atopic dermatitis (AD)	61	72	52	14	5	2	1	1	1				45
AD with asthma	94	62	12	6	2								78

Abbreviation: AHR, airway hyper-responsiveness.

### 3.4 Histoire démogénétique de la population canadienne anglaise

Le peuplement de ce qui est considéré aujourd'hui comme le « Canada anglophone » (correspondant aux provinces des maritimes ainsi que les provinces à l'ouest de l'Ontario) remonte à l'arrivée des réfugiés loyalistes après 1776 dans le territoire actuel de l'Ontario. En raison de différences culturelles et linguistiques, la population anglaise s'isole de la population québécoise et, ainsi, débute la migration des peuples anglais vers l'ouest canadien. Après la confédération canadienne de 1867, le Canada compte déjà deux colonies dans l'Ouest soit celle de Rivière rouge (Manitoba) et de la Colombie-Britannique [144]. En 1891, la population totale des provinces de l'ouest demeure significativement faible avec un maigre 350 000 habitants, représentant donc 7,2% de la population canadienne [145]. À la fin du siècle, avec l'essor de l'agriculture dans les prairies et une industrie forestière fleurissante en Colombie-Britannique, le peuplement des provinces de l'Ouest débute massivement. Le Canada entre dans une période d'immigration de masse et amorce le début du multiculturalisme canadien (Tableau 12). Aujourd'hui, les grands centres urbains canadiens ont une importante composante multiethnique. Ainsi, 19% de la population de Toronto et de sa région métropolitaine sont d'une autre origine que canadienne, 45% de Vancouver, 22% d'Edmonton et 20% de Winnipeg soulignant donc l'apport des populations migrantes sur le Canada contemporain [146].

**Tableau 12** Origine des immigrants venus s'établir au Canada au 20<sup>e</sup> siècle [tiré de 145]

<b>Origine</b>	<b>1900-1910</b>	<b>1911-1914</b>	<b>1919-1921</b>	<b>1921-1931</b>
Angleterre et Irlande	372 000	283 000	130 000	308 000
États-Unis	134 000	80 000	47 000	87 000
Terre-Neuve	6 000	4 000	3 000	8 000
Europe slave	119 000	83 000	8 000	201 000
Scandinavie	28 000	15 000	4 000	56 000
Italie	13 000	12 000	6 000	14 000
Allemagne	8 000	5 000	-	19 000
Chine	13 000	11 000	3 000	4 000
Autres*	49 000	31 000	14 000	52 000
<b>Total</b>	<b>742 000</b>	<b>523 000</b>	<b>215 000</b>	<b>749 000</b>

\*Autriche, Belgique, Grèce, Pays-Bas, France, Inde et Japon

### **3.4 Canadian Healthy Infant Longitudinal Development Study – Cohorte CHILD**

La cohorte *Canadian Healthy Infant Longitudinal Development Study* ([www.childstudy.ca](http://www.childstudy.ca)) abrégée CHILD est une cohorte de naissance basée sur des femmes enceintes dans quatre villes des provinces du Canada anglais soit Vancouver (Colombie-Britannique), Edmonton (Alberta), Winnipeg (Manitoba) et Toronto (Ontario) regroupant ainsi 3624 familles [147]. Les enfants sont suivis de leur naissance jusqu'à l'âge de 5 ans via des examens par des médecins affiliés ainsi que des questionnaires médicaux durant la grossesse et à 3, 6, 12, 18, 24, 30 mois ainsi qu'à 3, 4 et 5 ans [147]. Des échantillons sanguin, urinaire, nasal, fécal ainsi que de lait maternel ont été recueillis pour former une collection biologique de 3 542 enfants. L'environnement proximal dans lequel les enfants ont grandi a également été évalué par l'entremise de prélèvements effectués à la maison lorsqu'ils ont atteint l'âge de 3-4 mois. La cohorte CHILD possède également une forte composante multiculturelle avec 27% des mères et 26% des pères n'étant pas d'origine caucasienne. Le profil démographique, ethnique et socio-économique des parents est présenté dans le Tableau 13. La cohorte CHILD a ainsi contribué à mieux comprendre l'interaction entre les composantes génétiques et environnementales dans l'asthme et les maladies atopiques ainsi que dans le rôle du microbiote intestinal [148].

**Tableau 13** Caractéristiques des parents participants dans la cohorte CHILD [tiré de 147]

	<b>Éléments évalués (écart-type)</b>	<b>Côté maternel</b>	<b>Côté paternel</b>
	Âge de grossesse moyen	26,7 (6,3)	-
	Âge moyen	32,3 (4,7)	33,8 (5,6)
Ethnicité	Caucasien	2532 (72,9)	2536 (73,7)
	Asie du Sud-Est	428 (12,3)	343 (10,0)
	Asie du Sud	91 (2,6)	106 (3,1)
	Premières nations	177 (5,1)	145 (4,2)
	Noir	77 (2,2)	110 (3,2)
	Autres	161 (4,6)	174 (3,2)
	Inconnu	6 (0,2)	29 (0,8)
	Scolarité	Secondaire et moins	295 (8,7)
Collégial et universitaire		2443 (72,2)	2151 (63,1)
Postsecondaire		647 (19,1)	756 (22,2)
Statut de santé	Asthme	760/3424	547/3000
	Allergies	2832/3425	2094/3000
	Atopie	1768/3073	1657/2432
	Diabète gestationnel	115/3475	-



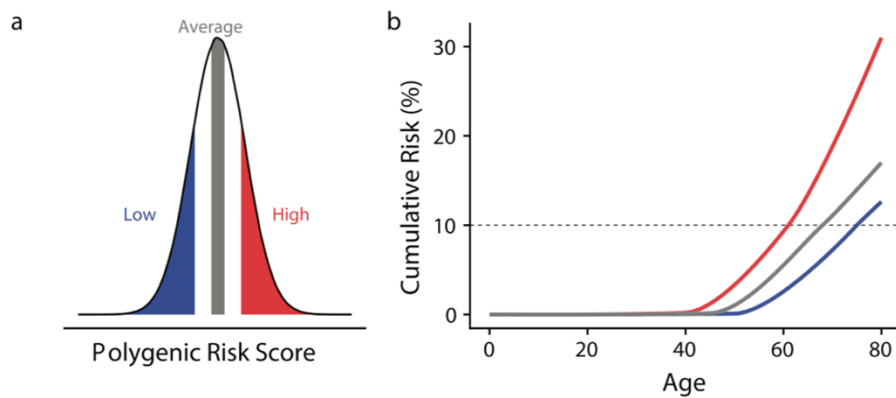
# CHAPITRE 4 – SCORE DE RISQUE POLYGÉNIQUE

## 4.1 Génétique à l'ère post-GWAS

Avec l'accroissement rapide des données génétiques des dernières années, une quantité substantielle d'informations est maintenant disponible et accessible par les chercheurs en génomique. Les études GWAS ont révolutionné la génétique médicale en mettant en évidence des éléments génomiques reliés à certaines maladies. Comme les études GWAS sont puissantes dans la détection des variations génomiques impliquées dans les traits quantitatifs chez plusieurs individus elles peuvent donc continuer de servir de point d'ancrage dans une ère dite post-GWAS (analyse de données brutes) [30]. Toutefois, l'héritabilité expliquée dans les études GWAS demeure relativement faible dans le cas des traits complexes partiellement en raison des diverses interactions gènes-environnement. En prenant en considération tous les loci susceptibles d'avoir un impact sur le phénotype clinique, il est possible d'établir un outil statistique faisant la somme pondérée de l'effet de chacun de ces locus [149]. Cette méthode est nommée score de risque polygénique et présente l'avantage de pouvoir être utilisée comme outil de prédiction pour des pathologies dans l'optique de définir la prédisposition qu'a un individu face à une maladie et d'en prévenir le développement pouvant affecter de manière considérable la vie des patients (maladies dégénératives, cancers héréditaires ou encore des maladies mentales envahissantes). Plus le nombre de loci inclus dans le PRS est important, plus la signification clinique de ce dernier augmente. La forme la plus communément utilisée pour l'établissement d'un PRS consiste à effectuer la somme des  $m$  allèles de susceptibilité multipliée par la taille de l'effet  $\beta$  ( $\ln(\text{OR})$ ) [150].

$$PRS_i = \sum_{j=1}^m x_{ij}\beta_j$$

Plusieurs considérations sont à prendre lors de l'utilisation et de l'établissement d'un PRS. En premier lieu, l'emploi de grand nombre de SNP dans la construction du modèle peut avoir un effet plus négatif que positif sur la prédictibilité en augmentant l'ampleur de la taille de l'effet induisant ainsi un « bruit de fond » dans le modèle. [150]. Un choix judicieux des SNP à être inclus dans le score s'impose en se basant sur l'effet plus important et significatif (p-value). En second lieu, le déséquilibre de liaison (**DL**) entre des SNP voisins résultant en une surreprésentation de certains loci dans le modèle ayant pour effet de réduire son pouvoir discriminatoire. Pour faire face à ces problèmes, un SNP aléatoire d'une région en DL est retiré du modèle ou le seuil de valeur p est augmenté pour prendre en compte moins de SNP en DL [150, 151]. L'engouement ressenti depuis quelques années pour le développement de PRS repose sans doute sur sa capacité à prédire le développement des traits complexes à prévalence faible mieux que l'histoire familiale qui réussit seulement à expliquer environ 4% de l'héritabilité des traits quantitatifs [152]. Un autre avantage des PRS est celui de la prédiction du risque de la maladie à long terme (*lifetime risk trajectory*). En effet, en utilisant une courbe de Kaplan-Meier à partir des données du PRS, il est possible de déterminer le risque cumulatif de développer la maladie en fonction de l'âge des individus (Figure 4, où la courbe grise représente le risque moyen, la courbe rouge le risque plus élevé et la courbe bleu le risque le moins élevé).



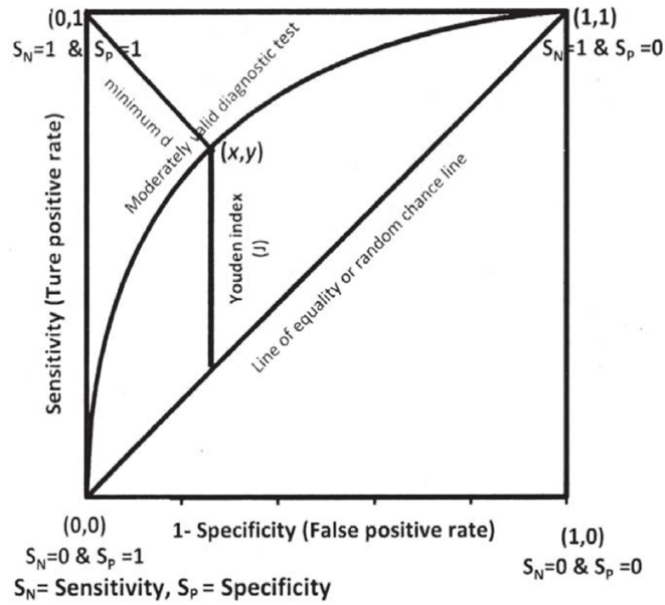
**Figure 5** Courbe de Kaplan-Meier du risque cumulatif (%) en fonction de l'âge [tiré de 150]

Avec les mouvements migratoires de plus en plus fréquents et facilités depuis les dernières décennies, les populations de certains pays présentent une composante multiculturelle de plus en plus prononcée. Cependant, en 2019, il est estimé qu'environ 79% de toutes les études GWAS sont basées sur des individus d'origine européenne en dépit du fait qu'il ne représente que 16% de la population mondiale réduisant ainsi le nombre de PRS appliqués à d'autres populations [153]. Les PRS développés de 2008 à 2017 surreprésentent les populations caucasiennes (459 études) tandis que seulement 140 études concernent les populations asiatiques, 15 études sur les populations africaines, 9 sur les populations latino-américaines et 4 sur celles du Moyen-Orient [154]. Dans l'optique d'adapter les méthodes de recherche pour les rendre plus universelles et accessibles à toutes les populations, l'établissement d'un PRS multiethnique s'avère donc essentiel. Pour augmenter le pouvoir de prédiction du score de risque polygénique, des chercheurs ont utilisé des données GWAS d'origine européenne combinées avec des données GWAS issues de populations latino-américaines pour augmenter la prédictibilité de plus de 70% pour le diabète de type II dans la population sud-américaine [155]. Une autre équipe de recherche américaine s'est penchée sur l'étude de la dermatite atopique dans deux cohortes et comportant plus de 90 000 individus d'origine afro-américaine établis aux États-Unis. Ils ont développé un score de risque polygénique basé sur 27 SNP pangénomiques identifiés dans la population caucasienne et ont conclu que le PRS développé était hautement prédictif de la dermatite atopique dans la population caucasienne mais demeurait très peu prédictif chez les afro-américains suggérant ainsi une architecture génétique différente pour la dermatite atopique dans les deux populations [156]. Aujourd'hui, de nombreux PRS ont été établis pour diverses maladies et plusieurs d'entre eux ont effectivement permis d'effectuer des mesures préventives pour contrer le développement de maladies invasives soulignant donc leur importance dans une ère de médecine dite « personnalisée » [157-160]. L'exemple particulier d'un PRS établi pour le cancer du sein au Royaume-Uni chez 67 000 femmes d'origine européenne a permis

d'identifier les femmes à risque de développer un cancer du sein dix ans avant la moyenne des femmes et donc dix ans avant le début des tests de dépistage standards [161]. Toutefois, des efforts sont toujours requis pour mieux adapter ces outils de prédiction et les rendre plus universels pour ainsi améliorer la prévention.

#### 4.2 Fonction d'efficacité du récepteur

Les courbes de fonction d'efficacité du récepteur ou plus communément appelées, en français, courbes **ROC** (*Receiver Operating Characteristic*) sont des outils graphiques mettant en relation la spécificité et la sensibilité d'un modèle phénotypique dichotomique (patients atteints ou non atteints) et continu dans l'optique de mesurer la performance. Une courbe ROC typique est séparée en deux portions, par la ligne d'égalité des chances, qui comprennent chacune une aire sous la courbe (**AUC**) de 0,5. Si la courbe se trouve sur la ligne d'égalité des chances, l'AUC est, par conséquent, de 0,5 ce qui signifie que le modèle prédictif développé discrimine les individus atteints des non atteints dans 50% des cas (Figure 6) [162]. La mesure d'AUC résume la position globale de la courbe et plus sa mesure est élevée (le plus près de 1), plus le modèle diagnostique fait la distinction entre patients et témoins [163]. Statistiquement, il est plus avantageux de retenir la mesure d'AUC comme seuil de prédiction que des points quelconques car elle représente la sensibilité moyenne pour toutes les valeurs de spécificité possibles [163, 164]. Plusieurs avantages sont reliés à l'emploi de courbes ROC d'un point de vue statistique et médical. Les courbes ROC sont indépendantes de la prévalence de la pathologie étudiée car elles se basent sur les mesures de sensibilité et de spécificité de l'outil de prédiction. Toutefois, si la prévalence est relativement faible, le taux de faux positifs ( $1 - \text{spécificité}$ , correspondant à l'axe des X) devra, par ailleurs, demeurer bas sinon tous les patients classés comme atteints seront en réalité non atteints et vice-versa [165].



**Figure 6** Structure d'une courbe ROC [tiré de 162]

La sensibilité, la spécificité ainsi que l'exactitude sont calculées à l'aide des équations suivantes (note : **NT** représente le nombre total d'individus; **NTA** le nombre de patients réellement atteints (donc diagnostiqués) et **NTN** le nombre de patients réellement non atteints) :

$$\text{Sensibilité} = \frac{\text{proportion de vrais positifs (VP)}}{\text{proportion de vrais positifs (VP)} + \text{proportion de faux négatifs (FN)}}$$

$$\text{Spécificité} = \frac{\text{proportion de vrais négatifs (VN)}}{\text{proportion de faux positifs (FP)} + \text{proportion de vrais négatifs (VN)}}$$

$$\text{Exactitude} = \frac{VP}{NTA} \times \frac{NTA}{NT} + \frac{VN}{NTN} \times \frac{NTN}{NT}$$

Plusieurs outils statistiques ont été développés en parallèle à la courbe ROC pour tenir compte et essayer d'expliquer la variance dans les modèles logistiques. Un peu à la manière du  $R^2$  dans une régression linéaire renseigne sur l'adaptabilité du modèle en fonction des variables choisies ou, autrement, à quel point le modèle décrit la tendance linéaire des données sur une échelle de 0 à 1 (0 correspondant à un modèle n'expliquant pas la tendance linéaire et 1 un modèle qui l'explique à 100%). L'équivalent du  $R^2$  de la régression linéaire dans un modèle logistique est le pseudo- $R^2$ . Néanmoins, il de mise de rester prudent quant à son utilisation car il ne correspond pas exactement au  $R^2$  des régressions linéaires mais bien à une probabilité, ce pourquoi il est nommé « pseudo »  $R^2$  [166]. De nombreux statisticiens ont établi des mesures de pseudo- $R^2$  dont McFadden, Cox et Snell et Cragg et Uhler (repris par Nagelkerke d'où l'ambiguïté sur le nom officiel de cette statistique). L'une de ces statistiques la plus employée lors d'études épidémiologiques est celle de Nagelkerke. Il permet de calculer le pourcentage de variance expliquée par le modèle. Le pseudo- $R^2$  de Nagelkerke correspond à celui de Cox-Snell corrigé par le maximum de la mesure du pseudo- $R^2$  de Cox-Snell [166]. Il se calcule à partir de l'équation suivante (note : L correspond au logarithme de la vraisemblance,  $M_0$  au modèle sans variables,  $M_1$  au modèle avec variables et N au nombre total d'observations) :

$$R_{Nagelkerke}^2 = \frac{R_{Snell-Cox}^2}{R_{Snell-Cox}^2 \text{ maximal}} = \frac{1 - \left[ \frac{L(M_0)}{L(M_1)} \right]^{\frac{2}{N}}}{1 - L(M_0)^{\frac{2}{N}}}$$

Il est de mise d'être prudent lorsque le  $R^2$  de Nagelkerke est employé car il a tendance à maximiser les valeurs de  $R^2$  comparativement aux autres mesures. Il a aussi tendance à prendre des valeurs d'ajustement plus grandes qu'en réalité [166]. Plus récemment, d'autres méthodes statistiques ont été développées pour évaluer l'efficacité du modèle de prédiction. Le *Net Reclassification Index (NRI)* compare un ancien modèle de prédiction avec un nouveau modèle prédictif. Lorsqu'un nouveau marqueur est ajouté au modèle le NRI mesure les patients qui seront reclassés comme non atteints s'ils étaient malades ou les témoins qui seront considérés comme malades si l'ajout de marqueur influence la maladie [167].

# CHAPITRE 5 – DÉVELOPPEMENT D'UN SCORE DE RISQUE POLYGÉNIQUE POUR LA DERMITE ATOPIQUE DANS LA POPULATION CANADIENNE

**Mathieu Simard<sup>1,2</sup>, Anne-Marie Madore<sup>1</sup>, Simon Girard<sup>1,2</sup>, Susan Waserman<sup>3</sup>, CHILD Cohort Study<sup>4</sup>, Qingling Duan<sup>5</sup>, Padjama Subbarao<sup>6</sup>, Malcolm Sears<sup>7</sup>, Charles Morin<sup>1,8</sup>, Philippe Bégin<sup>9,10</sup>, Catherine Laprise<sup>1,2,8</sup>**

<sup>1</sup>Département des sciences fondamentales, Université du Québec à Chicoutimi, Saguenay, Québec, Canada

<sup>2</sup>Centre Intersectoriel en Santé Durable, Université du Québec à Chicoutimi, Saguenay, Québec, Canada

<sup>3</sup>McMaster University, Hamilton, Ontario, Canada

<sup>4</sup>CHILD cohort study, Toronto, Ontario, Canada

<sup>5</sup>School of Computing and Dept of Biomedical & Molecular Sciences, Queen's University, Kingston, Ontario, Canada

<sup>6</sup>Department of Pediatrics, Hospital for Sick Children, University of Toronto, Toronto, Ontario, Canada

<sup>7</sup>Department of Medicine, McMaster University, Hamilton, Ontario, Canada

<sup>8</sup>Department of Pediatrics, Centre intégré universitaire de santé et de services sociaux du Saguenay–Lac-Saint-Jean, Saguenay, Québec, Canada

<sup>9</sup>CHUM, Hôpital Notre-Dame, Montréal, Québec, Canada

<sup>10</sup>CHU Sainte-Justine, Department of Allergy and Immunology, Montréal, Québec, Canada



## **AVANT-PROPOS**

Cet article a été soumis (décembre 2019) et accepté (mai 2020) dans le *Journal of Allergy and Clinical Immunology* (JACI) et sera publié prochainement.

### Contribution des auteurs à l'article :

Mathieu Simard a effectué les analyses bio-informatiques ainsi que la rédaction de l'article. Anne-Marie Madore a validé les résultats obtenus et a également contribué à la rédaction de l'article. Susan Wasserman a aussi révisé le manuscrit. Par la suite, Qingling Duan (Cohorte CHILD – responsable de l'accès aux données), Padmaja Subbarao (Cohorte CHILD – responsable du recrutement et de l'évaluation), Malcom R. Sears (Cohorte CHILD – principal investigateur), Theo J. Moraes (Cohorte CHILD – coordonnateur), Allan B. Becker (Cohorte CHILD – responsable du recrutement et de l'évaluation), Stuart E. Turvey (Cohorte CHILD – responsable de la gestion des données) et Piushkumar J. Mandhane (Cohorte CHILD – coordonnatrice) ont contribué à la réalisation de ce projet en nous procurant les données nécessaires ainsi qu'une révision du manuscrit. Ensuite, Charles Morin a joué un rôle dans le recrutement des participants de la cohorte d'asthme du SLSJ ainsi que dans la révision du manuscrit et Philippe Bégin a participé à la conception de l'étude et à la révision de l'article. Finalement, Catherine Laprise a contribué à la conception, la construction et à la gestion des données et des échantillons de la cohorte d'asthme du SLSJ. Elle est également responsable du design méthodologique, de la supervision des analyses, de la rédaction du manuscrit ainsi qu'à la participation au manuscrit et de l'approbation de la version finale à soumettre.

Source:

Simard M, Madore A-M, Girard S, Wasserman S, Duan Q, Subbarao P, Sears MR, Moraes TJ, Becker AB, Turvey SE, Mandhane PJ, Morin C, Bégin P, Laprise C, POLYGENIC RISK SCORE FOR ATOPIC DERMATITIS IN THE CANADIAN POPULATION, *Journal of Allergy and Clinical Immunology* (2020), doi: <https://doi.org/10.1016/j.jaci.2020.04.057>.

## **5.1 Résumé**

La dermatite atopique (DA) est la première étape d'une série de manifestations cliniques appelée marche atopique. Des études cliniques ont démontré des manières simples d'empêcher son développement mais, pour ce faire, des outils de prédiction sont nécessaires pour identifier les enfants à risque élevé. Dans l'optique de répondre à ce besoin clinique, un score de risque polygénique avec une aire sous la courbe de 85% et une explication de la variance de 36% a été développé à partir de loci issus d'études GWAS.

## **5.2 Abstract**

Atopic dermatitis (AD) is the first step of a process called the allergic march. Clinical trials have demonstrated simple ways to block its development but to benefit from these, predictive tools are needed to identify high risk children. To meet this need, a polygenic risk score having an area under the curve (AUC) of 85% and explaining 36% of AD variance was built with loci identified from GWAS results.

### 5.3 Method and results

Atopic dermatitis (AD) is characterized by a damaged skin barrier that allows allergens to penetrate the body, leading to a sensitization and a higher risk of developing food allergy (35%), asthma (50%), and/or allergic rhinitis (75%), all characterizing the allergic march.[168] Recent evidences have shown that the allergic march can be stopped in high-risk infants with early interventions directed at reestablishing and/or maintaining skin barrier function, using methods as simple as the use of emollient cream, and introducing food allergens early into the diet.[169-172] Based on this assumption, building a predictive tool to identify children at high risk of developing moderate to severe AD (msAD) would be of great interest. In this study, a polygenic risk score (PRS) with an area under the curve (AUC) of 85% and explaining 36% of msAD variance was established for the Canadian population.

Two PRS scenarios have been tested. One with GWAS loci identified through literature and another based on the strongest GWAS hits found in two Canadian cohorts (Supplementary Table 1; see Supplementary Method for the detailed methodology).

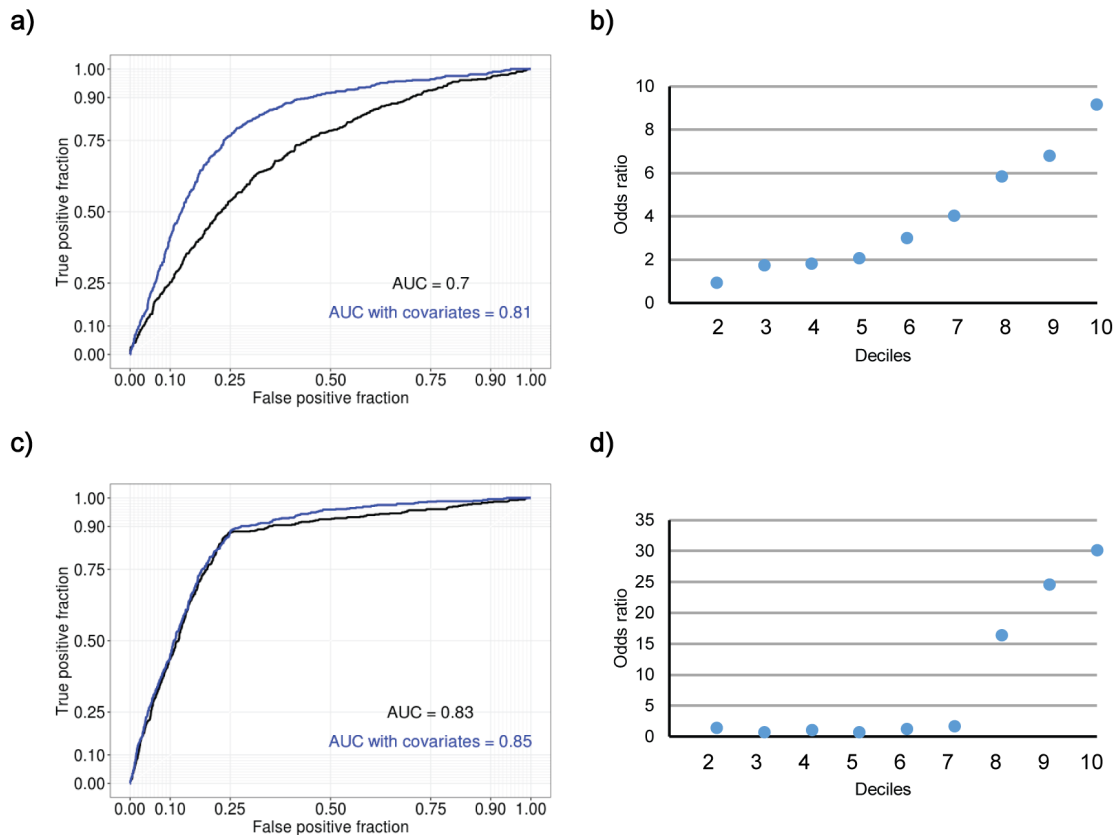
The first scenario was conducted to evaluate if the best associations in the literature are suitable to build a PRS for AD with a good discriminative value in a specific population. For this scenario, the 25 best associations documented in GWASs of AD (Annexe 1 – Tableau 15) were selected. For each of these, a region spanning  $\pm 100$  kb was tested for the best associated genetic variant with msAD in unrelated cases and controls from the two Canadian cohorts ( $n = 3,394$  individuals) using a general regression model (Supplementary Table 3).[173] The PRS was then built for each individual of both cohorts ( $n = 4,025$ ) considering the number of risk alleles of the genetic variants weighted by their  $\beta$

estimates ( $= \ln(\text{OR})$ ). The discriminative value of the PRS was assessed by a receiver operating characteristic (ROC) curve analysis and gave an AUC of 70% (Figure 1a) and a proportion of variance explained of 10% according to Nagelkerke's pseudo- $R^2$ . Once covariates were added to the model (sex, age and parents' ethnicity) an AUC of 81% was reached (Figure 1c) and the model explained 24% of msAD variance. These results highlight the dependence of the first scenario upon covariates to reach a good discriminative value (AUC between 80 and 89%).[174]

The second scenario took advantage of the two Canadian cohorts to build another PRS based on the best associations for AD in the Canadian population considering its ethnic diversity. The Saguenay–Lac-Saint-Jean (SLSJ) asthma familial cohort [137] includes individuals of French descent from the SLSJ region in northern Quebec, Canada, and the Canadian Healthy Infant Longitudinal Development (CHILD) birth cohort [147] includes principally individuals of English descent but also of other origins and living in Vancouver, Edmonton, Winnipeg and Toronto. The GWAS was performed on unrelated cases and controls from both cohorts using the MFQLS test [175] and a PRS was built with the 25 best associations (Annexe 1 – Tableau 15) after running a general regression model to extract their  $\beta$  estimates. The PRS obtained had an AUC of 83% (Figure 1c) and explained 33% of msAD variance. Adding the same covariates as for the first scenario, the AUC was of 85% (Figure 1c), corresponding to a sensitivity and specificity of 88% and 75%, and the PRS explained 36% of msAD variance. Both models of this second scenario showing a good discriminative value [174]. In comparison, AUCs derived from PRSs in the literature range from 53% to 76%,[176] and PRSs show relatively low percentages of explained variance (from 0.002% to 3% compared to 36% in this study) [177-179].

These results demonstrate that the second scenario, which used data from the targeted population to build the PRS, explains best the risk of developing msAD even without considering any covariate. It is interesting to note that only one locus was common between the 25 best associations of the two scenarios, the 5q33.3 region represented by the genetic variants rs2862717 and rs10600273 in the second scenario.[180] It confirms the need to characterize the genetic profile of each specific population before building a PRS in order to reach a good discriminative value. Moreover, if looking at possible PRSs from the second scenario but considering only genetic variants associated with  $p < 1e^{-20}$  ( $n = 13$ ) or  $p < 1e^{-40}$  ( $n = 6$ ), the discriminative value of the PRS was well preserved with corresponding AUCs of 85% and 84%. These results making the PRS even more interesting as it could fit more easily into a potential combined predictive genetic tool for Canadian infants.

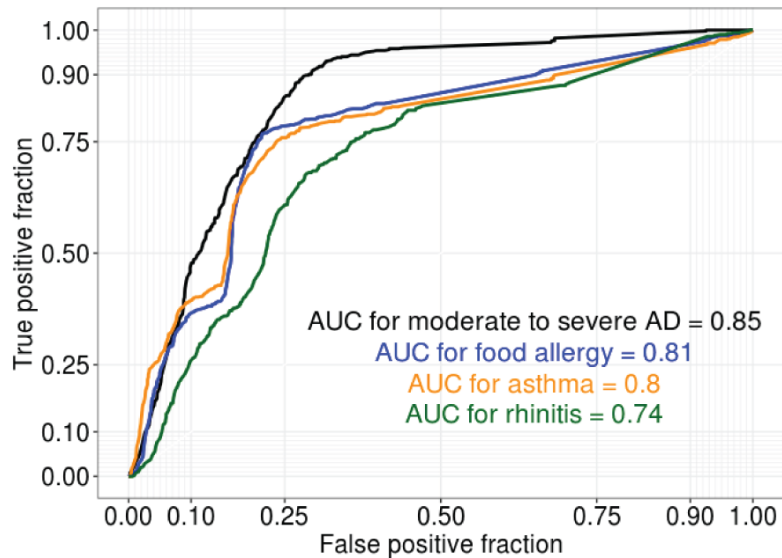
To be an efficient predictive tool, the PRS as to be modified into a binary value identifying low-risk and high-risk individuals. By examining the distribution of individuals' PRSs on a decile scale, a significant higher risk of developing msAD for the second scenario was found from the eighth decile (OR = 16.35) to the tenth decile (OR = 30.08; Figure 1d and Figure 2) with a clear demarcation between the seventh (OR = 1.68) and the eighth ones. Such a demarcation was not observed in the first scenario (Figure 1c). Patients were then classified in one of the two categories following the second scenario's PRS; the low-risk category (first to the seventh deciles) and the high-risk one (eighth to the tenth deciles; Figure 1d). The binary PRS was tested for association with diseases of the allergic march using a logistic binomial model with above-mentioned covariates and significant results were found with AD ( $p = 7.67e^{-07}$ ), msAD ( $p = 1.28e^{-13}$ ), food allergy ( $p = 2.84e^{-15}$ ), asthma ( $p = 3.60e^{-16}$ ) and rhinitis ( $p = 0.037$ ).



**Figure 1** ROC curves for the polygenic risk scores using SNPs associated in the literature or in two Canadian cohorts. This figure shows the ROC curves and the area under the curves (AUCs) for the polygenic risk scores (PRSs) calculated for SNPs associated in the literature (a) and in two Canadian cohorts (c) along with the comparison between deciles 2 to 10 on decile 1 for both PRSs in (b) and (d) respectively. Dark lines show the ROC curves for the models without covariate and blue lines the models with covariates.

Moreover, AUCs were calculated for these diseases giving 85% (msAD), 81% (food allergy), 80% (asthma) and 74% (allergic rhinitis; Figure 2). Comparing these results with the literature, the PRS built in this study showed a stronger association with diseases of the allergic march with ORs of 1.77 and 5.01 for AD or msAD compared to an OR of 1.22 for AD for a PRS built from four selected genes (*GSTP1*, *TNF*, *TLR2* and *TLR4*).[181] It also showed a better discriminative value compared to the

Predicting Asthma Risk in Children (PARC) tool based on preschool respiratory symptoms to predict asthma at school age (AUC = 80% compared to AUC = 77%).[182]



**Figure 2** ROC curves calculated with the binary polygenic risk score for four diseases that characterize the atopic march. This figure shows the ROC curve and the area under the curve (AUC) for moderate to severe atopic dermatitis (black), food allergy (blue), asthma (orange) and rhinitis (green). The PRS was calculated from the association results in two Canadian cohorts.

## 5.4 Conclusion

To conclude, the use of two Canadian cohorts allowed to establish a PRS with an AUC of 85% and explaining 36% of msAD variance. Considering the ease of treatment to avoid the development of AD, this PRS may have enough discriminative value to become an accurate clinical predictive tool to identify Canadian infants that are at a higher risk of developing msAD and then prevent the development of the allergic march.

## CHAPITRE 6 – AUTRES CONTRIBUTIONS

Ce chapitre présente d'autres contributions faites en parallèle à celle du développement d'un score de risque polygénique pour la dermatite atopique dans la population canadienne.

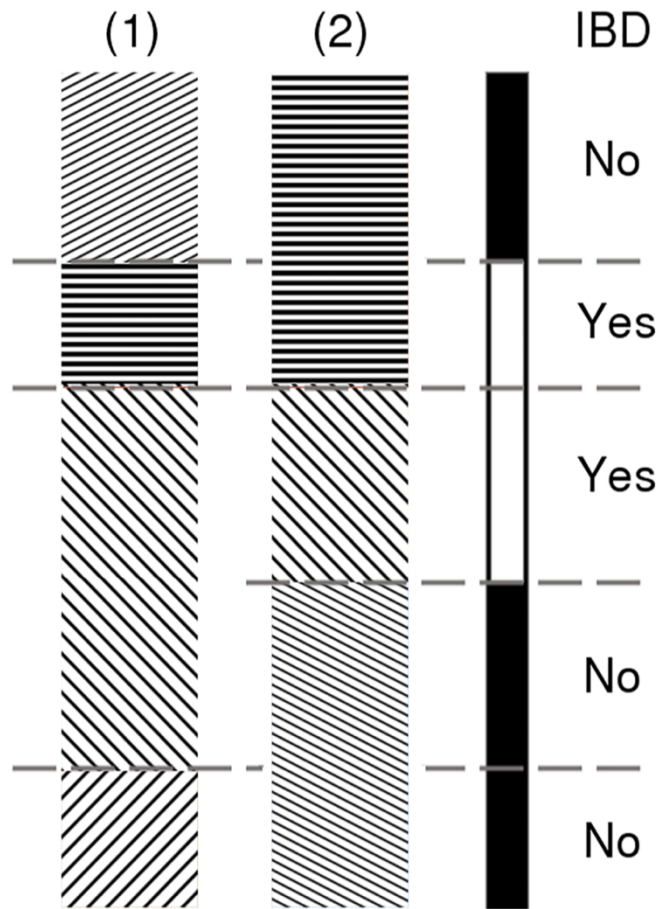
### 6.1 Identité par descendants

La notion d'identité par descendants (**IBD**) est un concept clef en génétique des populations car elle permet de retracer des relations familiales remontant à plus de 6 ou 7 générations [142]. Ainsi, un fragment d'ADN dupliqué ou copié (locus) à partir d'un fragment ancestral est dit identique par descendant [183]. Cette définition s'illustre par l'exemple de parents qui transmettent leur ADN dupliqué à leurs enfants. L'emploi des IBD dans l'analyse d'arbres généalogiques permet d'effectuer des études de liaison génétique. Ainsi, dans les familles à petits arbres généalogiques, les segments d'IBD ont tendance à être plus longs (plus de 10 cM) et plus facilement détectables à l'aide de marqueurs microsatellites ou de SNP [184]. Pour les arbres généalogiques contenant un très grand nombre d'individus et de générations provenant d'un même ancêtre, les segments d'IBD sont sensiblement courts et peuvent ultimement se résumer à un seul SNP dans certains cas car, en provenant d'un ancêtre beaucoup plus éloigné, il aura subi plus de recombinaisons diminuant donc drastiquement la longueur du segment [184]. La longueur des segments d'IBD entre des individus sur  $c$  chromosomes peut être quantifiée sur un nombre de générations à l'aide du nombre de méioses  $d$  séparant les individus (avec  $vd$  représentant le taux de recombinaisons et  $N$ , le nombre total des segments IBD partagés entre deux individus) [185, 186] :

$$N = \frac{1}{2a} (vd + c)$$



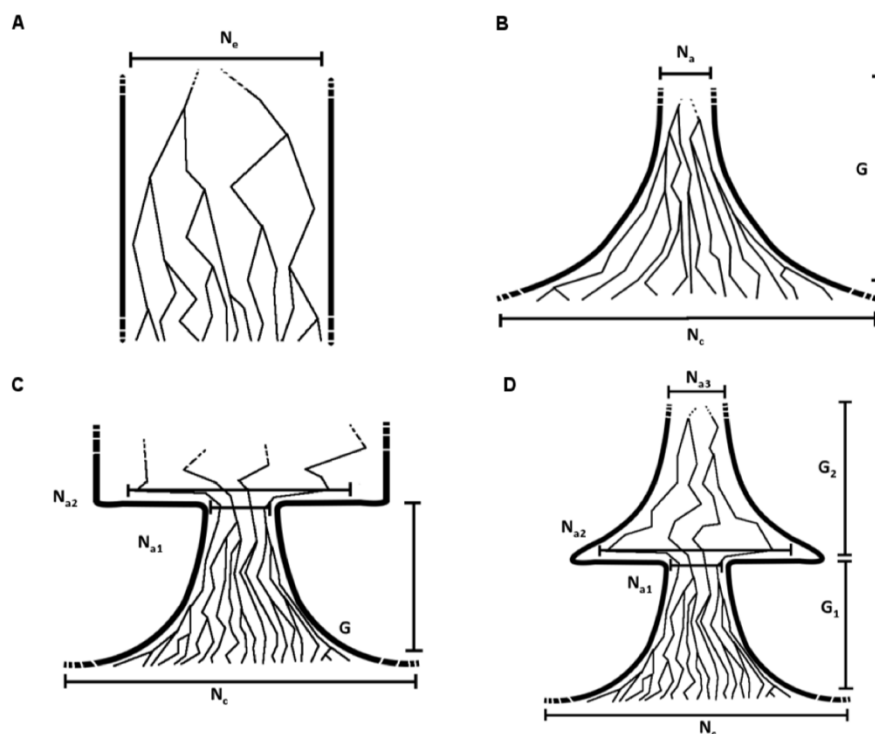
On distingue deux types de segments d'IBD dits multilocus : les IBD stricts et les IBD relâchés [187]. Les IBD stricts doivent avoir sur un nombre  $n$  de chromosomes ancestraux tous leurs loci provenant du même ancêtre tandis que les IBD relâchés ne partageront pas nécessairement tous leurs loci avec le même ancêtre (Figure 7) [187].



**Figure 7** Chromosomes homologues avec segments IBD partagés [tiré de 187]

Étant donné leur grande utilité en génétique des populations, les IBD peuvent également être appliqués à la démographie des populations dans un contexte géographique de mouvements et de croissance de population. En effet, il est possible d'illustrer la distribution des segments d'IBD dans une population en fonction de son histoire démogénétique comme à la Figure 8 [188]. Sur cette figure, le scénario A représente la distribution des IBD dans une population ayant subi une

croissance constante sans évènement ayant affecté sa structure comme dans le scénario B ou la population a subi une croissance exponentielle et où les IBD se sont rapidement dispersés. Dans le scénario C, un effet fondateur a affecté le patrimoine génétique des individus et a contribué à rapprocher les segments d'IBD et, par la suite, une croissance rapide de la population a contribué à enrichir de nouveau la structure génétique et ainsi séparer de plus en plus les segments d'IBD. Le dernier scénario (D) est celui qui représente le plus l'histoire de la génétique au Saguenay–Lac-Saint-Jean car on y retrouve 3 effets fondateurs successifs avec une croissance exponentielle de la population entre ceux-ci [133]. Il est même possible de remarquer que la première croissance de la population a contribué à éloigner les segments d'IBD partagés entre les individus mais que la diversité demeurait tout de même faible (en analogie avec les mouvements migratoires Québec-Charlevoix) comparativement à la croissance associée au deuxième effet fondateur (Charlevoix-Saguenay–Lac-Saint-Jean).



**Figure 8** Schématisation de la distribution des IBD en fonction de divers scénarios démographiques [tiré de 188]

## 6.2 Intérêt et apports des segments IBD dans l'asthme

Il peut sembler, à première vue, difficile de faire un lien entre la distribution tant génomique que géographique de segments IBD et l'intérêt que ces derniers peuvent apporter à l'étude d'une maladie polygénique et complexe comme l'asthme. Dans un premier lieu, l'évaluation du degré d'hétérogénéité des patients atteints d'asthme peut être effectuée à l'aide d'études impliquant des IBD. En effet, une étude française a permis d'identifier un facteur génétique sur le chromosome 7q impliqué dans l'asthme et qui jouait un rôle dans l'âge d'apparition de la maladie en utilisant des paires d'individus apparentés et en comparant les segments IBD partagés [189]. Une autre étude européenne, à l'aide des segments IBD d'individus atteints et d'autres sains, a identifié deux variantes génétiques rares (chromosomes 7 (fréquence de 0,8%) et 15 (fréquence de 0,7%)) étant fortement liées au phénotype d'asthme chez des individus ethniquement différents et cela a permis de retracer certains mouvements migratoires qui étaient jusqu'alors inconnus et peu documentés [190]. Des associations significatives entre les loci flanquant ces deux variants génétiques rares avec l'asthme sont d'ailleurs documentés dans la littérature scientifique. Une étude danoise portant sur 236 familles a associé le chromosome 3q13.31 à la rhinite à l'aide d'une étude des segments IBD révélant un partage des IBD de 63,9% entre les individus et pour laquelle le pourcentage d'IBD partagés est descendu sous la barre de 9% lorsque les familles avec des individus atteints de rhinite étaient retirés de l'analyse [191]. En considérant les apports que les études de segments IBD ont apporté dans la compréhension de la composante génétique sous-jacente à l'asthme, il est important de considérer leur rôle potentiel à contribuer à définir la part d'héritabilité manquante de l'asthme. Les segments IBD identifiés dans une population à effet fondateur pourraient également contribuer au développement de programme de prévention visant à identifier les individus à risque en bas âge à partir du pourcentage d'IBD partagés avec des ancêtres ayant déjà été diagnostiqués d'asthme.

### 6.3 Étude des haplotypes de l'asthme dans la population du Saguenay–Lac-Saint-Jean par identité par descendants<sup>1</sup>

**Mathieu Simard<sup>1,2</sup>, Claudia Moreau<sup>1,2</sup>, Catherine Laprise<sup>1,2</sup>, Simon Girard<sup>1,2</sup>**

<sup>1</sup>Département des sciences fondamentales, Université du Québec à Chicoutimi, Saguenay, Québec, Canada

<sup>2</sup>Centre Intersectoriel en Santé Durable, Université du Québec à Chicoutimi, Saguenay, Québec, Canada

#### **ABSTRACT**

Asthma is a respiratory disease characterized by chronic inflammation of the airways that results in coughing, shortness of breath, wheezing and difficult breathing. It is a complex disease which is affected by a genetic and environmental component. In order to better investigate the genetic component of asthma, an analytical method based on haplotypes was used using genetic data from individuals in the Saguenay–Lac-Saint-Jean asthma cohort. Ten genome-wide loci were selected for the haplotypic association analyzes to allow the identification of 55 haplotypes located in 8 genes. Although these 55 haplotypic blocks are not significantly associated with the asthma phenotype in this population, their identification has made it possible to enrich knowledge of the genetic component of this disease and thus better characterize this regional population.

---

<sup>1</sup> Cet article est en cours de révision et d'élaboration pour une soumission dans un journal scientifique dans un futur proche.

## INTRODUCTION

### Asthma and haplotypes

The Global Initiative for Asthma (GINA) defines asthma as “[...] a chronic inflammatory disorder of the airways in which many cells and cellular elements play a role. The chronic inflammation is associated with airway hyperresponsiveness that leads to recurrent episodes of wheezing, breathlessness, chest tightness, and coughing [...]” [192]. Its prevalence has been increasing since the 1950s in developed countries where industrialization and urbanization are higher [117-119, 193]. As of today, the worldwide prevalence of 4.3% in adults has been identified (for physician-diagnosed asthma) and forecasts show that it will increase by 50% in the next decade [194]. In Canada, asthma represents an economic burden through its direct and indirect costs on provincial and territorial health systems, with a global estimated cost in 2012 of \$2.1 billion [114]. As a multifactorial (complex) disease, asthma is influenced by a genetic component as well as an environmental component (such as dust and cigarette smoke). Numerous genes have been linked to asthma in various global populations through GWASs studies which have classified these genes into three categories of influence on the phenotype : 1) immune functions influence; 2) inflammatory pathways influence; 3) airways remodeling [83-88]. One of the ways to further study the genetic component of asthma in a population is to take haplotypes into account. A haplotype is defined as a chromosomal segment where the alleles are linked and transmitted together [195]. By taking into account the linkage disequilibrium (LD), consisting in a non-random genetic association between two loci (i.e., when the A allele is always associated with the b allele), we can define what a haplotypic block is [196]. As LD is reversed by recombination events (crossing-overs) that increase genetic variability among individuals within the population, it is possible to identify genomic regions with high LD named “blocks” as well as regions designated as “recombination hotspots” where LD is very low or almost in linkage equilibrium

[196, 197]. These regions of low LD where the haplotypic diversity is relatively low have a few common haplotypes which represent the majority of the allelic variation of the population are therefore called “haplotypic blocks” [196, 198]. The study and analysis of haplotypic blocks is useful when one wants to study complex traits (genetic and environmentally influenced diseases). Indeed, if a mutation in such a block is inherited in a population, by retracing this block, it will be possible to associate it with a trait. In addition, as they take into account the demogenetic history, haplotypic blocks are better markers than SNPs because they take into account LD in the various genomic regions thus increasing the identification of susceptibility genes likely to be associated with the disease [199].

#### Québec's demogenetic history

The European settlement of Québec began during the 17<sup>th</sup> century (1608) with the arrival of immigrants from the regions of northern France, mainly from Normandy, Poitou-Charentes and Brittany in present-day Québec city [124]. In a few decades, French immigration increased rapidly along the banks of the St. Lawrence River from Québec to Montréal. Following the English conquest of 1760, the province passed to the British putting almost exclusive end to French immigration to North America [125]. With an increasing number of settlers, the geographic dispersal of these settlers is increasing and so began the colonization of the interior regions of Québec (Saguenay–Lac-Saint-Jean, Côte Nord, Charlevoix, Outaouais and Mauricie) [127, 128]. The Saguenay–Lac-Saint-Jean population is characterized by a triple founding effect which has shaped its genetic heritage causing lower allelic diversity at some loci. The first founding effect corresponds to the first migrants from France since 1608. When some of them started to move in 1675 to the region of Charlevoix, the second wave of allelic loss took place with a rapidly increasing population passing from 599 migrants to more than 18,000 in just seven generation. Finally, the third founding effect corresponds to that of

migrants who passed from Charlevoix to Saguenay–Lac-Saint-Jean and who formed very large families in a short period of time [133].

## **METHOD AND RESULTS**

### Population and genotyping

The Saguenay–Lac-Saint-Jean (SLSJ) asthma cohort is made up of 1,394 individuals identified via asthmatic probands, divided into 271 independent families spread over the territory of SLSJ, a northeastern region of Québec [137]. The family size varies between 3 to 17 individuals with a number of asthmatic individuals from 1 to 11 per family [137]. The cohort also includes a collection of more than 75 phenotypic characteristics collected since 1997 as well as information from medical questionnaires [137]. DNA was extracted from 1282 blood samples collected during the patients' first clinical visit. For the 1214 samples having passed the quality control tests, the genotyping was carried out on an Illumina 610K Quad array at the Centre national de génotypage in Evry, France [137]. The genetic data resulting from the 1,214 blood samples underwent strict quality control analyzes using the genome-wide association tool PLINK version 1.9 (<http://pngu.mgh.harvard.edu/purcell/plink/>). The missing rate per SNP was set at 98%, the frequency of the minor allele (MAF) set at 99.999% and the exact Hardy-Weinberg equilibrium (HWE) to a threshold of  $1 \times 10^{-4}$ . Individuals with a missing phenotypic data were removed from analyzes. The 1,141 individuals who passed the quality control criteria were used for the rest of analyzes.

### Selection of genes of interest

The genes of interest were selected from a multiancestry meta-analysis from Demenais et al. [200]. This meta-analysis used genetic data from the Trans-National Asthma Genetic Consortium which gathers together individuals of African, European, Hispanic and Japanese ancestry, thus totaling 2.83 million genotyped SNPs. In the context of this study, only individuals of European ancestry were selected to make a suitable comparison with those of the Saguenay–Lac-Saint-Jean asthma cohort, a cohort based on an exclusivity of subjects with European origin. A total of 18 loci associated with asthma were retained for further analysis. Of these 18, only those with an odds ratio (OR) greater than 1.0 were selected, corresponding to 10 loci. These 10 loci are presented in Table 1.

**Table 1.** Associated loci in asthma in the SLSJ cohort

SNP at the associated loci	Alleles	Chromosome	Nearest gene(s)	Odds ratio (OR)	log (OR)	CI (95%)	p-value
rs1233578	G/A	6	<i>GPX5, TRIM27</i>	1.064	0.127	0.829-1.365	0.2405
rs2855812	T/G	6	<i>MICB, HCP5</i>	1.188	0.107	0.963-1.466	0.1068
rs2589561	A/G	10	<i>GATA3, CELF2</i>	1.067	0.111	0.858-1.327	0.5619
rs12534811	G/A	8	<i>TPD52, ZBTB10</i>	1.190	0.087	1.003-1.411	0.0454
rs1420101	T/C	2	<i>IL1RL1, IR1RL2</i>	1.256	0.084	1.065-1.480	0.0066
rs10455025	C/A	5	<i>SLC25A46, TSLP</i>	1.149	0.086	0.971-1.359	0.1053
rs20541	A/G	5	<i>IL13, RAD50, IL4</i>	1.205	0.094	1.002-1.448	0.0474
rs992969	A/G	9	<i>RANBP6, IL33</i>	1.209	0.090	1.013-1.443	0.0355
rs2033784	G/A	15	<i>SMAD3, SMAD6</i>	1.026	0.084	0.870-1.211	0.7605
rs2952156	A/G	17	<i>ERBB2, PGAP3</i>	1.207	0.086	1.020-1.428	0.0285



### Haplotype blocks identification and association analyzes

The haplotypic blocks were generated from imputed data filtered using Haploview software (<https://www.broadinstitute.org/haploview/haploview>) on the genotypes of the PED files of the previous PLINK analyzes. Two different association analyzes were performed on the filtered data: a family-based trios (assocTDT) and a case-control (assocCC) associations. In the presence of a binary trait (affected on unaffected by the disease), as proposed for the first time by Spielman et al. (1993), a transmission disequilibrium test (TDT) can be performed using affected offspring as well as its both parents to determine the offspring's susceptibility to developing the trait [201]. In the absence of an association between the phenotype and the risk allele, the offspring will inherit an allele from each of its parents in a Mendelian fashion (50% chance of obtaining one of the two alleles of each of the two parents). In addition, the test takes into account whether specific haplotypic blocks are transmitted to the offspring more often than in a scenario where there is no haplotype-phenotype association. In contrast, case-control (CC) association analysis tends to verify that the haplotypic block identified as being associated with the phenotype is more common in patients than in healthy individuals in the population. Being more frequent in patients, the haplotypic block is said to be linked to the phenotype (positive association).

From the association files (assocCC and assocTDT) generated by Haploview for each of the 18 genomic regions of interest, those with a p-value less than or equal to 0.01 were kept. The significance threshold (p-value) was corrected to 0.01 by a Bonferroni correction. Thus, to take into account the false-discovery rate (FDR) that could have been generated by multiple comparisons scenarios across the whole genome, Bonferroni's correction divides the p-value by the number of tests performed. This allows to consider the large number of polymorphic sites in the genome as well as the fact that the genetics of a disease is sometimes poorly understood and that it is very unlikely

that all of these variants are associated with the phenotype [202]. However, the Bonferroni correction does not always allow applied in situations where there is epistasis or in regions with rare variants with low LD [202]. The results are presented in the first part of table 2.

### Permutation tests

In order to confirm the associations previously made with Haploview, 1000 permutations were made between the patients and the controls (assocCC) as well as 1000 others between family trios where the child suffering from asthma was changed by an unaffected child. Permutations were done using a Python script (<https://www.python.org>). Permutations were performed to ensure that the haplotype-phenotype associations made previously were really and significantly associated to asthma. Data from the permutation test are in the second part of Table 2. A permutation test can be applied to: “evaluate the assumption that genetic variant combinations present in many patients are more likely to be significantly associated with the disorder than combinations found in few patients” [203]. If the data are permutable, the genetic variants likely to be associated with the phenotype are therefore not and the association is summarized by a random event rather than a real link [203].

**Table 2.** Haploview association analysis (CC and TDT)

Chromosome	Nearest genes	Number of significant haplotypes		Permutations p-value	
		CC	TDT	CC	TDT
17q21.33	<i>ZNF652</i>	3	2	0.131	0.376
15q22.2	<i>RORA</i>	4	9	0.191	0.025
10p14	<i>GATA3</i>	9	2	0.120	0.938
6p22.1	<i>GPX5-TRIM27</i>	9	0	0.046	1.000
8q21.13	<i>TPD52</i>	1	2	0.725	0.265
9p24.1	<i>RANBP6</i>	4	5	0.520	0.430
11q13.4	<i>EMSY</i>	4	1	0.049	0.473
6p21.31	<i>HLA-DRB1</i>	1	1	0.501	0.473

## DISCUSSION

Based on this bioinformatic analysis method, 55 haplotypes located in 8 genes that have been associated with asthma in the Saguenay–Lac-Saint-Jean population were identified. More specifically, 33 haplotypic blocks were linked with the asthma phenotype in the case-control association analyzes (CC) as well as 22 other haplotypic blocks in analyzes based on the family trios, although they were not significantly associated ( $p$ -value  $> 0.01$ ) with the phenotype in the SLSJ population. These results were obtained from 18 SNPs associated with asthma in a multiancestry European meta-analysis. Odds ratio (OR) were quantified from the genetic data of the 1394 individuals in the SLSJ asthma cohort. Only those greater than 1.0 were used for the association analyzes. Using Haploview software, case-control association and family trio analyzes were performed to establish whether markers and phenotypes were matched in the SLSJ population. The 55 haplotypic blocks thus identified by asthma via these two types of analysis were then subjected to a permutation test. The aim of the 1000 permutations was to compare the significant haplotypes from previous association analyzes to confirm whether they were really due to a real genetic association with the phenotype and not to an analysis error. Looking at the  $p$ -values of the permutation tests in CC and TDT, we notice that they are all above the corrected threshold of 0.01 (via Bonferonni correction) which means that when the patients and the controls are exchanged (permuted), the resulting statistic does not reach a sufficient level of significance which confirms that the association with the phenotype is indeed due to haplotypic blocks, although if it insignificant. By taking a closer look at the results of the TDT analyzes, we note that the  $p$ -values are slightly higher than those of the CC associations, suggesting that the parent-child genetic link is taken into account and causes less effects on the overall result as case-control associations. With the haplotypic blocks identified, we can reinforce the hypothesis of a link between chromosome 6 and asthma. With 10 haplotypes (CC) found on chromosome 6 in the genes *GPX5-TRIM27* and *HLA-DRB1*, which have already been

associated with asthma as well as childhood asthma and allergic asthma via GWAS studies carried out subsequently, we can therefore state that chromosome 6 influences the development of different types of asthma in the Saguenay – Lac-Saint-Jean population [200, 204-206]. In addition, since the HLA superlocus has already been identified as a recombination hotspot of the human genome, a small number of common haplotypes which represent the majority of the population are found there [207]. A more enhanced characterization of the only haplotype found in HLA-DRB1 would therefore be necessary to confirm that it is also a common haplotypic block outside the Saguenay–Lac-Saint-Jean region. The other 6 genes that comprise the other identified haplotypes have also all been associated with an asthma phenotype in at least one GWAS study (adult, childhood asthma or asthma in general). Thus, the method used was useful to prove the great allelic heterogeneity of the different asthma phenotypes using haplotypic blocks even if the results from analyzes of genetic associations were found to be insignificant ( $p$ -value  $> 0.01$ ). The same experimental method could be reapplied, but this time with a broader phenotypic classification which comes down to more than just a binary classification. As a study has also demonstrated, these methods of haplotypic associations with binary traits do not always allow the gene-environment interactions found in complex traits to be encompassed, which can therefore help to explain why no haplotypic blocks have been linked to asthma [208]. The newly identified haplotypic blocks, although not significant, could be used to study the genetic history of the regional population to better characterize the loci which are specific to it and which diverged during the three founding effects which forged its demogenetic history.

## **CONCLUSION**

The new method of analysis based on haplotypic blocks has identified 55 different haplotypes in the Saguenay–Lac-Saint-Jean population divided into 8 genes that have been previously linked to asthma. Unfortunately, the association of these 55 haplotypic blocks with asthma phenotype was inconclusive. This discovery will make it possible to enrich and increase the knowledge that we have on the particular genetic structure of the regional population. These haplotypes can be used to conduct a validation study on an external cohort such as the Epidemiological study of the Genetics and Environment of Asthma, bronchial hyperresponsiveness and atopy (EGEA) cohort to increase knowledge on the pathogenicity of these blocks in asthma and other atopic diseases. In addition, the use of these haplotypic blocks in the newly developed polygenic risk score in the same regional cohort could contribute to enriching its clinical predictivity so as to better identify infants (<3 months) that are at higher risk of developing asthma and atopic diseases.

## **ACKNOWLEDGEMENTS**

We would like to thank all the participants who took time to participate in this study. Also, Mathieu Simard was awarded a master's grant from the *Réseau en santé respiratoire du Québec* (RSR). Catherine Laprise (C.L) is the director of the *Centre intersectoriel en santé durable de l'UQAC*, is the co-director of the Environment / Genetics / Cancer Axis of the Quebec Respiratory Health research Network and Canada Research Chair in the Environment and Genetics of Respiratory Disorders and Allergies ([www.chairs.gc.ca](http://www.chairs.gc.ca)).

## DISCUSSION

Ce projet de recherche de maîtrise avait pour but de mieux documenter la distribution ainsi que la transmission de variants génétiques impliqués dans la marche atopique et l'asthme dans la population canadienne. Le développement d'un outil de prédiction sous forme d'un score de risque polygénique (SRP) a ainsi permis d'identifier les variants génétiques de ces phénotypes à partir de la structure génétique des Canadiens.

Dans un premier temps, deux cohortes canadiennes distinctes ont été employées pour identifier les variants génétiques les plus associées à la marche atopique et à l'asthme dans la population. Avec son histoire démogénétique relativement jeune et bien documentée, la population du Saguenay–Lac-Saint-Jean s'avère être fort utile dans les recherches en génétique clinique et fondamentale. Sa contribution conjointe à la cohorte canadienne anglaise CHILD a permis d'établir un score de risque polygénique prédictif de la dermatite atopique et des manifestations cliniques de la marche atopique (allergies alimentaires, asthme allergique et rhinite allergique) lui étant associées. Le modèle expliquant le plus le risque de développer la DA est celui qui prend en compte les 25 polymorphismes pangénomiques issus des deux cohortes canadiennes avec l'ajout des covariables (sexe, âge et ethnie parentale). Ce modèle discrimine correctement les patients dans 85% des cas et contribue à expliquer 36% de la variance. Cette contribution permettra sans doute de mieux identifier les nouveau-nés à risque de développer la dermatite atopique en se basant sur leur architecture génétique personnelle dans l'objectif ultime d'améliorer la qualité de vie de ces personnes. Ultimement, on espère que cet outil prédictif pourra devenir une application clinique directe et sera utilisée par les cliniciens canadiens. De plus, un second modèle prédictif binaire a été établi pour mieux mettre en évidence la pertinence de l'outil prédictif dans le contexte de la marche atopique. Il a

permis de discriminer les patients souffrant de dermite atopique modérée à sévère à 85% pour les allergies alimentaires, 81% pour l'asthme et 80% pour la rhinite allergique. Toutefois, pour qu'un SRP puisse être mis en pratique dans le domaine médical, les aires sous la courbe (AUC) obtenues doivent d'au moins 90% [209]. Pour contrer le faible pourcentage manquant dans la mesure de la AUC, une analyse de l'héritabilité des marqueurs épigénétiques pourraient être effectuées pour ainsi venir compléter le SRP développé en y ajoutant les marques épigénétiques les plus associées dans la population canadienne. L'ajout de ces marqueurs épigénétiques (méthylation, acétylation ou phosphorylation de l'ADN et/ou des histones voire encore les petits ARN non codants [miARN, siARN et/ou lncARN]) permettrait également de venir étudier plus en profondeur la composante environnementale de la DA [210]. Dans le SRP développé dans le cadre de cette maîtrise, cinq long ARN non codants (lncARN) [LINC02055, LINC01503, LINC00680, LINC00603 et LINC01653] ont déjà été associés à la DA dans la population canadienne ce qui vient donc renforcer le lien probant entre données génétiques et épigénétiques dans la modulation de cette maladie. Comme la DA est considérée comme un trait complexe, c'est-à-dire, une maladie qui est influencée par une composante génétique et environnementale, l'ajout d'informations issues de l'épigénome donnerait une vue d'ensemble plus large quant à la nature complexe de cette maladie affectant 9% des enfants canadiens [111]. Par contre, certains pourraient être tentés d'affirmer qu'un SRP qui comporte plusieurs marques génétiques/épigénétiques associées serait un meilleur outil de prédiction de la maladie car il prend en considération la somme de l'effet de plusieurs locus cumulés. En effet, l'emploi d'un nombre trop important de marqueurs génétiques/épigénétiques dans la construction du SRP a plus d'effets néfastes que bénéfiques car cela augmente considérablement l'ampleur de l'effet réel, d'où l'importance d'être judicieux dans le choix des marqueurs [211]. L'ajout de ces données, judicieusement sélectionnées, au SRP va également permettre de développer un autre outil prédictif novateur nommé *Comprehensive Genetic Testing for Severe Allergy* (CGTA) par l'équipe de la

Chaire de recherche du Canada en environnement et génétique des troubles respiratoires et de l'allergie du Dre Laprise.

Dans un second temps, dans l'optique de confirmer la valeur prédictive du nouvel outil développé, la réplication des données dans une cohorte familiale européenne externe à la population canadienne devrait être employée pour s'assurer que la variance expliquée ainsi que le pouvoir de discrimination du SRP soient réellement caractéristiques des Canadiens. De plus, cette réplication permettrait également de valider si les marqueurs génétiques associés à la DA dans la population canadienne possèdent la même somme pondérée (effet cumulé) pour chacun des locus sélectionnés. Une étude a documenté un phénomène nommé *flip-flop* dans laquelle des marqueurs génétiques associés au développement du cancer du sein chez les femmes de certaines populations se trouvaient protecteur chez des femmes d'ethnies différentes [212]. Compte tenu de la diversité ethnique de la population canadienne (et surtout canadienne anglaise), une telle étude de réplication s'avère donc pertinente dans le cadre de cette étude et pour des fins de validation de l'outil prédictif développé.

De plus, parallèlement au projet de développement du score de risque polygénique pour la dermite atopique et la marche atopique dans la population canadienne, une étude basée sur les haplotypes impliqués dans l'asthme dans la cohorte du Saguenay–Lac-Saint-Jean a été effectuée. Via deux cas d'analyses d'associations génétiques soit cas-témoins et trio familiaux, 55 haplotypes répartis dans 8 gènes ont été recensés dans la population régionale. Quoique que l'association ait été non significative avec le phénotype d'asthme, ces haplotypes nouvellement identifiés pourrait se voir être utiles dans une future étude de validation sur une cohorte externe tel la *Epidemiological Study on The Genetics and Environment of Asthma, Bronchial Hyperresponsiveness and Atopy* (EGEA) pour augmenter les connaissances sur la pathogénicité de ces blocs dans l'asthme et les autres maladies



atopiques. Une des raisons majeures qui pourrait expliquer ce pourquoi les associations génétiques aient été non significatives repose sur le fait que les interactions gènes-environnement ne sont pas prises en considération lorsque que le phénotype d'un trait complexe comme l'asthme (c'est-à-dire avec un ensemble de facteurs environnementaux et de facteurs génétiques) est classé selon une classification dite binaire (un groupe contenant les personnes affectées et l'autre les personnes non affectées) qui ne tient pas compte des phénotypes intermédiaires résultant ainsi en une perte de diversité phénotypique. De plus, l'emploi de ces blocs haplotypiques dans le score de risque polygénique développé à partir de la même cohorte régionale pourrait contribuer à enrichir sa prédictivité clinique pour ainsi mieux identifier en bas âge (<3 mois) les enfants à risque de développer de l'asthme et des maladies atopiques. Cela permettrait également d'établir un lien entre les deux contributions effectuées au cours de ce projet de maîtrise et pourrait peut-être même, dans le meilleur des scénarios, contribuer à améliorer la vie de plusieurs personnes.

## CONCLUSION

La mise en œuvre de ce SRP contribuera, avec un grand espoir, à diminuer la prévalence de la DA et de la marche atopique dans la population canadienne. L'identification des enfants de moins de 3 mois les plus à risque de développer ces maladies pourra être effectuée et, ainsi, permettre de les traiter le plus tôt possible pour contribuer à améliorer leur qualité de vie future. Des démarches ont déjà été effectuées pour amorcer l'identification des marqueurs épigénétiques dans la DA pour ainsi améliorer le SRP développé. En espérant que cet outil puisse être utilisé dans les hôpitaux du pays pour contribuer à aider les patients atteints ainsi qu'à diminuer les coûts sur les systèmes de santé provinciaux et territoriaux.

Dans des perspectives futures, comme il a déjà été mentionné dans ce mémoire, la définition clinique du phénotype de la DA est très variable d'un pays à l'autre et, parfois même, variable à l'intérieur d'un même pays. Pour contrer ce problème de définition très large, un outil diagnostique de caractérisation phénotypique basé sur la génétique devrait être développé pour ainsi mieux diagnostiquer les patients. En se basant sur les différences génétiques des individus, cet outil permettrait d'uniformiser la définition de la DA. De plus, comme le développement du SRP pour la DA qui discriminait correctement les nouveau-nés à risque dans 85% des cas tout en expliquant 36% du phénotype clinique a permis de mieux comprendre l'héritabilité expliquée par les meilleures associations géniques dans la population canadienne, une étude plus approfondie sur un des gènes expliquant le plus l'héritabilité de la DA serait requise. Ce gène particulier, la profilaggrine (*FLG*), explique plus de 10% de l'héritabilité de la DA [49]. Sa perte de fonction entraîne une formation de la barrière cutanée incomplète laissant donc les allergènes la pénétrer et initiant donc une réponse immunitaire et allergique. Comme ce gène explique une bonne partie de l'héritabilité de la maladie et

comme la DA représente la première manifestation de la marche atopique, la restauration de l'allèle muté vers l'allèle de type sauvage par la technologie d'édition génique CRISPR-Cas9 (ou encore par interférence à l'ARN [ARNi]) chez les enfants atteints pourraient s'avérer être une stratégie thérapeutique utile pour endiguer la progression des symptômes atopiques chez les enfants identifiés comme atopiques à un jeune âge. De plus, cette stratégie d'édition génétique pourrait s'appliquer à plus que la seule population canadienne et pourrait permettre de diminuer la prévalence mondiale de la DA ainsi que son fardeau économique.

# Annexe 1

Matériel supplémentaire de l'article : Développement d'un score de risque polygénique pour la dermatite atopique dans la population canadienne

**Tableau 14** Les 25 meilleures associations pangénomiques dans la littérature

Chromosome	SNP	HGVS nomenclature	Nearest gene distance (bp)	OR	p-value	Reference (PMID)
1	rs12130219	g.152162106A>G	<i>HRNR</i> / 22,446 (FLG / 112,545)	-	1E-23	25574825
	rs12144049	g.152440910C>T	<i>LCE5A</i> / 42,410 (FLG / 143,231)	1,54	3E-30	25574825, 25865352
	rs61813875	g.152536650C>G	<i>LCE3E</i> / 1,525	1,61	6E-29	26482879
	rs77199844	g.152757095_152757096del	<i>LCE1E</i> / 1,657	1,23	2E-17	25574825
2	rs13015714	g.102971865G>T	<i>IL18R1</i> / 391	1,27	8E-18	23042114, 26482879
	rs62176107	g.179300971G>A	<i>PRKRA</i>	-	1E-34	25574825
3	rs7613051	g.33065339G>A	<i>GLB1</i>	1,29	6E-21	23042114
	rs12634229	g.112376308T>C	<i>CCDC80</i> / 16,318	1,29	2E-19	23042114
5	rs6596086	g.131952222T>C	<i>RAD50</i>	-	2E-18	25574825
	rs17728338	g.150478318G>A	<i>ANXA6</i> / 1,949	-	4E-38	25574825
	rs11135056	g.158687281C>T	<i>UBLCP1</i> / 2,794	1,45	1E-25	25574825
	rs2546890	g.158759900A>G	<i>IL2B</i> / 2,419 (LOC285626)	1,39	3E-35	25574825
6	rs9368677	g.31272321G>A	<i>HLA-C</i> / 32,408	1,36	1E-17	23042114
	rs176095	g.32158319T>C	<i>GPSM3</i> / 224	1,40	8E-20	23042114, 25574825
	rs9469099	g.32308908G>A	<i>TSBP1-AS1</i> , <i>TSBP1</i>	1,61	5E-19	23042114
	rs240993	g.111673714T>C	<i>REV3L</i>	-	6E-18	25574825
	rs9481169	g.111929862G>T	<i>TRAF3IP2</i> / 2,385	1,58	1E-26	25574825
8	rs6473227	g.81285892C>A	<i>ZBTB10</i> / 111,962	-	5E-18	26482879
10	rs10995251	g.64398466C>T	<i>ZNF365</i>	1,28	6E-20	23042114
11	rs878860	g.7968359C>T	<i>NLRP10</i> / 12,797	1,31	2E-22	23042114
	rs10791824	g.65559266A>G	<i>OVOL1</i>	1,12	2E-19	26482879
	rs7127307	g.128187383T>C	<i>ETS1</i> / 141,273	-	1E-20	26482879
18	rs12458130	g.8667061G>A	<i>GACAT2</i> / 40,558	-	4E-16	25574825
19	rs380743	g.49241014G>A	<i>RASIP1</i>	-	4E-22	25574825
20	rs6020157	g.48591758G>A	<i>SNAI1</i> / 7,755	-	4E-17	25574825

**Tableau 15** Les 25 meilleures associations pangénomiques dans les cohortes canadiennes combinées

Chromosome	SNP	HGVS nomenclature	Nearest gene distance (bp)	Allele	Test	SE	p-value
1	rs35854925	g.101733553A>G	<i>SIP1R1</i> / 26,477	G	ADD	0,228	1,76E-06
					DOMDEV	0,262	0,1675
					GENO_2DF	NA	4,33E-10
	rs150104310	g.218161147A>T	<i>LINC01653</i> / 55,700	T	ADD	0,434	0,05251
					DOMDEV	0,465	0,3709
					GENO_2DF	NA	5,71E-12
2	rs62151071	g.81145487G>A	<i>CTNNA2</i> / 269,494	A	ADD	0,254	3,24E-05
					DOMDEV	0,302	0,9308
					GENO_2DF	NA	5,02E-12
	rs62151125	g.81163281T>G	<i>CTNNA2</i> / 287,288	G	ADD	0,254	3,27E-05
					DOMDEV	0,301	0,8887
					GENO_2DF	NA	2,99E-11
rs213549	g.234786586A>G	<i>MSL3P1</i> / 9,531	G	ADD	0,087	2,14E-54	
				DOMDEV	0,135	0,002586	
				GENO_2DF	NA	2,14E-55	
3	rs9825865	g.1865710T>C	<i>CNTN4</i> / 274,840	T	ADD	0,094	8,97E-12
					DOMDEV	0,128	0,09701
					GENO_2DF	NA	6,99E-11
	rs12633281	g.3278270T>G	<i>IL5RA</i> / 126,212	G	ADD	0,501	0,01775
					DOMDEV	0,534	0,4352
					GENO_2DF	NA	3,23E-16
	rs6778759	g.29956943G>T	<i>RBMS3</i>	T	ADD	0,147	7,52E-13
					DOMDEV	0,189	0,001137
					GENO_2DF	NA	1,55E-36
	rs6799780	g.37410704T>G	<i>GOLGA4</i> / 2,334	G	ADD	0,111	3,75E-25
					DOMDEV	0,157	0,0008466
					GENO_2DF	NA	1,14E-43
4	rs11737164	g.154970397G>A	<i>DCHS2</i> / 185,130	A	ADD	0,155	0,0003137
					DOMDEV	0,190	0,06937
					GENO_2DF	NA	8,41E-12
	rs1438119	g.166348346T>C	<i>CPE</i>	C	ADD	0,094	9,15E-21
					DOMDEV	0,135	0,2966
					GENO_2DF	NA	4,57E-23
rs10011527	g.184507881C>T	<i>RWDD4</i> / 52,907	T	ADD	0,176	0,03793	
				DOMDEV	0,242	0,0009011	
				GENO_2DF	NA	2,03E-10	
5	rs10053502	g.39979172C>T	<i>LINC00603</i> / 73,221	T	ADD	0,145	2,67E-16
					DOMDEV	0,188	0,2538
					GENO_2DF	NA	2,20E-31
	rs181100198	g.155582304G>T	<i>SGCD</i> / 171,463	T	ADD	0,383	0,000554
					DOMDEV	0,426	0,7867
					GENO_2DF	NA	1,38E-14
	rs2862717	g.156499054C>T	<i>HAVCR1</i> / 12,927	C	ADD	0,105	3,70E-10
					DOMDEV	0,147	0,1701
					GENO_2DF	NA	7,48E-10
	rs10600273	g.156530399_156530401del	<i>HAVCR2</i>	TAAA	ADD	0,110	4,05E-12
					DOMDEV	0,151	0,0561
					GENO_2DF	NA	7,42E-12
6	rs112464710	g.58757732A>G	<i>LINC00680-GUSBP4</i> / 470,008	G	ADD	0,225	1,07E-05
					DOMDEV	0,256	0,0367
					GENO_2DF	NA	1,33E-27
	rs2321443	g.79204960T>C	<i>IRAK1BP1</i> / 372,301	C	ADD	0,097	6,59E-12
					DOMDEV	0,141	0,2533
					GENO_2DF	NA	1,42E-14

**Tableau 15** Les 25 meilleures associations pangénomiques dans les cohortes canadiennes combinées (suite...)

7	rs854560	g.94946084A>T(p.Met55Leu)	PON1	A	ADD	0,086	1,85E-16
					DOMDEV	0,130	0,2339
					GENO_2DF	NA	8,70E-16
7	rs11984204	g.120352110T>G	KCND2	G	ADD	0,136	1,22E-14
					DOMDEV	0,176	0,6792
					GENO_2DF	NA	9,21E-24
8	rs4554481	g.19373265T>C	CSGALNACT1	C	ADD	0,359	0,004634
					DOMDEV	0,424	0,09308
					GENO_2DF	NA	7,43E-14
8	rs1841513	g.137811069T>C	LINC02055	T	ADD	0,093	1,81E-35
					DOMDEV	0,148	0,059
					GENO_2DF	NA	1,25E-42
9	rs17441316	g.132108247C>T	LINC01503	T	ADD	0,257	0,01402
					DOMDEV	0,351	0,001706
					GENO_2DF	NA	2,07E-12
13	rs4885693	g.80650117C>G	LINC01080 / 49,319	G	ADD	0,111	3,49E-08
					DOMDEV	0,145	0,07671
					GENO_2DF	NA	8,96E-13
16	rs16960052	g.83304230T>C	CDH13	C	ADD	0,184	0,0004515
					DOMDEV	0,228	0,1303
					GENO_2DF	NA	1,03E-11

**Tableau 16** Caractéristiques phénotypiques des participants de deux cohortes canadiennes

	SLSJ cohort (n=1,214)	CHILD cohort (n=3,081)
M :F ratio	1 :1.17	1 :0.91
Age, mean (range) <sup>a</sup>	38 (0-93)	1 (1-1)
Age, median	39	1
Asthma, n(%)	592 (49)	244 (9)
Allergy, n(%)	670 (57)	175 (7)
Rhinitis, n(%)	372 (31)	224 (9)
Atopic dermatitis, n(%)	478 (40)	330 (12)
Moderate to severe atopic dermatitis, n(%) <sup>b</sup>	478 (40)	74 (3)

<sup>a</sup>For the CHILD cohort, age at follow up for atopic dermatitis

<sup>b</sup>For the SLSJ cohort, all individuals with atopic dermatitis

# Annexe 2

Revue de littérature sur l'impact de l'âge du père à la  
conception sur la santé globale de l'enfant



# Impact of paternal age at conception on human health

Mathieu Simard<sup>1,2</sup>, Catherine Laprise<sup>1,2</sup> et Simon Girard<sup>1,2</sup>

<sup>1</sup>Département des sciences fondamentales, Université du Québec à Chicoutimi, Saguenay, Québec, Canada

<sup>2</sup>Centre Intersectoriel en Santé Durable, Université du Québec à Chicoutimi, Saguenay, Québec, Canada

Cet article, publié en janvier 2019 dans *Clinical Chemistry* (PMID: 30602478), répertorie les changements physiologiques associés au vieillissement des géniteurs de sexe masculin ainsi que l'évolution de la génétique des cellules germinales masculines dans la transmission d'allèles potentiellement délétères chez leurs progénitures [213].

## Contribution des auteurs :

Mathieu Simard a recensé la littérature scientifique et a écrit la première version du papier. Catherine Laprise et Simon Girard ont révisé l'article et ont également contribué à enrichir le texte original à l'aide de leur expertise respective.

## Source:

Simard, M., C. Laprise, and S.L. Girard, Impact of Paternal Age at Conception on Human Health. *Clin Chem*, 2019. 65(1): p. 146-152.

## **Résumé**

Dans plusieurs populations occidentales, l'âge à laquelle les parents ont leur premier enfant est en augmentation. L'effet de l'âge maternel à la conception sur la santé de leurs progénitures est un phénomène mieux documenté que celui de l'effet de l'âge paternel. Cette revue de littérature explore les changements physiologiques et morphologiques observés dans les cellules germinales des testicules des pères âgés ainsi que les effets associés sur les spermatozoïdes. Elle se poursuit par une étude des changements génétiques qui sont associés à l'âge paternel à la conception et leurs conséquences phénotypiques sur la santé des progénitures issus d'unions avec des pères plus âgés.

## **Abstract**

### Background

The effect of maternal age at conception on various aspects of offspring health is well documented and often discussed. We seldom hear about the paternal age effect on offspring health, although the link is now almost as solid as with maternal age. The causes behind this, however, are drastically different between males and females.

### Content

In this review article, we will first examine documented physiological changes linked to paternal age effect. We will start with all morphological aspects of the testis that have been shown to be altered with aging. We will then move on to all the parameters of spermatogenesis that are linked with paternal age at conception. The biggest part of this review will focus on genetic changes associated with paternal age effects. Several studies that established a strong link between paternal age at conception and de novo mutations rate will be reviewed. We will next discuss paternal age effects associated with telomere length and try to better understand the seemingly contradictory results.

Finally, severe diseases that affect brain functions and normal development have been associated with older paternal age at conception. In this context, we will discuss the cases of autism spectrum disorder, schizophrenia as well as several childhood cancers.

### Summary

In many western civilizations, the age at which parents have their first child has increased significantly in the last decades. It is important to summarize major health issues associated with an increased paternal age at conception in order to better model public health systems.

### **Introduction**

During the Dutch famine of 1944-45, a Nazi blockade shut off fuel and food shipments destined for large towns. Following this, several thousands of people were undernourished for many months. This is one of the most documented famines in history. Years later, scientists started to formulate hypothesis about the lack of food that mothers experienced during their pregnancy and the possible effects on their offspring [214]. These findings were put in relation with previous observations documenting an increase of the incidence of autosomal trisomy 21 with the mother's age at conception [215]. Despite the fact that this effect, now called maternal age effect, is well documented and well covered in mainstream media, we seldom hear about the effect of paternal age at conception on children's health. This is quite puzzling as scientists have been studying the Paternal Age Effect (PAE) for more than a century. In 1912, Wilhelm Weinberg, a German obstetrician, was confronted by a strange disease that was responsible for dwarf limbs in children. Intriguingly, these dwarf limbs seemed to appear only in the last-born child of a family [216]. At the time he made these observations, the new science of genetics had barely reached its infancy and Weinberg could not use it to explain these observations [217]. 41 years later, another scientist named Robert Krooth

published a study about the very same disease [218]. He was the first to use the expression paternal age effect. Throughout the last decade, it was finally established and acknowledged that the age at conception for both parents can significantly influence progeny through various pathways ranging from de novo mutations in their germ cells to complex physiological mechanisms. These observations are of crucial interest as families in Western civilizations tend to have their first child at a more advanced age compared to previous generations [219]. Throughout this review article, we will document and discuss several instances where the paternal age effect significantly impacts the health traits in children.

### **Testicular morphological changes**

One of the simplest ways to study changes occurring when men age is by examining the morphology of their testis. As for many organs in the human body, testis function changes over time. Additionally, even though ageing occurs at a slower pace in male reproductive organs compared to female, the male aging process eventually reaches a point where it can significantly impair the proper function of the testis. In a vast study conducted by Gunes et al. in Turkey, it was shown that the testicular volume progressively increased in men aged 11 to 30 years old and then reached a plateau between age 30 to 60 [220]. After 60 years of age the testicular volume starts to slowly decrease. The authors reported that the volume of the testis is 31% smaller in men over 75-year-old than those between 18-40 years of age. Mahmoud et al. reported that this decreasing testicular volume was associated with an increased level of the follicle-stimulating hormone (FSH) from the hypophysis and with a reduction of the number of Sertoli cells [221]. Later, it was shown that the increase of the follicle-stimulating hormone level could be explained by a reduction of the number of Leydig cells, the cells that secrete testosterone [222]. In the same study, Midzak et al. suggest that the luteinizing hormone (LH) no longer stimulates the Leydig cells [222]. This lack of Leydig cell stimulation alters the transfer of

cholesterol into the mitochondria and cause the inhibition of the metabolism of cholesterol to testosterone. In addition, paternal age is not only associated with testicle size, but also with the global number of spermatozoa. Levitas et al., in a study of 4604 males, recorded a diminution of 24.3% of the global number of spermatozoa per male between the age of 30 and 55 [223]. Not only was the number of spermatozoa decreased, but sperm motility was diminished by 53% over the same age period. Kumar et al. reported an overall negative association between age and semen parameters such as semen volume, total sperm count, sperm motility as well as morphology [224].

### **Mutations in spermatogonial cells**

During spermatogonia, the differentiation of germ cells into spermatozoa (a phenomenon called spermatogenesis) takes place in the basal membrane of the seminiferous tubules, located in the testis. Each testicle is composed of 200 to 300 testicular lobules and each of these lobules contains between 1 and 4 seminiferous tubules [225]. The wall of seminiferous tubules is composed of two types of cells: 1) Sertoli cells, which are feeder support cells for sperm maturation and 2) germ cells. Germ cells are basic spermatogonia cells responsible for maintaining the number and the integrity of spermatozoa. It is these spermatogonia that will later divide to form primary spermatocytes by mitotic division. These primary spermatocytes, still in the diploid stage, will subsequently be moved to the abluminal compartment where they will divide by meiosis into secondary spermatocytes to reach the haploid stage. These secondary spermatocytes will divide again to form the haploid spermatids. Finally, during spermiogenesis, spermatids will mature to form functional spermatozoa. Males produce billions of spermatozoa in their life. It is well documented that as the spermatogenesis chains up divisions by meiosis, mutations are expected to be incorporated into the genome of the male undifferentiated germ cells. Goriely and Wilkie reported that cells with mutated DNA are selected prior to those without mutations [226]. The authors also reported that these mutated cells expand

clonally. Over time, this type of mutation takes more importance and the number of mutated undifferentiated germ cells increases, more damageable consequences can be incorporated in the spermatozoa themselves. This mechanism will be discussed more in details in a later section.

### ***De novo mutations***

Mutations that appear sporadically in an individual are called de novo mutations. These mutations can occur in cells that are diploid ( $2n$  chromosomes) and undergo division by mitosis called somatic cells. Somatic cells constitute most of the cells in the human body except those that undergo meiosis to form haploid ( $n$  chromosome): oocyte and spermatozoa. When de novo mutations occur in meiotic cells they are defined as germline de novo mutations. These mutations can contribute to developmental disorders and traits notably by affecting the protein synthesis machinery. In a stunning breakthrough, Kong et al. studied de novo mutations rates in 219 Icelandic individuals affected with either autism spectrum disorder or schizophrenia (78 progenies and their two parents) and 1,859 unaffected individuals [227]. They have shown that the rate of de novo mutations was about  $1.20 \times 10^{-8}$  per nucleotide each generation. This means that for a diploid genome of about 6 billion base pairs such as the human genome, there are about 50-100 new mutations that will occur in the gamete cells of the parents and will be transmitted to one or multiple offspring. From this number, we would typically expect that between 0-2 mutations will be found in genomic sequences directly coding for proteins [228]. In that study, the average paternal age at conception was 29.7 years old. Although the rate of de novo mutations is consistent across all human populations, Kong et al. reported that the father age at conception was the main contributing factor to the variation of the de novo mutations rate in offspring. An interesting finding of this Icelandic population study was that the de novo mutations rate was increased by a factor of two for every 16.5 years increase in paternal age [227]. Our team also investigated the de novo mutations rate in a newborn twin cohort and found that the

offspring of fathers older than 40 years carried almost twice as much de novo mutations than those 20 years younger [229]. Kloosterman et al., in a study of 258 individuals in the Netherlands, have investigated indels and single nucleotide variants (SNV) prevalence [230]. They found that mutations with paternal origins were more consequential and frequent in progeny. They also reported 66.1% more indels and SNVs on paternal chromosomes than the maternal ones. We now know that germline mosaicism (or gonosomal mosaicism), the case where several germ cell (sperm or oocyte) carries a specific mutation, is more frequent than we once thought. Acuna-Hidalgo et al. have shown that they can constitute more than 5% of all germline mutations resulting in an increased gonosomal mosaicism in gametes [231].

### **Selfish sperm selection**

A basic population genetics concept is that harmful mutation will decrease an individual's fitness for survival unless it is counterbalanced by a beneficial effect at least as strong. In that regard, de novo mutations are unique as they escape this process for the generation in which they are studied. However, when transmitted to the next generation, de novo mutations will act as any mutations and will be affected by natural selection processes. However, recent findings show that at least some de novo mutations undergo another type of selection that can impact an individual's health. Selfish spermatogonial selection is a phenomenon that occurs in the testis. It has been described to be the result of mutations/variants dysregulating the RAS signalling pathway [232], a pathway involved in growth and cell division. When a germ cell develops such a mutation, it can confer a certain proliferative benefit to the spermatogonia, for example mutations in the RAS signalling pathway proteins may result in an uncontrolled growth [232]. Over time, these spermatogonia cells undergo meiosis and develop more and more of these mutant spermatozoa, increasing their fitness over the other germ cells. The risk of transmitting this specific mutation to the offspring thus increases as the

father gets older [232]. These spermatozoa are selected over others because the mutation gives them selective advantages and they can survive more easily. For this reason, selfish sperm selection can be a harmful mechanism that highly contributes to the transmission of mutations through future generations.

### **Telomere length**

Telomeres are small nucleotide sequences of TTAGGG that cap chromosomes arms. They prevent genomic instability by blocking neighbouring chromosomes from crossing-over or simply fusing together [233]. Through time, multiple studies have documented that telomere length decreases with ageing [234-236]. It is also known that once telomeres reach a critical threshold, the cell will stop its replication by stopping its cell-cycle processes or by activating apoptosis pathways [237]. The enzyme responsible for keeping chromosomes telomeres integrity is called telomerase. Telomerase is a ribonucleoprotein composed of a long non-coding RNA, the telomerase reverse transcriptase (TERT) and other proteins [238]. When the telomerase complex is formed, it is directed to a specific place within the cell's nucleus called the Cajal bodies. During the S phase of the cell cycle, the telomerase complex is bound to the telomeres of chromosomes allowing telomeres synthesis at the end of chromosomes [239]. Surprisingly, Eisenberg et al., in a Philippine cohort, reported that males with advanced age have longer telomeres and that their progeny will also inherit longer telomeres [240]. Other studies also observed that the telomeres of aged fathers were longer and that those of their progeny were also more elongated [241, 242]. This lengthening of telomeres in germ cells is opposed to the classical mechanism of shortening over time and cell replications. The telomeres in male germ cells are increasing in length at about 17 bp per year [243]. At first, it sounds counterintuitive that men sperms escape the shortening of their telomeres that all somatic cell telomeres undergo. Hjelmborg et al., in a study on monozygotic and dizygotic twins, hypothesized



that there are some germ stem cells (GSCs) in male testis that possess longer telomeres. Through replication cycles, GSCs with longer telomeres have greater chances of survival due to the protection of their telomeric ends [244]. This would provide them with a positive selection process. After a number of years, there are more GSCs with long telomeres than GSCs with small telomere. When spermatogonial cells replicate, they will tend to form sperm with long telomeres [244]. In summary, as men produce over millions of spermatozoa per gram of testis a day, the development of a mechanism to avoid the shortening of telomeres was essential [245]. Nördfjall et al. conducted a study on a Swedish population of 962 individuals by extracting gDNA (genomic DNA) and estimating the telomere length by real-time polymerase-chain reaction and concluded that telomere length is mainly paternally inherited [246]. It has been also reported by Ozturk that the telomerase activity is less regulated in the last stages of spermatogenesis and thus, lead to an increase in telomere length [247]. Further studies need to be conducted to identify genes or epigenetic factors that may be implied in telomeric expression in germ cells.

#### A link between molecular biology mechanisms and evolutionary biology

Telomere length of spermatozoa increases with paternal age. This apparent contradiction with known natural selection processes is called phenotypic plasticity [248]. Fusco and Minelli define phenotypic plasticity as: "the ability of individual genotypes to produce different phenotypes when exposed to different environmental conditions" [249]. Darwin's natural selection theory predicts that as an organism age, the cost of reproduction and livelihood cannot both be assumed at the same time because it would be too costly. To skirt this energy problem, organisms must invest in one or another in the appropriate situation. If it feels that its life is threatened it must invest in reproduction (with time if the organism is not life threatened, the alleles implied in livelihood will be favoured by natural selection). Phenotypic plasticity is the closest model of Lamarck's theory of acquired traits passed

through generations. This also translate to de novo mutations in germ cells and the selfish sperm selection that can be transmitted to future generations by the older fathers, as they are acquired in the father spermatozoa and passed to its progeny.

### **Chromosomal aneuploidies**

Aneuploidy defines an abnormal number of chromosomes, due to loss or gain of part of a chromosome set resulting in a chromosome number that is not a multiple of haploid chromosome number (23 in humans). Most cases of aneuploidy will end in spontaneous abortion but, in cases leading to a live birth, progeny show severe deformities, in most of the instances [250]. Only about 1% of chromosomal aneuploidies will end up in a living offspring and 6% of stillbirths show chromosomal abnormalities [250]. Autosomal aneuploidies tend to be more maternally inherited [251] with trisomy 16 exclusively maternal due to a non-disjunction in meiosis I [252]. Paternal inheritance of aneuploidies has a prevalence of 55% in sex chromosomes [251]. Lowe et al. discovered that the age of a man correlated with a higher rate of XY sperm (a spermatozoa with an X chromosome and a Y chromosome), with a 3.5% increase each year [253]. This may result in an XXY karyotype when the fecundation takes place between one XY sperm and an XX oocyte. This phenomenon corresponds to Klinefelter syndrome. However, another study showed that in 50% of cases, the Klinefelter phenotype was due to a non-disjunction in meiosis I of the father and the other 50% was maternal [254]. Emerging evidence in Klinefelter syndrome is conflicting and there is no consensus on the heritability as well as a potential paternal age effect [255]. Buwe et al. reported that 6% of XXX, 50% of XXY, and 100% of XYY aneuploidies were due to a non-disjunction in the paternal meiosis [256]. It is still not well understood whether paternal age plays a role in autosomal aneuploidies (aneuploidies that are not on sex chromosomes). For trisomy 21 (Down syndrome), only 5% of the cases would have a paternal origin. For other autosomal aneuploidies such as trisomy 22,

13, 15 and 17, the father influence seems to have more impact with respective paternal contribution of 11% (trisomy 22), 12% (trisomy 13 and 15) and 17% (trisomy 14) [256].

### **Paternal age effect disorders**

As discussed previously, the age of the father at conception has a drastic effect on the number of de novo mutations that will be transmitted by germ cells. Extrapolation of this concept could mean that all diseases caused by a genetic defect could be impacted by the age of that father at conception. While we cannot support that notion with our current state of knowledge, we present here several examples of diseases for which the prevalence is highly associated with paternal age at conception.

### **Autism Spectrum Disorders (ASD)**

According to the World Health Organization (WHO), autism spectrum disorder is characterised by impaired social behaviour, communication and language, and a narrow range of interests and activities that are both unique to the individual and carried out repetitively” [WHO, 2017]. In 2012, worldwide, up to 6-10/1,000 individuals are affected by ASD, representing around 67 million people [257]. Janecka et al. highlighted that the de novo mutations rate increasing with paternal age is over three times higher in male than in female [258]. So male de novo mutations mainly account for neurodevelopmental disorders in offspring. The same study shows that the number of de novo mutations is doubling every 16.5 years, meaning that fathers of 40 years old have more than twice the numbers of de novo mutations than those of 20 years old [258]. Gabis et al., in a 323 autistic children study, found that men who have children when they are aged 45 years and older have 1.4% greater chance of having an offspring suffering from ASD [259]. A three-generation study conducted by Frans et al., suggested that the risk of having an offspring affected by ASD increases with paternal

age. Surprisingly, they also showed that men who had fathered a son or a daughter when they were 50 years or older were more likely to have a grandchild affected by ASD [260].

## **Schizophrenia**

Schizophrenia (SCZ) is a psychotic disorder which gives rise to visual and auditory hallucinations, delirium and in a reduction of everyday activities. It has a huge social impact on those who are affected. The prevalence of SCZ is about 3-7/1000 affected individuals per year [261]. Malaspina et al. reported that the offspring's risk of developing schizophrenia when the father was more than 45 years old at conception was three times higher than those of 20-24 years old [262]. Sipos et al. in a cohort study of 754,330 individuals in Sweden concluded that the increasing number of de novo mutations in the father's sperm cells is the main contributing factor to schizophrenia when concerning PAE [263]. Jenkins et al. identified different DNA methylation patterns in some genes involved in schizophrenia and other mental disorders [264]. The genes DRD4 and TNXB, on chromosome 11 and chromosome 6 respectively, were hypomethylated (more active genes) with increasing father's age. However, the global involvement of the TNXB gene in schizophrenia is still unclear [265, 266].

## **Childhood Cancers**

In 2018, in the United States of America, pediatric cancers are the second most common cause of death in children [American Cancer Society]. Around 11,000 children will be diagnosed with cancer before they reach the age of 15 and about 80% of them survive for 5 years or over [American Cancer Society]. A Swedish study conducted by Yip et al. on 7 million registered children since 1932 reported that males older than 40 years at conception had 2.69 times more risk of having a progeny developing a central nervous system (CNS) or non-CNS tumour [267]. For men between 35-39 years at conception the risk of having a child developing a CNS cancer was also higher (1.70 times).

Accordingly, Nybo Andersen and Urhoj found out in a 2 million pediatric subjects' study in Denmark that the risk to develop acute lymphoblastic leukaemia was 1.13 higher for every 5-year increase in paternal age [268]. Another study, in Peru, by Rios et al. discovered that fathers and mothers  $\geq 35$  years at conception have higher risks of having children with retinoblastoma, tumours in gonadal cells and non-Hodgkin lymphoma [269].

## **Epigenetics**

As the field of epigenetics emerges and becomes refined, the nature-nurture relation will help us to understand how environment shapes our genome in a unique manner and affects our everyday life. Epigenetics is defined as the study of all non-DNA altering changes that affect gene expression. Such phenotypical variations in genome expression result notably from DNA methylation at CpG dinucleotides, histone tail modification that can remodel chromatin in the nucleus and miRNAs that can interfere during protein expression by targeting the 3'UTR end of mRNAs [270]. These chemical changes of the DNA are called epigenetics marks and are heritable and maintained through generations. However, there is a reprogramming stage that happens just after the fertilization of the oocyte by a sperm when the paternal histones are replaced by protamines [271]. Right after this, a hyperacetylation and a hypomethylation of the paternal genome occur and allow specific epigenetic factors to distinguish it from the mother's one. An interesting fact reported by Perrin et al. highlighted that an epimutation in an imprinted gene causative for Prader-Willis syndrome lies close to 15q13-14, a region suspected to be associated with schizophrenia [272]. This demonstrates the importance of studying parental imprinting processes in schizophrenia and other neurodevelopmental disorders. These processes could definitively be linked to PAE. Katz-Jaffe et al. have reported that male mice that were older than 15 months had alterations in the expression of 4 genes : Ace-1, Prm1, Prm2, Smcp [273]. The most significant of these genes, Prm1 and Prm2 (for Protamine 1 and 2,

respectively) are genes coding for protamines, proteins made to replace histones in male chromatin. Additionally, *Smcp* is a key gene in sperm motility (61). Finally, *Ace-1* is important for acrosomal reaction after fertilization. *Ace-1* is inhibited when DNA methyltransferase is inhibited [274]. Taken together, these findings suggest how ageing sperm could have decreased reproduction parameters. If genes are switched off by methylation marks, their expression is also switched off and they are more likely to not being reprogrammed in germ cells because this specific reprogramming is for marks that are normally transmitted and regulated for example by DNA methyltransferases. If we change the methylation of one cytosine and it is not regulated it will probably result in a chromatin remodelling and consequently, alter genes expression. More studies will have to investigate the following question: Does paternal age alters epigenetic marks and epimutations as well as their heritability pattern?

# BIBLIOGRAPHIE

1. Aronson, S.J. and H.L. Rehm, *Building the foundation for genomics in precision medicine*. Nature, 2015. **526**(7573): p. 336-42.
2. Kuleshov, V., et al., *A machine-compiled database of genome-wide association studies*. Nat Commun, 2019. **10**(1): p. 3341.
3. Morin, A., et al., *Exploring rare and low-frequency variants in the Saguenay-Lac-Saint-Jean population identified genes associated with asthma and allergy traits*. Eur J Hum Genet, 2019. **27**(1): p. 90-101.
4. Tran, M.M., et al., *Predicting the atopic march: Results from the Canadian Healthy Infant Longitudinal Development Study*. J Allergy Clin Immunol, 2018. **141**(2): p. 601-607 e8.
5. Taft, R.J., et al., *Non-coding RNAs: regulators of disease*. J Pathol, 2010. **220**(2): p. 126-39.
6. El-Hattab, A.W., W.J. Craigen, and F. Scaglia, *Mitochondrial DNA maintenance defects*. Biochim Biophys Acta Mol Basis Dis, 2017. **1863**(6): p. 1539-1555.
7. H. Lodish, A.B., C. A. Kaiser, M. Krieger, A. Bretscher, H. Ploegh, A. Amon, K. C. Martin, *Molecular Cell Biology* W. Freeman, Editor. 2016: États-Unis. p. 306.
8. Blackwood, E.M. and J.T. Kadonaga, *Going the distance: a current view of enhancer action*. Science, 1998. **281**(5373): p. 60-3.
9. Gagniuc, P. and C. Ionescu-Tirgoviste, *Eukaryotic genomes may exhibit up to 10 generic classes of gene promoters*. BMC Genomics, 2012. **13**: p. 512.
10. Yang, M.Q. and L.L. Elnitski, *Diversity of core promoter elements comprising human bidirectional promoters*. BMC Genomics, 2008. **9 Suppl 2**: p. S3.
11. Pickering, B.M. and A.E. Willis, *The implications of structured 5' untranslated regions on translation and disease*. Semin Cell Dev Biol, 2005. **16**(1): p. 39-47.
12. Lynch, M., *The origins of eukaryotic gene structure*. Mol Biol Evol, 2006. **23**(2): p. 450-68.
13. Deutsch, M. and M. Long, *Intron-exon structures of eukaryotic model organisms*. Nucleic Acids Res, 1999. **27**(15): p. 3219-28.
14. Sakharkar, M.K., V.T. Chow, and P. Kanguane, *Distributions of exons and introns in the human genome*. In Silico Biol, 2004. **4**(4): p. 387-93.
15. Jo, B.S. and S.S. Choi, *Introns: The Functional Benefits of Introns in Genomes*. Genomics Inform, 2015. **13**(4): p. 112-8.
16. Auton, A., et al., *A global reference for human genetic variation*. Nature, 2015. **526**(7571): p. 68-74.
17. Subramanian, S., R.K. Mishra, and L. Singh, *Genome-wide analysis of microsatellite repeats in humans: their abundance and density in specific genomic regions*. Genome Biol, 2003. **4**(2): p. R13.
18. Toth, G., Z. Gaspari, and J. Jurka, *Microsatellites in different eukaryotic genomes: survey and analysis*. Genome Res, 2000. **10**(7): p. 967-81.
19. Brookes, K.J., *The VNTR in complex disorders: the forgotten polymorphisms? A functional way forward?* Genomics, 2013. **101**(5): p. 273-81.
20. Nakamura, Y., K. Koyama, and M. Matsushima, *VNTR (variable number of tandem repeat) sequences as transcriptional, translational, or functional regulators*. J Hum Genet, 1998. **43**(3): p. 149-52.
21. Bennett, P., *Demystified ... microsatellites*. Mol Pathol, 2000. **53**(4): p. 177-83.
22. Gray, I.C., D.A. Campbell, and N.K. Spurr, *Single nucleotide polymorphisms as tools in human genetics*. Hum Mol Genet, 2000. **9**(16): p. 2403-8.
23. Ingman, M., et al., *Mitochondrial genome variation and the origin of modern humans*. Nature, 2000. **408**(6813): p. 708-13.
24. Andersen, M.M. and D.J. Balding, *How many individuals share a mitochondrial genome?* PLoS Genet, 2018. **14**(11): p. e1007774.

25. Ingman, M. and U. Gyllensten, *mtDB: Human Mitochondrial Genome Database, a resource for population genetics and medical sciences*. Nucleic Acids Res, 2006. **34**(Database issue): p. D749-51.
26. Ikegawa, S., *A short history of the genome-wide association study: where we were and where we are going*. Genomics Inform, 2012. **10**(4): p. 220-5.
27. Visscher, P.M., et al., *10 Years of GWAS Discovery: Biology, Function, and Translation*. Am J Hum Genet, 2017. **101**(1): p. 5-22.
28. M. J. Khoury, S.R.B., M. Gwinn, J. P. T. Higgins, J. P. A. Loannidis et J. Little, *Human Genome Epidemiology: A Scientific Foundation for Using Genetic Information to Improve Health and Prevent Disease - Second Edition*. 2004, Royaume-Uni: Oxford University Press.
29. Caballero, A., A. Tenesa, and P.D. Keightley, *The Nature of Genetic Variation for Complex Traits Revealed by GWAS and Regional Heritability Mapping Analyses*. Genetics, 2015. **201**(4): p. 1601-13.
30. Visscher, P.M., et al., *Five years of GWAS discovery*. Am J Hum Genet, 2012. **90**(1): p. 7-24.
31. Milani, L., L. Leitsalu, and A. Metspalu, *An epidemiological perspective of personalized medicine: the Estonian experience*. J Intern Med, 2015. **277**(2): p. 188-200.
32. "Global strategy for asthma management and prevention: GINA executive summary." E.D. Bateman, S.S. Hurd, P.J. Barnes, J. Bousquet, J.M. Drazen, J.M. FitzGerald, P. Gibson, K. Ohta, P. O'Byrne, S.E. Pedersen, E. Pizzichini, S.D. Sullivan, S.E. Wenzel and H.J. Zar. Eur Respir J 2008; 31: 143-178. Eur Respir J, 2018. **51**(2).
33. National Heart, L., and Blood Institute, *Guide-lines for the Diagnosis and Management of Asthma (EPR-3)*. 2007.
34. Johansson, S.G., et al., *A revised nomenclature for allergy. An EAACI position statement from the EAACI nomenclature task force*. Allergy, 2001. **56**(9): p. 813-24.
35. A. D. Irvine, P.H.H., A. C. Yan Harper's *Textbook of Pediatric Dermatology*, in *Harper's Textbook of Pediatric Dermatology*. 2011, Wiley-Blackwell: United States. p. 22.1-22.13.
36. Smith, S.M. and S.T. Nedorost, "*Dermatitis*" defined. Dermatitis, 2010. **21**(5): p. 248-50.
37. Williams, H.C., et al., *The U.K. Working Party's Diagnostic Criteria for Atopic Dermatitis. III. Independent hospital validation*. Br J Dermatol, 1994. **131**(3): p. 406-16.
38. Tham, E.H. and D.Y. Leung, *Mechanisms by Which Atopic Dermatitis Predisposes to Food Allergy and the Atopic March*. Allergy Asthma Immunol Res, 2019. **11**(1): p. 4-15.
39. Hill, D.A. and J.M. Spergel, *The atopic march: Critical evidence and clinical relevance*. Ann Allergy Asthma Immunol, 2018. **120**(2): p. 131-137.
40. Belgrave, D.C., et al., *Developmental profiles of eczema, wheeze, and rhinitis: two population-based birth cohort studies*. PLoS Med, 2014. **11**(10): p. e1001748.
41. Hill, D.A., et al., *The epidemiologic characteristics of healthcare provider-diagnosed eczema, asthma, allergic rhinitis, and food allergy in children: a retrospective cohort study*. BMC Pediatr, 2016. **16**: p. 133.
42. Bartnikas, L.M., et al., *Epicutaneous sensitization results in IgE-dependent intestinal mast cell expansion and food-induced anaphylaxis*. J Allergy Clin Immunol, 2013. **131**(2): p. 451-60 e1-6.
43. Marenholz, I., et al., *Meta-analysis identifies seven susceptibility loci involved in the atopic march*. Nat Commun, 2015. **6**: p. 8804.
44. Spergel, J.M., *From atopic dermatitis to asthma: the atopic march*. Ann Allergy Asthma Immunol, 2010. **105**(2): p. 99-106; quiz 107-9, 117.
45. Egawa, G. and K. Kabashima, *Multifactorial skin barrier deficiency and atopic dermatitis: Essential topics to prevent the atopic march*. J Allergy Clin Immunol, 2016. **138**(2): p. 350-358 e1.
46. Miajlovic, H., et al., *Effect of filaggrin breakdown products on growth of and protein expression by Staphylococcus aureus*. J Allergy Clin Immunol, 2010. **126**(6): p. 1184-90 e3.
47. Schultz Larsen, F., *Atopic dermatitis: a genetic-epidemiologic study in a population-based twin sample*. J Am Acad Dermatol, 1993. **28**(5 Pt 1): p. 719-23.
48. Morar, N., et al., *The genetics of atopic dermatitis*. J Allergy Clin Immunol, 2006. **118**(1): p. 24-34; quiz 35-6.



49. Bussmann, C., S. Weidinger, and N. Novak, *Genetics of atopic dermatitis*. J Dtsch Dermatol Ges, 2011. **9**(9): p. 670-6.
50. Bin, L. and D.Y. Leung, *Genetic and epigenetic studies of atopic dermatitis*. Allergy Asthma Clin Immunol, 2016. **12**: p. 52.
51. Weidinger, S., et al., *Loss-of-function variations within the filaggrin gene predispose for atopic dermatitis with allergic sensitizations*. J Allergy Clin Immunol, 2006. **118**(1): p. 214-9.
52. Palmer, C.N., et al., *Common loss-of-function variants of the epidermal barrier protein filaggrin are a major predisposing factor for atopic dermatitis*. Nat Genet, 2006. **38**(4): p. 441-6.
53. Brown, S.J. and W.H. McLean, *One remarkable molecule: filaggrin*. J Invest Dermatol, 2012. **132**(3 Pt 2): p. 751-62.
54. Weidinger, S., et al., *Atopic dermatitis*. Nat Rev Dis Primers, 2018. **4**(1): p. 1.
55. Liang, Y., C. Chang, and Q. Lu, *The Genetics and Epigenetics of Atopic Dermatitis-Filaggrin and Other Polymorphisms*. Clin Rev Allergy Immunol, 2016. **51**(3): p. 315-328.
56. Fortugno, P., et al., *The 420K LEKTI variant alters LEKTI proteolytic activation and results in protease deregulation: implications for atopic dermatitis*. Hum Mol Genet, 2012. **21**(19): p. 4187-200.
57. Kumar, H., T. Kawai, and S. Akira, *Toll-like receptors and innate immunity*. Biochem Biophys Res Commun, 2009. **388**(4): p. 621-5.
58. Ahmad-Nejad, P., et al., *The toll-like receptor 2 R753Q polymorphism defines a subgroup of patients with atopic dermatitis having severe phenotype*. J Allergy Clin Immunol, 2004. **113**(3): p. 565-7.
59. Kaesler, S., et al., *Toll-like receptor 2 ligands promote chronic atopic dermatitis through IL-4-mediated suppression of IL-10*. J Allergy Clin Immunol, 2014. **134**(1): p. 92-9.
60. Belderbos, M.E., et al., *Low neonatal Toll-like receptor 4-mediated interleukin-10 production is associated with subsequent atopic dermatitis*. Clin Exp Allergy, 2012. **42**(1): p. 66-75.
61. Zhang, Y., et al., *Analysis of the Association of Polymorphisms rs5743708 in TLR2 and rs4986790 in TLR4 with Atopic Dermatitis Risk*. Immunol Invest, 2019. **48**(2): p. 169-180.
62. Moller-Larsen, S., et al., *Association analysis identifies TLR7 and TLR8 as novel risk genes in asthma and related disorders*. Thorax, 2008. **63**(12): p. 1064-9.
63. Novak, N., et al., *Putative association of a TLR9 promoter polymorphism with atopic eczema*. Allergy, 2007. **62**(7): p. 766-72.
64. He, J.Q., et al., *Genetic variants of the IL13 and IL4 genes and atopic diseases in at-risk children*. Genes Immun, 2003. **4**(5): p. 385-9.
65. Curie, A.A.J.V.e.A., *Interleukin*. 2019.
66. A. D. Irvine, P.H.H., A. C. Yan *Harper's Textbook of Pediatric Dermatology*, in *Harper's Textbook of Pediatric Dermatology*. 2011, Wiley-Blackwell: United States. p. 23.11.
67. Hoffjan, S. and J.T. Epplen, *The genetics of atopic dermatitis: recent findings and future options*. J Mol Med (Berl), 2005. **83**(9): p. 682-92.
68. Tamari, M. and T. Hirota, *Genome-wide association studies of atopic dermatitis*. J Dermatol, 2014. **41**(3): p. 213-20.
69. Esparza-Gordillo, J., et al., *A functional IL-6 receptor (IL6R) variant is a risk factor for persistent atopic dermatitis*. J Allergy Clin Immunol, 2013. **132**(2): p. 371-7.
70. Hirota, T., et al., *Genome-wide association study identifies eight new susceptibility loci for atopic dermatitis in the Japanese population*. Nat Genet, 2012. **44**(11): p. 1222-6.
71. Namkung, J.H., et al., *An association between IL-9 and IL-9 receptor gene polymorphisms and atopic dermatitis in a Korean population*. J Dermatol Sci, 2011. **62**(1): p. 16-21.
72. Sohn, M.H., et al., *Association of interleukin-10 gene promoter polymorphism in children with atopic dermatitis*. J Pediatr, 2007. **150**(1): p. 106-8.
73. Lan, C.C., et al., *Distinct SPINK5 and IL-31 polymorphisms are associated with atopic eczema and non-atopic hand dermatitis in Taiwanese nursing population*. Exp Dermatol, 2011. **20**(12): p. 975-9.
74. Pua, H.H. and K.M. Ansel, *MicroRNA regulation of allergic inflammation and asthma*. Curr Opin Immunol, 2015. **36**: p. 101-8.

75. Sonkoly, E., et al., *MiR-155 is overexpressed in patients with atopic dermatitis and modulates T-cell proliferative responses by targeting cytotoxic T lymphocyte-associated antigen 4*. J Allergy Clin Immunol, 2010. **126**(3): p. 581-9 e1-20.
76. Mu, Z., Liu, Xiaojing, Chang, Christopher, Zhang, Jianzhong, *Molecular Biology of Atopic Dermatitis*. Clinical Reviews in Allergy & Immunology, 2014. **47**: p. 193-218.
77. Ma, L., V.B. Bajic, and Z. Zhang, *On the classification of long non-coding RNAs*. RNA Biol, 2013. **10**(6): p. 925-33.
78. Wang, X., et al., *Integrative Analysis of lncRNAs, miRNAs, and mRNA-Associated ceRNA Network in an Atopic Dermatitis Recurrence Model*. Int J Mol Sci, 2018. **19**(10).
79. Paternoster, L., et al., *Multi-ancestry genome-wide association study of 21,000 cases and 95,000 controls identifies new risk loci for atopic dermatitis*. Nat Genet, 2015. **47**(12): p. 1449-1456.
80. Weidinger, S. and N. Novak, *Atopic dermatitis*. Lancet, 2016. **387**(10023): p. 1109-1122.
81. Thomsen, S.F., et al., *Estimates of asthma heritability in a large twin sample*. Clin Exp Allergy, 2010. **40**(7): p. 1054-61.
82. Thomsen, S.F., *Genetics of asthma: an introduction for the clinician*. Eur Clin Respir J, 2015. **2**.
83. Moffatt, M.F., et al., *A large-scale, consortium-based genomewide association study of asthma*. N Engl J Med, 2010. **363**(13): p. 1211-1221.
84. Yan, Q., et al., *A meta-analysis of genome-wide association studies of asthma in Puerto Ricans*. Eur Respir J, 2017. **49**(5).
85. Hirota, T., et al., *Genome-wide association study identifies three new susceptibility loci for adult asthma in the Japanese population*. Nat Genet, 2011. **43**(9): p. 893-6.
86. Ferreira, M.A., et al., *Genome-wide association analysis identifies 11 risk variants associated with the asthma with hay fever phenotype*. J Allergy Clin Immunol, 2014. **133**(6): p. 1564-71.
87. Vercelli, D., *Discovering susceptibility genes for asthma and allergy*. Nat Rev Immunol, 2008. **8**(3): p. 169-82.
88. Kumar, A. and B. Ghosh, *Genetics of asthma: a molecular biologist perspective*. Clin Mol Allergy, 2009. **7**: p. 7.
89. Ober, C., et al., *Genome-wide search for asthma susceptibility loci in a founder population. The Collaborative Study on the Genetics of Asthma*. Hum Mol Genet, 1998. **7**(9): p. 1393-8.
90. Nicolaides, N.C., et al., *Interleukin 9: a candidate gene for asthma*. Proc Natl Acad Sci U S A, 1997. **94**(24): p. 13175-80.
91. Noguchi, E., et al., *Association of asthma and the interleukin-4 promoter gene in Japanese*. Clin Exp Allergy, 1998. **28**(4): p. 449-53.
92. Heinzmann, A., et al., *Genetic variants of IL-13 signalling and human asthma and atopy*. Hum Mol Genet, 2000. **9**(4): p. 549-59.
93. Humbert, M., et al., *Relationship between IL-4 and IL-5 mRNA expression and disease severity in atopic asthma*. Am J Respir Crit Care Med, 1997. **156**(3 Pt 1): p. 704-8.
94. Drazen, J.M., J.P. Arm, and K.F. Austen, *Sorting out the cytokines of asthma*. J Exp Med, 1996. **183**(1): p. 1-5.
95. T., B.P.A.e.H.S., *Asthma Genetics*. Allergo 2002(11): p. 246-252.
96. Slager, R.E., et al., *Genetics of asthma susceptibility and severity*. Clin Chest Med, 2012. **33**(3): p. 431-43.
97. Movahedi, M., et al., *Association of HLA class II alleles with childhood asthma and Total IgE levels*. Iran J Allergy Asthma Immunol, 2008. **7**(4): p. 215-20.
98. Nicolae, D., et al., *Fine mapping and positional candidate studies identify HLA-G as an asthma susceptibility gene on chromosome 6p21*. Am J Hum Genet, 2005. **76**(2): p. 349-57.
99. Ober, C., *HLA-G: an asthma gene on chromosome 6p*. Immunol Allergy Clin North Am, 2005. **25**(4): p. 669-79.
100. Lasky-Su, J., et al., *HLA-DQ strikes again: genome-wide association study further confirms HLA-DQ in the diagnosis of asthma among adults*. Clin Exp Allergy, 2012. **42**(12): p. 1724-33.

101. Kontakioti, E., et al., *HLA and asthma phenotypes/endotypes: a review*. Hum Immunol, 2014. **75**(8): p. 930-9.
102. Saba, N., et al., *Single nucleotide polymorphisms in asthma candidate genes TBXA2R, ADAM33 FCER1B and ORMDL3 in Pakistani asthmatics a case control study*. Asthma Res Pract, 2018. **4**: p. 4.
103. Rava, M., et al., *Serum levels of Clara cell secretory protein, asthma, and lung function in the adult general population*. J Allergy Clin Immunol, 2013. **132**(1): p. 230-2.
104. Hao, K., et al., *Lung eQTLs to help reveal the molecular underpinnings of asthma*. PLoS Genet, 2012. **8**(11): p. e1003029.
105. Miller, M. and D.H. Broide, *Why Is ORMDL3 on Chromosome 17q21 Highly Linked to Asthma?* Am J Respir Crit Care Med, 2019. **199**(4): p. 404-406.
106. Das, S., M. Miller, and D.H. Broide, *Chromosome 17q21 Genes ORMDL3 and GSDMB in Asthma and Immune Diseases*. Adv Immunol, 2017. **135**: p. 1-52.
107. Sleiman, P.M., et al., *ORMDL3 variants associated with asthma susceptibility in North Americans of European ancestry*. J Allergy Clin Immunol, 2008. **122**(6): p. 1225-7.
108. Galanter, J., et al., *ORMDL3 gene is associated with asthma in three ethnically diverse populations*. Am J Respir Crit Care Med, 2008. **177**(11): p. 1194-200.
109. Zhao, C.N., et al., *The Association of GSDMB and ORMDL3 Gene Polymorphisms With Asthma: A Meta-Analysis*. Allergy Asthma Immunol Res, 2015. **7**(2): p. 175-85.
110. Li, H.F., et al., *Association between ADAM33 polymorphisms and asthma risk: a systematic review and meta-analysis*. Respir Res, 2019. **20**(1): p. 38.
111. Barbeau, M. and H.L. Bpharm, *Burden of Atopic dermatitis in Canada*. Int J Dermatol, 2006. **45**(1): p. 31-6.
112. Williams, H., et al., *Worldwide variations in the prevalence of symptoms of atopic eczema in the International Study of Asthma and Allergies in Childhood*. J Allergy Clin Immunol, 1999. **103**(1 Pt 1): p. 125-38.
113. Garg, N. and J.I. Silverberg, *Epidemiology of childhood atopic dermatitis*. Clin Dermatol, 2015. **33**(3): p. 281-8.
114. L. Thériault, G.H., D. Goldfarb, C. Stoneridge et F. Bounajm *Cost Risk Analysis for Chronic Lung Disease in Canada*. 2012. 62.
115. Lai, C.K., et al., *Global variation in the prevalence and severity of asthma symptoms: phase three of the International Study of Asthma and Allergies in Childhood (ISAAC)*. Thorax, 2009. **64**(6): p. 476-83.
116. Québec, M.d.l.S.e.d.S.s.d., *Statistiques de santé et de bien être selon le sexe - Tout le Québec*. 2013-2014.
117. Lundback, B., et al., *Is asthma prevalence still increasing?* Expert Rev Respir Med, 2016. **10**(1): p. 39-51.
118. Beasley, R., et al., *Prevalence and etiology of asthma*. J Allergy Clin Immunol, 2000. **105**(2 Pt 2): p. S466-72.
119. Woolcock, A.J. and J.K. Peat, *Evidence for the increase in asthma worldwide*. Ciba Found Symp, 1997. **206**: p. 122-34; discussion 134-9, 157-9.
120. von Hertzen, L. and T. Haahtela, *Signs of reversing trends in prevalence of asthma*. Allergy, 2005. **60**(3): p. 283-92.
121. Sembajwe, G., et al., *National income, self-reported wheezing and asthma diagnosis from the World Health Survey*. Eur Respir J, 2010. **35**(2): p. 279-86.
122. Manolio, T.A., *Cohort studies and the genetics of complex disease*. Nat Genet, 2009. **41**(1): p. 5-6.
123. Manolio, T.A., J.E. Bailey-Wilson, and F.S. Collins, *Genes, environment and the value of prospective cohort studies*. Nat Rev Genet, 2006. **7**(10): p. 812-20.
124. H. Charbonneau, B.D., J. Légaré et H. Denis, *A Population History of North America* C.U. Press, Editor. 2000, Michael H., Richard S.: Cambridge, UK. p. 99-142.

125. H. Vézina, M.T., B. Desjardins et L. Houde, *Origines et contributions génétiques des fondatrices et des fondateurs de la population québécoise*. Cahiers québécois de démographie, 2005. **32**(2): p. 235-258.
126. Wilhelmy, J.-P., *Les mercenaires allemands au Québec, 1776-1783*. Revue d'Histoire de l'Amérique française, 1998. **51**(3): p. 327-482.
127. Houdaille, J., *L'évolution démographique du Québec depuis 1710*. Population - Édition française, 1975(30-1): p. 253-256.
128. Bouchard, G., *Population Studies and Genetic Epidemiology in Northeast Québec*. Canadian Studies in Population 1989. **16**(1): p. 61-86.
129. Desjardins, B., *La contribution différentielle des immigrants français à la souche canadienne-française*. Annales de Normandie, 2008. **3-4**(58-3-4): p. 69-79.
130. Hamel, M., *De Charlevoix au Saguenay : caractéristiques des familles émigrantes au XIXe siècle*. Revue d'Histoire de l'Amérique française, 1993. **47**(1): p. 5-25.
131. Scriver, C.R., *Human genetics: lessons from Quebec populations*. Annu Rev Genomics Hum Genet, 2001. **2**: p. 69-101.
132. R. Jetté, D.G., M. Guérin, J. Morissette et M. Hamel, *Histoire d'un génôme*, G.B.e.M.D. Braekeleer, Editor. 1990, Presses de l'Université du Québec. p. 77-148.
133. Lee, N., et al., *A genomewide linkage-disequilibrium scan localizes the Saguenay-Lac-Saint-Jean cytochrome oxidase deficiency to 2p16*. Am J Hum Genet, 2001. **68**(2): p. 397-409.
134. De Braekeleer, M., et al., *Complete identification of cystic fibrosis transmembrane conductance regulator mutations in the CF population of Saguenay Lac-Saint-Jean (Quebec, Canada)*. Clin Genet, 1998. **53**(1): p. 44-6.
135. Grompe, M., et al., *A single mutation of the fumarylacetoacetate hydrolase gene in French Canadians with hereditary tyrosinemia type I*. N Engl J Med, 1994. **331**(6): p. 353-7.
136. Laberge, A.M., et al., *Population history and its impact on medical genetics in Quebec*. Clin Genet, 2005. **68**(4): p. 287-301.
137. Laprise, C., *The Saguenay-Lac-Saint-Jean asthma familial collection: the genetics of asthma in a young founder population*. Genes Immun, 2014. **15**(4): p. 247-55.
138. Mukaneza, Y., et al., *mTORC1 is required for expression of LRPPRC and cytochrome-c oxidase but not HIF-1alpha in Leigh syndrome French Canadian type patient fibroblasts*. Am J Physiol Cell Physiol, 2019. **317**(1): p. C58-C67.
139. Morin, A., et al., *Combining omics data to identify genes associated with allergic rhinitis*. Clin Epigenetics, 2017. **9**: p. 3.
140. Berube, J.C., et al., *Identification of Susceptibility Genes of Adult Asthma in French Canadian Women*. Can Respir J, 2016. **2016**: p. 3564341.
141. Al Tuwaijri, A., et al., *Local genotype influences DNA methylation at two asthma-associated regions, 5q31 and 17q21, in a founder effect population*. J Med Genet, 2016. **53**(4): p. 232-41.
142. Gauvin, H., et al., *Genome-wide patterns of identity-by-descent sharing in the French Canadian founder population*. Eur J Hum Genet, 2014. **22**(6): p. 814-21.
143. Madore, A.M., et al., *Distribution of CFTR mutations in Saguenay- Lac-Saint-Jean: proposal of a panel of mutations for population screening*. Genet Med, 2008. **10**(3): p. 201-6.
144. R. L. Gentilcore, D.M., R. H. Walder, B. Moldofsky et G. J. Matthews, *Historical Atlas of Canada : Volume II: The Land Transformed, 1800-1891*, D.M.e.R.H. Walder, Editor. 1993, University of Toronto Press. p. 43-78.
145. McInnis, M., *A Population History of North America*, M.R.H.e.R.H. Steckel, Editor. 2000, Cambridge University Press. p. 530-559.
146. Canada, S. *Immigration and Ethnocultural Diversity in Canada*. 2011; Available from: <https://www12.statcan.gc.ca/nhs-enm/2011/as-sa/99-010-x/99-010-x2011001-eng.cfm>.
147. Subbarao, P., et al., *The Canadian Healthy Infant Longitudinal Development (CHILD) Study: examining developmental origins of allergy and asthma*. Thorax, 2015. **70**(10): p. 998-1000.

148. Takaro, T.K., et al., *The Canadian Healthy Infant Longitudinal Development (CHILD) birth cohort study: assessment of environmental exposures*. J Expo Sci Environ Epidemiol, 2015. **25**(6): p. 580-92.
149. Dudbridge, F., *Power and predictive accuracy of polygenic risk scores*. PLoS Genet, 2013. **9**(3): p. e1003348.
150. SA Lambert, G.A.e.I.M., *Towards clinical utility of polygenic risk scores*. Human Molecular Genetics, 2019(ddz187): p. 1-10.
151. Vilhjalmsón, B.J., et al., *Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk Scores*. Am J Hum Genet, 2015. **97**(4): p. 576-92.
152. Do, C.B., et al., *Comparison of family history and SNPs for predicting risk of complex disease*. PLoS Genet, 2012. **8**(10): p. e1002973.
153. Martin, A.R., et al., *Clinical use of current polygenic risk scores may exacerbate health disparities*. Nat Genet, 2019. **51**(4): p. 584-591.
154. Duncan, L., et al., *Analysis of polygenic risk score usage and performance in diverse human populations*. Nat Commun, 2019. **10**(1): p. 3328.
155. Marquez-Luna, C., P.R. Loh, and A.L. Price, *Multiethnic polygenic risk scores improve risk prediction in diverse populations*. Genet Epidemiol, 2017. **41**(8): p. 811-823.
156. Abuabara, K., et al., *Genetic ancestry does not explain increased atopic dermatitis susceptibility or worse disease control among African American subjects in 2 large US cohorts*. J Allergy Clin Immunol, 2019.
157. Udler, M.S., et al., *Genetic Risk Scores for Diabetes Diagnosis and Precision Medicine*. Endocr Rev, 2019. **40**(6): p. 1500-1520.
158. Axelrud, L.K., et al., *Polygenic Risk Score for Alzheimer's Disease: Implications for Memory Performance and Hippocampal Volumes in Early Life*. Am J Psychiatry, 2018. **175**(6): p. 555-563.
159. Escott-Price, V., et al., *Polygenic risk score analysis of pathologically confirmed Alzheimer disease*. Ann Neurol, 2017. **82**(2): p. 311-314.
160. Shieh, Y., et al., *Breast cancer risk prediction using a clinical risk model and polygenic risk score*. Breast Cancer Res Treat, 2016. **159**(3): p. 513-25.
161. Mavaddat, N., et al., *Prediction of breast cancer risk based on profiling with common genetic variants*. J Natl Cancer Inst, 2015. **107**(5).
162. Kumar, R. and A. Indrayan, *Receiver operating characteristic (ROC) curve for medical researchers*. Indian Pediatr, 2011. **48**(4): p. 277-87.
163. Hajian-Tilaki, K., *Receiver Operating Characteristic (ROC) Curve Analysis for Medical Diagnostic Test Evaluation*. Caspian J Intern Med, 2013. **4**(2): p. 627-35.
164. Wray, N.R., et al., *The genetic interpretation of area under the ROC curve in genomic profiling*. PLoS Genet, 2010. **6**(2): p. e1000864.
165. Metz, C.E., *Basic principles of ROC analysis*. Semin Nucl Med, 1978. **8**(4): p. 283-98.
166. Wolf, H.B.e.C., *The SAGE Handbook of Regression Analysis and Causal Inference*. 2014. p. 160-164.
167. Kerr, K.F., et al., *Net reclassification indices for evaluating risk prediction instruments: a critical review*. Epidemiology, 2014. **25**(1): p. 114-21.
168. Carlsten, C., et al., *Atopic dermatitis in a high-risk cohort: natural history, associated allergic outcomes, and risk factors*. Ann Allergy Asthma Immunol, 2013. **110**(1): p. 24-8.
169. Du Toit, G., et al., *Effect of Avoidance on Peanut Allergy after Early Peanut Consumption*. N Engl J Med, 2016. **374**(15): p. 1435-43.
170. Horimukai, K., et al., *Application of moisturizer to neonates prevents development of atopic dermatitis*. J Allergy Clin Immunol, 2014. **134**(4): p. 824-830 e6.
171. Natsume, O., et al., *Two-step egg introduction for prevention of egg allergy in high-risk infants with eczema (PETIT): a randomised, double-blind, placebo-controlled trial*. Lancet, 2017. **389**(10066): p. 276-286.

172. Simpson, E.L., et al., *Emollient enhancement of the skin barrier from birth offers effective atopic dermatitis prevention*. J Allergy Clin Immunol, 2014. **134**(4): p. 818-23.
173. Dizier, M.H., F. Demenais, and F. Mathieu, *Gain of power of the general regression model compared to Cochran-Armitage Trend tests: simulation study and application to bipolar disorder*. BMC Genet, 2017. **18**(1): p. 24.
174. Carter, J.V., et al., *ROC-ing along: Evaluation and interpretation of receiver operating characteristic curves*. Surgery, 2016. **159**(6): p. 1638-1645.
175. Thornton, T. and M.S. McPeck, *Case-control association testing with related individuals: a more powerful quasi-likelihood score test*. Am J Hum Genet, 2007. **81**(2): p. 321-37.
176. Do, C.B., et al., *Comparison of family history and SNPs for predicting risk of complex disease*. PLoS Genet. Vol. 8. 2012, United States. e1002973.
177. Bigdeli, T.B., et al., *Genome-wide association study reveals greater polygenic loading for schizophrenia in cases with a family history of illness*. Am J Med Genet B Neuropsychiatr Genet, 2016. **171B**(2): p. 276-89.
178. Bush, W.S., et al., *Evidence for polygenic susceptibility to multiple sclerosis--the shape of things to come*. Am J Hum Genet, 2010. **86**(4): p. 621-5.
179. Markota, M., et al., *Association of schizophrenia polygenic risk score with manic and depressive psychosis in bipolar disorder*. Transl Psychiatry, 2018. **8**(1): p. 188.
180. Baurecht, H., et al., *Genome-wide comparative analysis of atopic dermatitis and psoriasis gives insight into opposing genetic mechanisms*. Am J Hum Genet, 2015. **96**(1): p. 104-20.
181. Huls, A., et al., *Atopic dermatitis: Interaction between genetic variants of GSTP1, TNF, TLR2, and TLR4 and air pollution in early life*. Pediatr Allergy Immunol, 2018. **29**(6): p. 596-605.
182. Pedersen, E.S.L., et al., *The Simple 10-Item Predicting Asthma Risk in Children Tool to Predict Childhood Asthma-An External Validation*. J Allergy Clin Immunol Pract, 2019. **7**(3): p. 943-953 e4.
183. Thompson, E.A., *Identity by descent: variation in meiosis, across genomes, and in populations*. Genetics, 2013. **194**(2): p. 301-26.
184. Browning, S.R. and B.L. Browning, *High-resolution detection of identity by descent in unrelated individuals*. Am J Hum Genet, 2010. **86**(4): p. 526-39.
185. Thomas, A., M.H. Skolnick, and C.M. Lewis, *Genomic mismatch scanning in pedigrees*. IMA J Math Appl Med Biol, 1994. **11**(1): p. 1-16.
186. Buffalo, V., S.M. Mount, and G. Coop, *A Genealogical Look at Shared Ancestry on the X Chromosome*. Genetics, 2016. **204**(1): p. 57-75.
187. Turet, M. and F. Hospital, *Blocks of chromosomes identical by descent in a population: Models and predictions*. PLoS One, 2017. **12**(11): p. e0187416.
188. Palamara, P.F., et al., *Length distributions of identity by descent reveal fine-scale demographic history*. Am J Hum Genet, 2012. **91**(5): p. 809-22.
189. Dizier, M.H., et al., *Indication of linkage and genetic heterogeneity of asthma according to age at onset on chromosome 7q in 107 French EGEA families*. Eur J Hum Genet, 2001. **9**(11): p. 867-72.
190. Yunusbaev, U., et al., *Reconstructing recent population history while mapping rare variants using haplotypes*. Sci Rep, 2019. **9**(1): p. 5849.
191. Brasch-Andersen, C., et al., *Highly significant linkage to chromosome 3q13.31 for rhinitis and related allergic diseases*. J Med Genet, 2006. **43**(3): p. e10.
192. Bateman, E.D., et al., *Global strategy for asthma management and prevention: GINA executive summary*. Eur Respir J, 2008. **31**(1): p. 143-78.
193. Holgate, S.T., et al., *Asthma*. Nat Rev Dis Primers, 2015. **1**: p. 15025.
194. Braman, S.S., *The global burden of asthma*. Chest, 2006. **130**(1 Suppl): p. 4S-12S.
195. Gabriel, S.B., et al., *The structure of haplotype blocks in the human genome*. Science, 2002. **296**(5576): p. 2225-9.
196. Cardon, L.R. and G.R. Abecasis, *Using haplotype blocks to map human complex trait loci*. Trends Genet, 2003. **19**(3): p. 135-40.

197. Spencer, C.C., et al., *The influence of recombination on human genetic diversity*. PLoS Genet, 2006. **2**(9): p. e148.
198. Liu, N., et al., *Haplotype block structures show significant variation among populations*. Genet Epidemiol, 2004. **27**(4): p. 385-400.
199. Silverman, E.K., *Haplotype thinking in lung disease*. Proc Am Thorac Soc, 2007. **4**(1): p. 4-8.
200. Demenais, F., et al., *Multiancestry association study identifies new asthma risk loci that colocalize with immune-cell enhancer marks*. Nat Genet, 2018. **50**(1): p. 42-53.
201. Spielman, R.S., R.E. McGinnis, and W.J. Ewens, *Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM)*. Am J Hum Genet, 1993. **52**(3): p. 506-16.
202. Yi, N., et al., *Multiple comparisons in genetic association studies: a hierarchical modeling approach*. Stat Appl Genet Mol Biol, 2014. **13**(1): p. 35-48.
203. Mellerup, E. and G.L. Moller, *Combinations of Genetic Variants Occurring Exclusively in Patients*. Comput Struct Biotechnol J, 2017. **15**: p. 286-289.
204. Yao, Y., et al., *Association of HLA-DRB1 Gene Polymorphism with Risk of Asthma: A Meta-Analysis*. Med Sci Monit Basic Res, 2016. **22**: p. 80-6.
205. Mishra, M.N., P. Dudeja, and R.K. Gupta, *Association of HLA-Class II and IgE serum levels in pediatric asthma*. Iran J Immunol, 2014. **11**(1): p. 21-8.
206. Hanchard, N.A., et al., *An assessment of the association between childhood asthma and HLA DRB1\*03 using extended haplotype analysis*. Tissue Antigens, 2010. **76**(6): p. 491-4.
207. Arnheim, N., P. Calabrese, and M. Nordborg, *Hot and cold spots of recombination in the human genome: the reason we should find them and how this can be achieved*. Am J Hum Genet, 2003. **73**(1): p. 5-16.
208. Wen, S.H. and M.Y. Tsai, *Haplotype association analysis of combining unrelated case-control and triads with consideration of population stratification*. Front Genet, 2014. **5**: p. 103.
209. Fullerton, J.M. and J.I. Nurnberger, *Polygenic risk scores in psychiatry: Will they be useful for clinicians?* F1000Res, 2019. **8**.
210. Potaczek, D.P., et al., *Epigenetics and allergy: from basic mechanisms to clinical applications*. Epigenomics, 2017. **9**(4): p. 539-571.
211. Janssens, A. and M.J. Joyner, *Polygenic Risk Scores That Predict Common Diseases Using Millions of Single Nucleotide Polymorphisms: Is More, Better?* Clin Chem, 2019. **65**(5): p. 609-611.
212. Wang, S., et al., *Genetic variants demonstrating flip-flop phenomenon and breast cancer risk prediction among women of African ancestry*. Breast Cancer Res Treat, 2018. **168**(3): p. 703-712.
213. Simard, M., C. Laprise, and S.L. Girard, *Impact of Paternal Age at Conception on Human Health*. Clin Chem, 2019. **65**(1): p. 146-152.
214. Tobi, E.W., et al., *DNA methylation signatures link prenatal famine exposure to growth and metabolism*. Nat Commun, 2014. **5**: p. 5592.
215. Allen, E.G., et al., *Maternal age and risk for trisomy 21 assessed by the origin of chromosome nondisjunction: a report from the Atlanta and National Down Syndrome Projects*. Hum Genet, 2009. **125**(1): p. 41-52.
216. Crow, J.F., *The origins, patterns and implications of human spontaneous mutation*. Nat Rev Genet, 2000. **1**(1): p. 40-7.
217. Lorentz, C.P., et al., *Primer on medical genomics part I: History of genetics and sequencing of the human genome*. Mayo Clin Proc, 2002. **77**(8): p. 773-82.
218. Krooth, R.S., *Comments on the estimation of the mutation rate for achondroplasia*. Am J Hum Genet, 1953. **5**(4): p. 373-6.
219. Bray, I., D. Gunnell, and G. Davey Smith, *Advanced paternal age: how old is too old?* J Epidemiol Community Health, 2006. **60**(10): p. 851-3.
220. Gunes, S., et al., *Effects of aging on the male reproductive system*. J Assist Reprod Genet, 2016. **33**(4): p. 441-54.

221. Mahmoud, A.M., et al., *Testicular volume in relation to hormonal indices of gonadal function in community-dwelling elderly men*. J Clin Endocrinol Metab, 2003. **88**(1): p. 179-84.
222. Midzak, A.S., et al., *Leydig cell aging and the mechanisms of reduced testosterone synthesis*. Mol Cell Endocrinol, 2009. **299**(1): p. 23-31.
223. Levitas, E., et al., *Relationship between age and semen parameters in men with normal sperm concentration: analysis of 6022 semen samples*. Andrologia, 2007. **39**(2): p. 45-50.
224. Kumar, N., A.K. Singh, and A.R. Choudhari, *Impact of age on semen parameters in male partners of infertile couples in a rural tertiary care center of central India: A cross-sectional study*. Int J Reprod Biomed (Yazd), 2017. **15**(8): p. 497-502.
225. Cheng, C.Y. and D.D. Mruk, *The biology of spermatogenesis: the past, present and future*. Philos Trans R Soc Lond B Biol Sci, 2010. **365**(1546): p. 1459-63.
226. Goriely, A. and A.O. Wilkie, *Missing heritability: paternal age effect mutations and selfish spermatogonia*. Nat Rev Genet, 2010. **11**(8): p. 589.
227. Kong, A., et al., *Rate of de novo mutations and the importance of father's age to disease risk*. Nature, 2012. **488**(7412): p. 471-5.
228. Gilissen, C., et al., *Genome sequencing identifies major causes of severe intellectual disability*. Nature, 2014. **511**(7509): p. 344-7.
229. Girard, S.L., et al., *Paternal Age Explains a Major Portion of De Novo Germline Mutation Rate Variability in Healthy Individuals*. PLoS One, 2016. **11**(10): p. e0164212.
230. Kloosterman, W.P., et al., *Characteristics of de novo structural changes in the human genome*. Genome Res, 2015. **25**(6): p. 792-801.
231. Acuna-Hidalgo, R., et al., *Post-zygotic Point Mutations Are an Underrecognized Source of De Novo Genomic Variation*. Am J Hum Genet, 2015. **97**(1): p. 67-74.
232. Goriely, A. and A.O. Wilkie, *Paternal age effect mutations and selfish spermatogonial selection: causes and consequences for human disease*. Am J Hum Genet, 2012. **90**(2): p. 175-200.
233. Blackburn, E.H., *Switching and signaling at the telomere*. Cell, 2001. **106**(6): p. 661-73.
234. Tsuji, A., et al., *Estimating age of humans based on telomere shortening*. Forensic Sci Int, 2002. **126**(3): p. 197-9.
235. Cronkhite, J.T., et al., *Telomere shortening in familial and sporadic pulmonary fibrosis*. Am J Respir Crit Care Med, 2008. **178**(7): p. 729-37.
236. Harley, C.B., A.B. Futcher, and C.W. Greider, *Telomeres shorten during ageing of human fibroblasts*. Nature, 1990. **345**(6274): p. 458-60.
237. Sharma, R., et al., *Effects of increased paternal age on sperm quality, reproductive outcome and associated epigenetic risks to offspring*. Reprod Biol Endocrinol, 2015. **13**: p. 35.
238. Wang, Y. and J. Feigon, *Structural biology of telomerase and its interaction at telomeres*. Curr Opin Struct Biol, 2017. **47**: p. 77-87.
239. Jafri, M.A., et al., *Roles of telomeres and telomerase in cancer, and advances in telomerase-targeted therapies*. Genome Med, 2016. **8**(1): p. 69.
240. Eisenberg, D.T., M.G. Hayes, and C.W. Kuzawa, *Delayed paternal age of reproduction in humans is associated with longer telomeres across two generations of descendants*. Proc Natl Acad Sci U S A, 2012. **109**(26): p. 10251-6.
241. Unryn, B.M., L.S. Cook, and K.T. Riabowol, *Paternal age is positively linked to telomere length of children*. Aging Cell, 2005. **4**(2): p. 97-101.
242. De Meyer, T., et al., *Paternal age at birth is an important determinant of offspring telomere length*. Hum Mol Genet, 2007. **16**(24): p. 3097-102.
243. Wiener-Megnazi, Z., R. Auslender, and M. Dirnfeld, *Advanced paternal age and reproductive outcome*. Asian J Androl, 2012. **14**(1): p. 69-76.
244. Hjelmborg, J.B., et al., *Paternal age and telomere length in twins: the germ stem cell selection paradigm*. Aging Cell, 2015. **14**(4): p. 701-3.
245. Amann, R.P. and S.S. Howards, *Daily spermatozoal production and epididymal spermatozoal reserves of the human male*. J Urol, 1980. **124**(2): p. 211-5.



246. Nordfjall, K., et al., *Large-scale parent-child comparison confirms a strong paternal influence on telomere length*. Eur J Hum Genet, 2010. **18**(3): p. 385-9.
247. Ozturk, S., *Telomerase activity and telomere length in male germ cells*. Biol Reprod, 2015. **92**(2): p. 53.
248. Eisenberg, D.T. and C.W. Kuzawa, *Commentary: The evolutionary biology of the paternal age effect on telomere length*. Int J Epidemiol, 2013. **42**(2): p. 462-5.
249. Fusco, G. and A. Minelli, *Phenotypic plasticity in development and evolution: facts and concepts. Introduction*. Philos Trans R Soc Lond B Biol Sci, 2010. **365**(1540): p. 547-56.
250. Templado, C., et al., *Advanced age increases chromosome structural abnormalities in human spermatozoa*. Eur J Hum Genet, 2011. **19**(2): p. 145-51.
251. Fonseka, K.G. and D.K. Griffin, *Is there a paternal age effect for aneuploidy?* Cytogenet Genome Res, 2011. **133**(2-4): p. 280-91.
252. Hassold, T.J., et al., *Molecular studies of non-disjunction in trisomy 16*. J Med Genet, 1991. **28**(3): p. 159-62.
253. Lowe, X., et al., *Frequency of XY sperm increases with age in fathers of boys with Klinefelter syndrome*. Am J Hum Genet, 2001. **69**(5): p. 1046-54.
254. Lorda-Sanchez, I., et al., *Reduced recombination and paternal age effect in Klinefelter syndrome*. Hum Genet, 1992. **89**(5): p. 524-30.
255. Jacobs, P.A., et al., *Klinefelter's syndrome: an analysis of the origin of the additional sex chromosome using molecular probes*. Ann Hum Genet, 1988. **52**(2): p. 93-109.
256. Buwe, A., M. Guttenbach, and M. Schmid, *Effect of paternal age on the frequency of cytogenetic abnormalities in human spermatozoa*. Cytogenet Genome Res, 2005. **111**(3-4): p. 213-28.
257. Newschaffer, C.J., et al., *The epidemiology of autism spectrum disorders*. Annu Rev Public Health, 2007. **28**: p. 235-58.
258. Janecka, M., et al., *Advanced paternal age effects in neurodevelopmental disorders-review of potential underlying mechanisms*. Transl Psychiatry, 2017. **7**(1): p. e1019.
259. Gabis, L., R. Raz, and Y. Kesner-Baruch, *Paternal age in autism spectrum disorders and ADHD*. Pediatr Neurol, 2010. **43**(4): p. 300-2.
260. Frans, E.M., et al., *Autism risk across generations: a population-based study of advancing grandpaternal and paternal age*. JAMA Psychiatry, 2013. **70**(5): p. 516-21.
261. Saha, S., et al., *Incidence of schizophrenia does not vary with economic status of the country: evidence from a systematic review*. Soc Psychiatry Psychiatr Epidemiol, 2006. **41**(5): p. 338-40.
262. Malaspina, D., C. Gilman, and T.M. Kranz, *Paternal age and mental health of offspring*. Fertil Steril, 2015. **103**(6): p. 1392-6.
263. Sipos, A., et al., *Paternal age and schizophrenia: a population based cohort study*. BMJ, 2004. **329**(7474): p. 1070.
264. Jenkins, T.G., et al., *Age-associated sperm DNA methylation alterations: possible implications in offspring disease susceptibility*. PLoS Genet, 2014. **10**(7): p. e1004458.
265. Liu, L.L., et al., *Lack of a genetic association between the TNXB locus and schizophrenia in a Chinese population*. Neurosci Lett, 2004. **355**(1-2): p. 149-51.
266. Wang, J., et al., *Further study of genetic association between the TNXB locus and schizophrenia*. Psychiatr Genet, 2011. **21**(4): p. 216.
267. Yip, B.H., Y. Pawitan, and K. Czene, *Parental age and risk of childhood cancers: a population-based cohort study from Sweden*. Int J Epidemiol, 2006. **35**(6): p. 1495-503.
268. Nybo Andersen, A.M. and S.K. Urhoj, *Is advanced paternal age a health risk for the offspring?* Fertil Steril, 2017. **107**(2): p. 312-318.
269. Rios, L., et al., *Advancing Parental Age and Risk of Solid Tumors in Children: A Case-Control Study in Peru*. J Oncol, 2018. **2018**: p. 3924635.
270. Ferreira, H.J. and M. Esteller, *Non-coding RNAs, epigenetics, and cancer: tying it all together*. Cancer Metastasis Rev, 2018. **37**(1): p. 55-73.

271. Cantone, I. and A.G. Fisher, *Epigenetic programming and reprogramming during development*. Nat Struct Mol Biol, 2013. **20**(3): p. 282-9.
272. Perrin, M.C., A.S. Brown, and D. Malaspina, *Aberrant epigenetic regulation could explain the relationship of paternal age to schizophrenia*. Schizophr Bull, 2007. **33**(6): p. 1270-3.
273. Lewis, M.I., M.J. Belman, and L. Dorr-Uyemura, *Nutritional supplementation in ambulatory patients with chronic obstructive pulmonary disease*. Am Rev Respir Dis, 1987. **135**(5): p. 1062-8.
274. Riviere, G., et al., *Epigenetic regulation of somatic angiotensin-converting enzyme by DNA methylation and histone acetylation*. Epigenetics, 2011. **6**(4): p. 478-89.