

Direction des bibliothèques

AVIS

Ce document a été numérisé par la Division de la gestion des documents et des archives de l'Université de Montréal.

L'auteur a autorisé l'Université de Montréal à reproduire et diffuser, en totalité ou en partie, par quelque moyen que ce soit et sur quelque support que ce soit, et exclusivement à des fins non lucratives d'enseignement et de recherche, des copies de ce mémoire ou de cette thèse.

L'auteur et les coauteurs le cas échéant conservent la propriété du droit d'auteur et des droits moraux qui protègent ce document. Ni la thèse ou le mémoire, ni des extraits substantiels de ce document, ne doivent être imprimés ou autrement reproduits sans l'autorisation de l'auteur.

Afin de se conformer à la Loi canadienne sur la protection des renseignements personnels, quelques formulaires secondaires, coordonnées ou signatures intégrées au texte ont pu être enlevés de ce document. Bien que cela ait pu affecter la pagination, il n'y a aucun contenu manquant.

NOTICE

This document was digitized by the Records Management & Archives Division of Université de Montréal.

The author of this thesis or dissertation has granted a nonexclusive license allowing Université de Montréal to reproduce and publish the document, in part or in whole, and in any format, solely for noncommercial educational and research purposes.

The author and co-authors if applicable retain copyright ownership and moral rights in this document. Neither the whole thesis or dissertation, nor substantial extracts from it, may be printed or otherwise reproduced without the author's permission.

In compliance with the Canadian Privacy Act some supporting forms, contact information or signatures may have been removed from the document. While this may affect the document page count, it does not represent any loss of content from the document.

Université de Montréal

Analyse de descendance : Une approche bio-informatique
pour estimer le risque d'hypertension et d'obésité

Par
François Gauthier

Programme de bio-informatique
Faculté de médecine

Mémoire présenté à la Faculté des études supérieures et postdoctorales
en vue d'obtention du grade de M.Sc.
en bio-informatique

Août 2007

© François Gauthier, 2007



Université de Montréal
Faculté des études supérieures

Ce mémoire intitulé :

Analyse de descendance : Une approche bio-informatique
pour estimer le risque d'hypertension et d'obésité

présenté par :

François Gauthier

a été évalué par un jury composé des personnes suivantes :

Christine Maugard, M.D. Ph.D.

Président-rapporteur

Pavel Hamet, M.D. Ph.D.

Directeur de recherche

Ettore Merlo, Ph.D.

Co-directeur

Claude Laberge, M.D. Ph.D.

Membre du jury

Abstract

The Saguenay-Lac St-Jean population of Quebec has genealogical records up to 17th century French settlers. In 120 extended families phenotype data were gathered from 900 subjects. A novel approach to estimate hypertension/obesity risk (high, medium or low) based on genealogical data is presented. In opposition to previous studies, we identified subgroups of founders more related to affected or healthy individuals. Furthermore, results show that descendants of the former have up to 56% more chances to suffer from these pathologies than those of the latter. Simulations confirmed the statistical association between descendant's phenotype values and genealogical transmission of ancestral traits. Once validated on a larger cohort, we are convinced that our method could be a major tool in public health as part of studies on high-risk populations. It could also be used in genome-wide association studies aiming to compare the genotypes of groups of individuals with highly polarized phenotypes.

Keywords : Graph Theory, Pedigree, Hypertension, Obesity, Multiple Testing Corrections

Résumé

Les registres généalogiques de la population du Saguenay-Lac-Saint-Jean de la province de Québec sont disponibles jusqu'aux premiers colons Français. Des données phénotypiques et génotypiques ont aussi été collectées auprès de neuf cents sujets provenant de cent vingt familles étendues. Nous présentons ici une nouvelle approche pour estimer le risque (élevé, moyen ou faible) d'hypertension et d'obésité basée sur la généalogie. Contrairement aux études précédentes, nous avons été en mesure d'identifier des sous-groupes de fondateurs fortement apparentés à des individus affectés ou sains. Les résultats montrent que les descendants du premier type de fondateur ont jusqu'à 56% plus de chance d'être atteints de ces pathologies que les descendants de fondateurs plus apparentés à des individus sains. Des simulations ont confirmé statistiquement l'association entre les phénotypes des descendants et la transmission de traits ancestraux. Une fois validée sur de plus grandes cohortes, nous sommes convaincu que notre méthode pourrait constituer un outil majeur en santé publique, dans le cadre d'études sur des populations à risque. Elle pourrait aussi être utilisée pour des études de génotypage à haut débit visant à comparer les génotypes de groupes d'individus présentant des phénotypes très polarisés.

Mots clés : Théorie des Graphes, Généalogie, Hypertension, Obésité, Corrections pour Tests Multiples

Table des matières

Table des figures	vi
Liste des tableaux	viii
I Introduction	1
1 Mise en contexte	2
2 Fondation du Saguenay-Lac-Saint-Jean	4
2.1 Effet fondateur au Saguenay-Lac-Saint-Jean	4
2.1.1 Projet BALSAC-RETRO	5
2.2 Saguenay-Lac-Saint-Jean et maladies génétiques	5
3 Historique familial et risque de maladie	7
3.1 Définition de l'historique familial de maladie	7
3.1.1 Stratification de l'historique	8
3.1.2 Stratification du risque	8
3.1.3 Évaluation du risque réel	8
3.2 Historique d'hypertension et risques	8
3.3 Historique d'obésité et risques	10
3.4 Prévention et prise de conscience	11
4 Hypertension	13
4.1 Définition de l'hypertension	13
4.1.1 Pression systolique vs. diastolique	13
4.1.2 Risques associés à l'hypertension	15
4.2 Facteurs de risque de l'hypertension	16
4.2.1 Âge	16
4.2.2 Ethnicité	17
4.2.3 Sodium et potassium	17
4.2.4 Poids	17
4.2.5 Alcool	18
4.2.6 Dyslipidémie	18
4.2.7 Prédispositions génétiques	18
4.2.8 Autres facteurs	19
5 Obésité	20
5.1 Indice de masse corporelle	20
5.1.1 Risques associés à l'obésité	22

5.2	Facteurs de risque de l'obésité	23
5.2.1	Apport énergétique	23
5.2.2	Sédentarisation	23
5.2.3	Ethnicité	23
5.2.4	Prédispositions génétiques	24
5.2.5	Autres facteurs	24
6	Héritabilité des phénotypes	25
6.1	Études d'héritabilité	26
6.1.1	Étude de jumeaux	26
6.1.2	Étude d'adoption et de jumeaux séparés	27
6.1.3	Études familiales	27
6.2	Héritabilité de la pression sanguine	27
6.3	Héritabilité de l'indice de masse corporelle	29
7	Obésité et hypertension	31
7.1	Association des phénotypes	31
7.2	Corrélation entre pression sanguine et indice de masse corporelle	32
II	Méthodologie et Expérimentations	33
8	Méthodologie	34
8.1	Nécessité d'une nouvelle approche	34
8.2	Recrutement des familles	35
8.3	Phénotypage	36
8.4	Génotypage	37
8.5	Traitement de la généalogie	37
8.5.1	Couches généalogiques	39
8.5.2	Contribution génétique des ancêtres	40
8.6	Conception de l'étude	44
8.6.1	Définition des variables	44
8.6.2	Classification des descendants	45
8.6.3	Interprétation de la spécificité génétique	46
8.7	Étude comparative des descendance	47
8.7.1	Procédure de comparaison	48
8.7.2	Analyse statistique	50
9	Expérimentations	53
9.1	Base de données et confidentialité	53
9.2	Applications et systèmes informatiques	54
III	Résultats	56
10	Résultats	57
10.1	Analyses ANOVA	58
10.2	Ensembles phénotypiques et seuils de risque	59

10.2.1	Évaluation du risque phénotypique	59
10.3	Ensembles phénotypiques et hypertension	62
10.3.1	Évaluation du risque d'hypertension	62
10.4	Simulations et validité statistique	65
10.4.1	Stratification sur l'obésité	65
10.4.2	Stratification sur l'âge	66
10.4.3	Stratification sur le sexe	68
10.5	Temps d'exécutions	69
IV	Discussion et conclusions	71
11	Discussion	72
11.1	Analyse comparative des descendance	72
11.2	Évaluation du risque	74
11.2.1	Risque phénotypique	75
11.2.2	Risque d'hypertension	75
11.3	Ensembles phénotypiques et couches généalogiques	77
11.4	Simulations et validité statistique	80
11.4.1	Indice de masse corporelle	80
11.4.2	Pression sanguine diastolique corrigée	81
11.4.3	Pression sanguine systolique corrigée	82
11.4.4	HDL	83
11.5	Validité de l'étude	84
11.5.1	Validité interne	85
11.5.2	Validité externe	85
11.5.3	Validité de construction	86
11.5.4	Validité des conclusions statistiques	86
12	Conclusions et perspectives	87
	Bibliographie	89

Table des figures

4.1	Distribution des valeurs de pressions sanguine systolique en fonction de différents groupes d'âge aux États-Unis.	14
4.2	Valeurs moyennes de pression sanguine systolique (SBP) et diastolique (DBP) en fonction de l'âge et du sexe selon deux études américaine (NHANES III) et canadienne (CHHS) [32].	16
4.3	Divers facteurs pouvant influencer les mesures de pression sanguine. Figure adaptée de [55]	19
5.1	Distribution des valeurs d'indice de masse corporelle et classes pondérales des adultes Canadiens de 18 ans et plus.	21
5.2	Distribution des valeurs d'indice de masse corporelle moyennes en fonction de l'âge et du sexe selon deux études américaines (NHANES et BRFSS) et deux périodes de temps (1988-1994 et 1999-2002). [67]	22
8.1	Pedigree familial fictif.	39
8.2	Généalogie fictive encodée sous forme de graphe généalogique	41
8.3	Graphique d'identité par descendance des ancêtres pour deux classes de descendants.	43
8.4	Comparaison des descendance en fonction de la spécificité des ancêtres.	49
9.1	Schéma de la répartition des opérations. Les chiffres indiquent l'ordre dans lequel les analyses étaient effectuées.	55
10.1	Analyse ANOVA de l'indice de masse corporelle (IMC) des descendance en fonction de la spécificité des ancêtres pour la première couche généalogique.	60
10.2	Analyse ANOVA de l'indice de masse corporelle (IMC) des descendance en fonction de la spécificité des ancêtres pour la quatrième couche généalogique.	60
10.3	Analyse ANOVA de l'indice de masse corporelle (IMC) des descendance en fonction de la spécificité des ancêtres pour la septième couche généalogique.	61
10.4	Analyse ANOVA de l'indice de masse corporelle (IMC) des descendance en fonction de la spécificité des ancêtres pour la douzième couche généalogique.	61
10.5	p-values ANOVA simulées de la première à la dernière couche généalogique pour les trois phénotypes investigués.	66
10.6	p-values ANOVA simulées de la première à la dernière couche généalogique pour la pression systolique corrigée avec et sans stratification sur l'obésité.	67

10.7	p-values ANOVA simulées de la première à la dernière couche généalogique pour la pression diastolique corrigée avec et sans stratification sur l'obésité.	67
10.8	p-values ANOVA simulées de la première à la dernière couche généalogique pour la pression systolique corrigée avec et sans strates d'âge. . .	68
10.9	p-values ANOVA simulées de la première à la dernière couche généalogique pour la concentration sanguine d'HDL avant et après séparations des sexes.	69

Liste des tableaux

IV.I Classification des pressions sanguines pour les adultes de 18 ans et plus ¹ [35]	15
V.I Classification du surpoids et de l'obésité par indices de masse corporelle et risque de maladie associé à la circonférence de la taille pour les adultes de 18 ans et plus [63]	21
IX.I Spécifications techniques des différents postes informatiques utilisés.	55
X.I Proportion (%) d'individus à risque et cardinalité (taille) des classes de descendants pour toutes les couches généalogiques et phénotypes investigués.	63
X.II Proportion (%) d'individus hypertendus et cardinalité (taille) des classes de descendants pour toutes les couches généalogiques et phénotypes investigués.	64
X.III Temps de calculs moyen pour les différentes étapes de l'analyse de descendance sur une couche généalogique pour trois phénotypes investigués.	70

Liste des abréviations

- AVC : Accident vasculaire cérébral
- BMI : Body mass index (Indice de masse corporelle)
- BRFSS : Behavioral risk factor surveillance system
- CHHS : Canadian Heart Health Surveys
- CORAMH : Corporation de recherche et d'action sur les maladies héréditaires
- DAG : Directed acyclic graph (Graphe dirigé acyclique)
- DBP : Diastolic blood pressure (Pression diastolique)
- FHRP : Family high risk program
- IMC : Indice de masse corporelle
- NHANES : National Health and Nutrition Examination Survey (États-Unis)
- NHES : National Health Examination Survey (États-Unis)
- NIH : National Institute of Health (États-Unis)
- OMS : Organisation mondiale de la santé
- SBP : Systolic blood pressure (Pression systolique)
- SLSJ : Saguenay-Lac-Saint-Jean
- WHO : World health organization

Dédicace

À mes parents, pour leur support et leurs encouragements

Remerciements

Je tiens d'abord à remercier les familles participantes de la région du Saguenay-Lac-Saint-Jean ainsi que toute l'équipe du projet BALSAC-RETRO. J'aimerais ensuite remercier tout ceux et celles qui m'ont aidé de près ou de loin au cours des années, particulièrement Pierre-Luc Brunelle et Alexandru Gurau pour leur support informatique, Louis-Philippe Lemieux-Perrault pour son expertise en $\text{\LaTeX} 2_{\epsilon}$, Audrey Noël pour son support moral et bien sûr Dr. Ettore Merlo et Dr. Pavel Hamet pour avoir si bien guidé mes recherches.

Je tiens aussi à remercier le Dr. Ondrej Šeda pour ses précieuses analyses de corrélation.

Finalement, je tiens à exprimer ma gratitude à l'endroit de Mélanie Robitaille, correctrice assidue, pour son soutien indéfectible.

Première partie

Introduction

1. Mise en contexte

Les maladies multifactorielles telles l'hypertension et l'obésité sont en partie causées par des facteurs génétiques, environnementaux et culturels. Alors que les traits génétiques sont clairement transmis de générations en générations, certains traits culturels et environnementaux font aussi l'objet d'une aggrégation familiale. Par exemple, une étude a récemment révélé que l'obésité se propage à travers les réseaux sociaux [1]. Plutôt que de tenter de départager l'influence de la génétique, de la culture et de l'environnement, les études d'historique familial visent à considérer tous ces facteurs en combinaison afin d'estimer le risque de maladie d'un individu.

Alors que les études d'historique familial se concentrent sur la famille nucléaire voire étendue, nous présentons ici une étude d'historique ancestral d'hypertension et d'obésité de la population du Saguenay-Lac-Saint-Jean (SLSJ). Après avoir présenté les faits saillants de la fondation du SLSJ et les études d'historique familial, nous ferons un bref survol de l'hypertension, de l'obésité et des études d'héritabilité.

Nous présenterons ensuite une nouvelle approche visant à estimer le risque d'hypertension et d'obésité d'un individu selon son ancestralité. Pour ce faire, il est d'abord nécessaire d'identifier les ancêtres susceptibles d'avoir transmis des traits pathologiques à leurs descendants. À cette fin, nous avons développé une nouvelle mesure : la spécificité génétique. Celle-ci permet de quantifier la contribution génétique différentielle des ancêtres pour des ensembles de descendants affectés et sains. Nous avons ainsi identifié des sous-groupes de fondateurs fortement apparentés à des descendants hypertendus

ou obèses, alors qu'une étude similaire sur la même population concluait «qu'il n'y a pas de sous-groupe de fondateurs davantage apparenté à n'importe quel groupe de descendants affectés par une maladie monogénique» [2].

Une fois ces sous-groupes identifiés, nous avons voulu vérifier si leur spécificité génétique pour des individus hypertendus ou obèses se traduisait en une augmentation significative de ces traits chez leurs descendants. Nous sommes même allés plus loin en classifiant les fondateurs selon leur spécificité génétique afin de vérifier à quel point celle-ci est proportionnelle aux taux d'hypertension et d'obésité chez les descendants. Finalement, tous nos résultats ont été validés statistiquement par une méthode de correction pour tests multiples par simulations.

2. Fondation du Saguenay-Lac-Saint-Jean

Cette étude s'inscrit dans un vaste projet de caractérisation de l'hypertension et de l'obésité chez la population du Saguenay-Lac-Saint-Jean (SLSJ). Afin de bien saisir le contexte de l'étude, il est impératif de remonter le temps, jusqu'au XIX siècle, au moment de la fondation du SLSJ. Le peuplement de cette région a débuté au milieu des années 1800 par des individus provenant majoritairement de Charlevoix. On estime aujourd'hui que 90% des individus nés entre 1950 et 1970 descendent de ces fondateurs [3]. Une connaissance approfondie de la fondation du Saguenay par Charlevoix est donc nécessaire pour bien saisir la situation génétique qui y prévaut actuellement.

2.1 Effet fondateur au Saguenay-Lac-Saint-Jean

La région de Charlevoix a été peuplée par des habitants de la ville de Québec vers la fin du XVII siècle. Il y avait alors environ 200 habitants. En 1831, la région comprenait plus de 8000 habitants et la migration vers le SLSJ débuta. De 1838 à 1911, près de 14 000 personnes ont émigrés de Charlevoix vers le SLSJ [4]. Le patrimoine génétique observé dans les régions de Charlevoix et du SLSJ provient donc, en grande partie, d'un nombre restreint d'ancêtres de la région de Québec qui ont produit un effet fondateur [5, 6]. On désigne sous l'appellation effet fondateur le phénomène par lequel un groupe migre d'une population mère vers un nouveau territoire où ils peuvent se reproduire et fonder une nouvelle population. L'effet fondateur a pour conséquence de

provoquer une certaine homogénéité génétique dans la nouvelle population pour certains loci. Dans le cas des régions de Charlevoix et du SLSJ, l'homogénéité génétique observée est d'autant plus marquée que les fondateurs venus de la régions de Québec étaient eux-mêmes parents de premier degré (frère et soeurs) à plus de 50% [4]. On estime aujourd'hui que 65% du patrimoine génétique du SLSJ provient des fondateurs de Charlevoix [7].

2.1.1 Projet BALSAC-RETRO

Ces estimations sont rendues possibles, en partie, grâce à la base de données BALSAC-RETRO [8, 9]. Elle répertorie la généalogie des habitants du SLSJ qui comporte plus de 40 000 personnes, s'étend sur 17 générations et retrace jusqu'aux premiers colons Français, établis au début du XVII^e siècle. Les données qui nourrissent la base de données sont tirées principalement des actes de l'état civil (naissances, mariages, sépultures). L'objectif est de recouvrir l'ensemble de la population du Québec, depuis le début du peuplement au XVII^e siècle jusqu'à la période actuelle. Cette base de données généalogiques nous est accessible.

2.2 Saguenay-Lac-Saint-Jean et maladies génétiques

Au cours des 20 dernières années, plusieurs études sur la population du SLSJ ont enrichies nos connaissances sur diverses maladies monogéniques [10]. Le taux de natalité élevé et l'effet fondateur au SLSJ sont jugés responsables de la fréquence anormalement élevée de maladies génétiques récessives dans cette population [10, 3].

Bien que le lien entre l'effet fondateur et les maladies polygéniques soit plus difficile à démontrer, des chercheurs se sont récemment intéressés à l'investigation, dans cette population, de maladies multifactorielles telles l'Alzheimer, la maniaque-dépression et

les lymphomes [11, 2].

Dans le cadre de cette étude, nous nous sommes plutôt concentré sur l'investigation de la transmission de deux maladies multifactorielles et polygéniques [12, 13], l'hypertension et l'obésité, à travers les descendance. Des traits génétiques, épigénétiques, culturels et environnementaux sont en effet transmis au fil des générations et influencent le risque de maladie d'un individu. Bien qu'il y ait des différences au niveau de la méthodologie employée, notre analyse s'apparente grandement aux études d'historique familial, par lesquelles on peut aussi estimer le risque de maladie d'un individu.

3. Historique familial et risque de maladie

L'historique familial est couramment utilisée en santé publique comme un outil simple et peu coûteux pour le dépistage et la prévention à grande échelle. Il permet, en effet, d'estimer le risque de maladie d'un individu sans recours à des techniques onéreuses de génotypage/phénotypage. De plus, l'historique familial est en mesure de capturer la complexité d'interactions gènes-gènes, gènes-environnement, et autres, en identifiant des groupes de personnes présentant une combinaison de facteurs de risque menant à l'expression de la maladie [14]. L'historique familial est donc tout indiqué pour l'étude de maladies complexes telles l'hypertension et l'obésité pour lesquelles les mécanismes génétiques sont encore peu connus [15].

Dans le cadre de cette étude, nous estimerons le risque d'hypertension/obésité à partir de «l'historique ancestral» d'un individu. Par analogie à l'historique familial, permettant d'estimer le risque de maladie d'un individu selon sa parenté, l'historique ancestral permet de qualifier le risque d'un individu en fonction de ses ancêtres.

3.1 Définition de l'historique familial de maladie

L'historique familial de maladie est un terme très général ; dépendamment du type de maladie ou des données accessibles, les définitions varieront. Certaines constantes se dégagent toutefois des différentes études.

3.1.1 Stratification de l'historique

Étant donné un pedigree familial quelconque pour lequel l'historique de maladie est connu, on classera souvent les individus selon le nombre et la proximité de parents affectés [16]. On fera par exemple la distinction entre un individu dont un parent de premier degré (frère, soeur, parents) est affecté et un autre dont seulement un parent de second degré est malade.

3.1.2 Stratification du risque

La stratification de l'historique mène à l'estimation du risque. Le risque de maladie est généralement classé en trois catégories distinctes soient : élevé, modéré et bas [17, 18]. Les critères de classification sont variables et prennent la forme de conditions binaires (Par exemple, une personne sera considérée à risque élevé si elle possède au moins un parent de premier degré affecté par la maladie, à risque modéré si elle n'a que des parents de second degré affecté et à risque bas sinon.)

3.1.3 Évaluation du risque réel

Une fois les catégories de risque déterminées, on estimera le risque réel de maladie associé à l'appartenance à l'une ou l'autre de ces classes. Les méthodes d'estimation sont diverses. Certains compareront la prévalence de la maladie en fonction des classes de risque [17], d'autres calculeront des *odds ratio* (OR) [19, 18] ou encore des risques relatifs [15] d'être affecté en fonction de la classe de risque.

3.2 Historique d'hypertension et risques

De manière plus spécifique, diverses études ont montrés une association positive entre l'historique familial d'hypertension artérielle ou de haute pression sanguine et un risque accru pour différentes maladies (démence, maladies coronarienne, rétinopathie, dysfonction rénale et autres). Les résultats varient toutefois passablement d'une étude

à l'autre, dépendamment de la méthodologie utilisée et des covariables considérées, nous ne proposons donc ici qu'un bref survol. Pour plus de détails, voir Goldstein *et al.* qui ont fait une revue détaillée des différents types d'études et des facteurs influençant les résultats [20].

Goldstein *et al.*, après avoir classifiés les sujets selon leur nombre de parents hypertendus (0, 1 ou 2), ont mis en évidence une association significative entre l'historique d'hypertension et l'existence d'une pression sanguine élevée. L'effet de l'historique familial s'est révélé significatif (p-value ANOVA < 0.05) pour toutes les mesures de pression sanguine systolique et diastolique testées (éveillé, au repos et ponctuelle). Les sujets ayant 2 parents hypertendus présentaient les mesures de pression les plus élevées suivi de ceux n'ayant qu'un et aucun parent hypertendu respectivement [20].

Suivant un modèle de classification similaire, Vandenberg *et al.* ont mis en lumière la corrélation entre les valeurs de pression sanguine chez les parents et les enfants. Les sujets étaient classifiés en fonction du tertile de pression sanguine de leur parent (0, 1 ou 2 parents dans le plus haut tertile). L'étude longitudinale sur 27 ans a montré qu'à partir de 15 ans les pressions sanguines systolique et diastolique des enfants étaient systématiquement plus élevées chez ceux dont les 2 parents étaient dans le plus haut tertile suivi de ceux dont 1 et 0 parents étaient dans le premier tertile [21].

Lascaux-Lefebvre *et al.*, ont estimé que les chances d'être hypertendu étaient environ deux fois plus élevées chez les enfants dont au moins un parent présentait un historique d'hypertension alors que chez les enfants dont les deux parents étaient hypertendus, les chances étaient environ trois fois plus élevées [22]

Kondo *et al.* ont, quant à eux, montré l'existence d'une coaggrégation entre les historiques d'hypertension et d'accident vasculaire cérébral (AVC). L'historique était

défini selon le nombre de parents, de frères et de soeurs atteints d'hypertension et/ou ayant subi un AVC. Des OR permettaient de montrer les corrélations entre l'hypertension et les AVC en fonction de l'historique familial [19].

3.3 Historique d'obésité et risques

L'environnement, la génétique et plusieurs facteurs sociaux influencent le développement de l'obésité. L'historique familial permet de capturer l'effet de la combinaison de ces facteurs et des études ont montré une association entre l'obésité parentale et infantile. Nous en proposons ici un bref survol.

L'indice de masse corporelle (IMC ou BMI en anglais) est, selon le NIH, «une mesure de gras corporel basée sur la grandeur et le poids qui s'applique aux hommes et aux femmes» et constitue le critère par excellence d'évaluation de l'obésité/surpoids. La distribution de l'IMC dans la population et les seuils d'obésité/surpoids seront présentés à la section 5.1.

Une étude de Park *et al.* a montré qu'un historique d'obésité parentale favorisait l'obésité chez les enfants. Leur étude, portant sur 260 familles nucléaires coréennes, a révélé que les enfants dont les deux parents sont en surpoids ont significativement plus de chances (OR entre 5.2 et 6.1) d'être aussi en surpoids par rapport à ceux dont les parents sont de poids normal [23].

Neutzling *et al.* ont aussi mis en lumière l'influence de l'historique familial dans une étude de cas-contrôle au Brésil portant sur 264 adolescents de poids normaux et 264 adolescents en surpoids/obèses (IMC \geq 85^{ième} percentile). Ils ont montré que les rapports de cotes d'être en surpoids/obèse étaient respectivement de 2.51 et 2.48 lorsque le père ou la mère étaient obèses. Cependant, aucune mesure de risque n'a été

estimée pour les sujets dont les deux parents sont obèses [24]. Hajian-Tilaki et Hei-dari ont aussi mis en relief l'augmentation du risque d'obésité chez les sujets avec un historique d'obésité parentale. Leur étude portant sur 1800 hommes et 1800 femmes Iraniens, a montré que les sujets dont au moins un parent est obèse ont 2.14 fois plus de chances d'être eux-même obèse que les sujets sans historique familial d'obésité [25].

Lors d'une étude parrainée par l'OMS portant sur 5389 adultes gambiens, Van Der Sande *et al.* ont montré qu'un historique familial d'obésité (au moins un parent de premier degré affecté) entraînait une hausse significative de l'IMC moyen des sujets. Ils ont aussi calculé que les chances d'être obèse étaient environ quatre fois plus élevées chez les sujets ayant un historique familial d'obésité (OR de 3.95) [26].

3.4 Prévention et prise de conscience

L'évaluation du risque de maladie par historique familial est généralement utilisée dans le cadre de programmes de prévention en santé publique. Deux courants de pensée peuvent être distingués : la prévention à large échelle auprès de la population et les interventions ciblées auprès de petits groupes à haut risque [14]. L'historique familial a l'avantage de s'appliquer dans les deux cas. En effet, une fois le risque estimé, les professionnels de la santé peuvent déterminer quelle est la meilleure approche à adopter.

Bien qu'il soit reconnu comme un facteur de risque pour plusieurs maladies, l'historique familial est malheureusement sous-utilisé en médecine préventive. Henderson et Scheuner ont d'ailleurs montré que l'historique familial n'était noté au dossier que dans 36% des cas où le sujet la fournissait [27] et une étude d'Acheson *et al.* a révélé que seulement 11% des dossiers des sujets contenaient un arbre familial [28]. Il reste donc à démontrer que l'historique familial est un outil de prévention efficace afin que les médecins l'intègre à leur pratique quotidienne.

Un projet pilote en Utah, aux États-Unis, a justement mis en lumière l'efficacité de cet outil peu coûteux pour la prévention auprès d'une population à haut risque. La situation de l'Utah est particulière puisque 70% des citoyens sont de confession mormonne et forment une population relativement isolée sur le plan social. Affectée par un effet fondateur, elle présente une prévalence élevée de maladies cardio-vasculaires et autres maladies tout comme la population du SLSJ. De plus, l'Église de Jésus-Christ des Saints des Derniers Jours, encourage aussi la tenue de registres généalogiques, ce qui permet souvent de retracer le pedigree des familles jusqu'à des ancêtres du 17^e siècle [29].

C'est dans ce contexte que fut implanté le *Family High Risk Program* (FHRP). Ce programme fut créé dans le but «d'identifier les familles à haut risque de développer des maladies à déclenchement tardif qui peuvent être prévenues, retardées ou traitées efficacement à l'aide d'interventions préventives.» [29]. De 1983 à 1999, 151 188 familles ont été questionnées sur leur historique pour diverses maladies. Un suivi médical était offert aux familles à haut risque et, selon les auteurs, «les familles [ayant participé au FHRP] étaient motivées à modifier leurs habitudes de vie à long terme par le simple fait de connaître leur historique familial.» [29]

Des programmes similaires sont depuis longtemps en place au SLSJ. La corporation de recherche et d'action sur les maladies héréditaires (CORAMH), créée en 1980 vise, entre autre, à informer les individus, principalement des étudiants du secondaire, sur les maladies génétiques héréditaires communes au SLSJ et sur les mesures à adopter lorsqu'on en est atteint [30]. Associé à la CORAMH, le programme de recherche ECOGENE-21 *de l'ADN à la communauté* étudie présentement la population du SLSJ et a pour mission l'acquisition, l'intégration et le transfert de connaissances sur le génome humain vers les professionnels de la santé, les individus et les communautés [31].

4. Hypertension

Selon l'Organisation Mondiale de la Santé (OMS), l'hypertension artérielle, ou haute pression sanguine, affecte environ 1 milliard d'individus dans le monde et serait responsable de 7,1 millions de décès par année. On estime que de 20% à 25% de la population nord-américaine serait affectée [32, 33]. Bien que des améliorations significatives aient été faites dans le traitement et la prévention de l'hypertension au cours des dernières décennies, des efforts doivent encore être investis.

4.1 Définition de l'hypertension

La pression sanguine, mesurée en mmHg, a une distribution continue gaussienne (voir figure 4.1). Toutes tentatives de définitions de seuils d'hypertension sont donc inévitablement biaisées. Il n'existe pas réellement de groupes hypertendus et normotendus, mais bien un intervalle de pressions sanguines, de la plus basse à la plus élevée avec la majorité des individus situés quelque part au milieu. Toutefois, pour les besoins d'études cliniques, on classifie souvent les sujets selon leur pression sanguine (voir le tableau IV.I). On nomme pression systolique la pression maximale au moment de la contraction du coeur (systole). La pression diastolique est définie comme la pression minimale au moment du relâchement du coeur (diastole).

4.1.1 Pression systolique vs. diastolique

Plusieurs études suggèrent que la hauteur de la pression systolique (en mmHg) est un meilleur indicateur d'infarctus et de crise cardiaque que celle de la pression diastolique [36, 35]. Ainsi, différents types d'hypertension sont définis selon qu'on considère

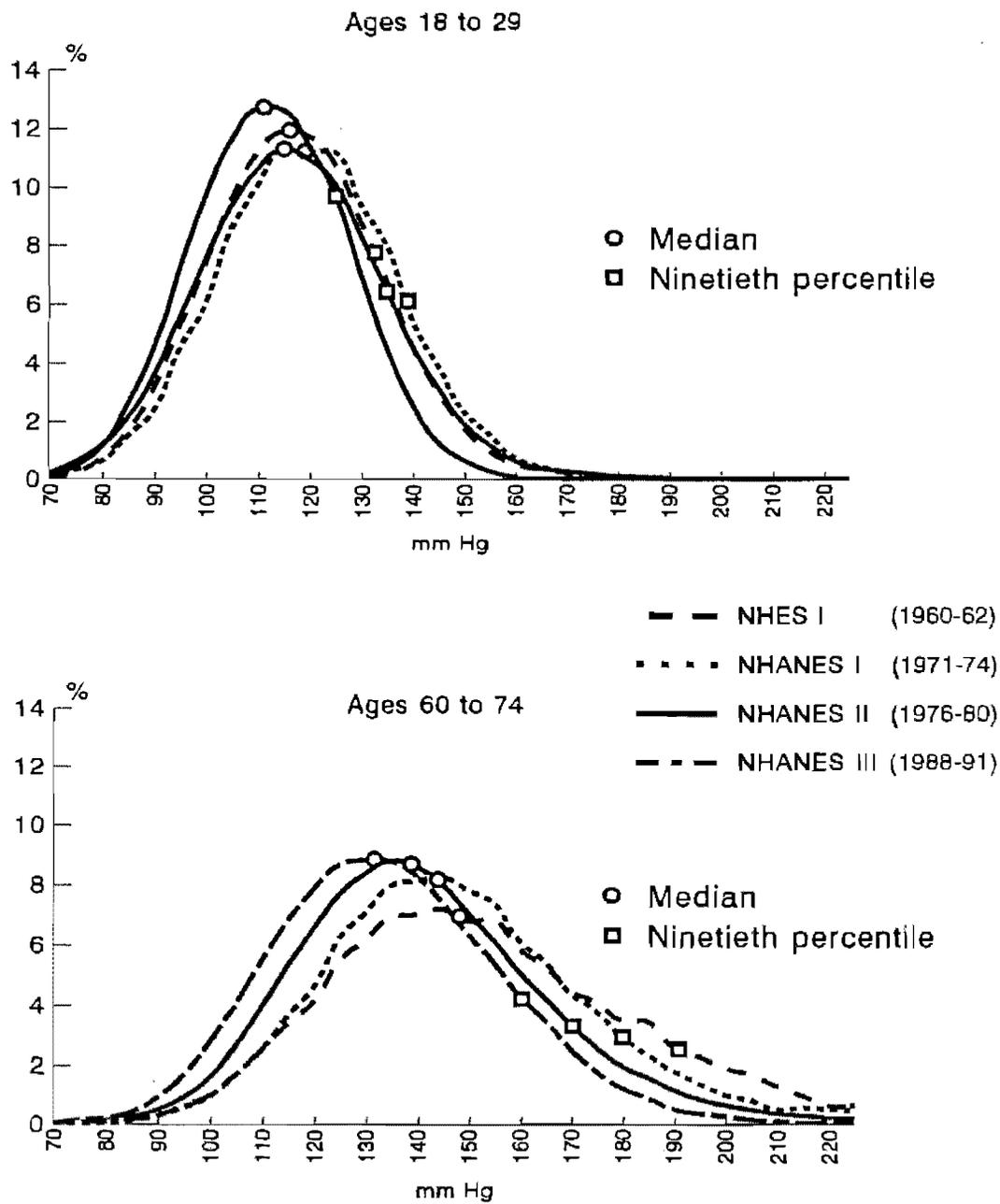


Fig. 4.1: Distribution des valeurs de pressions sanguine systolique en fonction de différents groupes d'âge aux États-Unis. Le premier graphique représente la distribution des valeurs de pressions sanguine d'individus entre 18 et 29 ans, le second de personnes âgées entre 60 et 74 ans. Les cercles représentent les valeurs médianes et les carrés le 90^e percentile. Les distributions sont normales dans les deux cas [34].

Tab. IV.I: Classification des pressions sanguines pour les adultes de 18 ans et plus¹ [35]

Catégorie	Pression sanguine (mmHg)		
	Systolique		Diastolique
Normal	<120	et	<80
Préhypertension	120-139	ou	80-89
Hypertension			
Stade 1	140-159	ou	90-99
Stade 2	≥160	ou	≥100

¹ Ne prenant pas de médicaments anti-hypertensifs et qui ne sont pas malades. Lorsque les pressions systoliques et diastoliques ne sont pas dans la même catégorie, la catégorie la plus élevée est choisie pour classer le sujet.

seulement la pression systolique (ISH, *Isolated systolic hypertension*), la pression diastolique (IDH, *Isolated diastolic hypertension*) ou les deux à la fois (SDH, *Systolic diastolic hypertension*). Comme la pression diastolique décroît généralement aux alentours de 50 ans et que la pression systolique ne cesse d'augmenter (voir figure 4.2), les cas d'IDH sont plus courant chez les personnes de 50 ans et moins alors qu'on observe davantage d'hypertension systolique chez les plus de 50 ans.

4.1.2 Risques associés à l'hypertension

L'hypertension en soi est un état de santé et n'est pas mortelle en elle-même. Elle peut toutefois entraîner de multiples complications et maladies qui sont des causes majeures de mortalité en Amérique du Nord. Les principales sont [36] :

- Infarctus du myocarde
- Démence due à de multiples accidents vasculaires cérébraux asymptomatiques.
- Maladie coronarienne
- Hypertrophie du ventricule gauche
- Dysfonction rénale
- Rétinopathie (maladie affectant les vaisseaux sanguins de la rétine)

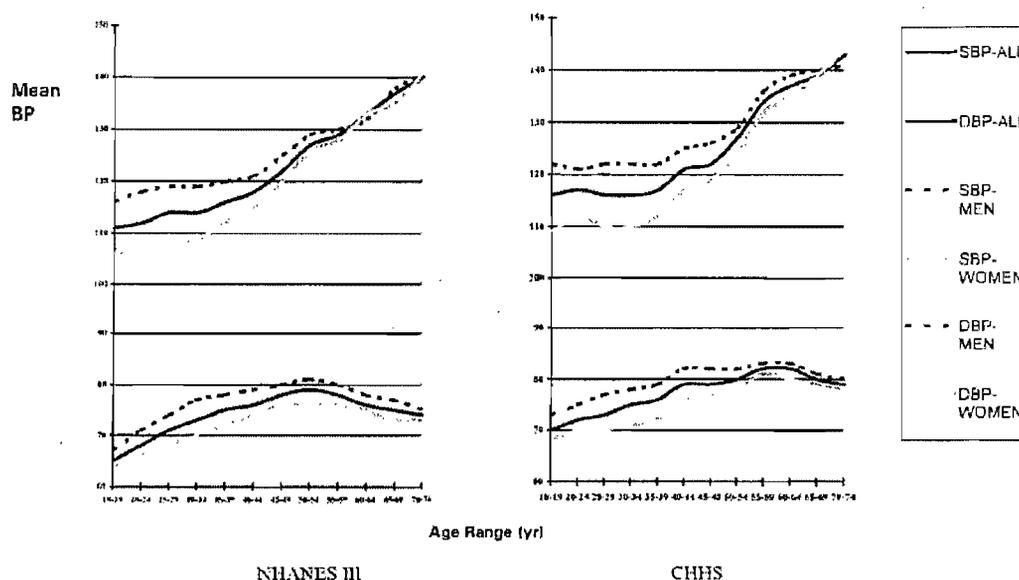


Fig. 4.2: Valeurs moyennes de pression sanguine systolique (SBP) et diastolique (DBP) en fonction de l'âge et du sexe selon deux études américaine (NHANES III) et canadienne (CHHS) [32].

4.2 Facteurs de risque de l'hypertension

L'hypertension n'est pas une maladie rare, surtout en Amérique du Nord, mais n'est toutefois pas distribuée également dans la population. Aux États-Unis, on estime qu'environ 24% de la population souffre d'hypertension et que de ces 24%, les trois quarts sont âgés de 50 ans et plus. Outre l'âge, plusieurs autres facteurs environnementaux ou génétiques ont une influence sur la prévalence de l'hypertension dans la population (voir figure 4.3).

4.2.1 Âge

Alors que la pression systolique ne cesse de croître avec l'âge, la pression diastolique commence à décroître aux alentours de 50 ans. Ce phénomène pourrait expliquer que la pression systolique est considérée comme un meilleur indicateur d'hypertension. Cependant, les raisons de la baisse de pression diastolique demeurent inexpliquées. Une hypothèse serait que les gens atteints d'hypertension diastolique, ou IDH, décèderaient

prématurément. Une autre explication plausible serait l'incapacité à maintenir une pression diastolique plus élevée à cause du déclin des fonctions cardiaques avec l'âge. Quoiqu'il en soit, la prévalence d'hypertension augmente avec l'âge [36].

4.2.2 Ethnicité

La plupart des études portant sur l'hypertension dans différents groupes ethniques ont révélé une plus forte prévalence de l'hypertension chez les personnes de race noire par rapport à celles de race blanche. Cette différence demeure significative même après correction pour divers facteurs environnementaux et socio-économiques [35, 37].

4.2.3 Sodium et potassium

De plus en plus d'études suggèrent que l'apport en sel a un effet direct sur la pression sanguine [38, 39]. Le projet *Intersalt* a permis la comparaison de données sur 52 populations provenant de 32 pays. Les résultats de l'étude suggèrent que les hausses de pression sanguine, avec l'âge, sont plus accentuées en milieu urbain que rural et que cette différence était attribuable à une consommation élevée de sel en milieu urbain. Le projet *Intersalt* a aussi montré qu'un apport élevé de potassium génèrait une prévalence plus basse d'hypertension [40].

4.2.4 Poids

Les individus obèses ou souffrant d'embonpoint ont tendance à présenter des mesures de pression sanguine plus élevées que les individus sveltes. L'étude de Framingham a d'ailleurs montré qu'une augmentation de la masse corporelle de 4,5kg était associée à une augmentation de 4 mmHg chez les hommes et les femmes. [41] D'autres études ont aussi montré une association entre l'embonpoint/obésité et une augmentation des mesures de pression sanguine. [42, 43]

4.2.5 Alcool

Des études, dont le projet *Intersalt* ont aussi montré une association entre la consommation d'alcool et la pression sanguine. De manière générale, il semble que plus l'apport en alcool augmente, plus la pression sanguine est élevée. Il a aussi été montré que cet effet est «réversible», puisqu'une diminution de la consommation d'alcool contribue à diminuer la pression sanguine [44]. Hamet *et al.* ont quant à eux mis en lumière les effets combinés des apports de sodium, de calcium et d'alcool sur la pression sanguine. Ils ont en effet révélé que l'augmentation de la pression sanguine attribuable à un apport élevé de sodium était amplifiée par une consommation importante d'alcool [45, 46].

4.2.6 Dyslipidémie

Non seulement communément associée à l'hypertension [47], des études suggèrent que la dyslipidémie pourrait être impliquée dans la pathophysiologie de l'hypertension en endommageant les parois endothéliales des vaisseaux sanguins [48]. Des études ont aussi relevé une augmentation significative du risque d'hypertension chez les individus atteints de dyslipidémie [49, 50, 51]. Nous verrons d'ailleurs à la section 8.2 que la dyslipidémie (cholestérol plasmatique ≥ 5.2 mmol/L et/ou cholestérol HDL ≤ 0.9 mmol/L ou prise de médicaments hypolipémiants) constituait, avec l'hypertension, un des critères de sélection des familles du SLSJ pour la participation à notre étude.

4.2.7 Prédispositions génétiques

Lorsqu'il est question de la génétique de l'hypertension, on doit faire la distinction entre deux formes de la maladie : l'hypertension monogénique et l'hypertension essentielle. Tel que son nom l'indique, l'hypertension monogénique résulte d'une mutation dans un gène affectant une voie métabolique particulière et provoquant l'hypertension chez le sujet. L'autre forme, l'hypertension essentielle, constitue 95% de tous les cas d'hypertension [52]. Cette forme de la maladie, à déclenchement tardif, est partiellement

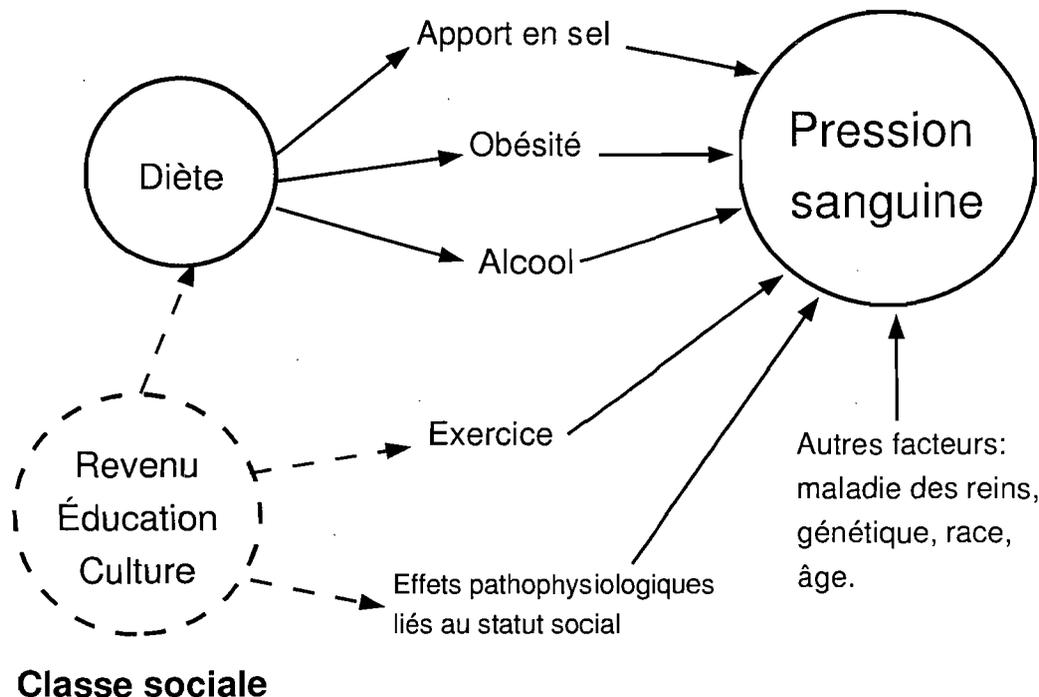


Fig. 4.3: Divers facteurs pouvant influencer les mesures de pression sanguine. Figure adaptée de [55]

influencée par de multiples facteurs environnementaux et génétiques [13]. L'identification d'éléments génétiques est toutefois une tâche colossale. Bien qu'il n'existe encore aucune preuve formelle, l'hypothèse la plus répandue suggère que l'hypertension essentielle est causée, en partie, par une combinaison de multiples facteurs génétiques de faible influence [53, 54].

4.2.8 Autres facteurs

Bien que difficilement quantifiables, une multitude d'autres facteurs peuvent influencer la pression sanguine. Le niveau de stress, la sédentarité, la culture, le revenu moyen, le tabagisme et l'éducation sont autant d'éléments qui peuvent avoir des conséquences sur le niveau de pression sanguine [55].

5. Obésité

L'obésité est en voie de devenir une épidémie mondiale et on estime qu'elle est la cause d'environ 2,6 millions de décès annuels à travers le monde. [56, 57, 58]. C'est toutefois en Amérique du Nord que la crise est la plus intense. Au Canada comme aux États-Unis, le taux d'individus obèses ou atteint de surpoids n'a cessé d'augmenter au cours de la dernière décennie [59, 60]. Aujourd'hui, on estime que 23% des Canadiens sont obèses et qu'un 36% additionnel sont en surpoids. Aux États-Unis ces chiffres atteignent respectivement 30% et 35% [61, 60, 62].

5.1 Indice de masse corporelle

Tel que mentionné précédemment, l'IMC constitue le critère par excellence d'évaluation de l'obésité/surpoids. La formule utilisée pour calculer l'IMC est :

$$\text{Métrique : } \quad IMC = \text{masse}(kg) / \text{grandeur}(mètres)^2$$

$$\text{Impérial : } \quad IMC = \text{masse}(livres) / \text{grandeur}(pouces)^2 \times 703$$

Tout comme la pression sanguine, l'IMC a une distribution continue gaussienne (voir figure 5.1). Toute classification est donc arbitraire et doit être considérée avec précautions. Toutefois, pour les besoins d'études cliniques, des seuils de surpoids et d'obésité ont été définis (voir tableau V.I). Comme la pression diastolique, l'IMC croît avec l'âge jusqu'aux alentours de 60 ans et décroît par la suite. (voir figure 5.2).

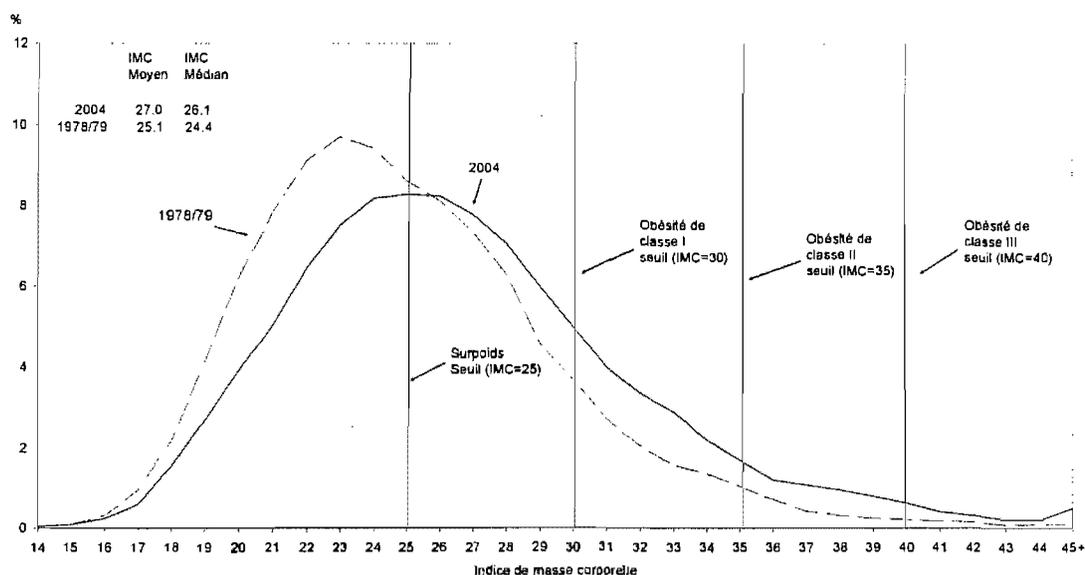


Fig. 5.1: Distribution des valeurs d'indice de masse corporelle et classes pondérales des adultes Canadiens de 18 ans et plus. Les données sont tirées des CHHS de 1978/1979 et 2004. Les distributions sont grossièrement normales dans les deux cas. [60]

Tab. V.I: Classification du surpoids et de l'obésité par indices de masse corporelle et risque de maladie associé à la circonférence de la taille pour les adultes de 18 ans et plus [63]

Catégorie	IMC kg/m ²	Risque de maladie ¹ par rapport aux individus de poids normal en fonction de la circonférence de la taille.	
		Hommes ≤102 cm	Hommes >102 cm
		Femmes ≤88 cm	Femmes >88 cm
Insuffisance pondérale	<18.5
Normal	18.5-24.9
Surpoids	25.0-29.9	Augmenté	Élevé
Obésité, classe			
I	30.0-34.9	Élevé	Très élevé
II	35.0-39.9	Très élevé	Très élevé
III (Obésité morbide)	≥40	Extrêmement élevé	Extrêmement élevé

¹ Risque de maladie pour le diabète de type 2, l'hypertension et les maladies cardio-vasculaires.

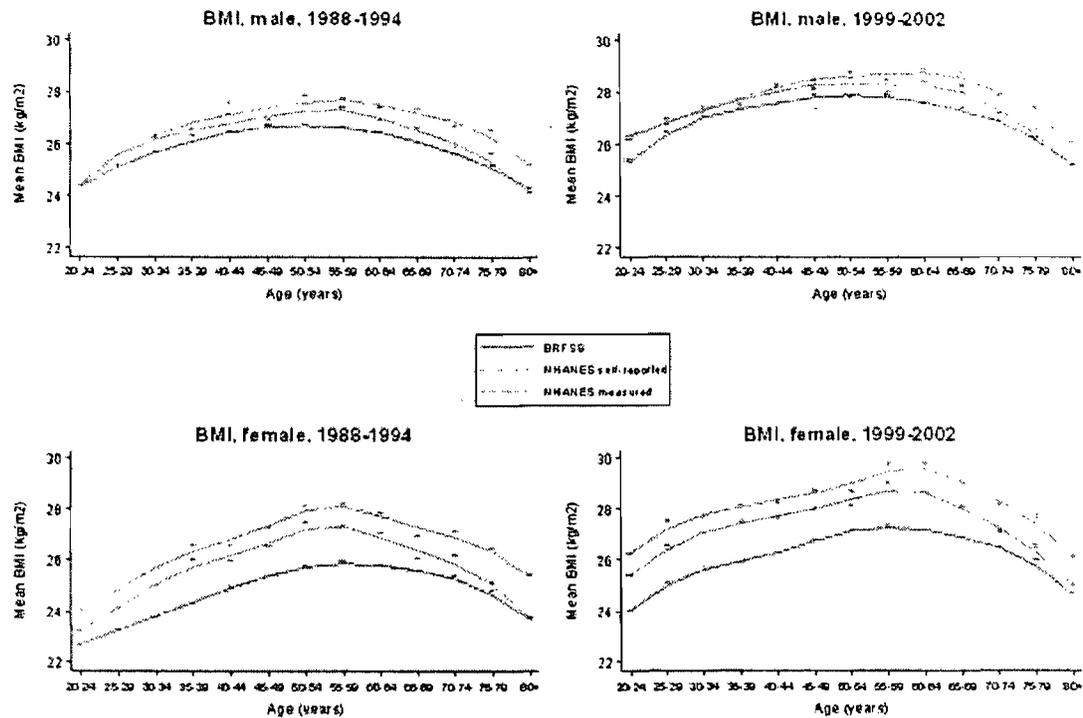


Fig. 5.2: Distribution des valeurs d'indice de masse corporelle moyennes en fonction de l'âge et du sexe selon deux études américaines (NHANES et BRFSS) et deux périodes de temps (1988-1994 et 1999-2002). [67]

5.1.1 Risques associés à l'obésité

L'obésité est associée à un risque plus élevé de morbidité et mortalité en comparaison des individus de poids normal [64, 65]. L'obésité est, en effet, un facteur de risque reconnu pour plusieurs maladies telles [66] :

- Diabète de type II
- Thromboses veineuses
- Accident vasculaire cérébral
- Arythmie cardiaque
- Certains cancers
- Hypertension
- Apnée du sommeil

5.2 Facteurs de risque de l'obésité

Tout comme l'hypertension, la prévalence d'obésité est plus forte en Amérique du Nord. La localisation géographique n'est toutefois pas le seul facteur influençant la prévalence d'obésité dans une population. Une multitude d'autres facteurs sont impliqués et expliqués ci-dessous.

5.2.1 Apport énergétique

L'augmentation de l'apport calorique quotidien explique en partie, l'augmentation continue des taux d'obésité dans le monde. Un rapport de l'OMS [68] estimait que l'apport énergétique per capita mondial était de 2300 kcal en 1963, 2440 kcal en 1971, 2720 kcal en 1992 et serait de 2900 kcal en 2010.

5.2.2 Sédentarisation

L'augmentation de la sédentarisation dans les pays développés a été proposée comme une des principales raisons de l'augmentation de la prévalence d'obésité, surtout en Amérique du Nord. La mécanisation de plusieurs types d'emplois et les habitudes de transport y sont assurément pour quelque chose. Par exemple, aux Pays-Bas, 30% des courts déplacements sont effectués en bicyclette et 18% à pied. En Angleterre, ces pourcentages sont de 8% en vélo et 9% à pied. Aux États-Unis, ils sont de 1% et 9% respectivement [69].

5.2.3 Ethnicité

L'ethnicité est souvent associée à une localisation géographique et à une culture particulière. Même en excluant l'influence de ces facteurs, il n'en demeure pas moins que la prévalence d'obésité varie d'une ethnie à l'autre. Aux États-Unis par exemple, des études ont montré que la prévalence d'obésité était plus forte chez les personnes de race noire que de race blanche et encore plus élevée chez les hispaniques. [56, 70]

5.2.4 Prédispositions génétiques

Tout comme pour l'hypertension, l'obésité existe sous deux formes : monogénique et polygénique, la seconde forme étant la plus courante. Des études d'association et de liaison à large échelle ont suggéré l'implication de plusieurs gènes dans la pathogénèse de l'obésité, dont celui de la leptine. Ce gène, identifié comme le «gène de l'obésité», est impliqué dans le contrôle de l'appétit [71]. Pausova *et al.* ont quant à eux montré l'effet du gène «tumor necrosis factor- α » sur la distribution des gras et l'obésité [72]. L'obésité n'est toutefois expliquée qu'en partie par les facteurs génétiques et leur grande diversité rend la tâche d'identification très ardue [73, 74, 75].

5.2.5 Autres facteurs

Bien qu'ils soient plus difficiles à quantifier, plusieurs autres facteurs ont une influence sur la prévalence d'obésité. Le comportement, l'éducation, le revenu familial, la taille des portions alimentaires, le manque d'exercice physique, etc. sont autant de facteurs qui peuvent mener à l'obésité [76, 77]

6. Héritabilité des phénotypes

Dans le cadre de cette étude, nous étudierons la transmission de traits phénotypiques à travers les descendances. La pression sanguine et l'indice de masse corporelle sont, comme mentionné plus haut, en partie influencés par des facteurs génétiques. Or, comme les membres d'une descendance partagent, par définition, des traits génétiques communs, les estimations d'héritabilité permettront d'estimer l'influence de la génétique des ancêtres sur le phénotype de leurs descendants.

L'héritabilité est définie comme : «la part de la variance d'un phénotype qui est expliquée par le génotype.» [78]. Formellement, étant donné un modèle statistique de la forme :

$$Var(P) = Var(G) + Var(E) + 2 \cdot Cov(G, E) \quad (6.1)$$

Où P représente le phénotype, G le génotype et E l'environnement. La covariance du génotype et de l'environnement est souvent fixée à 0 pour faciliter les calculs. L'héritabilité est définie comme :

$$H^2 = \frac{Var(G)}{Var(P)} \quad (6.2)$$

H^2 représente l'héritabilité au sens large ; elle reflète toutes les contributions possibles de la génétique à la variance phénotypique d'une population. Elles incluent entre autre : la présence d'allèles multiples, la dominance/récessivité et l'épistasie (interactions gènes-gènes). Différentes combinaisons d'allèles peuvent en effet influencer l'expression du phénotype. On parlera alors d'effet additif et on dira que les allèles contribuent à la variance génétique additive (A). De plus, un gène dominant peut aussi inhiber l'expression d'un gène récessif. De tels gènes contribueront à la variance génétique de

dominance (D). Finalement, l'épistasie contribue quant à elle à la variance génétique d'interaction (I). Lorsque ces paramètres sont considérés séparément, la variance génétique est partitionnée et le modèle statistique devient :

$$Var(P) = \underbrace{Var(A) + Var(D) + Var(I)}_{Var(G)} + Var(E) + 2 \cdot Cov(G, E) \quad (6.3)$$

Dans ce contexte, on peut définir l'héritabilité au sens restreint comme :

$$h^2 = \frac{Var(A)}{Var(P)} \quad (6.4)$$

h^2 représente la proportion de la variance phénotypique qui est due à l'effet additif des différents allèles [79].

6.1 Études d'héritabilité

On distinguera différents types d'études selon les candidats qui sont recrutés. Chaque modèle comporte ses avantages et ses inconvénients et les méthodes d'estimation varieront en fonction du type d'étude.

6.1.1 Étude de jumeaux

Les études de jumeaux classiques impliquent l'identification des paires de jumeaux monozygotiques (MZ, provenant d'un même ovule) et dizygotiques (DZ, provenant de deux ovules différents). En s'assurant que les paires jumeaux MZ et DZ partagent des facteurs environnementaux semblables, on peut estimer l'héritabilité d'un trait. Comme les jumeaux MZ sont deux fois plus semblables, d'un point de vue génétique, que les jumeaux DZ, il est possible d'approximer l'héritabilité comme la différence de corrélation du trait entre jumeaux MZ et DZ [78].

6.1.2 Étude d'adoption et de jumeaux séparés

Les études d'adoption visent à séparer les effets des gènes et de l'environnement en comparant des individus ne partageant aucun trait génétique, mais un environnement commun. Inversement, les études de jumeaux séparés visent à comparer des individus partageant un patrimoine génétique commun, mais des environnements différents.

6.1.3 Études familiales

Les études familiales peuvent impliquer plusieurs familles nucléaires, étendues ou encore des pedigree larges et complexes. Dans ce contexte, les méthodes d'estimation d'héritabilité doivent prendre en compte le fait que les données ne sont pas indépendantes. Des méthodes de régression seront souvent utilisées pour modéliser l'effet des gènes et de l'environnement des équations 6.1 et 6.3.

6.2 Héritabilité de la pression sanguine

De multiples études ont suggéré que les pressions systolique et diastolique sont génétiquement héritées dans diverses proportions [80, 81, 82, 83, 84]. Les estimations d'héritabilité varient de 22% à 62% pour la pression systolique et de 38% à 63% pour la pression diastolique.

Mongeau *et al.* ont suggéré, après une analyse des pressions sanguines de 756 enfants adoptés, 445 enfants naturels et 1176 parents, que les facteurs génétiques partagés expliquaient respectivement 61% et 58% des corrélations de pression sanguine systolique et diastolique observées entre les parents et les enfants [85].

Une étude plus récente, faite aux Pays-Bas portant sur 230 jumeaux monozygotiques, 305 jumeaux dizygotiques et 257 singletons provenant de 339 familles suggère que l'héritabilité des pressions ambulatoires systolique et diastolique se situe entre

44% et 63%. Plusieurs modèles statistiques ont été testés sur différents sous-groupes hypertendus et normotendus, d'où la plage d'estimations. Le modèle le plus plausible comprenait des facteurs génétiques additifs (héritabilité au sens restreint, voir équations 6.3 et 6.4) et des facteurs environnementaux uniques aux individus en laissant de côté les facteurs environnementaux partagés entre membres d'une même famille [86]. Une autre étude, portant aussi sur les pressions systolique et diastolique ambulatoires, chez des familles Suédoises cette fois, a plutôt estimé leur héritabilité à 30% et 29% respectivement [87].

Hottenga *et al.* ont analysé les données de quatre études de jumeaux aux Pays-Bas et ont estimé que l'héritabilité des pressions systolique et diastolique étaient de 48% à 60% et de 34% à 67% respectivement. Le modèle retenu, celui modélisant le mieux les données des quatre études ne comprenait aussi que des facteurs génétiques additifs et des facteurs environnementaux uniques aux individus [88].

En août 2006, une étude portant sur 6 148 Sardes, de pedigrees connus, a plutôt suggéré que l'héritabilité au sens restreint se chiffrerait à 16% pour la pression systolique et 12% pour la pression diastolique. Leurs mesures d'héritabilité au sens large (voir équations 6.1 et 6.2) sont plus semblables à celles mentionnées plus haut ; 65% et 45% pour les pressions systolique et diastolique. [89]

Finalement, les auteurs d'une étude portant sur 314 habitants des Îles Seychelles ont quant à eux estimé que l'héritabilité au sens restreint des pressions systolique et diastolique étaient de 37% et 24% respectivement. Leur modèle ne tenait compte que des facteurs génétiques additifs et des facteurs environnementaux [90].

Il semble donc y avoir un consensus dans la communauté scientifique sur le fait que la pression sanguine est un trait héritable. On remarque toutefois que les estima-

tions d'héritabilité de la pression sanguine sont très variables en fonction des études et des populations. Il semble aussi y avoir une certaine confusion entre héritabilité au sens large et restreint ; l'héritabilité calculée ne semble pas toujours correspondre à la méthodologie présentée. La prudence est donc de mise lors de l'interprétation de ces résultats.

6.3 Héritabilité de l'indice de masse corporelle

Plusieurs études suggèrent une influence non négligeable des traits génétiques sur l'obésité et plus particulièrement sur l'indice de masse corporelle. On estime que l'héritabilité de l'indice de masse corporelle se situe entre 25% et 73% [91, 92, 93, 94, 95]. Des études plus récentes indiquent toutefois que l'ethnicité semble influencer fortement l'héritabilité de ce trait.

Bouchard *et al.* ont montré, lors d'une étude portant sur 12 paires de jumeaux monozygotiques mâles sur-nourris durant 84 jours, que les variations de poids et la répartition des gras étaient influencées par des facteurs génétiques. Les variations de poids et de répartition des gras étaient en effet beaucoup plus marquées entre les paires de jumeaux qu'à l'intérieur de celles-ci, suggérant une forte influence des facteurs génétiques sur ces traits [96].

Li *et al.* [97], lors d'une étude sur 478 sujets provenant de 105 familles nucléaires Mexicano- américaine, ont suggéré une héritabilité au sens large de 59% pour l'indice de masse corporelle. Ils ont inclus dans leur modèle, l'âge, le sexe et « toute variable associée avec une p-value inférieure à 0.05 ». Ces autres variables n'étaient toutefois pas spécifiées.

Lors d'une récente étude portant sur 1277 Arabes provenant de cinq familles étendues, hautement cosanguines, Bayoumi *et al.* [98] ont calculé que l'héritabilité au sens

large de l'indice de masse corporelle s'élevait à 68%. Ils n'ont considéré que l'âge et le sexe comme facteurs environnementaux dans leur modèle.

Une autre étude portant sur 1030 enfants Hispaniques de 4 à 19 ans provenant de 319 familles comportant au moins un individu atteint de surpoids a suggéré une héritabilité, au sens large, de l'IMC de l'ordre de 39%. Les covariables environnementales considérées ne sont encore ici que l'âge et le sexe [99].

Enfin, une étude de Luke *et al.* [100] a permis de mettre en relief les différences inter ethniques de l'héritabilité de l'IMC. Ils ont en effet étudié conjointement 1868 familles du Nigéria, 623 de la Jamaïque et 2132 familles de race noire de Chicago aux États-Unis. L'héritabilité au sens large a été calculée chez les populations entières et chez les adultes (plus de 19 ans) seulement. L'héritabilité de l'IMC était de 49% et 48% dans la population nigérienne et les adultes respectivement. Chez les Jamaïquains, ces taux s'élevaient à 53% et 58% respectivement alors qu'aux États-Unis ils étaient de 57% et 53%.

Tout comme pour la pression sanguine, il semble donc y avoir un consensus dans la communauté scientifique à savoir que l'IMC est un trait héritable. Bien que l'héritabilité varie d'un article à l'autre, les études les plus récentes font état de taux oscillant entre 39% et 68%. Encore une fois, il est impératif de considérer ces résultats avec précautions, puisque les estimations varient selon les modèles, les populations et les logiciels utilisés.

7. Obésité et hypertension

7.1 Association des phénotypes

Dans le cadre de cette étude, l'obésité et l'hypertension seront à la fois considérée comme facteurs de risque indépendants et en association. En effet, bien que l'obésité et l'hypertension prédisposent indépendamment à diverses maladies cardio-vasculaire, ces deux conditions sont aussi fréquemment associées. Plusieurs études ont en effet montré une corrélation entre la masse corporelle et la pression sanguine [43, 101, 102, 103].

La *National Heart Association of Australia Risk Factor Prevalence Study* a d'ailleurs permis d'estimer que l'obésité était la cause première d'hypertension chez un peu plus de 30% des individus [101]. Une autre étude, portant sur 1 millions d'Américains a montré que que les chances d'être hypertendu était significativement plus élevées chez les obèses en comparaison aux non-obèses. Ils ont calculé que l'OR d'être hypertendu était de 2,42 chez les sujets âgés de 20 à 39 ans et de 1,54 chez les 40 à 64 ans [104].

Les raisons expliquant cette association sont encore méconnues. Certaines hypothèses de mécanismes physiologiques ont été avancées incluant une plus grande résistance à l'insuline, une rétention accrue du sodium et le syndrome métabolique. Des recherches sont en cours à savoir si la résistance à l'insuline est impliquée dans la pathogénèse de l'hypertension. Le syndrome métabolique (parfois appelé syndrome X) est quant lui défini par la présence d'une combinaison de facteurs de risque dont font partie une pression sanguine élevée et le surpoids/obésité. Le syndrome métabolique est asso-

cié à une forte augmentation du risque de maladies cardio-vasculaires. [105, 106, 107].

7.2 Corrélation entre pression sanguine et indice de masse corporelle

Non seulement associés, des études d'héritabilité ont permis d'estimer les corrélations génétiques et environnementales entre l'obésité et l'hypertension. Selon le Grand Dictionnaire Terminologique de l'Office québécois de la langue française, la corrélation se définit comme une : «Liaison entre des données quantitatives ou qualitatives, mise en évidence par une analyse statistique». Le *coefficient de corrélation de Pearson* sera souvent utilisé à cette fin dans le cadre d'études médicales. Il est formellement défini comme :

$$\rho_{X,Y} = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y} \quad (7.1)$$

où X sera le premier trait, Y le second, σ représentant l'écart-type et *cov* la covariance. Les valeurs de corrélation, en valeur absolue, s'étendent sur l'intervalle [0,1]; 0 indiquant l'absence de corrélation et 1 une corrélation parfaite.

Encore une fois les estimations varient d'une étude à l'autre. Cui *et al.*, dans une étude portant sur 767 adultes, ont estimé que la corrélation entre la pression systolique et l'IMC s'élevait à 0,36 et qu'elle était de 0,41 entre la pression diastolique et l'IMC [108]. Une autre étude de Choh *et al.*, sur 1342 Samoa adultes a quant à elle estimé une corrélation de 0,585 entre la pression systolique et l'IMC et une corrélation de 0,503 entre pression diastolique et IMC [109].

De manière plus qualitative, une étude de Pausova *et al.* sur la population du SLSJ a montré que les fratries hypertendues sont significativement plus obèses ($p < 0.001$) que les fratries normotendues, suggérant aussi une forte corrélation entre pression sanguine et IMC élevés [110].

Deuxième partie

Méthodologie et Expérimentations

8. Méthodologie

Dans ce chapitre, nous expliquerons le cadre théorique de notre étude, les données à notre disposition et les applications qui ont été développées. Dans un premier temps, nous présenterons les procédures de recrutement des familles, de phénotypage et génotypage et nous couvrirons ensuite les aspects théoriques du traitement de la généalogie et des descendances. Nous présenterons, dans le chapitre suivant, les différents aspects de la base de données médicales, les dispositions éthiques mises en place et les applications développées.

8.1 Nécessité d'une nouvelle approche

Sans nier la pertinence des études d'historique familial, nous croyons qu'elles présentent quelques limites lorsqu'il est question d'étudier la population du SLSJ. Les données généalogiques du projet BALSAC-RETRO ont révélé la présence d'un effet fondateur [5, 6]; la vaste majorité des individus partagent donc des ancêtres communs. Dans ce contexte, le concept d'historique familial peut être amélioré. En effet, plusieurs études de ce genre considèrent seulement les parents de premier voire de second degré (Voir section 3). Dans le cadre d'une étude où la population est suffisamment grande et diversifiée, on peut assumer que les familles sont pratiquement indépendantes les unes des autres sans introduire de distorsion statistique. Dans le contexte du SLSJ toutefois, il n'est pas rare que des individus issus de familles nucléaires distinctes partagent en fait un ancêtre commun à aussi peu que 2 à 3 générations plus anciennes. À titre indicatif, Braekeleer et Gauthier ont montré que le coefficient de parentalité (*kinship*) moyen était 5.57 fois plus élevé dans la population du SLSJ que dans un groupe contrôle et

que 45.5% des individus étaient cousins au troisième degré [111]. Dans un tel contexte, nous ne pouvons clairement pas traiter les familles comme des entités indépendantes.

L'étude proposée ici se concentre sur les descendance plutôt que sur les familles. Ce faisant, nous sommes en mesure d'exploiter toute l'information généalogique à notre disposition en plus d'éliminer certaines ambiguïtés liées à l'interdépendance des familles. Bien que des distorsions subsistent, l'approche proposée constitue un complément intéressant aux études familiales classiques lorsque de vastes généalogies sont disponibles.

8.2 Recrutement des familles

Toutes les familles retenues pour l'étude étaient originaires du Saguenay-Lac-Saint-Jean et possédaient au moins une paire de germains atteints d'hypertension et de dyslipidémie. Plus précisément, les critères d'inclusion des germains affectés étaient : être atteint d'hypertension essentielle (pression systolique > 140 mmHg et/ou pression diastolique > 90 mmHg à deux occasions ou prise de médicaments anti-hypertensifs), de dyslipidémie, (cholestérol plasmatique ≥ 5.2 mmol/L et/ou cholestérol HDL ≤ 0.9 mmol/L ou prise de médicaments hypolipidémiant), avoir un IMC < 35 kg/m², être âgé entre 18 et 55 ans et être d'origine canadienne-française. Une fois les paires de germains sélectionnées, tous leurs parents de premier et second degré âgés de plus de 18 ans étaient invités à participer à l'étude, indépendamment de leur état de santé. Les structures familiales ont été validées en calculant les probabilités de parenté selon les données génotypiques. Suite aux analyses de paternité, deux sujets ont été retirés de l'étude [112].

8.3 Phénotypage

Toutes les mesures de phénotypes ont été prises par du personnel qualifié qui a suivi les procédures standards [113, 80, 114, 72]. Le phénotypage restreint au jour 0 incluait des prélèvements sanguins à des fins d'extraction d'ADN et de mesure des concentrations de lipides plasmatiques, de glycémie à jeun, d'acide urique, de pression sanguine et de bio-impédance pour déterminer la masse de gras corporel. Les données anthropométriques recueillies incluaient 3 mesures globales d'obésité (IMC et pourcentage de gras corporel dérivés des plis cutanés et de la bio-impédance) et 11 mesures locales : 6 circonférences et 5 plis cutanés. Les sujets sans contre-indication étaient invités le jour suivant pour un phénotypage étendu. Leurs médicaments anti-hypertensifs leurs étaient alors retirés pour une semaine et leurs médicaments hypolipidémiants retirés pour un mois. Au total, 294 individus normotendus et hypertendus ont été retenus pour un phénotypage étendu.

Au jour 1 du phénotypage étendu, un cathéter intraveineux était inséré. Des échantillons à jeun ont été pris pour mesurer les concentrations de lipides, glucose et insuline. Les sujets étaient ensuite soumis à un test postural : leurs fonctions cardiaques (par pléthysmographie d'impédance) et leur pression sanguine étaient enregistrés durant 60 min en position allongée suivi de 10 min debout et 20 min assis. L'impédance et la pression sanguine ont aussi été mesurées après un stress mental induit par des tests mathématiques [115]. À la fin du jour 1, le monitoring de la pression sanguine pour une durée de 24h était initié.

Au jour 2, la pression sanguine, le débit sanguin rénal et le taux de filtration glomérulaire ont été mesurés avant et après des infusions successives de norépinéphrine de 0.01 à 0.05 $\mu\text{g}/\text{kg}/\text{min}$ [80]. Les sujets ont ensuite reçu une charge aqueuse intraveineuse de 15ml/kg en plus de leur soluté habituel sur une période de 6h. Les concentrations d'arginine-vasopressine, de facteur atrial natriurétique, de cGMP et cAMP et l'activité

de la rénine ont été quantifiés avant et après la surcharge hydrique [112].

8.4 Génotypage

Les échantillons d'ADN des 500 premiers sujets recrutés, originaires de 97 familles, ont été génotypés au Broad Institute, par l'utilisation de 377 marqueurs microsatellites provenant d'une version modifiée du *Cooperative Human Linkage Center Screening Set* (version 6.0) qui incluait des marqueurs Genethon et assurait une couverture moyenne de 9.1 cM de tout le génome humain. Les détails du système de génotypage sont décrits ailleurs [116]. Des marqueurs additionnels (Genethon) à une densité de 5 cM ont été utilisés pour les chromosomes 1q et 3q chez un total de 810 sujets. Des analyses à *posteriori* ont estimé un taux d'erreur de génotypage d'environ 0.016% [112].

8.5 Traitement de la généalogie

Le pedigree du SLSJ est stocké dans la base de données BALSAC-RETRO [8, 9]. Pour les besoins de notre étude (traitement des descendances, calculs d'identité par descendance), la généalogie a été encodée sous forme de graphe dirigé. La théorie des graphes est couramment utilisée pour le traitement de généalogies [117, 118] et dans plusieurs autres domaines de la biologie et de la bio-informatique [119, 120, 121, 122]. Dans le cas présent, chaque noeud représente une personne et les parents sont reliés aux enfants par un arc du parent vers l'enfant. Voir la figure 8.2 pour un exemple graphique. Les noeuds ayant un degré entrant égal à zéro, les sources, sont considérés comme étant des fondateurs alors que ceux ayant un degré sortant égal à zéro, les puits, sont généralement les individus les plus récents. Les individus phénotypés et/ou génotypés sont situés près des puits. Le terme fondateur doit toutefois être nuancé puisque les sources du graphe ne correspondent pas nécessairement aux fondateurs de Charlevoix, arrivés à la fin du 17^e siècle. Certaines données manquantes ou l'absence de connaissances sur la généalogie des immigrants font en sorte que des sources sont parfois rencontrées à aussi

peu que trois générations des individus les plus récents. Le traitement de données manquantes ou erronée constitue un domaine de recherche en soi et nous présenterons, à la section 8.5.1, les mesures qui ont été appliquées dans ces situations. De plus, bien que les puits soient généralement des individus d'aujourd'hui, des personnes sans enfants répertoriés sont rencontrés à différentes générations moins récentes. Ces observations suggèrent une possible distorsion numérique dans l'application et l'implantation des équations conventionnelles de contribution génétique des fondateurs due aux données généalogiques disponibles.

Une autre dimension à considérer est la «forme» de la généalogie du SLSJ. On s'imagine souvent une généalogie sous forme d'arbre, d'où le terme «arbre généalogique». Cependant, cette idée est définitivement erronée dans le cas qui nous concerne. Un arbre, au sens de la théorie des graphes, est un graphe *non dirigé*, *acyclique* et *connexe* ce qui implique qu'il ne doit exister qu'un seul chemin entre n'importe quelle paire de noeuds. Dans le cas qui nous concerne, les chemins généalogiques se séparent à chaque génération, mais se fusionnent souvent quelques générations plus tard. Il en résulte alors des situations contre-intuitives. Par exemple, deux ancêtres, situés à exactement n générations d'un individu peuvent être parent et enfant et il arrive qu'un ancêtre puisse faire partie de deux générations différentes à la fois. L'idée même de génération devient alors ambiguë. Cette situation est illustrée à la figure 8.1. La généalogie n'est donc pas un arbre, mais bien un graphe *dirigé acyclique* (DAG). Dirigé car les arrêtes sont orientées du parent vers l'enfant et acyclique car il ne peut exister une arrête partant d'un enfant vers le parent. Nous utiliserons donc une approche basée sur la théorie des graphes qui nous permettra d'effectuer une étude rigoureuse de la transmission de traits ancestraux à travers les descendance tout en tenant compte des distorsions soulevées précédemment.

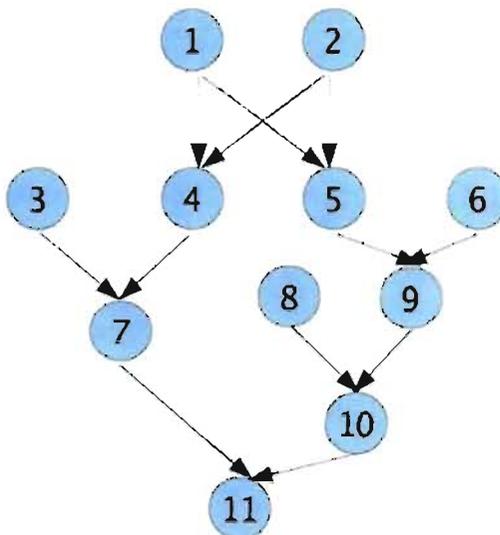


Fig. 8.1: Pedigree familial fictif. Les ancêtres 1 et 5 sont situés à trois générations de l'individu 11 et sont donc ses arrière-grands-parents. Toutefois, ils sont aussi parent et enfant. Il existe deux chemins de 1 vers 11 soient (1,4,7,11) et (1,5,9,10,11). Non seulement 1 et 5 sont-ils parent et enfant, mais 1 appartient à la fois à la troisième et à la quatrième génération en partant de 11.

8.5.1 Couches généalogiques

Le concept de couches généalogiques a été défini dans cette optique [112]. L'idée sous-jacente est de comparer les ancêtres à une distance commune d'un ensemble S d'individus phénotypés. Les ancêtres partageant une distance commune constituent ce que nous appelons une couche généalogique. Étant donné un ensemble de descendants, la première couche sera constituée de leurs parents, la deuxième de leurs grands-parents et ainsi de suite jusqu'au tout début de la colonie. Deux ancêtres v_i et v_j feront partie de la même couche généalogique C_k s'ils sont situés à la même distance k d'un ensemble de descendants S :

$$v_i, v_j \in C_k \leftrightarrow \text{distance}(v_i) = \text{distance}(v_j) = k \quad (8.1)$$

Les couches sont donc dépendantes de la définition de distance entre ancêtres et descendants. Plusieurs définitions de distances ont été développées, analysées et discu-

tées dans les travaux de Merlo *et al.* [123]. La définition utilisée dans le cadre de cette étude est celle de *couches contraintes de plus courte distance*. La couche C_k comprendra les ancêtres dont le plus court chemin les reliant à n'importe quel descendant de l'ensemble S est de longueur k en plus des ancêtres sans parent répertorié (sources) dont la longueur du plus court chemin est d'au plus k . Le choix d'inclure ces sources dans les couches subséquentes n'est pas gratuit. En effet, tous les individus d'aujourd'hui n'ont pas nécessairement d'ancêtres connus à tous les niveaux généalogiques. Les sources propagées permettent donc d'assurer que tous les descendants possèdent au moins un ancêtre dans chaque couche généalogique. La propriété *contrainte* nous assure de plus qu'aucun ancêtre n'est descendant d'un autre ancêtre de la même couche. Les probabilités de transmission génétique des ancêtres d'une même couche sont donc indépendantes. La figure 8.2 présente un exemple détaillé de la division en couches ; elles y sont représentées par différentes couleurs.

8.5.2 Contribution génétique des ancêtres

Dans l'optique de caractériser la transmission de traits ancestraux, étudions la contribution génétique d'ancêtres appartenant à la même couche généalogique. La définition de contribution génétique pour un ensemble de descendants S telle que fournie dans [2] utilise la notion de chemins entre un descendant et un ancêtre. Les équations 8.2 et 8.3 définissent respectivement un chemin dans le DAG généalogique G avec V_G l'ensemble des noeuds et E_G l'ensemble des arrêtes. Alors, la contribution génétique d'un ancêtre v pour un ensemble de descendants S se traduit par :

$$\begin{aligned} \text{Chemins}(v_i, v_j) = \{c = (w_0, w_1, \dots, w_n) \mid (\forall k \in [0, n], w_k \in V_G) \wedge (w_0 = v_i) \wedge (w_n = v_j) \\ \wedge (\forall l \in [1, n], (w_{l-1}, w_l) \in E_G)\} \end{aligned} \quad (8.2)$$

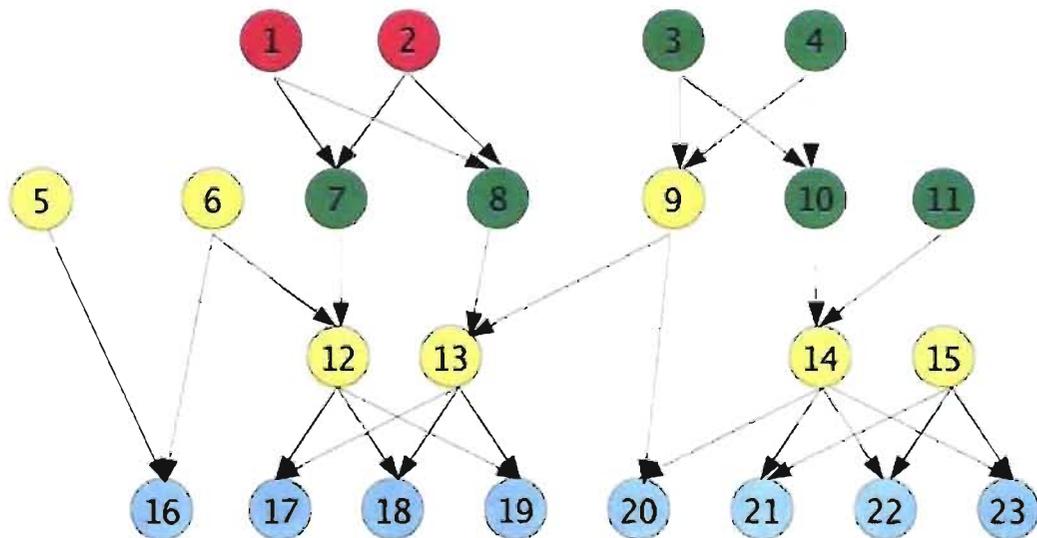


Fig. 8.2: Généalogie fictive encodée sous forme de graphe généalogique. Chaque parent possède un arc vers chacun de ses enfants. Les individus 1,2,3,4,5,6 et 11 sont les sources du graphe et sont considérés comme des fondateurs. Les individus 16 à 23 constituent les puits du graphe et sont les individus les plus récents de la généalogie. Les individus en bleu représentent l'ensemble de descendants à partir duquel les couches généalogiques seront définies. Les personnes en jaune, vert et rouge font respectivement partie de la première, deuxième et troisième couche généalogique. Il existe deux chemins généalogiques reliant l'individu 9 à l'ensemble de descendants (9,20 et 9,13,19). Comme les couches sont définies selon la distance minimale entre l'ancêtre et l'ensemble de descendants, 9 fait partie de la première couche généalogique.

$\text{Chemins}(v_i, v_j)$ est l'ensemble de toutes les séquences finies d'arrêtes du graphe G qui commence en v_i et se termine en v_j .

$$\text{Contribution génétique}(v, S) = \sum_{i=1}^d \sum_{j=1}^c \left(\frac{1}{2}\right)^{|c|} \quad (8.3)$$

$\text{Contribution génétique}(v, S)$ est la contribution génétique de l'ancêtre v à l'ensemble S où $d =$ le nombre de descendants de l'ensemble S qui sont apparentés à v ; c le nombre de chemins, tels que définis à l'équation 8.2 entre l'ancêtre et le descendant; $|c| =$ la longueur du chemin.

La probabilité qu'un individu quelconque d'un ensemble de descendant S partage un gène identique à celui d'un ancêtre v , appartenant à la couche généalogique C_k , peut être calculée à partir de la contribution génétique pour autant que les probabilités de transmission des ancêtres soient indépendantes. Cette probabilité est appelée probabilité d'identité par descendance (pIPD) entre un ancêtre et un ensemble de descendants et est définie comme suit :

$$IPD(v, S, C_k) = \frac{\text{Cont. génétique}(v, S)}{\sum_{w \in C_k} \text{Cont. génétique}(w, S)} \quad (8.4)$$

La probabilité d'identité par descendance d'un ancêtre v appartenant à la couche C_k peut aussi être calculée pour deux classes de descendants S_A et S_B dérivées de S :

$$IPD(v, S_A, S_B, C_k) = \{(x, y) \mid (x = IPD(v, S_A, C_k)) \wedge (y = IPD(v, S_B, C_k))\} \quad (8.5)$$

Spécificité des ancêtres

Étant donné deux classes de descendants S_A et S_B , tous les fondateurs d'une même couche généalogique n'y contribuent pas également. Certains ancêtres ne contribueront qu'à S_A , d'autres qu'à S_B et la majorité contribuera aux deux classes selon différentes proportions. Afin de mieux illustrer notre propos, reportons les probabilités d'IPD des

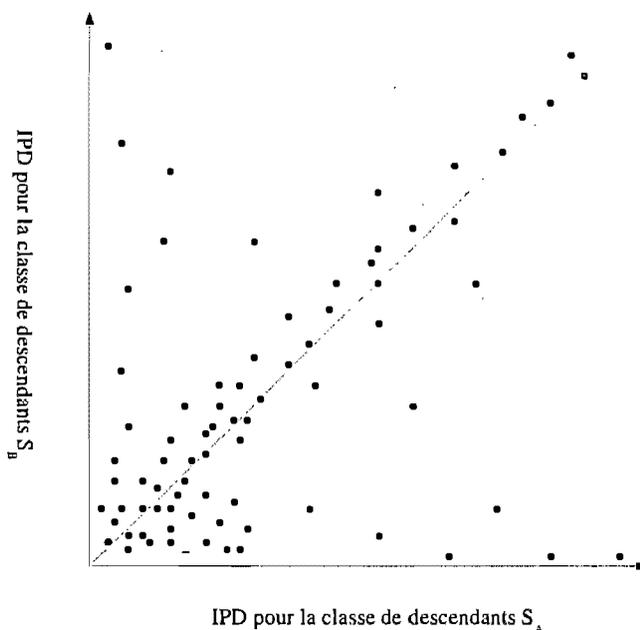


Fig. 8.3: Graphique d'identité par descendance des ancêtres pour deux classes de descendants. L'axe des X représente la probabilité d'identité par descendance d'un ancêtre pour la classe de descendants S_A et l'axe des Y correspond à la probabilité d'identité par descendance d'un ancêtre pour la classe de descendants S_B . Les fondateurs contribuant également aux deux classes se situent sur la diagonale passant par l'origine.

ancêtres sur un graphique où l'abscisse représente la probabilité d'IPD pour la classe S_A et l'ordonnée celle pour la classe S_B . La figure 8.3 montre un exemple typique de la distribution des probabilités d'IPD des ancêtres dans la population du SLSJ. Les fondateurs ne contribuant qu'à une seule classe se situent sur un des axes alors que ceux contribuant également aux deux classes sont plutôt situés sur la diagonale passant par l'origine. Afin de caractériser la contribution des fondateurs aux deux classes, nous avons utilisé le concept de *spécificité des ancêtres* [112] :

$$\text{Spécificité}(v, S_A, S_B, C_k) = \begin{cases} 1 - \frac{4}{\pi} \cdot \arctan\left(\frac{IPD(v, S_A, C_k)}{IPD(v, S_B, C_k)}\right) & \text{si } IPD(v, S_B, C_k) \neq 0 \\ 1 & \text{sinon} \end{cases}$$

La spécificité est une mesure indiquant la probabilité qu'un ancêtre partage davan-

tage de gènes avec une classe de descendants plutôt que l'autre. Elle sera égale à 1 si l'ancêtre partage uniquement ses gènes avec la classe S_A , 0 s'il contribue également aux deux classes et -1 s'il ne partage ses gènes qu'avec la classe S_B . Notez qu'il est impossible que $IPD(v, S_A, C_k)$ et $IPD(v, S_B, C_k)$ soient toutes deux égales à zéro par construction des couches généalogiques [112]. Le degré de spécificité n'est pas lié à l'ampleur des probabilités d'identité par descendance puisqu'un ancêtre peut être très spécifique à une classe et n'y contribuer que très peu. Inversement, d'autres ancêtres peuvent être de grands contributeurs aux deux classes et seront très peu spécifiques. La spécificité permet de mettre en relief la contribution différentielle des ancêtres pour deux classes de descendants.

8.6 Conception de l'étude

Jusqu'à maintenant, nous avons présenté une méthode de traitement de la généalogie où la spécificité génétique doit être interprétée en fonction de classes génériques de descendants S_A et S_B . C'est toutefois à partir du moment où ces classes sont définies selon un critère biologique pertinent que l'analyse de la spécificité génétique des ancêtres prend tout son sens.

8.6.1 Définition des variables

Au cours de cette étude, nous utiliserons différents termes, parfois semblables, pour qualifier les variables investiguées. Afin d'éviter toute confusion, nous proposons une brève définition des termes qui seront utilisés.

- Classe phénotype élevé : Utilisée afin de calculer la spécificité génétique, correspond au terme S_A dans l'équation 8.5.2
- Classe phénotype bas : Utilisée afin de calculer la spécificité génétique, correspond au terme S_B dans l'équation 8.5.2

- Ensemble phénotype bas : Désigne les individus descendant uniquement d'ancêtres spécifiques à la classe phénotype bas
- Ensemble phénotype moyen : Désigne les individus descendant à la fois d'ancêtres spécifiques à la classe phénotype bas et à la classe phénotype élevé
- Ensemble phénotype élevé : Désigne les individus descendant uniquement d'ancêtres spécifiques à la classe phénotype élevé

Il est donc crucial de noter que l'appartenance d'un individu à l'un ou l'autre des trois ensembles phénotypique (bas, moyen ou élevé) n'est influencée que par la spécificité de ses ancêtres et non par sa classe phénotypique. Un individu de la classe phénotype bas peut donc appartenir à l'ensemble phénotype élevé et vice-versa.

8.6.2 Classification des descendants

Dans le cadre de cette étude, les classes S_A et S_B sont définies sur la base des mesures phénotypiques des descendants. Étant donné l'ensemble de descendants phénotypés S , S_A contiendra les individus de phénotype élevé et S_B les individus de phénotype bas. Deux critères de classification ont été développés et testés soient par seuil médical et par régression linéaire corrigée pour l'âge et le sexe.

La classification par seuil médical se résume en une procédure très simple. Étant donné le phénotype à investiguer, elle permet de déterminer au delà de quelle valeur le sujet est considéré à risque et de procéder à la classification des individus. S_A contiendra les descendants dont les mesures phénotypiques sont au-delà du seuil médical et S_B les autres. Ce procédé présente des avantages évidents au niveau de l'interprétation des résultats. On sera en mesure d'identifier les ancêtres ayant davantage contribué génétiquement aux descendants considérés médicalement à risque pour un phénotype donné. Cette approche présente par contre quelques limites. Bien qu'il existe parfois des seuils médicaux différents pour les hommes et les femmes, l'âge n'est généralement

pas pris en ligne de compte. La régression linéaire permettra de contrer ces problèmes.

Une régression linéaire des valeurs de phénotype en fonction de l'âge et du sexe des sujets produit un plan de régression. Les individus dont la mesure de phénotype se situe au-dessus du plan sont classifiés dans S_A et les autres dans S_B . Le logiciel de programmation statistique R [124] a été utilisé pour le calcul des régressions linéaires. Cette méthode permet donc d'identifier les ancêtres spécifiques aux descendants présentant des mesures phénotypiques plus élevées que celles des autres individus de même âge et sexe. L'avantage majeur de cette méthode est son impartialité. Alors qu'il existe souvent plusieurs seuils médicaux différents pour un même phénotype, le résultat de la régression linéaire est unique. Les analyses préliminaires ont montrés que les deux approches produisaient des résultats très similaires. La régression linéaire corrigée pour l'âge et le sexe fut adoptée pour la suite de l'étude. Pour plus de clarté, nous désignerons désormais les classes S_A et S_B comme étant les classes phénotype élevé et phénotype bas respectivement.

8.6.3 Interprétation de la spécificité génétique

La spécificité génétique nous permet donc de quantifier la différence de contribution des ancêtres pour des classes de descendants de phénotype élevé et bas. Un ancêtre de spécificité égale à -1 contribue uniquement à la classe phénotype bas alors qu'un autre de spécificité égale à 1 contribue uniquement à la classe phénotype élevé.

Considérons d'abord les ancêtres dont la spécificité tend vers 1. Ce sont ceux pour qui la différence de contribution en faveur des descendants de phénotype élevé est la plus accentuée. Il est donc probable que ces derniers aient transmis des gènes, des traits culturels ou environnementaux à leurs descendants qui auraient influencés leurs mesures phénotypiques. Inversement, les ancêtres dont la spécificité tend vers -1 ont

peut-être transmis des gènes protecteurs, de saines habitudes de vie ou quelconque autre facteur non mesurable favorisant un phénotype bas. Les descendances d'ancêtres situés aux deux extrêmes du spectre de spécificité devraient donc présenter des phénotypes très polarisés.

De manière plus générale, le taux de descendants présentant une mesure de phénotype élevée devrait être d'autant plus grand que la spécificité de leurs ancêtres tend vers 1. En conséquence, la spécificité des ancêtres devrait être garante du phénotype des descendants. Si tel est le cas, il serait possible d'estimer le «risque phénotypique» d'un individu à partir de sa généalogie, à la manière de l'approche «back to the future» présentée dans [112]. Cette hypothèse constitue le coeur de notre étude et c'est par une analyse comparative des descendances en fonction de la spécificité des ancêtres que nous l'avons vérifiée.

8.7 Étude comparative des descendances

Les descendances constituent l'unité de base de notre étude. La descendance d'un individu v est aisément transposée en théorie des graphes comme l'ensemble des noeuds pour lesquels il existe un chemin partant de v et se terminant au noeud en question. L'équation 8.2 définit formellement un chemin dans le graphe généalogique et la descendance d'un ancêtre A est définie comme :

$$Descendance(A) = \{d \in V_G \mid \exists p \in Chemins(A, d)\} \quad (8.6)$$

$Descendance(A)$ est l'ensemble de tous les noeuds v pour lesquels il existe au moins un chemin partant de l'ancêtre A et se terminant au descendant d .

Les descendances ont été générées à l'aide d'une recherche en profondeur sur chaque individu de la généalogie [125]. L'algorithme se résume à visiter le noeud ancestral et

tous ses descendants itérativement. Comme la généalogie est fixe, cette opération peut n'être exécutée qu'une seule fois si les résultats sont stockés en mémoire. Notez qu'une personne peut avoir plusieurs descendants et qu'un individu peut appartenir à plusieurs descendances. Toute personne du graphe généalogique, à l'exception des puits, est donc la source d'une descendance à laquelle est attribué un numéro arbitraire. La correspondance source \leftrightarrow numéro de descendance est stockée dans un fichier et les numéros de descendances auxquelles appartient un individu sont stockés séparément. Étant donné un graphe généalogique de n individus, cette procédure s'effectue en $O(n^2)$.

8.7.1 Procédure de comparaison

Une fois générées, les descendances sont comparées en fonction de la spécificité de l'ancêtre, une couche généalogique à la fois. Comme, selon notre hypothèse, la spécificité est garante du phénotype des descendants, les descendances d'ancêtres situés aux extrêmes du spectre de spécificité devraient présenter les phénotypes les plus polarisés. Logiquement, cette polarisation devrait aussi s'atténuer à mesure que nous comparons les descendances d'ancêtres moins spécifiques.

Étant donné une séquence d'ancêtres triée en ordre croissant de spécificité, les descendances sont comparées et cumulées en débutant par les ancêtres situés aux extrêmes du spectre de spécificité. Concrètement, les phénotypes des descendants de l'ancêtre le plus spécifique à la classe phénotype élevé sont d'abord comparés à ceux des descendants de l'ancêtre le plus spécifique à la classe phénotype bas. On ajoute ensuite les deuxièmes ancêtres les plus spécifiques et ainsi de suite jusqu'à ce que tous les ancêtres soient considérés. La figure 8.4 illustre l'ordre de comparaison des descendances.

Les descendances ne sont toutefois pas des unités indépendantes. En effet, plusieurs ancêtres d'une même couche généalogique partagent des descendants communs. Un

individu peut donc descendre à la fois d'ancêtres spécifiques à la classe phénotype élevé et phénotype bas (voir figure 8.4). Trois ensembles de descendants indépendants ont donc été construits :

1. Ensemble d'individus descendant uniquement d'ancêtres spécifiques à la classe phénotype élevé
2. Ensemble d'individus descendant à la fois d'ancêtres spécifiques à la classe phénotype élevé et à la classe phénotype bas
3. Ensemble d'individus descendant uniquement d'ancêtres spécifiques à la classe phénotype bas

Les individus de chaque nouvelle descendance sont donc classifiés dans l'un ou l'autre de ces ensembles. Pour plus de clarté, le premier ensemble sera désormais désigné comme l'ensemble phénotype élevé, le second comme l'ensemble phénotype moyen et le troisième comme l'ensemble phénotype bas. Il est important de souligner que les descendants ne sont pas classifiés *à priori*. Une personne peut donc passer de l'ensemble phénotype élevé à l'ensemble phénotype moyen ou de l'ensemble phénotype bas à l'ensemble phénotype moyen en cours d'analyse. Des analyses statistiques permettront ensuite d'attester de la significativité des différences phénotypiques entre les trois ensembles de descendants.

8.7.2 Analyse statistique

Le *t-test* permet d'estimer la probabilité que les moyennes de deux groupes de données soient significativement différentes. La généralisation du *t-test* à plusieurs ensembles de données est le test ANOVA (*Analysis of variance*). Contrairement à ce que son nom semble indiquer, le test ANOVA permet d'estimer la probabilité, sous forme de p-value, que les moyennes de tous les groupes de données soient significativement différentes les unes des autres. Les tests ANOVA ont été faits à l'aide du logiciel de

programmation statistique R [124]

Tests d'hypothèse ANOVA

Avec chaque nouvelle descendance considérée, les phénotypes des individus des trois ensembles décrits plus haut sont comparés par un test ANOVA. La p-value obtenue indiquera laquelle de l'hypothèse nulle H_0 (les moyennes de phénotype des trois ensembles de descendants ne sont pas significativement différentes) ou de l'hypothèse alternative H_1 (les moyennes de phénotype des trois ensembles de descendants sont significativement différentes) sera retenue en fonction de l'erreur α . La p-value représente, dans le cas présent, la probabilité de l'hypothèse nulle et l'erreur α indique, quant à elle, le seuil de p-value en deçà duquel on rejettera l'hypothèse nulle. Nous avons utilisé une erreur α de 5% tout au long de l'étude, un seuil généralement accepté en sciences de la santé. Nous rejetterons donc faussement l'hypothèse nulle dans 5% des cas.

Notons toutefois que la p-value obtenue n'est représentative que de la significativité de la différence entre les ensembles de données. Dans le cas qui nous concerne, nous faisons aussi l'hypothèse que les moyennes de phénotypes seront cohérentes avec les ensembles de descendants. On s'attend effectivement à ce que l'ensemble phénotype élevé présente la moyenne la plus élevée, l'ensemble phénotype bas, la moyenne la plus basse et l'ensemble phénotype moyen, une moyenne intermédiaire. Par contre, une p-value significative et des valeurs de moyennes cohérentes ne sont pas suffisantes pour confirmer la validité des résultats. En effet, une analyse ANOVA distincte est faite pour chaque descendance d'une couche généalogique et la probabilité d'obtenir des résultats significatifs par chance augmente avec le nombre de tests effectués. Il était donc impératif de corriger les p-valeurs pour les tests multiples.

Corrections des p-values pour tests multiples

Il existe différentes méthodes de corrections des p-values, certaines très sévères et irréalistes comme Bonferroni, où l'on suppose l'indépendance des données et où la p-value est simplement multipliée par le nombre de tests qui ont été effectués, d'autres plus permissives comme le «False Discovery Rate» de Benjamini et Hochberg et d'autres plus exactes, mais plus coûteuses en temps de calcul comme les méthodes de simulation. La simulation consiste à reproduire la même expérience avec des données aléatoires et observer si la nouvelle p-value est plus ou moins significative que l'originale. De manière formelle, la simulation est un test d'hypothèse. Dans le cadre de cette étude, les simulations permettront de conserver ou de rejeter l'hypothèse nulle H_0 selon laquelle les moyennes de phénotype des trois ensembles de descendants ne sont pas significativement différentes.

Les données comparées ici sont les valeurs de phénotype. Ce sont donc ces valeurs qui ont été randomisées. Une valeur de phénotype aléatoire, sélectionnée avec remplacement parmi la distribution de mesures phénotypiques disponible, a été attribuée à chaque descendant. L'analyse comparative des descendance a alors été effectuée et la p-value simulée stockée. Ce processus a été répété 1000 fois pour chaque phénotype et chaque couche généalogique. La p-value corrigée était calculée selon la formule suivante :

$$\text{p-value corrigée} = \frac{\text{Nombre de p-values simulées} < \text{p-value réelle}}{\text{Nombre total de p-values simulées}} \quad (8.7)$$

9. Expérimentations

Nous présenterons, dans cette section, les aspects techniques de notre expérimentation : utilisation de la base de données médicales, spécifications des systèmes informatiques utilisés et applications développées.

9.1 Base de données et confidentialité

L'accès et l'utilisation de données médicales soulèvent plusieurs questions d'éthiques et de confidentialité. Voici la description de la base de données et des mesures de sécurité qui ont été mises en place afin d'assurer la confidentialité des données.

L'accès aux données généalogiques du projet BALSAC-RETRO [8, 9] est soumise aux procédures de confidentialité en vigueur à l'Hôtel-Dieu du CHUM, au centre hospitalier de Chicoutimi et à l'Université du Québec à Chicoutimi. Bien que cette généalogie comporte plus de 40 000 individus, les données médicales ne sont disponibles que pour les sujets les plus récents, soit environ 1200 sujets. Toutes les mesures phénotypiques et génotypiques ont été prises par du personnel qualifié qui a suivi les procédures standards [112]. Tous les sujets ont signé un formulaire de consentement à l'utilisation des données recueillies pour fin de recherches. Afin d'assurer l'intégrité de leur identité, un numéro unique leur a été attribué dans la base de données. La correspondance entre les noms des sujets et les numéros n'est connue que des médecins assurant leur suivi.

La base de données est encodée en PostGreSQL et recense plus de 300 mesures phénotypiques différentes sur 1200 sujets en plus de données génétiques (marqueurs

micro-satellites, SNPs et haplotypes) pour un peu plus de 450 sujets. Tous ces individus font partie de la généalogie du projet BALSAC-RETRO [8, 9] et la base de données permet de faire la correspondance entre les numéros de sujets et les numéros d'individus dans la généalogie.

9.2 Applications et systèmes informatiques

Différents systèmes informatiques ont été requis pour le développement des applications nécessaires à cette étude. Nous présenterons brièvement les applications développées et les systèmes informatiques utilisés.

Les descendances ont été générées une seule fois, lors d'un pré-traitement de la généalogie, et stockées dans deux fichiers distincts tel que décrit à la section 8.7. L'interface de traitement de généalogies [123] a permis le développement des applications nécessaires pour générer les descendances. Elle était stockée sur un serveur IRIX64 version 6.5 avec processeur MIPS R14000 de 4500 MHz et 2048 Mo de mémoire vive.

Les autres étapes de l'analyse : la classification des descendants par régression linéaire, le calcul des couches généalogiques et de la spécificité génétique des ancêtres, l'analyse comparative des descendances et la correction pour tests multiples par simulations ont été effectuées sur différents postes informatiques. Le tableau IX.1 précise les spécifications techniques des postes informatiques utilisés et la figure 9.1 montre le schéma de la répartition des opérations entre ceux-ci.

Tab. IX.I: Spécifications techniques des différents postes informatiques utilisés.

Système d'exploitation	Processeur	Mémoire vive
IRIX64 6.5	MIPS R14000 4.5 GHz	2048 Mo
Red Hat Linux 6.0	Pentium 4 2.4GHz	768 Mo
Linux Fedora 5	Pentium 4 3,0 GHz	1024 Mo

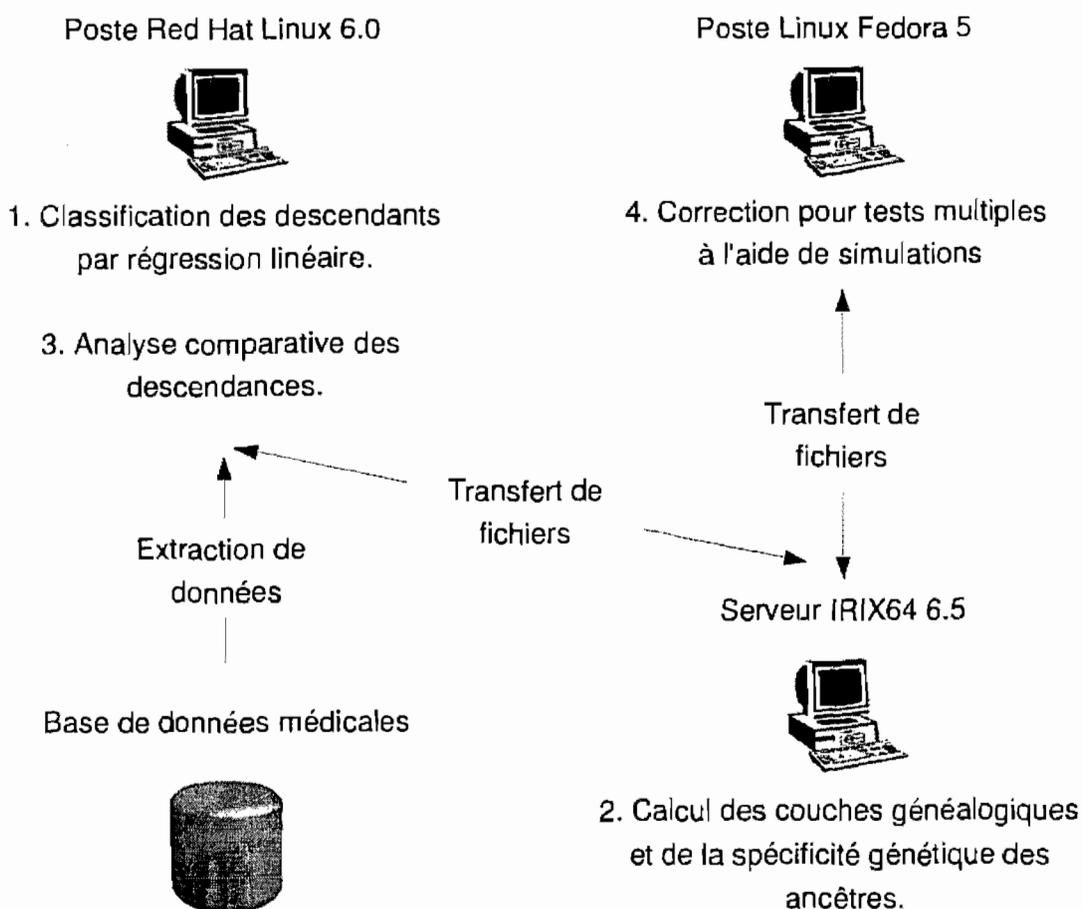


Fig. 9.1: Schéma de la répartition des opérations. Les chiffres indiquent l'ordre dans lequel les analyses étaient effectuées.

Troisième partie

Résultats

10. Résultats

Rappelons d'abord les questions de recherche auxquelles les analyses de descendance ont permis de répondre :

1. Est-il possible d'identifier des sous-groupes de fondateurs génétiquement spécifiques à des descendants de pression sanguine élevée ou d'IMC élevé ?
2. La spécificité génétique est-elle garante du phénotype des descendants ?
3. La spécificité génétique des fondateurs pour des individus de pression sanguine ou d'IMC élevé est-elle proportionnelle aux taux d'hypertension et d'obésité observés chez leurs descendants ?
4. Les analyses demeurent-elles statistiquement significatives après corrections des p-values pour tests multiples par simulations ?

Les analyses de descendance ont été effectuées sur trois phénotypes distincts soient : l'indice de masse corporelle (IMC) et les pressions systoliques et diastoliques corrigées pour la prise de médicaments [108, 126]. Une analyse a été effectuée pour chaque couche généalogique et chaque phénotype. Vu le nombre imposant d'analyses effectuées, la majeure partie des graphiques est présentée en annexe. La concentration sanguine de HDL, utilisée pour déterminer le statut dyslipidémique des paires de germains (voir section 8.2), a aussi été analysée à des fins de validation statistique.

Les figures 10.1, 10.2, 10.3 et 10.4 montrent les résultats de l'analyse de descendance pour l'indice de masse corporelle sur les couches généalogiques 1,4,7 et 12. Ces couches ont été sélectionnées afin d'illustrer l'évolution des résultats selon le niveau

généalogique des ancêtres. La première couche correspond aux parents des individus phénotypés, la quatrième à leurs arrière-arrière-grands-parents et ainsi de suite. Les graphiques des autres couches et phénotypes sont présentés aux annexes I, II et III.

10.1 Analyses ANOVA

Les analyses débutent toujours par la comparaison des descendance des ancêtres les plus spécifiques à la classe phénotype élevé versus celles des ancêtres les plus spécifiques à la classe phénotype bas. Les p-values et moyennes de phénotypes sont rapportées en fonction de la spécificité des ancêtres. Une spécificité de 100% correspond aux ancêtres contribuant uniquement à la classe phénotype élevé ou à la classe phénotype bas alors qu'une spécificité de 0% correspond à une contribution génétique égale aux deux classes. L'échelle à gauche des graphiques représente les valeurs moyenne de phénotype, celle à droite, les p-values ANOVA. Les cercles correspondent aux p-values d'un test ANOVA effectué après l'ajout d'une nouvelle descendance. Les lignes marquées par des «+», des «*» et des «x» représentent respectivement les valeurs moyennes des phénotypes des individus des ensembles phénotype haut, moyen, et bas. La couche généalogique est indiquée au bas de chaque graphique.

On observe une plus grande variabilité des moyennes de phénotype en début d'analyse. Cela s'explique aisément par le fait que les trois ensembles de descendants contiennent alors peu d'individus. L'ajout de nouvelles descendance influence donc facilement les moyennes. On peut d'ailleurs constater que cette variabilité décroît au fur et à mesure de l'analyse alors que davantage d'individus sont considérés.

10.2 Ensembles phénotypiques et seuils de risque

Tout au long de l'analyse, les valeurs moyennes de phénotype de l'ensemble phénotype haut demeurent au-dessus des seuils «à risque» définis par le NIH pour toutes les couches généalogiques et ce, pour les trois phénotypes investigués. De plus, l'écart entre les moyennes des ensembles phénotype haut et bas est d'au moins 20% et atteint parfois jusqu'à 35%.

On remarque aussi que les p-values décroissent conjointement avec la spécificité des ancêtres, tel qu'attendu. Les p-values demeurent toutefois très significatives même pour de faibles valeurs de spécificité.

Le tableau X.I rapporte les proportions d'individus à risque selon les seuils du NIH et la cardinalité (taille) de chaque ensemble de descendants pour tous les phénotypes investigués et toutes les couches généalogiques. Les p-values sont calculées grâce à un test χ^2 permettant d'estimer la significativité des différences de proportions d'individus à risque entre les ensembles de descendants. Les seuils de risque sont indiqués sous le tableau. Les données ont été collectées à la fin de l'analyse, lorsque tous les fondateurs étaient considérés.

10.2.1 Évaluation du risque phénotypique

On peut constater qu'il existe une distinction claire entre les proportions d'individus à risque des différents ensembles phénotypiques, confirmée par les p-values. Toutefois, le nombre d'individus dans l'ensemble phénotype moyen augmente avec le niveau des couches généalogiques. Il y a donc de moins en moins d'individus qui descendent uniquement d'un seul type d'ancêtre. Ceci concorde avec l'intuition selon laquelle la spécificité devrait décroître avec l'augmentation du niveau des couches généalogiques. La taille des descendance augmentant de manière exponentielle avec le nombre de générations,

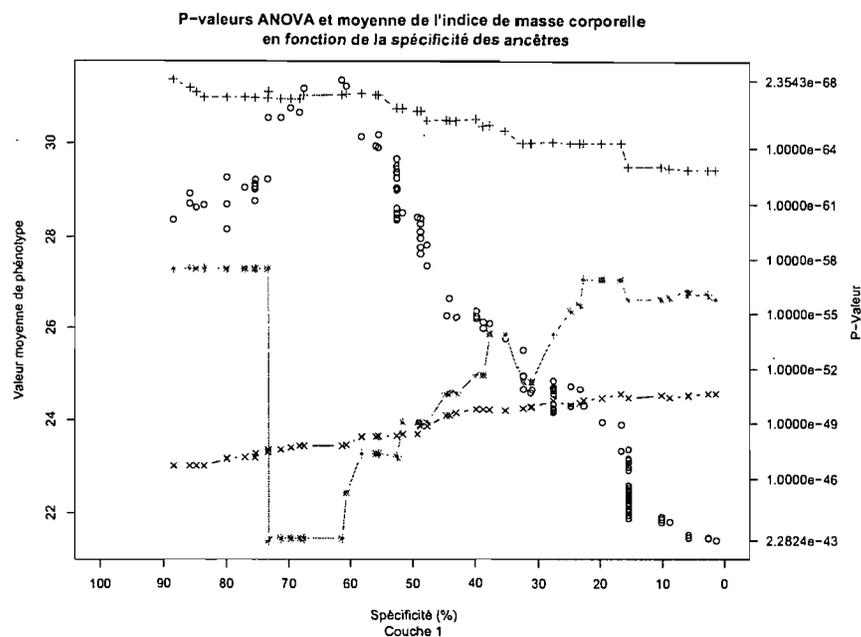


Fig. 10.1: Analyse ANOVA de l'indice de masse corporelle (IMC) des descendance en fonction de la spécificité des ancêtres pour la première couche généalogique.

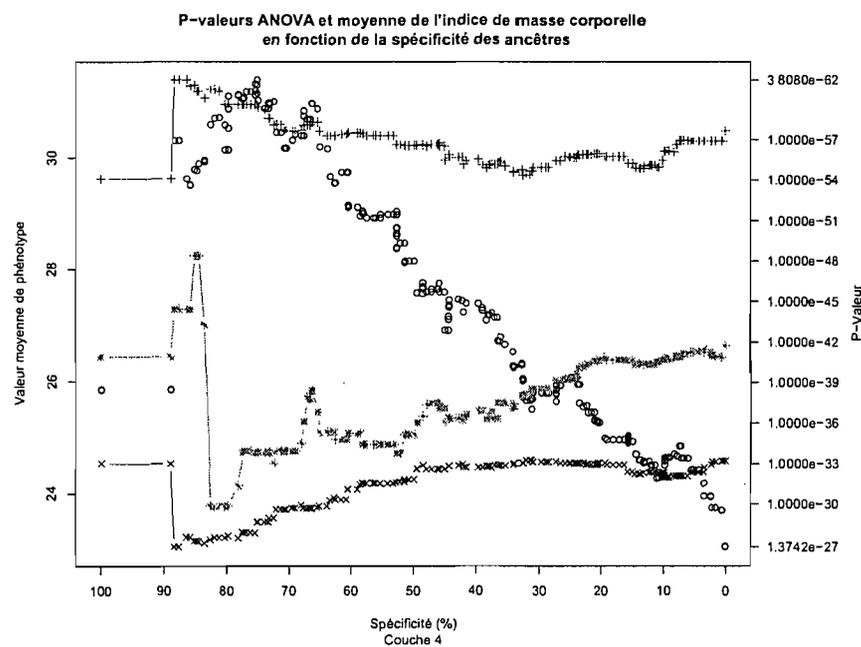


Fig. 10.2: Analyse ANOVA de l'indice de masse corporelle (IMC) des descendance en fonction de la spécificité des ancêtres pour la quatrième couche généalogique.

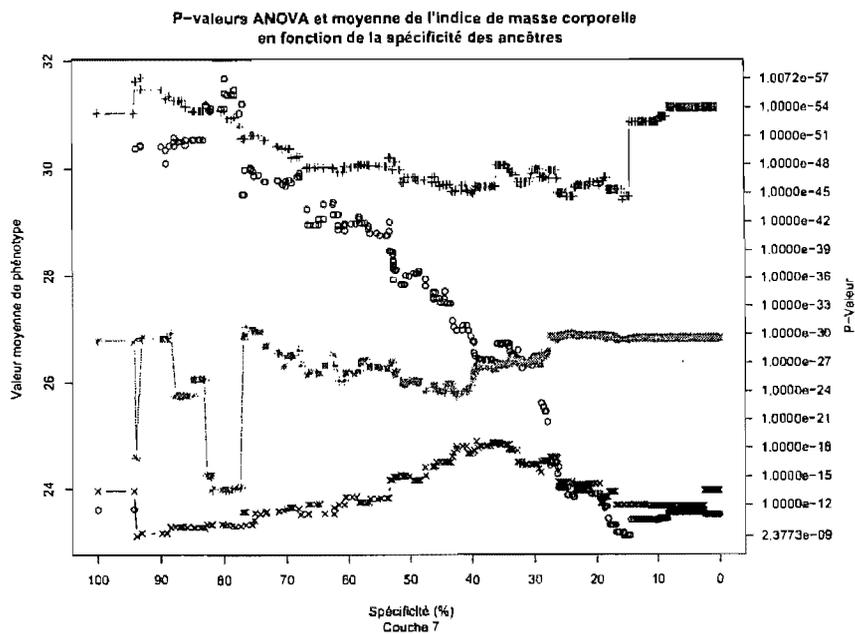


Fig. 10.3: Analyse ANOVA de l'indice de masse corporelle (IMC) des descendance en fonction de la spécificité des ancêtres pour la septième couche généalogique.

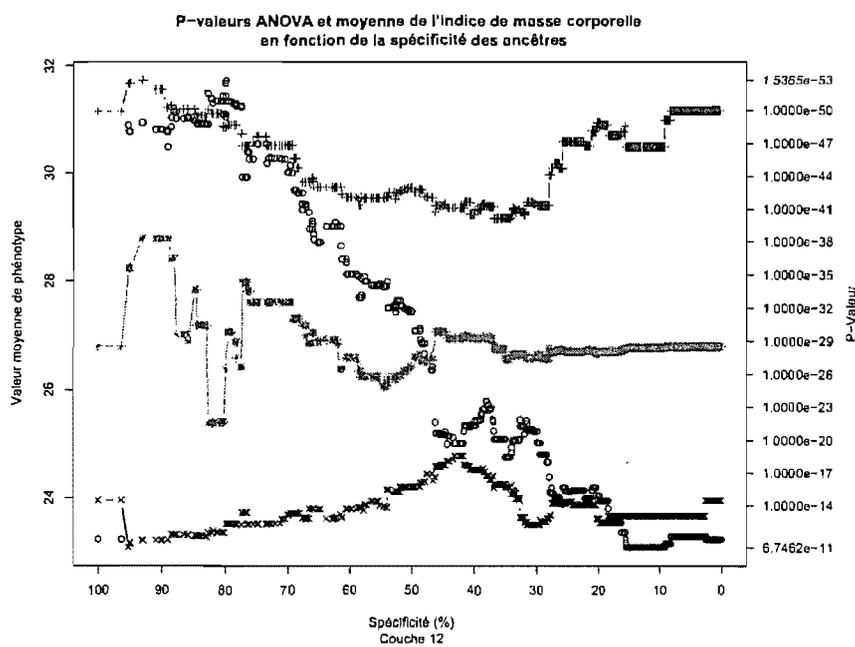


Fig. 10.4: Analyse ANOVA de l'indice de masse corporelle (IMC) des descendance en fonction de la spécificité des ancêtres pour la douzième couche généalogique.

nous nous attendons effectivement à ce que les ancêtres plus anciens soient de moins en moins spécifiques à une classe particulière de descendants.

10.3 Ensembles phénotypiques et hypertension

Le tableau X.II rapporte les proportions d'individus hypertendus et la cardinalité (taille) de chaque ensemble phénotypique pour tous les phénotypes investigués et toutes les couches généalogiques. Le statut *hypertendu* était attribué aux personnes prenant des médicaments anti-hypertensifs ou pour lesquelles la pression systolique dépassait 140 mmHg et/ou la pression diastolique s'élevait au-dessus de 90 mmHg à deux occasions [112]. Les p-values sont issues d'un test χ^2 permettant d'estimer la significativité des différences de proportions d'individus hypertendus entre les ensembles de descendants. Les données ont été collectées à la fin de l'analyse, lorsque tous les fondateurs étaient considérés.

10.3.1 Évaluation du risque d'hypertension

Bien que les différences de proportions soient significatives, certains résultats contre-intuitifs sont observés. Pour les trois premières couches généalogiques basées sur l'IMC et les deux premières basées sur les pressions systoliques et diastoliques, les ensembles phénotype bas contiennent une proportion d'hypertendus supérieure à celle de l'ensemble phénotype moyen. Outre ces exceptions, les résultats sont cohérents ; plus le pourcentage d'individus à risque (voir tableau X.I) est élevé, plus la proportion d'hypertendus de l'ensemble de descendants augmente, suggérant une corrélation entre l'hypertension et le risque phénotypique.

Tab. X.I: Proportion (%) d'individus à risque et cardinalité (taille) des classes de descendants pour toutes les couches généalogiques et phénotypes investigués.

Phénotype	Couche	Proportion (%) à risque			Cardinalité			P-Value	
		Classe phéno. bas	Classe phéno. moyen	Classe phéno. élevé	Classe phéno. bas	Classe phéno. moyen	Classe phéno. élevé		
IMC ¹	1	9.0	21.6	41.5	410	97	357	8.56e-25	
	2	8.6	21.0	42.2	383	138	344	3.33e-25	
	3	8.7	21.4	42.6	334	228	303	1.03e-22	
	4	7.0	22.4	52.4	200	518	147	6.64e-22	
	5	4.5	21.4	57.6	66	707	92	1.09e-16	
	6	3.4	23.3	58.3	58	759	48	1.94e-10	
	7	3.4	23.4	61.9	58	765	42	7.03e-11	
	⋮	⋮	⋮	⋮	⋮	⋮	⋮		
	14	3.4	23.4	61.9	58	765	42	7.03e-11	
	Pression systolique corrigée ²	1	26.8	17.5	67.5	369	80	326	3.33e-31
		2	26.8	30.0	66.9	362	110	305	3.26e-26
		3	26.6	46.3	69.3	346	177	254	3.15e-20
		4	23.4	46.4	69.4	239	414	124	6.94e-17
		5	22.9	42.6	73.0	83	631	63	9.77e-09
6		21.7	42.9	69.6	69	652	56	5.18e-07	
7		16.4	42.9	69.6	55	666	56	1.05e-07	
⋮		⋮	⋮	⋮	⋮	⋮	⋮		
15		16.4	42.9	69.6	55	666	56	1.05e-07	
Pression diastolique corrigée ³		1	9.3	18.6	50.9	279	70	426	5.67e-31
	2	10.5	24.8	51.9	287	105	385	2.36e-29	
	3	9.8	31.0	51.9	264	174	339	1.26e-26	
	4	7.2	34.6	51.9	166	422	189	2.86e-18	
	5	1.5	33.7	57.6	67	644	66	3.26e-11	
	6	0.0	33.9	56.9	61	658	58	1.46e-10	
	7	0.0	33.4	56.9	53	664	60	1.03e-09	
	8	0.0	33.5	56.9	53	666	58	1.14e-09	
	⋮	⋮	⋮	⋮	⋮	⋮	⋮		
	15	0.0	33.5	56.9	53	666	58	1.14e-09	

¹ Un IMC > 31 est considéré à risque

² Une pression systolique > 140 est considérée à risque

³ Une pression diastolique > 90 est considérée à risque

Les points verticaux indiquent que les valeurs ne changent pas pour les couches intermédiaires

Tab. X.II: Proportion (%) d'individus hypertendus et cardinalité (taille) des classes de descendants pour toutes les couches généalogiques et phénotypes investigués.

Phénotype	Couche	Proportion (%) d'hypertendus			Cardinalité			P-Value
		Classe phéno. bas	Classe phéno. moyen	Classe phéno. élevé	Classe phéno. bas	Classe phéno. moyen	Classe phéno. élevé	
Indice de masse corporelle	1	54.6	30.0	63.6	410	97	357	2.13e-08
	2	56.9	30.4	64.2	383	138	344	9.83e-11
	3	54.8	47.4	62.7	334	228	303	0.002
	4	51.5	54.6	64.6	200	518	147	0.041
	5	39.4	56.6	59.8	66	707	92	0.019
	6	34.5	57.4	52.1	58	759	48	0.003
	7	34.5	57.4	52.4	58	765	42	0.003
	∴	∴	∴	∴	∴	∴	∴	
	14	34.5	57.4	52.4	58	765	42	0.003
Pression systolique corrigée	1	36.3	31.2	73.0	369	80	326	4.51e-24
	2	35.9	41.8	73.1	362	110	305	1.18e-21
	3	35.0	54.2	71.7	346	177	254	5.04e-18
	4	31.0	57.2	71.0	239	414	124	9.24e-15
	5	30.1	52.5	68.3	83	631	63	1.31e-05
	6	27.5	52.8	64.3	69	652	56	4.69e-05
	7	21.8	52.7	64.3	55	666	56	8.13e-06
	∴	∴	∴	∴	∴	∴	∴	
	15	21.8	52.7	64.3	55	666	56	8.13e-06
Pression diastolique corrigée	1	35.5	28.6	65.3	279	70	426	3.79e-17
	2	36.2	35.2	67.0	287	105	385	5.28e-17
	3	34.8	43.7	68.1	264	174	339	3.57e-16
	4	30.7	52.1	67.7	166	422	189	2.71e-11
	5	34.3	52.1	60.6	67	644	66	0.006
	6	32.8	52.7	55.1	61	658	58	0.010
	7	30.2	52.6	56.7	53	664	60	0.005
	8	30.2	52.7	55.1	53	666	58	0.006
∴	∴	∴	∴	∴	∴	∴		
	15	30.2	52.7	55.1	53	666	58	0.006

10.4 Simulations et validité statistique

La généalogie comporte un peu plus de 40 000 individus et nous effectuons un test ANOVA pour chaque fondateur. Il était donc impératif de corriger les p-values pour les tests multiples afin d'assurer la validité statistique de l'étude. Les méthodes de simulation, telles que décrites à la section 8.7.2, visent à corriger les p-values en reproduisant les analyses un très grand nombre de fois avec des valeurs randomisées. La figure 10.5 montre les résultats de simulations pour toutes les couches généalogiques et tous les phénotypes. Les analyses ANOVA ont été effectuées 1000 fois avec permutation des valeurs de phénotype des descendants. Pour l'indice de masse corporelle (IMC), les p-values demeurent significatives (p-value < 0.05) de la première à la septième couche. Hormis la huitième couche, les p-values simulées pour la pression diastolique corrigée sont toutes significatives et aucune ne l'est pour la pression systolique corrigée. Dans le cas où les p-values ne sont pas significatives, nous ne pouvons pas rejeter l'hypothèse nulle selon laquelle la spécificité des ancêtres n'influence pas significativement les phénotypes des descendants.

10.4.1 Stratification sur l'obésité

Comme nous l'avons vu à la section 7.1, l'obésité et l'hypertension sont non seulement fortement corrélés, ils constituent des éléments clés du syndrome métabolique. Plusieurs études distinguent d'ailleurs l'hypertension associée à l'obésité des autres formes d'hypertension. Il semble en effet que les mécanismes pathophysiologiques sous-jacents diffèrent selon le cas [12]. Nous avons donc procédé à des simulations où les pressions systoliques et diastoliques des individus obèses (IMC > 31) étaient analysées séparément de celles des sujets non-obèses (IMC ≤ 31). Les figures 10.6 et 10.7 montrent les résultats de simulations avec et sans stratification sur l'obésité pour les pressions systolique et diastolique respectivement. On remarque d'abord que les résultats ne sont jamais significatifs pour les sujets obèses. Les p-values sont par contre

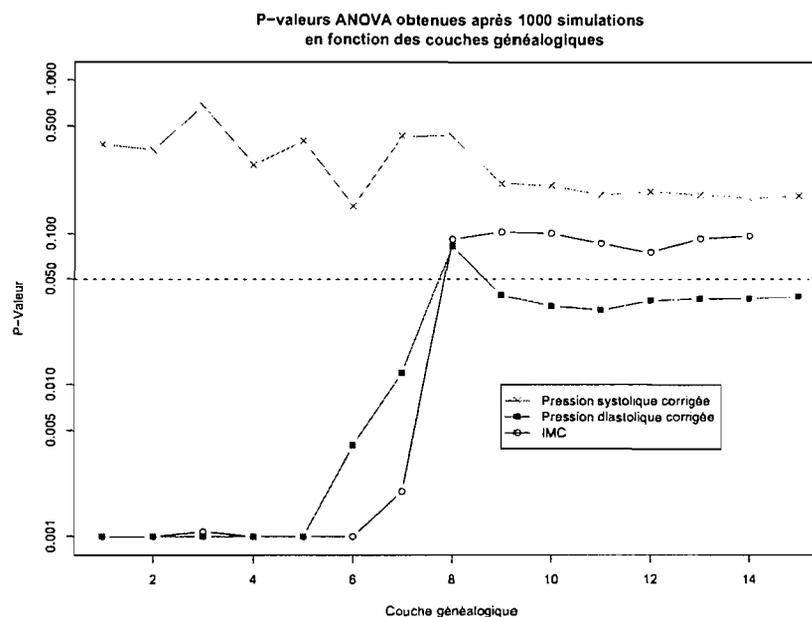


Fig. 10.5: p-values ANOVA simulées de la première à la dernière couche généalogique pour les trois phénotypes investigués.

généralement plus significatives pour les sujets non-obèses que lors des analyses sans stratification.

10.4.2 Stratification sur l'âge

Comme la pression systolique est fortement corrélée à l'âge des sujets [127], des analyses supplémentaires, où les descendants de 55 ans et moins étaient considérés séparément des individus de plus de 55 ans, ont été effectuées. La figure 10.8 rapporte les résultats de simulations avant et après stratification sur l'âge pour la pression systolique corrigée et toutes les couches généalogiques. Les p-values sont significatives de la première à la troisième couche et de la neuvième à la dernière couche généalogique pour les descendants de 55 ans et moins. Aucune influence significative de la spécificité des ancêtres n'est observée pour les descendants de plus de 55 ans.

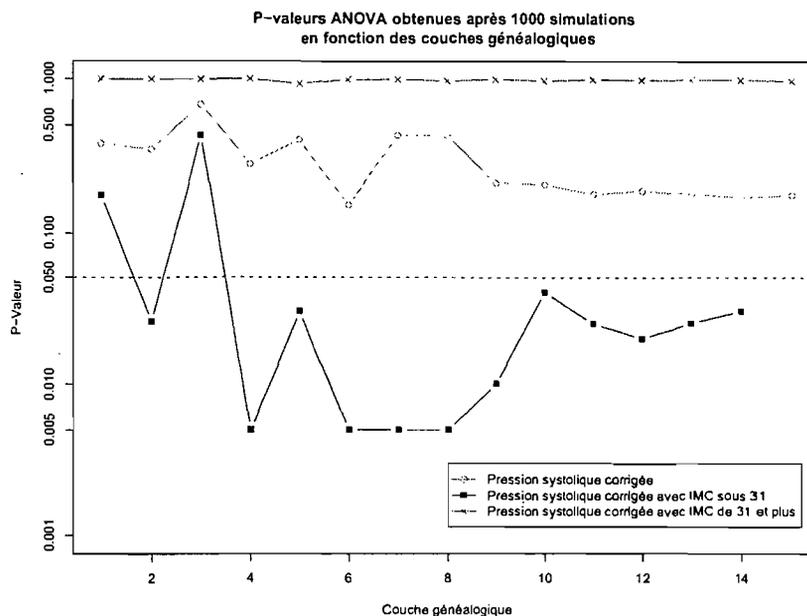


Fig. 10.6: p-valeurs ANOVA simulées de la première à la dernière couche généalogique pour la pression systolique corrigée avec et sans stratification sur l'obésité.

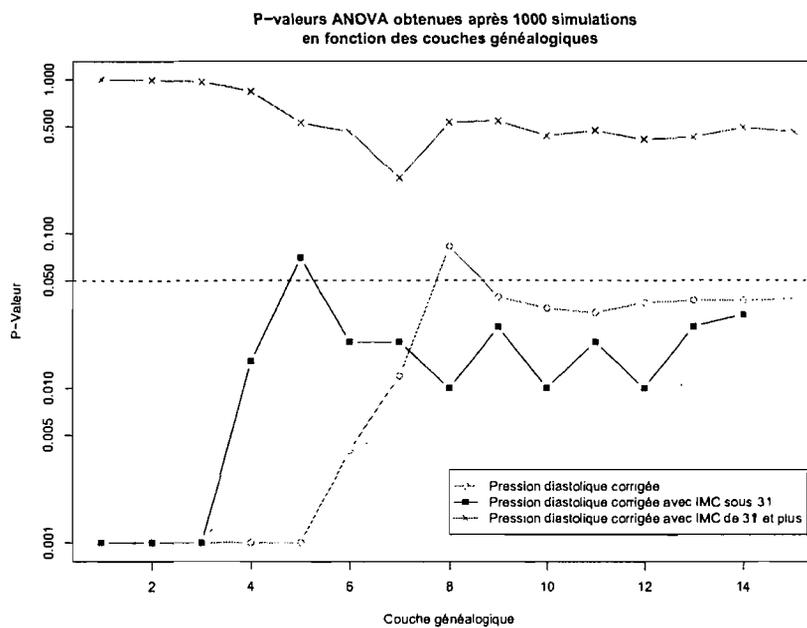


Fig. 10.7: p-valeurs ANOVA simulées de la première à la dernière couche généalogique pour la pression diastolique corrigée avec et sans stratification sur l'obésité.

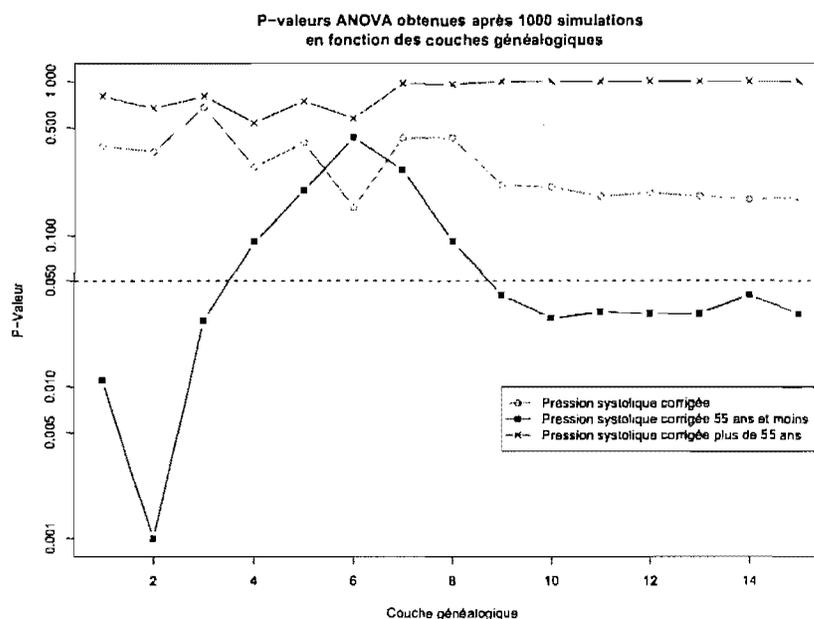


Fig. 10.8: p-values ANOVA simulées de la première à la dernière couche généalogique pour la pression systolique corrigée avec et sans strates d'âge.

10.4.3 Stratification sur le sexe

Comme nous l'avons mentionné plus haut, la concentration sanguine de HDL, utilisée pour déterminer le statut dyslipidémique des patients (voir section 8.2), a été investiguée à des fins de validation statistique. Étant donné les succès obtenus grâce la stratification sur l'âge et l'obésité pour les pressions sanguines, nous avons voulu vérifier l'effet de la stratification sur un autre facteur et un autre phénotype afin de confirmer la validité de notre approche. Une étude du Dr. Ondrej Šeda [127] a révélé que la concentration sanguine de HDL semble être fortement corrélée au sexe des individus. Nous avons donc procédé à des simulations d'analyses de descendance avec et sans séparation des sexes. La figure 10.9 rapporte les résultats de simulations sur le phénotype HDL pour toutes les couches généalogiques. Alors que les p-values sont significatives pour les premières couches des simulations pour sexes confondus et hommes seulement, elles ne le sont jamais pour les femmes seulement. Il semble donc que la stratification soit un moyen efficace pour capturer non seulement l'effet de l'âge et de l'obésité, mais

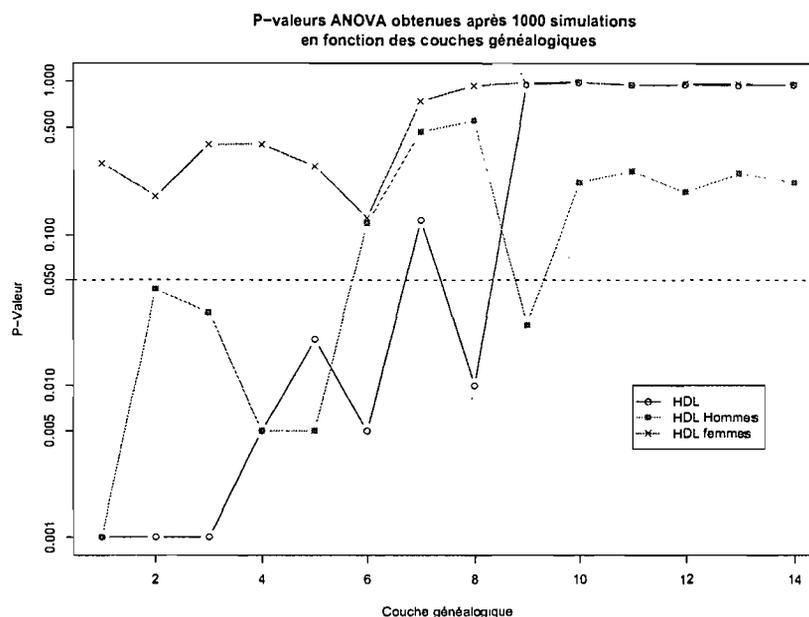


Fig. 10.9: p-values ANOVA simulées de la première à la dernière couche généalogique pour la concentration sanguine d'HDL avant et après séparations des sexes.

aussi du sexe sur le phénotype des individus.

10.5 Temps d'exécutions

Le tableau X.III présente les temps de calcul requis pour les différentes étapes de notre analyse. Les calculs de la spécificité des ancêtres et des couches généalogiques ont été effectués sur un serveur IRIX64 et les analyses comparatives des descendance sur un poste Red Hat Linux 6.0 (pour les spécifications techniques, voir le tableau IX.I). Les simulations se déroulaient en trois étapes :

1. Générer un ensemble de descendants de phénotypes aléatoires
2. Calculer les couches généalogiques et la spécificité des ancêtres à partir de l'ensemble de descendants
3. Procéder à l'analyse comparative des descendance basée sur la spécificité des ancêtres à l'aide de tests ANOVA.

Tab. X.III: Temps de calculs moyen pour les différentes étapes de l'analyse de descendance sur une couche généalogique pour trois phénotypes investigués.

	IMC	Pression systolique corrigée	Pression diastolique corrigée
Calcul d'une couche généalogique et de la spécificité des ancêtres	1.24 sec	1.40 sec	1.25 sec
Comparaison des descendances et tests ANOVA	8.82 sec	9.02 sec	8.61 sec
Temps de calcul moyen d'une simulation	19.45 sec	18.25 sec	20.46 sec

Les calculs de couches généalogiques et de spécificité nécessitent des bibliothèques disponibles uniquement sur le serveur IRIX64. Les étapes 1 et 3 ont donc été réalisées sur le poste Fedora 5 alors que l'étape 2 s'effectuait sur le serveur. Les résultats intermédiaires étaient transférés automatiquement entre le poste local et le serveur via une connexion SSH. Les temps de simulation sont donc dépendants de l'affluence sur le serveur et de la qualité de la connexion SSH, affectant la vitesse de transfert des données et la vitesse de calcul des couches généalogiques et de la spécificité.

Quatrième partie

Discussion et conclusions

11. Discussion

Nous avons présenté une nouvelle méthode pour estimer le risque d'hypertension et d'obésité d'un sujet selon son ascendance, validée statistiquement après corrections pour tests multiples par simulations. Notre méthode se distingue grandement des études familiales classiques, qui visent aussi à estimer le risque de maladie d'un sujet selon sa généalogie. Alors que celles-ci se limitent souvent à l'historique de familles nucléaires, notre méthode tire pleinement profit de l'information généalogique en investiguant des descendances d'ancêtres spécifiques séparés en couches généalogiques. Nous croyons fermement que notre méthode peut s'avérer utile en santé publique, dans le cadre d'études prospectives sur des populations à risque par exemple.

11.1 Analyse comparative des descendances

Notre méthodologie capture donc davantage d'informations que les études familiales classiques. Les résultats sont par contre plus complexes et leur interprétation nécessite une réflexion plus poussée. En étudiant les analyses ANOVA présentées à la section 10.1 et aux annexes I, II et III, on arrive à dégager un schéma général pour les trois phénotypes investigués. Les graphiques peuvent tous être divisés en trois sections approximatives :

1. Première section [100% spécificité, 70% spécificité] : On compare alors les descendances des ancêtres les plus spécifiques aux descendants de phénotype élevé aux ancêtres les plus spécifiques aux descendants de phénotype bas. Les sujets constituant la classe phénotype moyen (individus provenant des deux types d'an-

cêtres) sont donc peu nombreux et la moyenne des valeurs de phénotype varie beaucoup. Tel qu'attendu, les ancêtres situées aux deux extrêmes du spectre de spécificité ne partagent pas beaucoup de descendants communs. Les p-values atteignent aussi toujours leur significativité maximale dans cette région, suggérant qu'une forte polarisation de la spécificité génétique des ancêtres est garante d'une forte polarisation des phénotypes de leurs descendants.

2. Deuxième section [70% spécificité, 30% spécificité] : La significativité des p-values ANOVA décroît avec chaque nouvel ancêtre considéré. Cette observation confirme la cohérence de notre système de classification ; les différences entre les valeurs de phénotypes des trois ensembles de descendants s'atténuent à mesure que des ancêtres moins spécifiques sont considérés.
3. Troisième section [30% spécificité, 0% spécificité] : Alors que les p-values ANOVA décroissent, les valeurs moyennes de phénotype ne varient pratiquement plus. Les descendants des ancêtres les moins spécifiques, bien qu'ils n'aient que peu d'influence sur les moyennes phénotypiques, atténuent donc les différences entre les trois ensembles de descendants.

On remarque aussi que, sauf exceptions, les moyennes de phénotypes des trois ensembles de descendants sont cohérentes : les descendants de l'ensemble phénotype haut présentent la moyenne la plus élevée, ceux de l'ensemble phénotype bas la plus basse et ceux de l'ensemble phénotype moyen, une moyenne intermédiaire. Les résultats soutiennent donc notre hypothèse de départ selon laquelle la spécificité génétique des ancêtres est garante, dans une certaine mesure, du phénotype des descendants.

Nous ne pouvons toutefois passer sous silence certains résultats surprenants. Les analyses portant sur les pressions systolique et diastolique corrigées présentent une

particularité intéressante pour les deux premières couches généalogiques. La moyenne de phénotype des patients présentant un phénotype moyen est inférieure à celle des patients de phénotype bas. Ces situations se produisent généralement en début d'analyse (spécificité entre 100% et 70%), alors que l'ensemble phénotype moyen contient peu d'individus. Par conséquent, la moyenne de l'ensemble phénotype moyen est facilement affectée par une valeur de phénotype aberrante. Cette explication n'est toutefois pas suffisante dans les cas où la moyenne de l'ensemble phénotype moyen demeure sous celle de l'ensemble phénotype bas peu importe la spécificité des fondateurs.

Rappelons d'abord que les individus de l'ensemble phénotype moyen descendent à la fois d'ancêtres spécifiques phénotype haut et phénotype bas. Comment expliquer que le fait de descendre de ces deux types d'ancêtres confère aux individus une moyenne de phénotype inférieure pour les premières couches généalogiques? Comme il semble qu'aucune hypothèse biologique ne puisse expliquer ce phénomène, nous suggérons l'existence possible d'un biais lors de la sélection des familles, qui pourrait introduire une distorsion dans la distribution des phénotypes. Les biais à notre étude seront étudiés en profondeur à la section 11.5.

11.2 Évaluation du risque

Cette étude visait principalement à vérifier l'hypothèse selon laquelle il est possible d'estimer le risque d'hypertension/obésité d'un individu à partir de sa généalogie. Conformément aux procédures établies dans les études d'historique familial [17, 18], trois classes de risque (élevé, moyen et bas) ont été créées. Les proportions d'individus à risque ont ensuite été estimées pour chacune de ces classes, selon des seuils de risque reconnus dans la communauté scientifique [63, 35]. Les proportions d'individus hypertendus [112] ont aussi été comparées. Nous avons finalement vérifié la significativité des différences de proportions par un test χ^2 .

11.2.1 Risque phénotypique

Les proportions d'individus à risque dans chacune des classe sont cohérentes, hormis pour la première couche généalogique de la pression systolique corrigée où la proportion d'individus à risque est plus basse dans l'ensemble phénotype moyen que phénotype bas. Cela n'est en rien surprenant étant donné qu'une incohérence similaire a été observée au niveau des moyennes de ces même ensembles (voir section 11.1). De plus, les différences de proportion présentée dans le tableau X.I, très significative pour tous les phénotypes et toutes les couches généalogique, confirment la validité de notre système de stratification du risque et son utilisation potentielle dans le cadre d'un programme de prévention.

11.2.2 Risque d'hypertension

La comparaison des proportions d'individus hypertendus en fonction des classes de risque est aussi très intéressante. En effet, les classes de risque ont été définies pour trois phénotypes distincts : IMC, pression systolique et pression diastolique selon des seuils reconnus dans la littérature. Dans le cadre de cette étude, le statut «hypertendu» était conféré à des individus dont la pression systolique et/ou diastolique dépassait 140 mmHg et 90 mmHg respectivement à deux occasions ou qui prenaient des médicaments anti-hypertensifs [112]. La prise de médicaments anti-hypertensifs dépend évidemment du diagnostic du médecin traitant, qui est en mesure de prendre en compte une multitude de facteurs autres que les seuils de pression sanguine. Les différences entre les proportions d'individus hypertendus et considérés à risque selon les seuils de pressions sanguine illustre entre autre l'écart entre le risque théorique d'hypertension et le diagnostic réel posé par le médecin traitant. De plus, les taux d'hypertendus en fonction des ensembles d'IMC nous permettent d'évaluer l'association entre l'IMC et l'hypertension.

Hypertension et pressions sanguines corrigées

On remarque d'abord que les taux d'hypertendus sont cohérents entre les classes de risque pour les phénotypes de pression sanguine et que les différences de proportion entre les classes de risque sont significativement différentes. Toutefois, on observe encore quelques exceptions au niveau des premières couches généalogiques, tel qu'attendu en observant les résultats précédents. Il est aussi intéressant de noter que les taux de sujets hypertendus sont généralement plus élevés que les taux d'individus à risque pour une classe de risque et une couche généalogique données. Cela montre bien que l'hypertension ne peut être entièrement caractérisée par une seule mesure de pression sanguine. Les différences entre les taux de sujets à risque et les taux d'hypertendus sont toutefois plus marquées pour la pression sanguine diastolique que systolique. Il semble donc que la pression systolique soit un meilleur indicateur d'hypertension que la pression diastolique, tel que suggéré par d'autres études [36, 35].

Hypertension et indice de masse corporelle

Les différences entre les taux d'obésité et d'hypertension pour une classe de risque donnée sont très intéressantes. On remarque d'abord que les taux d'hypertendus ne sont pas cohérents avec les classes de risque. Bien que le taux d'hypertendus soit toujours plus élevé dans la classe phénotype élevé que phénotype bas, le taux d'hypertendus de la classe phénotype moyen varie énormément. De plus, la différence entre les proportions de sujets hypertendus et obèses est plus marquée pour la classe phénotype bas que pour la classe phénotype élevé. Un fort taux d'obésité semble donc garantir d'un fort taux d'hypertendus, en accord avec les études précédentes [109, 108, 110], suggérant une corrélation de l'ordre de 36% à 58% entre obésité et hypertension. Notre étude suggère toutefois que la corrélation entre minceur et diminution du risque d'hypertension est beaucoup moins forte chez les familles hypertendues et dyslipidémiques

11.3 Ensembles phénotypiques et couches généalogiques

L'évolution de la cardinalité des classes de risque au fil des couches généalogiques nous informe sur la structure du pedigree. On remarque d'abord que la taille des ensembles phénotypiques varient jusqu'à l'atteinte d'un plateau vers la septième ou huitième couche. Au-delà de ce niveau, les résultats ne varient pratiquement plus. Les fondateurs des couches supérieures apportent donc peu d'information supplémentaire dans le cadre de notre analyse et ce, pour les trois phénotypes investigués. Il est possible que la structure généalogique soit responsable de cet état de fait. Nous savons *à priori* que le SLSJ présente un effet fondateur [5, 6] et qu'un nombre restreint d'ancêtres est responsable d'une grande partie de la variabilité génétique qui y est observée aujourd'hui. Les descendances de ces quelques fondateurs incluent donc une très grande partie des sujets d'aujourd'hui et déterminent pratiquement à elles seules la constitution des ensembles phénotypiques lors de l'analyse des couches généalogiques supérieures. Heyer *et al.* [2] ont d'ailleurs rapporté que : «15% des fondateurs étaient responsables de 90% de la contribution génétique totale de tous les fondateurs».

De plus, il est important de rappeler que les couches généalogiques incluent les ancêtres situés à une distance minimale k de l'ensemble de descendants **en plus** des ancêtres sans parent répertoriés (sources du graphe généalogique) situés à une distance $< k$ de l'ensemble de descendants. On retrouve des sources à aussi peu que trois générations des individus d'aujourd'hui. Ces sources «propagées» aux couches supérieures influencent d'autant plus fortement la distribution des sujets parmi les classes de risque qu'elles sont de plus en plus nombreuses au fil des couches généalogiques.

Par contre, le choix de propager ces sources à travers les couches généalogiques n'est pas gratuit. D'abord, tous les individus d'aujourd'hui n'ont pas nécessairement

d'ancêtres à tous les niveaux généalogiques. Les sources propagées permettent donc d'assurer que tous les descendants sont représentés tout au long de notre analyse. Aussi, nous ne pouvons déterminer si ces sources sont apparentées à d'autres fondateurs Canadien-Français ou si elles proviennent d'une population extérieure dont nous ne connaissons pas l'ancestralité. Dans ce dernier cas, les exclure reviendrait à ignorer l'apport génétique d'une population extérieure.

On remarque aussi que la cardinalité de l'ensemble phénotype moyen augmente au fil des couches généalogiques alors que la taille des ensembles phénotype haut et bas diminuent. Tel qu'attendu, la spécificité génétique des ancêtres s'amenuise au fil des générations et le nombre de sujets descendant uniquement d'ancêtres spécifiques phénotype haut ou bas diminue. Heyer *et al.* [2], ont d'ailleurs conclu, après avoir calculé la contribution génétique des fondateurs du SLSJ pour des groupes d'individus atteints de diverses maladies, «qu'il n'y a pas de sous-groupe de fondateurs qui soit davantage apparenté à n'importe quel groupe d'individus». Bien que nous ayons aussi observé une diminution de spécificité, au niveau des couches généalogiques supérieures, nous ne pouvons pas entériner leur conclusion. Nous avons effectivement identifié des sous-groupes d'ancêtres fortement apparentés à des descendants présentant des mesures phénotypiques significativement basses ou élevées et ce, pour toutes les couches généalogiques. Les raisons expliquant ces différences sont multiples. Premièrement, la généalogie est traitée comme un graphe dirigé acyclique (DAG). Ainsi, à l'opposé de Heyer *et al.*, nous avons été en mesure de décomposer le DAG en couches généalogiques indépendantes et de les traiter de la plus récente à la plus ancienne. De plus, nous classifions les fondateurs selon leur spécificité génétique, qui met en relief leur contribution génétique *différentielle* pour deux groupes de descendants plutôt que leur contribution génétique *absolue*. Enfin, les fondateurs sont comparés par ordre décroissant de spécificité génétique afin de mettre en lumière la polarisation des phénotypes de leurs descendants.

Finalement, il est intéressant de constater que les quatrième et cinquième couche généalogique semblent être des points charnières de notre analyse. En effet, c'est lors de la transition vers ces couches généalogiques que la cardinalité des ensembles phénotypiques varie le plus ; les ensembles phénotype élevé et bas perdent alors énormément d'individus au profit de l'ensemble phénotype moyen. Il semble que l'effet fondateur soit en partie responsable de cette perte de spécificité génétique. En effet, une étude de Hamet, Merlo *et al.* [112] a permis de «quantifier» l'effet fondateur dans la population du SLSJ pour les différentes couches généalogiques. Plus précisément, ils ont calculé l'écart entre la distribution de contributions génétiques observée par rapport à la distribution uniforme pour chaque couche généalogique. L'effet fondateur calculé s'étendait sur l'intervalle $[0,1]$ et était égal à 0 lorsque les contributions génétiques des fondateurs étaient uniformément distribuées et égal à 1 lorsqu'un seul fondateur contribuait au pool génétique des descendants. Leur étude a révélé que l'effet fondateur, dans la population du SLSJ, connaissait une croissance significative de la quatrième à la septième couche inclusivement et demeurait alors relativement constant dans les couches subséquentes. Historiquement, les fondateurs des couches 4 à 7 proviennent, du moins en partie, de la vague migratoire de 1838-1911, au cours de laquelle près de 14 000 personnes ont émigré de Charlevoix vers le SLSJ (voir section 2.1). Il semble donc que cet apport massif d'immigrants ait contribué à atténuer l'effet fondateur observé dans les couches généalogiques supérieures, uniformisant par le fait même la distribution de contributions génétiques.

Outre l'effet fondateur, un avantage sélectif en faveur de l'hétérozygotie peut aussi expliquer les changements observés aux quatrième et cinquième couches généalogiques. Rappelons que les sujets de phénotype moyen descendent à la fois d'ancêtres spécifiques phénotype haut et bas. En supposant que ces deux types d'ancêtres portent des allèles distincts pour un gène donné, leurs descendants communs (sujets de phé-

notype moyen) seraient potentiellement hétérozygotes pour ce trait. Une sélection en faveur de l'hétérozygotie pourrait donc expliquer l'augmentation de la taille de l'ensemble phénotype moyen à partir de la quatrième couche généalogique. Un des cas les plus célèbres de sélection en faveur de l'hétérozygotie concerne l'anémie falciforme, une maladie génétique récessive et la malaria, une maladie infectieuse transmise par les moustiques. En effet, des études ont montré que le statut d'hétérozygote pour le gène de l'anémie falciforme conférait une protection accrue contre la malaria et donc, de meilleures chances de survie [128]. Ce phénomène a aussi été suggéré pour expliquer les fréquences d'autres maladies génétiques telles la phénylcétonurie et la fibrose kystique [129, 130, 131, 132]. Il semble toutefois difficile de départager la sélection en faveur des hétérozygotes de l'effet fondateur lorsqu'il est question d'expliquer la fréquence des maladies génétiques dans des populations fondatrices [133, 134, 135].

11.4 Simulations et validité statistique

Les résultats de simulations, présentés à la figure 10.5 permettent d'évaluer la validité statistique de notre analyse comparative des descendances. Rappelons qu'un test ANOVA est effectué pour chaque fondateur d'une couche généalogique et que la probabilité d'obtenir des résultats significatifs par chance augmente avec le nombre de tests effectués. Les simulations permettent de corriger les p-values pour les tests multiples. Étant donné la rapidité de nos procédés, nous avons été en mesure de reproduire les analyses 1000 fois pour chaque phénotype et couches généalogiques. Nous évaluons ainsi la probabilité réelle que la spécificité des ancêtres influence significativement le phénotype des descendants.

11.4.1 Indice de masse corporelle

Les p-values simulées sont significatives pour les sept premières couches généalogiques et non significatives pour les couches subséquentes. Nous attribuons cette perte

de significativité à la forte augmentation de l'effet fondateur de la quatrième à la septième couche généalogique, tel que rapporté par Hamet, Merlo *et al.* [112]. L'augmentation de la concentration de contribution génétique dans les couches supérieures entraînerait donc une baisse de spécificité génétique. En effet, une augmentation de l'effet fondateur indique qu'un nombre plus restreint d'ancêtres contribuent à la majeure partie du pool génétique des descendants. Ces ancêtres sont donc fort probablement très peu spécifiques et il semble que les petits contributeurs n'ont pas une influence suffisamment importante pour compenser la perte de significativité statistique observée. Rappelons en effet que la spécificité génétique est représentative de la contribution différentielle d'un ancêtre pour deux ensembles de descendants. Un ancêtre contribuant fortement au pool génétique de tous les descendants a donc peu de chances d'être très spécifique à un sous-ensemble de descendants. En conséquence, la spécificité des ancêtres n'a plus d'influence significative sur l'IMC des descendants au-delà de la septième couche généalogique.

11.4.2 Pression sanguine diastolique corrigée

Tout comme pour l'IMC, les p-values simulées sont significatives pour les sept premières couches généalogiques et une perte de significativité est observée à la huitième couche généalogique ; les p-values corrigées se situent légèrement sous le seuil de 0.05 pour les couches subséquentes. Les p-values sont toutefois nettement plus significatives pour les premières couches généalogiques. Tel que mentionné plus haut, la perte de significativité graduelle à partir de la sixième couche généalogique peut être attribuée à l'augmentation de l'effet fondateur ou encore à une sélection en faveur des hétérozygotes.

Les analyses avec stratification sur l'obésité ont montré que la spécificité génétique des ancêtres n'avait une influence significative que sur la pression diastolique des des-

cendants non-obèses. Il est donc probable que la pression diastolique chez les individus obèses soit davantage influencée par des facteurs environnementaux non mesurés par notre approche que chez les sujets présentant un $IMC < 31$. En effet, comme nous l'avons mentionné à la section 7.1 l'hypertension associée à l'obésité semble être régie par des mécanismes pathophysiologiques différents de l'hypertension non-obèse. Il est donc probable que notre approche ne soit pas suffisamment puissante pour caractériser pleinement l'hypertension associée à l'obésité.

11.4.3 Pression sanguine systolique corrigée

Les p-values simulées ne sont pas significatives pour toutes les couches généalogiques lors des analyses sur la pression systolique corrigée pour la prise de médicaments anti-hypertensifs. La pression systolique est toutefois un phénotype fortement corrélé à l'âge des sujets [127]. Des analyses supplémentaires ont en effet révélé une influence significative de la spécificité des ancêtres pour les trois premières couches généalogiques chez les 55 ans et moins. Par contre, la spécificité des ancêtres ne semble avoir aucune influence sur la pression systolique des plus de 55 ans et ce pour toutes les couches généalogiques. Les différences observées chez les deux groupes d'âge s'expliquent peut-être par les mécanismes de transmission génétique sous-jacents.

En effet, nous avons vu à la section 4.2, qu'environ trois quarts des individus atteints d'hypertension, aux États-Unis, étaient âgés de 50 ans et plus. Nous savons aussi que l'hypertension essentielle, la forme la plus courante d'hypertension (environ 95% des cas), est une maladie à déclenchement tardif, influencée par une combinaison de facteurs génétiques et environnementaux [52, 13]. À l'opposé, l'hypertension monogénique est une maladie à déclenchement précoce, souvent plus sévère que l'hypertension essentielle et associée à des taux plus élevés de décès prématurés [136, 137, 138, 139]. Il est donc fort probable que le taux d'hypertension monogénique soit nettement plus élevé chez les moins de 55 ans. L'hérédité monogénique pourrait donc mieux carac-

tériser la transmission de la pression systolique chez les moins de 55 ans alors qu'un modèle d'hérédité multigénique, davantage influencé par l'environnement, serait mieux adapté aux 55 ans et plus.

Bien que certains facteurs environnementaux soient partagés au sein d'une descendance, ils n'entrent explicitement dans le calcul de la spécificité génétique. En conséquence, il est possible que notre approche ne soit pas suffisamment puissante pour caractériser la transmission de la pression systolique chez les 55 ans et plus en raison de l'importante influence des facteurs environnementaux dans ce groupe d'âge. Par contre, elle semble appropriée pour caractériser la transmission de la pression systolique chez les moins de 55 ans, chez qui les facteurs génétiques ont probablement une influence prépondérante.

Les analyses de stratification sur l'obésité pour la pression systolique ont de plus révélé que la transmission génétique ancestrale n'influencait pas significativement le phénotype des sujets obèses. Hormis pour les première et troisième couches généalogiques, elle semble par contre avoir un impact indéniable sur la pression systolique des sujets non-obèses. Tout comme pour la pression diastolique, nous suggérons que notre approche n'est pas assez puissante pour modéliser correctement la transmission de l'hypertension associée à l'obésité.

11.4.4 HDL

Les analyses de descendance sur la concentration sanguine de HDL ont été réalisées à titre de contrôle. Étant donné les résultats positifs obtenus avec la stratification sur l'âge et l'obésité pour la pression sanguine, nous avons voulu vérifier s'il était aussi possible de capturer l'effet d'autres facteurs environnementaux en stratifiant la population. Des analyses ont montré que la concentration sanguine de HDL est fortement dépendante du sexe des individus [127]. Les résultats de simulations montrent en ef-

fet que la spécificité des ancêtres a une influence significative sur les cinq premières couches généalogiques pour les hommes et non pour les femmes. Il semble donc que la stratification soit un moyen efficace pour capturer non seulement l'effet de l'âge et de l'obésité, mais aussi du sexe sur le phénotype.

Les différences observées entre les sexes sont peut-être dues à des mécanismes d'hérédité liés au chromosome X. En effet, alors que les femmes possèdent deux copies du chromosome X, les hommes n'en ont qu'une seule. Une femme porteuse d'un gène de maladie récessif lié à l'X peut donc être totalement asymptomatique, alors qu'un homme, n'ayant qu'une seule copie du gène en question, sera forcément atteint. Bien que la réalité soit plus nuancée (des femmes porteuses peuvent présenter des symptômes de maladie à divers degrés), l'hérédité récessive liée à l'X est caractérisée par un taux plus élevé d'hommes atteints que de femmes atteintes [140]. Une étude de O'Meara *et al.* portant sur 2356 sujets hypertendus a d'ailleurs montré que le taux dyslipidémie, entre autre caractérisée par une concentration de cholestérol HDL ≤ 0.9 mmol/L, était significativement plus élevé chez les hommes que chez les femmes [141]. De plus, Freedman *et al.*, ont déterminé, après l'analyse d'échantillons sanguins de 1574 hommes et 1692 femmes de l'étude de Framingham, que les femmes avait une concentration sanguine de HDL deux fois supérieure à celle des hommes [142]. Un gène récessif sur le chromosome X, dont l'effet serait masqué par une copie dominante chez les femmes porteuses, pourrait donc expliquer, du moins en partie, le fait que la concentration sanguine de HDL soit fortement liée au bagage génétique ancestral des hommes et non des femmes.

11.5 Validité de l'étude

Nous discuterons ici des menaces à la validité pouvant affecter nos résultats. Il s'agit des menaces à la validité *interne*, *externe*, *de construction* et *des conclusions* [143, 144, 145, 146].

11.5.1 Validité interne

Une étude possédera une validité interne si elle est en mesure de montrer que les variables dépendantes sont seulement le résultat des manipulations sur les variables indépendantes. Dans le cadre de cette étude, les variables indépendantes sont les spécificités des ancêtres et les variables dépendantes, les phénotypes des descendants. Tout facteur, autre que la spécificité génétique, qui influence les phénotypes des descendants constitue donc une menace à la validité interne de l'étude. Nous avons vu aux sections 4.2 et 5.2 qu'il existe en fait une multitude de facteurs environnementaux et sociaux influençant la pression sanguine et l'IMC qui ne sont pas explicitement mesurés avec la spécificité génétique ni pris en compte par la stratification sur l'âge et le sexe. Des erreurs ont aussi pu se glisser dans la base de données ou encore dans les applications informatiques développées, biaisant les résultats. Finalement, bien que deux candidats aient été exclus suite à des analyses de paternité (voir section 8.2), il n'est pas exclu que certaines erreurs subsistent, créant de fausses descendance paternelles.

11.5.2 Validité externe

La validité externe est représentative du caractère généralisable d'une étude, de son potentiel à être transposée dans différents contextes avec d'autres participants. Plusieurs facteurs menacent la validité externe de notre étude. D'abord, la population du SLSJ est particulière en soit. Affectée par un effet fondateur, on y retrouve une prévalence élevée de maladies génétiques. L'ataxie spastique récessive de Charlevoix-Saguenay se retrouve d'ailleurs uniquement dans cette population. Les résultats obtenus ne sont donc pas applicables à d'autres populations sans vérifications supplémentaires. Notre étude nécessite aussi l'accès à des données généalogiques, rarement disponibles pour de vastes populations. De plus, il existe un biais de recrutement clair en faveur des familles hypertendues. La présence d'au moins une fratrie hypertendue et dyslipidémique constituait le critère principal de sélection des familles pour la participation à l'étude (voir section 8.2 pour plus de détails). Les parents de premier et second degré

des fratries hypertendues étaient ensuite appelés à participer. Les taux d'hypertension observés ne sont donc probablement pas représentatifs de ceux de la population du SLSJ.

11.5.3 Validité de construction

La validité de construction représente la capacité des variables dépendantes et indépendantes à répondre correctement à l'hypothèse de recherche. Nous devons donc nous demander si l'étude des phénotypes des descendants en fonction de la spécificité de leurs ancêtres est un moyen valable pour caractériser le risque phénotypique d'un individu selon son ascendance. Il semble que ce soit clairement le cas puisque trois ensembles indépendants de descendants ont été construits en fonction de la spécificité des ancêtres avant d'évaluer les taux de personnes à risque de chaque ensemble. Nous qualifions donc le risque phénotypique des individus en fonction de leur ascendance. De plus, les définitions de spécificité génétique et de descendance ont été présentées. Nos procédures sont donc totalement impartiales et permettent de vérifier notre hypothèse de recherche.

11.5.4 Validité des conclusions statistiques

Une étude sera valide au plan de ses conclusions statistiques si des mesures ont été prises afin de contrer les erreurs de types I (rejet erroné de l'hypothèse nulle) et de type II (échec à rejeter l'hypothèse nulle). Les analyses ANOVA sont paramétriques, en ce sens qu'elles requièrent des données distribuées normalement. Nous avons vu aux sections 4.1 et 5.1 que la pression sanguine et l'indice de masse corporelle respectent cette condition dans la population générale. Toutefois, notre étude porte sur un échantillon de la population du SLSJ et la normalité des données n'a pas été vérifiée. Les p-values ont par contre été corrigées pour les tests multiples par simulations. Les analyses ont été répétées 1000 fois pour chaque phénotype et couche généalogique. L'hypothèse nulle était rejetée lorsque la p-value corrigée était inférieure à 0.05.

12. Conclusions et perspectives

Cette étude montre qu'il est possible d'estimer le risque d'hypertension et d'obésité des individus selon leur ascendance. Alors que les études d'historique familial, se limitent à l'étude des parents de premier et second degré seulement, l'analyse de descendances est en mesure de considérer les ancêtres de toutes les couches généalogiques. Des tests de simulation ont montré la pertinence de notre analyse en confirmant la significativité de l'influence des ancêtres au-delà des deux premières couches généalogiques.

Trois classes de risques (élevé, moyen et bas) ont été construites pour chaque phénotype investigué. L'appartenance à l'une ou l'autre de ces classes était déterminée selon la spécificité génétique des ancêtres. Les taux d'individus à risque et hypertendus en fonction des classes de risque ont aussi été rapportés. Un test χ^2 permettait de confirmer la significativité statistique des différences de proportions d'individus à risque entre les classes. Des tests de simulation ont permis de confirmer l'impact de la spécificité génétique des ancêtres sur les descendants et nous avons montré que nous sommes en mesure de considérer d'autres facteurs tels l'âge et le sexe dans nos analyses.

Nous sommes convaincus que notre approche constitue un complément intéressant à l'historique familial qui mériterait d'être validée sur de grandes cohortes, sans biais de sélection, afin de contrer les menaces à la validité soulevées à la section 11.5. Elle pourrait alors constituer un outil majeur en santé publique, dans le cadre d'études sur des populations à risque ou visant à comparer deux groupes d'individus, l'un à risque

élevé et l'autre à risque faible par exemple.

Notre méthode pourrait aussi être appliquée dans le cadre d'études de génotypage à haut débit, en vue de comparaisons à grande échelle d'haplotypes ou de SNPs. Contraster les génotypes d'individus de phénotype élevé et bas pourrait en effet révéler d'intéressantes différences au niveau de certains traits génétiques et mener à des investigations plus poussées.

Bibliographie

- [1] Nicholas A. Christakis and James H. Fowler. The Spread of Obesity in a Large Social Network over 32 Years. *N Engl J Med*, 357(4) :370–379, 2007.
- [2] Heyer E and Tremblay M. Variability of the genetic contribution of Quebec population founders associated to some deleterious genes. *Am J Hum Genet*, 56(4) :970–978, 1995.
- [3] Bouchard G, Laberge C, and Scriver C. Reproduction démographique et transmission génétique dans le nord-est de la province du Québec (18ième-20ième siècle). *Eur J Population*, 4 :39–67, 1988.
- [4] Bouchard G and De Braekeleer M. *Histoire d'un génome*. Presses de l'Université du Québec, 1991.
- [5] Labuda M, Labuda D, Korab-Laskowska M, Cole D, Zietkiewicz E, Weissenbach J, Popowska E, Pronicka E, Root A, and Glorieux F. Linkage disequilibrium analysis in young populations : Pseudo-vitamin d-deficiency rickets and the founder effect in French Canadians. *Am J Hum Genet*, 59 :633–643, 1996.
- [6] Labuda D, Zietkiewicz E, and Labuda M. The genetic clock and the age of the founder effect in growing populations : A lesson from French Canadians and Ashkenazim. *Am J Hum Genet*, 61 :768–771, 1997.
- [7] Tremblay M, Lavoie E M, Houde L, and Vézina H. Demogenetic study of three populations within a region with strong founder effect. *Eur J Hum Genet*, 10 :183, 2002.
- [8] Vézina H Jomphe M, Tremblay M. Analyses généalogiques à partir du fichier RETRO. Document 1-c-204, Projet BALSAC, Chicoutimi, 2001.
- [9] Bouchard G, Roy R, Casgrain B, and Hubert M. [Population files and database management : the BALSAC database and the INGRES/INGRID system]. *Hist Mes*, 4 :39–57, 1989.
- [10] Scriver CR. Human genetics : lessons from Quebec populations. *Ann Rev Genomics Hum Genet*, 2 :69–101, 2001.
- [11] Heyer E, Toupance B, Perri C, De Vito O, Foncin J-F, and Bruni AC. Manic depressive illness in a founder population. *Eur J Hum Genet*, 11(8) :597–602, 2003.
- [12] Z. Pausova, D. Gaudet, F. Gossard, M. Bernard, M.L. Kaldunski, M. Jomphe, J. Tremblay, T.J. Hudson, G. Bouchard, T.A. Kotchen, et al. Genome-Wide Scan for Linkage to Obesity-Associated Hypertension in French Canadians. *Hypertens*, 46 :1280–1285, 2005.
- [13] Hamet P, Pausova Z, Adarichev S, and Tremblay J. Hypertension : genes and environment. *J Hypertens*, 16 :379–418, 1998.

- [14] Hunt S, Gwinn M, and Adams T. Family history assessment : Strategies for prevention of cardiovascular disease. *Am J Preventive Medicine*, 24(2) :136–142, 2003.
- [15] P.W. Yoon, M.T. Scheuner, K.L. Peterson-Oehlke, M. Gwinn, A. Faucett, and MJ Khoury. Can family history be used as a tool for public health and preventive medicine. *Genet Med*, 4(4) :304–10, 2002.
- [16] M.T. Scheuner, W.C. Whitworth, H. McGruder, P.W. Yoon, and M.J. Khoury. Expanding the definition of a positive family history for early-onset coronary heart disease. *Genet Med*, 8(8) :491–501, 2006.
- [17] Scheuner M, Wang S-J, Raffel L, Larabell S, and Rotter J. Family history : A comprehensive genetic risk assessment method for the chronic conditions of adulthood. *Am J Med Genet*, 71 :315–324, 1997.
- [18] Murff H, Peterson N, Greevy R, and Zheng W. Impact of patient age on family cancer history. *Genet in Medicine*, 8(7) :438–442, 2006.
- [19] T. Kondo, H. Toyoshima, Y. Tsuzuki, Y. Hori, H. Yatsuya, K. Tamakoshi, A. Tamakoshi, Y. Ohno, et al. Familial aggregation and coaggregation of history of hypertension and stroke. *J Hum Hypertens*, 19 :119–125, 2005.
- [20] I.B. Goldstein, D. Shapiro, and D. Guthrie. Ambulatory Blood Pressure and Family History of Hypertension in Healthy Men and Women. *Am J Hypertens*, 19(5) :486–491, 2006.
- [21] A.P.M. van den Elzen, M.A.J. de Ridder, D.E. Grobbee, A. Hofman, J.C.M. Witteman, and C.S.P.M. Uiterwaal. Families and the natural history of blood pressure A 27-year follow-up study. *Am J Hypertens*, 17(10) :936–940, 2004.
- [22] V. Lascaux-Lefebvre, JB Ruidavets, D. Arveiler, P. Amouyel, B. Haas, D. Cottel, A. Bingham, P. Ducimetiere, and J. Ferrieres. Influence of parental history of hypertension on blood pressure. *J Hum Hypertens*, 13(9) :631–636, 1999.
- [23] HS Park, KS Yim, and SI Cho. Gender differences in familial aggregation of obesity-related phenotypes and dietary intake patterns in Korean families. *Annals of Epidemiol*, 14(7) :486–91, 2004.
- [24] M.B. Neutzling, J.A.A.C. Taddei, and D.P. Gigante. Risk factors of obesity among Brazilian adolescents : a case–control study. *Public Health Nutrition*, 6(08) :743–749, 2003.
- [25] KO Hajian-Tilaki and B. Heidari. Prevalence of obesity, central obesity and the associated factors in urban population aged 20–70 years, in the north of Iran : a population-based study and regression approach. *Obesity Reviews*, 8(1) :3–10, 2006.
- [26] M.A.B. van der Sande, G.E.L. Walraven, P.J.M. Milligan, W.A.S. Banya, S.M. Ceesay, O.A. Nyan, and K.P.W.J. McAdam. Family history : an opportunity for early interventions and improved control of hypertension, obesity and diabetes. *Bulletin of the World Health Organization*, 79 :321–328, 2001.
- [27] LB Henderson and MT Scheuner. Suboptimal family history collection : an obstacle to genetic risk assessment and prevention. *Am J Hum Genet*, 63 :A240, 1998.

- [28] L.S. Acheson, G.L. Wiesner, S.J. Zyzanski, M.A. Goodwin, and K.C. Stange. Family history-taking in community family practice : implications for genetic screening. *Genet Med*, 2(3) :180–185, 2000.
- [29] Johnson J, Giles RT, Larsen L, Ware J, Adams T, and Hunt SC. Utah’s family high risk program : bridging the gap between genomics and public health. *Preventing Chronic Disease [serial online]*, 2(2) :1–7, 2005.
- [30] A. Vigneault, D. Brisson, C. BÃ©langer, and D. Gaudet. Community Genetics in Eastern Quebec : The Experience of the Corporation for Research and Action on Hereditary Diseases. *Community Genetics*, 3(3) :151–155, 2000.
- [31] D. Gaudet, CM. Laberge, and J. Simard. *Community genetics : Past and future.*, pages 49–59. Utrecht University, 2002.
- [32] Joffres M, Hamet P, MacLean D, L’Italien G, and Fodor G. Distribution of blood pressure and hypertension in canada and the united states. *Am J Hypertens*, 14 :1099–1105, 2001.
- [33] Franklin S, Jacobs M, Wong N, L’Italien G, and Lapuerta P. Predominance of isolated systolic hypertension among middle-aged and elderly us hypertensives : Analysis based on national health and nutrition examination survey (nhanes) iii. *Hypertens*, 37 :869–874, 2001.
- [34] Burt VL, Whelton P, Roccella EJ, Brown C, Cutler JA, and Higgins M. Prevalence of hypertension in the us adult population. results from the third national health and nutrition examination survey, 1988-1991. *Hypertens*, 25 :305–313, 1995.
- [35] Chobanian A, Bakris G, Black H, Cushman W, Green L, Izzo J, Jones D, Materson B, Oparil S, Wright J, and Roccella E. Seventh report of the joint national committee on prevention, detection, evaluation, and treatment of high blood pressure. *Hypertens*, 42 :1206–1252, 2003.
- [36] Beevers G, Lip G, and O’Brien E. *ABC of Hypertension*. BMJ Publishing Group, fourth edition edition, 2001.
- [37] Grim CE, Luft FC, and Millar JZ. Racial differences in blood pressure in evans county, georgia : relationship to sodium and potassium intake and plasma renin activity. *J Chronic Disease*, 33 :87–94, 1980.
- [38] VM Campese. Salt sensitivity in hypertension. Renal and cardiovascular implications [clinical conference]. *Hypertens*, 23(4) :531–550, 1994.
- [39] F. Cuzzola, F. Mallamaci, G. Tripepi, S. Parlongo, S. Cutrupi, A. Cataliotti, B. Stancanelli, L. Malatino, I. Bellanuova, C. Ferri, et al. Urinary adrenomedullin is related to ET-1 and salt intake in patients with mild essential hypertension. Salt Sensitivity Group of Italian Society of Hypertension. *Am J Hypertens*, 14(3) :224–30, 2001.
- [40] J Stamler. The INTERSALT Study : background, methods, findings, and implications [published erratum appears in *Am J Clin Nutr* 1997 Nov ;66(5) :1297]. *Am J Clin Nutr*, 65(2) :626S–642, 1997.
- [41] Higgins AL, Kannel W, Garrison R, Pinsky J, and Stokes JJ. Hazards of obesity : the Framingham experience. *Acta medica scandinavica supplement*, 723 :23–36, 1988.

- [42] Staessen J, Fagard R, and Amery A. The relationship between body weight and blood pressure. *J Hum Hypertens*, 2 :207–219, 1988.
- [43] Huang Z, Willett WC, Manson JE, Rosner B, Stampfer MJ, Speizer FE, and Colditz GA. Body weight, weight change and risk for hypertension in women. *Annals of internal medicine*, 128 :81–88, 1998.
- [44] Shaper AG, Wannamethee G, and Whincup PH. Alcohol and blood pressure in middle aged british men. *J Hum Hypertens*, 2 :71–78, 1988.
- [45] P. Hamet, E. Mongeau, J. Lambert, F. Bellavance, M. Daignault-Gelinas, M. Ledoux, and L. Whissell-Cambiotti. Interactions among calcium, sodium, and alcohol intake as determinants of blood pressure. *Hypertens*, 17(1 Suppl) :I150–4, 1991.
- [46] P. Hamet, M. Daignault-Gelinas, J. Lambert, M. Ledoux, L. Whissell-Cambiotti, F. Bellavance, and E. Mongeau. Epidemiological evidence of an interaction between calcium and sodium intake impacting on blood pressure. A Montreal study. *Am J Hypertens*, 5(6 Pt 1) :378–85, 1992.
- [47] GM. Reaven, H. Lithell, and L. Landsberg. Hypertension and associated metabolic abnormalities : the role of insulin resistance and the sympathoadrenal system. *NEJM*, 334 :374–381, 1996.
- [48] S. Oparil, M.A. Zaman, and D.A. Calhoun. Pathogenesis of Hypertension. *Ann Intern Med*, 139(9) :761–776, 2003.
- [49] R.O. Halperin, H.D. Sesso, J. Ma, J.E. Buring, M.J. Stampfer, and J. Michael Gaziano. Dyslipidemia and the Risk of Incident Hypertension in Men. *Hypertens*, 47(1) :45–50, 2006.
- [50] SM Haffner, H. Miettinen, SP Gaskill, and MP Stern. Metabolic precursors of hypertension. The San Antonio Heart Study. *Arch Intern Med*, 156(17) :1994–2001, 1996.
- [51] SC Hunt, SH Stephenson, PN Hopkins, and RR Williams. Predictors of an increased risk of future hypertension in Utah. A screening analysis. *Hypertens*, 17(6) :969–976, 1991.
- [52] Cowley AW. The genetic dissection of essential hypertension. *Nature reviews Genet*, 7 :829–840, 2006.
- [53] Gong M and Hubner N. Molecular genet of human hypertension. *Clinical science*, 110 :315–326, 2006.
- [54] Ruppert V and Maisch B. Genet of human hypertension. *Herz*, 8 :655–662, 2003.
- [55] Harrison W and Marshall T. The epidemiology of blood pressure in East Asia. *J Hum Hypertens*, 20 :97–99, 2006.
- [56] Eckel RH, York DA, Rossner S, Hubbard V, Caterson I, St Jeor ST, Hayman LL, Mullis Rm, and Blair SN. American heart association. prevention conference vii : obesity, a worldwide epidemic related to heart disease and stroke : executive summary. *Circulation*, 110 :2968–2975, 2004.
- [57] World health organization. Obesity : Preventing and managing the global epidemic. WHO Technical report series 894, World health organization, 2000.

- [58] Ezzati M, Lopez AD, Rodgers A, Vander HS, Murray CJL, and Comparative risk assesment collaborative group. Selected major risks factors and global and regional burden of disease. *Lancet*, 360 :1347–1360, 2002.
- [59] Flegal KM, Carroll MD, Kuczmarski RJ, and Johnson CL. Overweight and obesity in the United States : prevalence and trends. *Intl J Obes Related Metabolic Disorders*, pages 39–47, 1998.
- [60] Tjepkema M. Adult obesity in canada : measured height and weight. In *Component of Statistics Canada Catalogue*, volume 1 of *Nutrition : Findings from the Canadian community health survey*, pages 1–32. Statistics Canada, 2005.
- [61] Cynthia L. Ogden, Margaret D. Carroll, Lester R. Curtin, Margaret A. McDowell, Carolyn J. Tabak, and Katherine M. Flegal. Prevalence of Overweight and Obesity in the United States, 1999-2004. *JAMA*, 295(13) :1549–1555, 2006.
- [62] Flegal KM, Carroll MD, Ogden CL, and Johnson CL. Prevalence and trends in obesity among us adults. *J Am Med Assoc*, 288 :1723–1727, 2002.
- [63] National institutes of health. Clinical guidelines on the identification, evaluation and treatment of overweight and obesity in adults : the evidence report. *Obesity Reasearch*, Suppl 2 :51S–209S, 1998.
- [64] Olshansky SJ, Passaro DJ, Hershov RC, Layden J, Cames BA, Brody J, Hayflick L, Butler RN, Allison DB, and Ludwig DS. A potential decline in life expectancy in the United States in the 21st century. *NEJM*, 352 :1138–1145, 2005.
- [65] Fontaine KR, Redden DT, Wang C, Westfall AO, and Allison DB. Years of life lost due to obesity. *JAMA*, 289 :187–193, 2003.
- [66] Paul Poirier, Thomas D. Giles, George A. Bray, Yuling Hong, Judith S. Stern, F. Xavier Pi-Sunyer, and Robert H. Eckel. Obesity and cardiovascular disease : Pathophysiology, evaluation, and effect of weight loss : An update of the 1997 american heart association scientific statement on obesity and heart disease from the obesity committee of the council on nutrition, physical activity, and metabolism. *Circulation*, 113(6) :898–918, 2006.
- [67] Ezzati M, Martin H, Skjold S, Hoorn SV, and Murray CJL. Trends in national and state-level obesity in the usa after correction for self-report bias : analysis of health surveys. *J royal society of medicine*, 99 :250–257, 2006.
- [68] WHO. The world health report 1998. life in the 21st century. a vision for all. Technical report, WHO, Genève, 1998.
- [69] Bray GA and Bouliard C. *Handbook of obesity : Etiology and pathophysiology*. Marcel Dekker, second edition edition, 2004.
- [70] Seidell JC and Flegal KM. Assessing obesity : classification and epidemiology. *British Med bulletin*, 53 :238–252, 1997.
- [71] Lubrano-Berthelie C and Clément K. Génétique de l'obésité. *La revue de médecine interne*, 26 :802–811, 2005.
- [72] Z. Pausova, B. Deslauriers, D. Gaudet, J. Tremblay, T.A. Kotchen, P. Larochelle, A.W. Cowley, and P. Hamet. Role of Tumor Necrosis Factor- α Gene Locus in Obesity and Obesity-Associated Hypertension in French Canadians. *Hypertens*, 36(1) :14–19, 2000.

- [73] Barsh GS, Farooqi IS, and O’Rahilly S. Genet of body-weight regulation. *Nature*, 404 :644–651, 2000.
- [74] Clement K, Vaisse C, Lahlou N, Cabrol S, Pelloux V, and Xassuto D. A mutation in the human leptin receptor gene causes obesity and pituitary dysfunction. *Nature*, 392 :398–401, 1998.
- [75] Montague CT, Farooqi IS, Whitehead JP, Soos MA, Rau H, and Wareham NJ. Congenital leptin deficiency is associated with severe early-onset obesity in humans. *Nature*, 387 :903–908, 1997.
- [76] Swinburn B and Egger G. Preventive strategies against weight gain and obesity. *Obesity reviews*, 3 :289–301, 2002.
- [77] Egger G and Swinburn B. An ecological approach to the obesity pandemic. *British Med journal*, 315 :477–480, 1997.
- [78] Duncan C. Thomas. *Statistical methods in genetic epidemiology*. Oxford university press, 198 Madison avenue New York, 2004.
- [79] Kempthorne O. *An introduction to genetic statistics*. Wiley publications in statistics, New York, 1957.
- [80] Kotchen TA, Kotchen JM, Grim CE, George V, Kaldunski ML, Cowley AW, Hamet P, and Chelius TH. Genetic determinants of hypertension : identification of candidate phenotypes. *Hypertens*, 36(1) :7–13, 2000.
- [81] R. Fagard, J. Brguljan, J. Staessen, L. Thijs, C. Derom, M. Thomis, and R. Vlietinck. Heritability of Conventional and Ambulatory Blood Pressures A Study in Twins. *Hypertens*, 26(6) :919–924, 1995.
- [82] GW Somes, GA Harshfield, BS Alpert, MM Goble, and RM Schieken. Genetic influences on ambulatory blood pressure patterns. The Medical College of Virginia Twin Study. *Am J Hypertens*, 8(5 Pt 1) :474–8, 1995.
- [83] W.J. Vinck, R.H. Fagard, R. Loos, and R. Vlietinck. The impact of genetic and environmental influences on blood pressure variance across age-groups. *J Hypertens*, 19(6) :1007–1013, 2001.
- [84] R.H. Fagard, RJ Loos, G. Beunen, C. Derom, and R. Vlietinck. Influence of chorionicity on the heritability estimates of blood pressure : a study in twins. *J Hypertens*, 21(7) :1313–8, 2003.
- [85] JG Mongeau, P. Biron, and CF Sing. The influence of genetics and household environment upon the variability of normal blood pressure : the Montreal Adoption Survey. *Clin Exp Hypertens A*, 8(4-5) :653–60, 1986.
- [86] Kupper N, Willemsen G, Riese H, Posthuma D, Boomsma DL, and de Geus EJC. Heritability of daytime ambulatory blood pressure in an extended twin design. *Hypertens*, 45 :80–85, 2005.
- [87] C. Fava, P. Burri, P. Almgren, L. Groop, U.L. Hulthen, and O. Melander. Heritability of ambulatory and office blood pressure phenotypes in Swedish families. *J Hypertens*, 22(9) :1717–1721, 2004.
- [88] Hottenga J-J, Boomsma DI, Kupper N, Posthuma D, Snieder H, Willemsen G, and de Geus EJC. Heritability and stability of resting blood pressure. *Twin research and Hum Genet*, 8(5) :499–508, 2005.

- [89] Pilia G, Chen WM, Scuteri A, Orru M, Albai G, Dei M, Lai S, Usala G, Lai M, Loi P, Mameli C, Vacca L, Deiana M, Olla N, Masala M, Cao A, Najjar SS, Terracciano A, Nedorezov T, Sharov A, Zonderman AB, Abecasis GR, Costa P, Lakatta E, and Schlessinger D. Heritability of cardiovascular and personality traits in 6,148 Sardinians. *Plos Genet*, 2(8) :1207–1223, 2006.
- [90] Bochud M, Bovet P, Elston RC, Paccaud F, Falconnet C, Maillard M, Shamlaye C, and Burnier M. High heritability of ambulatory blood pressure in families of East African descent. *Hypertens*, 45 :445–450, 2005.
- [91] T. Rice, L. Perusse, C. Bouchard, and DC Rao. Familial aggregation of body mass index and subcutaneous fat measures in the longitudinal Quebec family study. *Genet Epidemiol*, 16(3) :316–334, 1999.
- [92] GP Vogler, TI Sorensen, AJ Stunkard, MR Srinivasan, and DC Rao. Influences of genes and shared family environment on adult body mass index assessed in an adoption study by a comprehensive path model. *Int J Obes Relat Metab Disord*, 19(1) :40–5, 1995.
- [93] IB Borecki, M. Higgins, PJ Schreiner, DK Arnett, E. Mayer-Davis, SC Hunt, and MA Province. Evidence for multiple determinants of the body mass index : the National Heart, Lung, and Blood Institute Family Heart Study. *Obes Res*, 6(2) :107–14, 1998.
- [94] MW Knuiaman, ML Divitini, TA Welborn, and HC Bartholomew. Familial correlations, cohabitation effects, and heritability for cardiovascular risk factors. *Ann Epidemiol*, 6(3) :188–94, 1996.
- [95] Stunkard, AJ and Pedersen, NL and McClearn, GE. The body-mass index of twins who have been reared apart. *NEJM*, 322 :1483–1487, 1990.
- [96] C. Bouchard, A. Tremblay, JP Despres, A. Nadeau, PJ Lupien, G. Theriault, J. Dussault, S. Moorjani, S. Pinault, and G. Fournier. The response to long-term overfeeding in identical twins. *NEJM*, 322(21) :1477–1482, 1990.
- [97] X. Li, M.J. Quinones, D. Wang, I. Bulnes-Enriquez, X. Jimenez, R. De La Rosa, G.L. Aurea, K.D. Taylor, W.A. Hsueh, J.I. Rotter, et al. Genetic Effects on Obesity Assessed by Bivariate Genome Scan : The Mexican-American Coronary Artery Disease Study. *Obesity Research*, 14(7) :1192, 2006.
- [98] R.A. Bayoumi, S.A.S. Al-Yahyaee, S.A. Albarwani, S.G. Rizvi, S. Al-Hadabi, F.F. Al-Ubaidi, A.T. Al-Hinai, M.N. Al-Kindi, H.T. Adnan, H.S. Al-Barwany, et al. Heritability of Determinants of the Metabolic Syndrome among Healthy Arabs of the Oman Family Study. *Obesity Research*, 15(3) :551, 2007.
- [99] N.F. Butte, G. Cai, S.A. Cole, and A.G. Comuzzie. Viva la Familia Study : genetic and environmental contributions to childhood obesity and its comorbidities in the Hispanic population. *Am J Clinical Nutrition*, 84(3) :646, 2006.
- [100] A. Luke, X. Guo, AA Adeyemo, R. Wilks, T. Forrester, W. Lowe Jr, AG Comuzzie, LJ Martin, X. Zhu, CN Rotimi, et al. Heritability of obesity-related traits among Nigerians, Jamaicans and US black people. *Intl J Obes*, 25 :1034–1041, 2001.
- [101] Kannel WB, Brand N, Skinner JJ, Dawber TR, and McNamara PM. The relation of adiposity to blood pressure and development of hypertension. the Framingham study. *Annals of internal medicine*, 67 :48–59, 1967.

- [102] Joffres MR, Hamet P, Rabkin SW, Gelskey D, Hogan K, and Fodor G. Prevalence, control and awareness of high blood pressure among Canadian adults. *J Canadian Med Assoc*, 146 :1997–2005, 1992.
- [103] Kotchen TA, Kotchen JM, and Boegehold MA. Nutrition and hypertension prevention. *Hypertens*, 18(suppl I) :115I–120I, 1991.
- [104] Stamler R, Stamler J, Riedlinger WF, Algera G, and Roberts R. Weight and blood pressure : findings in hypertension screening of 1 million americans. *JAMA*, 240 :1607–1609, 1978.
- [105] Sharma AM and Grassi G. Obesity and hypertension : cause or consequence? *J Hypertens*, 19 :2125–2156, 2001.
- [106] Meigs JB. Insulin resistance syndrome? syndrome x? multiple metabolic syndrome? a syndrome at all? factor analysis reveals patterns in the fabric of correlated metabolic risk factors. *Am J Epidemiol*, 152 :908–911, 2000.
- [107] Shmulewitz D, Auerbach SB, Lehner T, Bundell ML, Winick JD, Youngman LD, Skilling V, Heath SC, Ott J, Stoffel M, Breslow JL, and Friedman JM. Epidemiology and factor analysis of obesity, type ii diabetes, hypertension and dyslipidemia (syndrome x) on the island of krosrae, Federated States of Micronesia. *Hum Heredity*, 51 :8–19, 2001.
- [108] Cui J, Hopper J L, and Harrap S B. Genes and family environment explain correlations between blood pressure and body mass index. *Hypertens*, 40 :7–12, July 2002.
- [109] Choh AC, Gage TB, ST McGarvey, and Comuzzie AG. Genetic and environmental correlations between various anthropometric and blood pressure traits among adults Samoans. *Am J physical anthropology*, 115 :304–311, 2001.
- [110] Pausova, Z. and Gossard, F. and Gaudet, D. and Tremblay, J. and Kotchen, T.A. and Cowley, A.W. and Hamet, P. Heritability Estimates of Obesity Measures in Siblings With and Without Hypertension. *Hypertens*, 38 :41–47, 2001.
- [111] M. De Braekeleer and S. Gauthier. Autosomal recessive disorders in Saguenay-Lac-St-Jean, Quebec : study of kinship. *Hum Biol*, 68(3) :371–81, 1996.
- [112] Hamet P, Merlo E, Seda O, Tremblay J, Kaldunsky M, Gaudet D, Bouchard G, Deslauriers B, Gagnon F, Antoniol G, Kotchen T. A, Pausova Z, Labuda M, Jomphe M, Gossard F, Kirova R, Tonellato P, Orlov S. N., Pintos J, Rioux J. D, Platko J, Hudson T. J, and Cowley A. W. Jr. Quantitative founder effect analysis of french-canadian families identifies specific alleles contributing to metabolic phenotypes of hypertension. *Am J Hum Genet*, 2005.
- [113] T.A. Kotchen, U. Broeckel, C.E. Grim, P. Hamet, H. Jacob, M.L. Kaldunski, J.M. Kotchen, N.J. Schork, P.J. Tonellato, and A.W. Cowley. Identification of Hypertension-Related QTLs in African American Sib Pairs. *Hypertens*, 40(5) :634–639, 2002.
- [114] Z. Pausova, M. Jomphe, L. Houde, H. Vezina, S.N. Orlov, F. Gossard, D. Gaudet, J. Tremblay, T.A. Kotchen, A.W. Cowley, et al. A Genealogical Study of Essential Hypertension with and without Obesity in French Canadians. *Obesity Research*, 10(6) :463–470, 2002.

- [115] WG McAdoo, MH Weinberger, JZ Miller, NS Fineberg, and CE Grim. Race and gender influence hemodynamic responses to psychological and physical stimuli. *J Hypertens*, 8(10) :961–7, 1990.
- [116] Rioux JD, Stone VA, Daly MJ, Cargill M, Green T, Nguyen H, Nutman T, Zimmerman PA, Tucker MA, Hudson T, Goldstein AM, Lander E, and Lin AY. Familial eosinophilia maps to the cytokine gene cluster on human chromosomal region 5q31-q33. *Am J Hum Genet*, 63 :1086–1094, 1998.
- [117] Pirinen M and Gasbarra D. Finding consistent gene transmission patterns on large and complex pedigrees. *IEEE/ACM transactions on computational biology and bioinformatics*, 3(3) :252–262, 2006.
- [118] O’Connell J R and Weeks D E. An optimal algorithm for automatic genotype elimination. *Am J Hum Genet*, 65 :1733–1740, 1999.
- [119] Barabási A-L and Oltvai Z. Network biology : Understanding the cell’s functional organization. *Nat Rev Gen*, 5 :101–113, 2004.
- [120] Stewart I. Networking opportunity. *Nature*, 427 :601–604, 2004.
- [121] Subdø J, Bankfalvi A, Bryne M, Marcelpoil R, Boysen M, Piffko J, Hemmer J, Kraft K, and Reith A. Prognostic value of graph theory-based tissue architecture analysis in carcinomas of the tongue. *Lab Invest*, 80(12) :1881–1889, 2000.
- [122] Lindorff-Larsen K, Vendruscolo M, Paci E, and Dobson CM. Transition states for protein folding have native topologies despite high structural variability. *Nat Struct Molec Biol*, 11(5) :443–449, 2004.
- [123] Merlo E., Deslauriers B., Antoniol G., Brunelle P. L., Jomphe M., Bouchard G., Tremblay J., and Hamet P. Layered Founders : A Novel Approach to Investigate the Ancestral Transmission of Complex Traits. Technical Report EPM-RT-2004-05, École Polytechnique de Montréal, June 2004.
- [124] R Development Core Team. *R : A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2006. ISBN 3-900051-07-0.
- [125] Cormen TH, Leiserson CE, and Rivest RL. *Introduction to Algorithms*. The MIT Press, Cambridge, 1990.
- [126] Neaton J D, Grimm R H, Prineas R J, Stamler J, Grandits G A, Elmer P J, Cutler J A, Flack J M, Schoenberger J A, and McDonald R. Treatment of mild hypertension study. final results. *JAMA*, 270 :713–724, 1993.
- [127] Hamet P., Šeda O., Gaudet D., Brunelle P.-L., Gurau A., Merlo E., Pilote L., Kotchen T., Cowley A., and Tremblay J. Systematic search for sex-specific genomic determinants of 608 cardiovascular traits in French Canadians reveals novel loci/haplotypes. In *Abstract book*, page 111, Montreal, 2007. Human Genome Meeting.
- [128] M. Aidoo, D.J. Terlouw, M.S. Kolczak, P.D. McElroy, F.O. ter Kuile, S. Kariuki, B.L. Nahlen, A.A. Lal, and V. Udhayakumar. Protective effects of the sickle cell gene against malaria morbidity and mortality. *Lancet*, 359(9314) :1311–1312, 2002.
- [129] M. Krawczak and J. Zschocke. A role for overdominant selection in phenylketonuria? Evidence from molecular data. *Hum Mutation*, 21(4) :394–397, 2003.

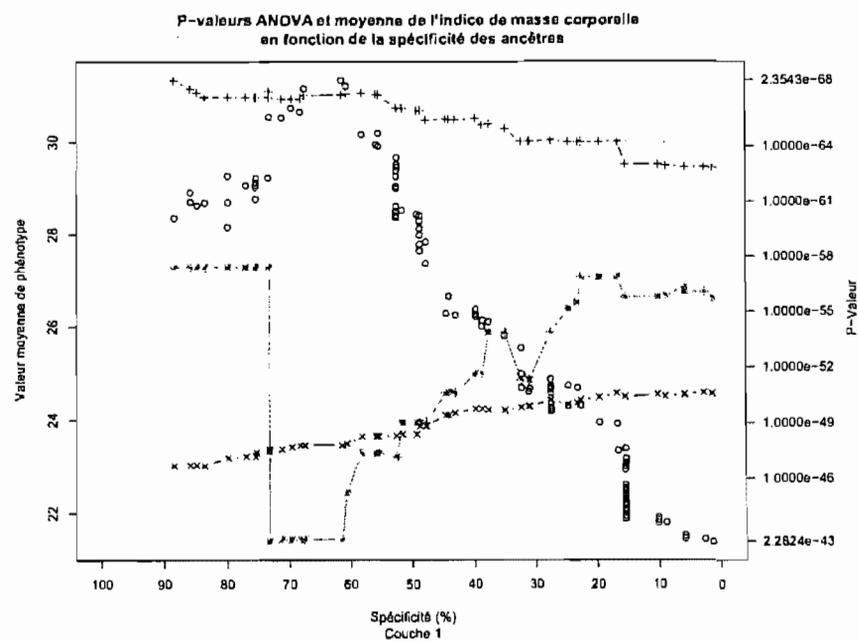
- [130] J. Zschocke. Phenylketonuria mutations in Europe. *Hum Mutation*, 21(4) :345–356, 2003.
- [131] SE Gabriel, KN Brigman, BH Koller, RC Boucher, and MJ Stutts. Cystic fibrosis heterozygote resistance to cholera toxin in the cystic fibrosis mouse model. *Science*, 266(5182) :107, 1994.
- [132] J.L. Bobadilla, M. Macek, J.P. Fine, and P.M. Farrell. Cystic fibrosis : A worldwide analysis of CFTR mutations - Correlation with incidence data and application to screening. *Hum Mutation*, 19(6) :575–606, 2002.
- [133] F. Austerlitz, L. Kalaydjieva, and E. Heyer. Detecting Population Growth, Selection and Inherited Fertility From Haplotypic Data in Humans. *Genet*, 165(3) :1579–1586, 2003.
- [134] M. Slatkin. A Population-Genetic Test of Founder Effects and Implications for Ashkenazi Jewish Diseases. *Am J Hum Genet*, 75(2) :282–293, 2004.
- [135] J. Charrow. Ashkenazi Jewish genetic disorders. *Familial Cancer*, 3(3) :201–206, 2004.
- [136] G.T. McMahon and R.G. Dluhy. Glucocorticoid-remediable aldosteronism. *Cardiol Rev*, 12(1) :44–8, 2004.
- [137] M.I. New, D.S. Geller, F. Fallo, and R.C. Wilson. Monogenic low renin hypertension. *Trends Endocrinol Metab*, 16(3) :92–7, 2005.
- [138] W.R. Litchfield, B.F. Anderson, R.J. Weiss, R.P. Lifton, and R.G. Dluhy. Intracranial Aneurysm and Hemorrhagic Stroke in Glucocorticoid-remediable Aldosteronism. *Hypertens*, 31(1) :445–450, 1998.
- [139] GM Rich, S. Ulick, S. Cook, JZ Wang, RP Lifton, and RG Dluhy. Glucocorticoid-remediable aldosteronism in a large kindred : clinical spectrum and diagnosis using a characteristic biochemical phenotype. *Ann Intern Med*, 116(10) :813–20, 1992.
- [140] R.L. Nussbaum, R.R. McInnes, and H.F. Willard. *Thompson & Thompson Genetics in Medicine*. WB Saunders Co, 2004.
- [141] J.G. O'Meara, S.L.R. Kardia, J.J. Armon, C.A. Brown, E. Boerwinkle, and S.T. Turner. Ethnic and Sex Differences in the Prevalence, Treatment, and Control of Dyslipidemia Among Hypertensive Adults in the GENOA Study. *Arch Intern Med*, 164(12) :1313–1318, 2004.
- [142] D.S. Freedman, J.D. Otvos, E.J. Jeyarajah, I. Shalurova, L.A. Cupples, H. Parise, R.B. D'Agostino, P.W.F. Wilson, and E.J. Schaefer. Sex and Age Differences in Lipoprotein Subclasses Measured by Nuclear Magnetic Resonance Spectroscopy : The Framingham Study. *Clinical Chemistry*, 50(7) :1189–1200, 2004.
- [143] Wohlin C and Runeson P and Höst M and Ohlsson M and Regnell B and Wesslén A. *Experimentation in Software Engineering - An Introduction*. Kluwer Academic Publishers, 2000.
- [144] Dewayne E. Perry, Adam A. Porter, and Lawrence G. Votta. Empirical studies of software engineering : a roadmap. In *ICSE '00 : Proceedings of the Conference on The Future of Software Engineering*, pages 345–355, New York, NY, USA, 2000. ACM Press.

- [145] T.D. Cook, WR Shadish, and DT Campbell. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Houghton Mifflin Co. Boston, MA, 2002.
- [146] D.T. Campbell and J.C. Stanley. *Experimental and quasi-experimental designs for research*. Rand McNally Chicago, 1963.

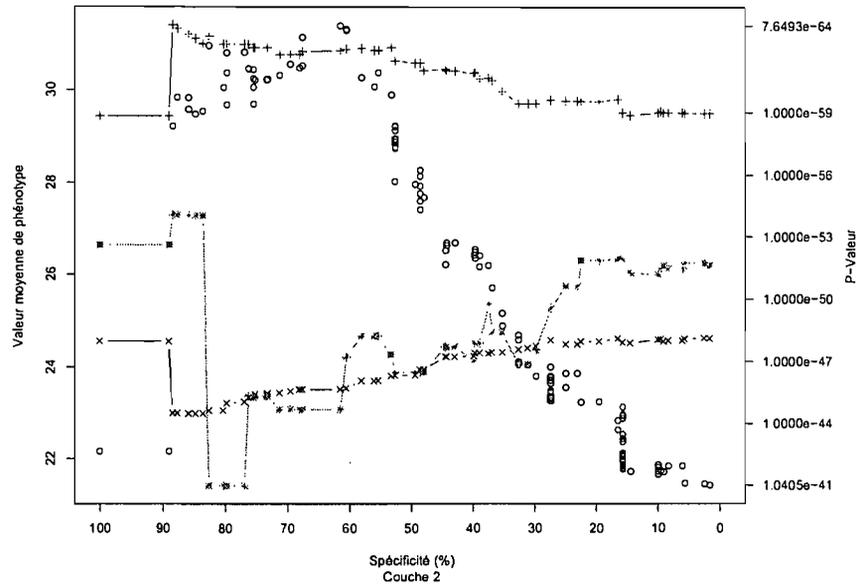
Annexe I

Graphiques d'analyses de descendance supplémentaires pour l'indice de masse corporelle

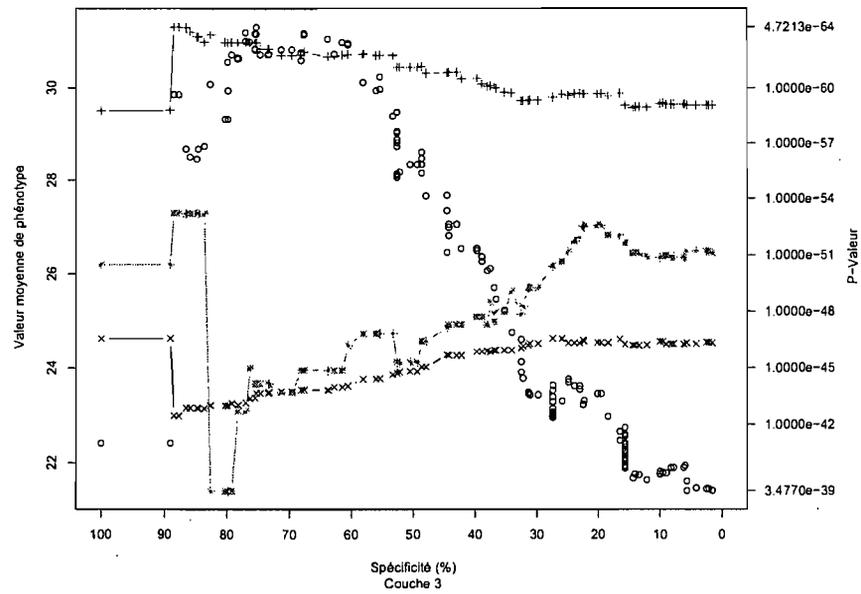
Analyses ANOVA de l'indice de masse corporelle (IMC) des descendance en fonction de la spécificité des ancêtres pour toutes les couches généalogiques. Voir la section 10.1 pour une explication détaillée des graphiques.



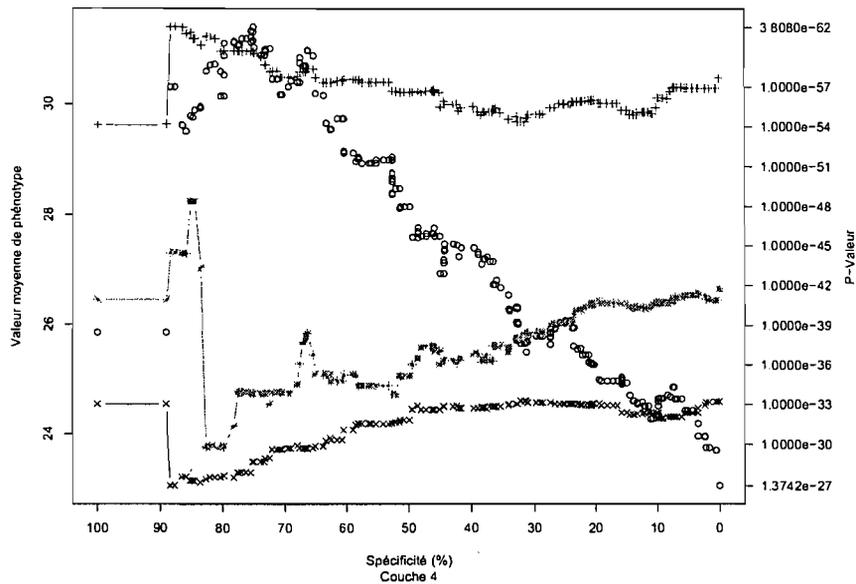
P-valeurs ANOVA et moyenne de l'indice de masse corporelle en fonction de la spécificité des ancêtres



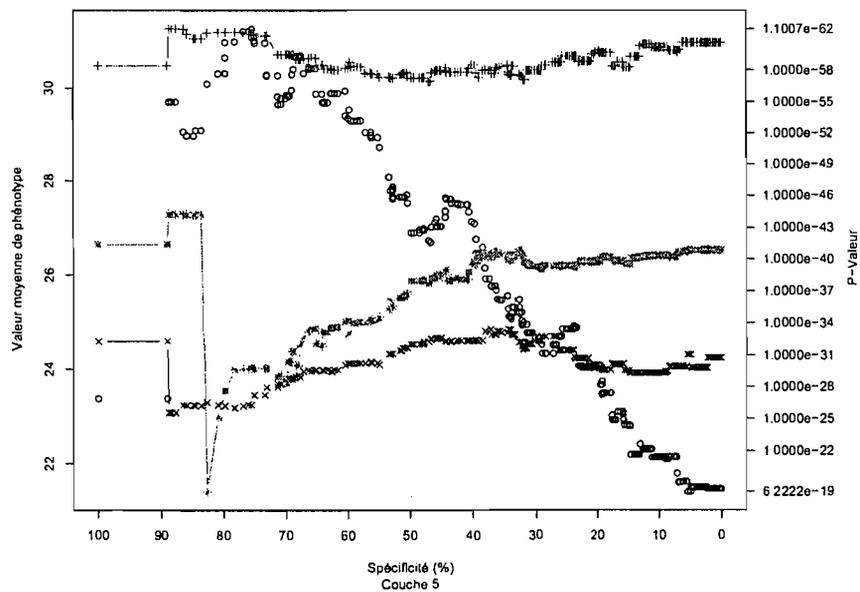
P-valeurs ANOVA et moyenne de l'indice de masse corporelle en fonction de la spécificité des ancêtres



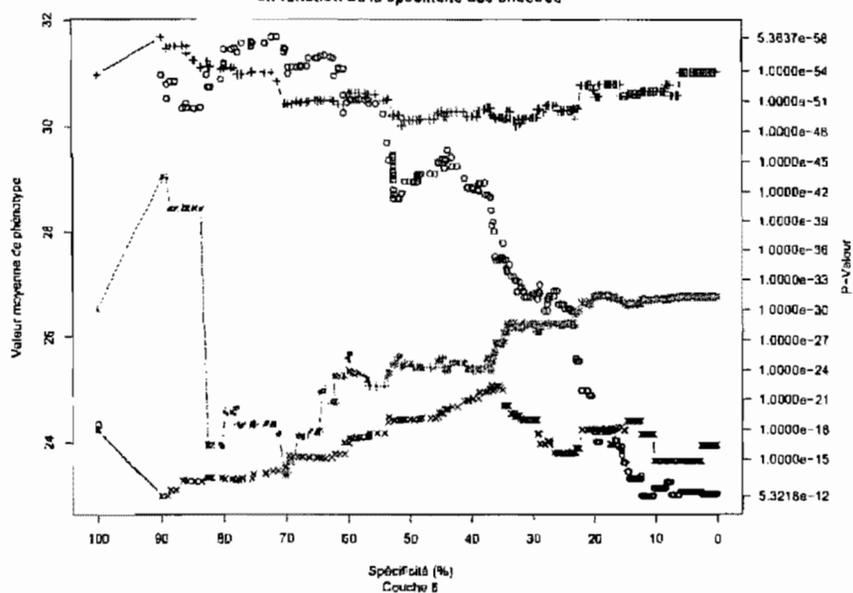
P-valeurs ANOVA et moyenne de l'indice de masse corporelle en fonction de la spécificité des ancêtres



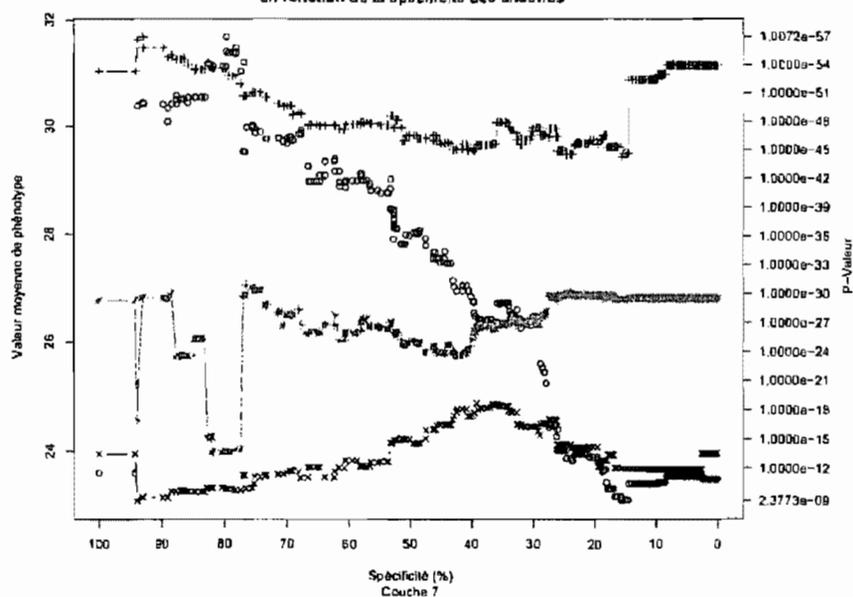
P-valeurs ANOVA et moyenne de l'indice de masse corporelle en fonction de la spécificité des ancêtres

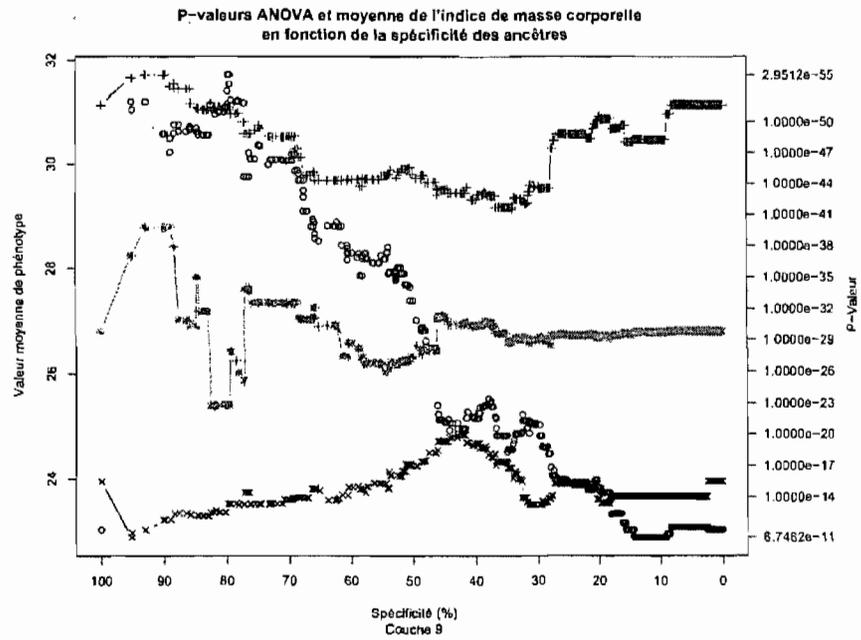
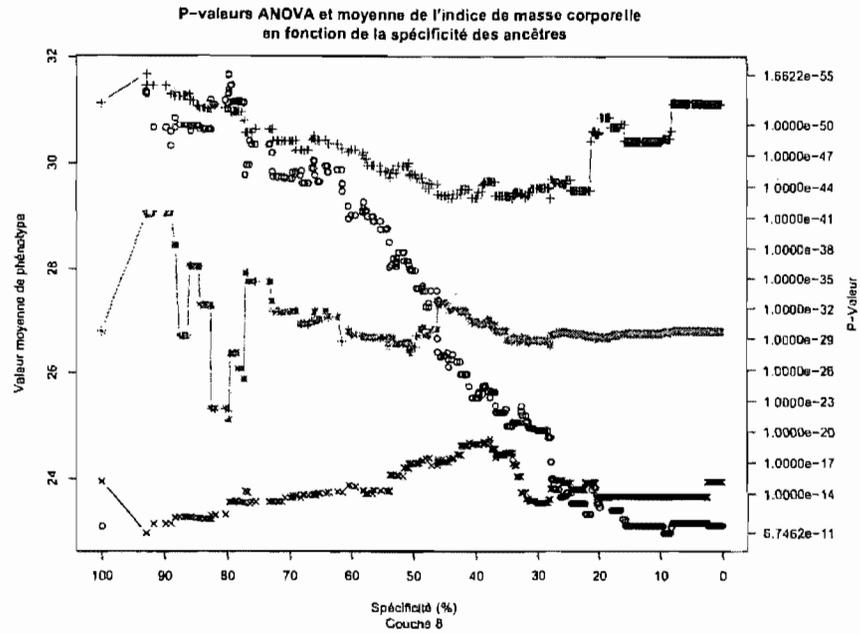


P-valeurs ANOVA et moyenne de l'indice de masse corporelle en fonction de la spécificité des ancêtres

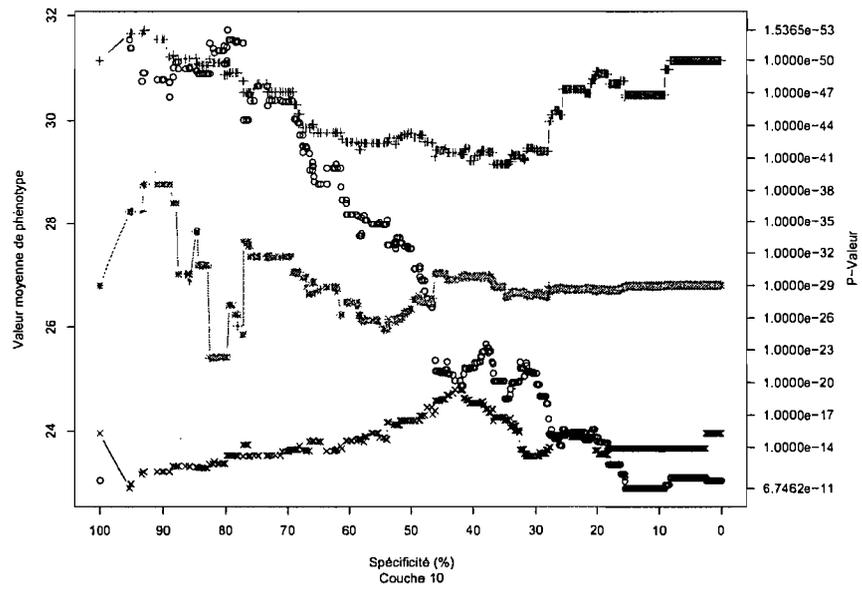


P-valeurs ANOVA et moyenne de l'indice de masse corporelle en fonction de la spécificité des ancêtres

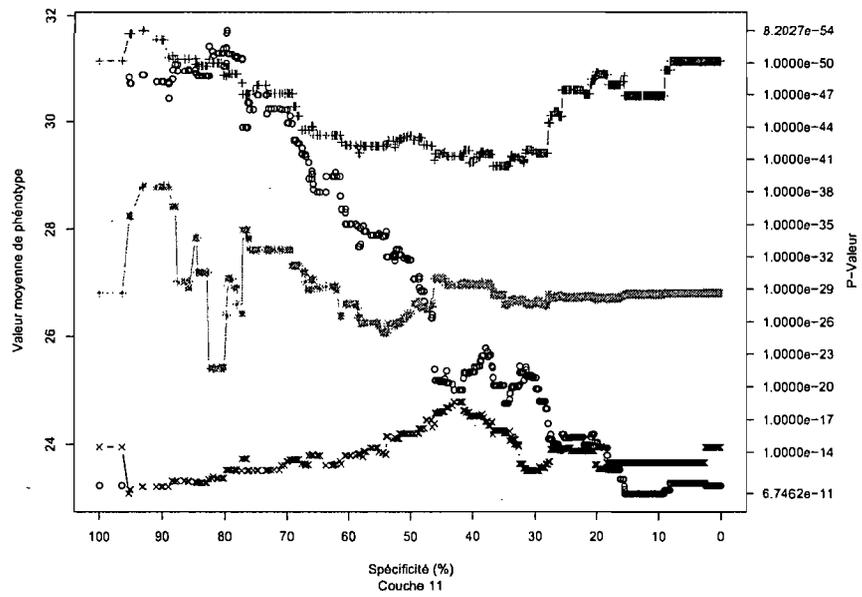




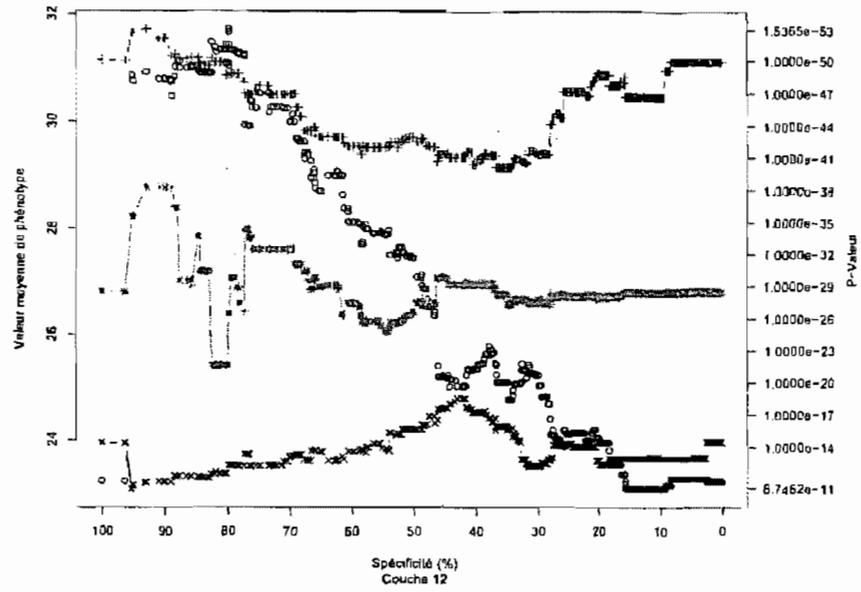
P-valeurs ANOVA et moyenne de l'indice de masse corporelle en fonction de la spécificité des ancêtres



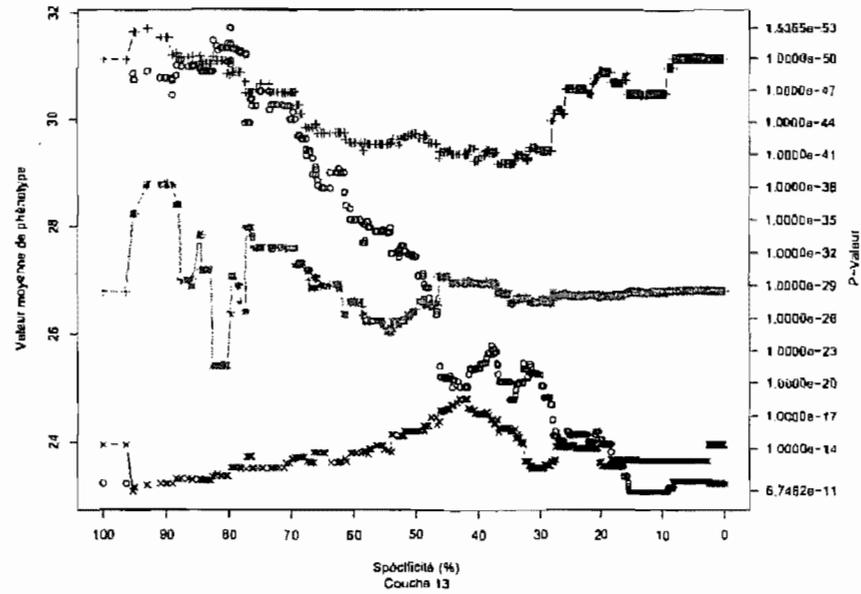
P-valeurs ANOVA et moyenne de l'indice de masse corporelle en fonction de la spécificité des ancêtres



P-valeurs ANOVA et moyenne de l'indice de masse corporelle en fonction de la spécificité des ancêtres



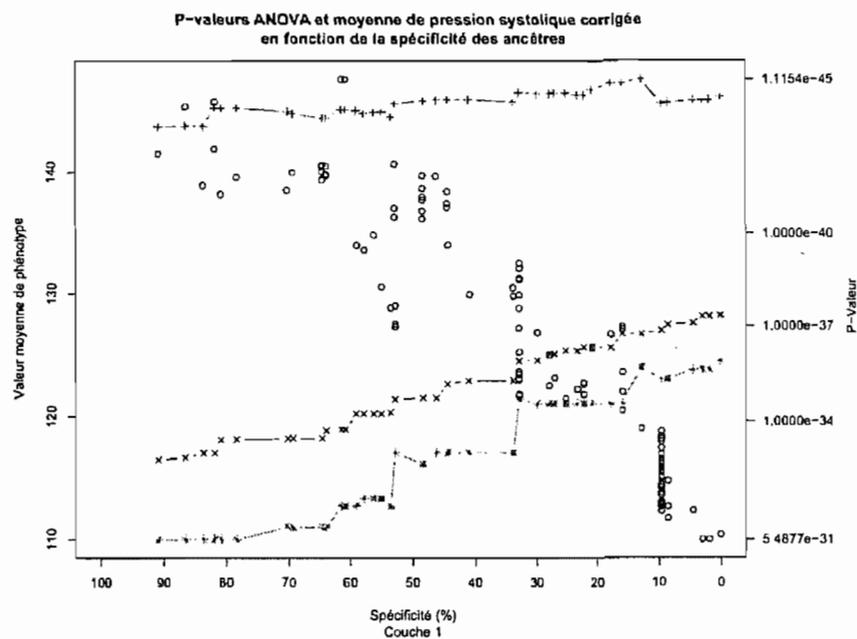
P-valeurs ANOVA et moyenne de l'indice de masse corporelle en fonction de la spécificité des ancêtres

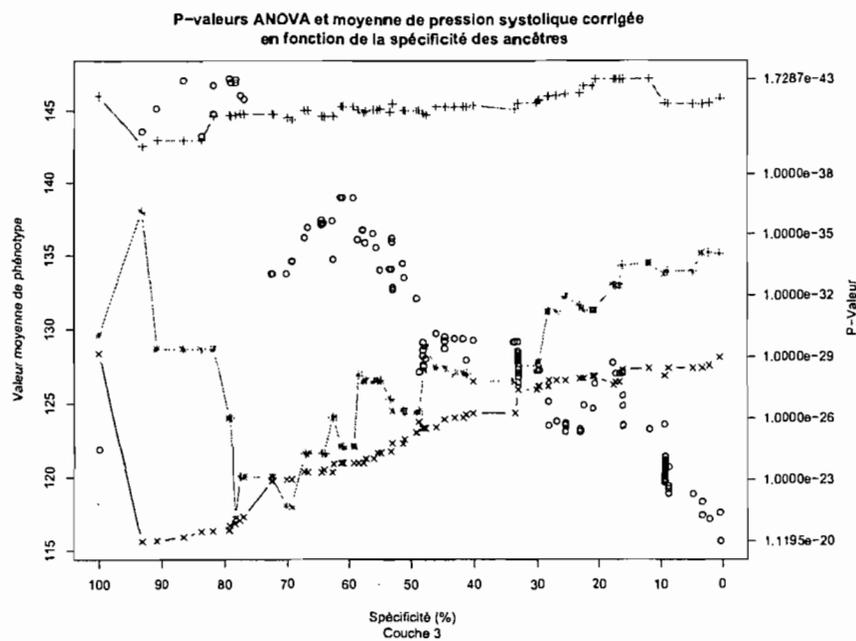
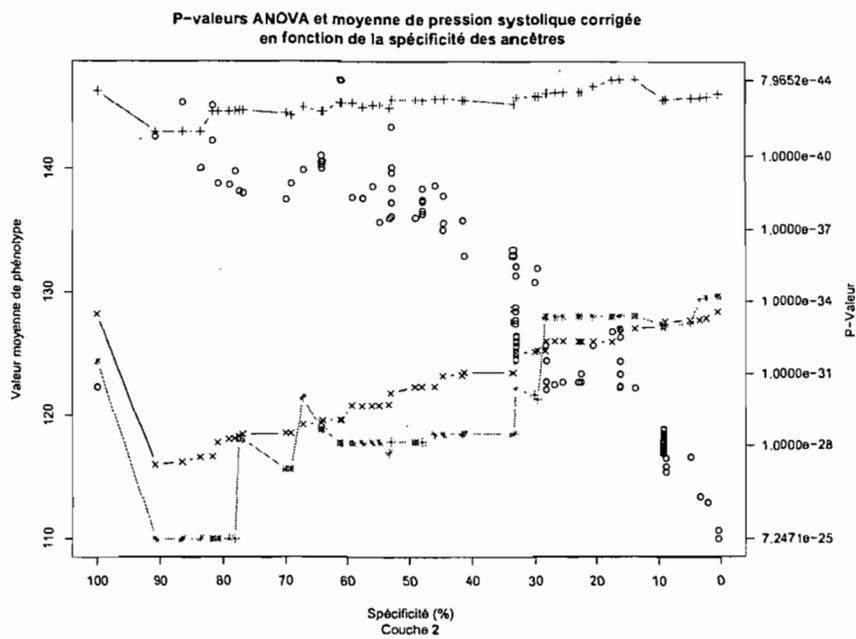


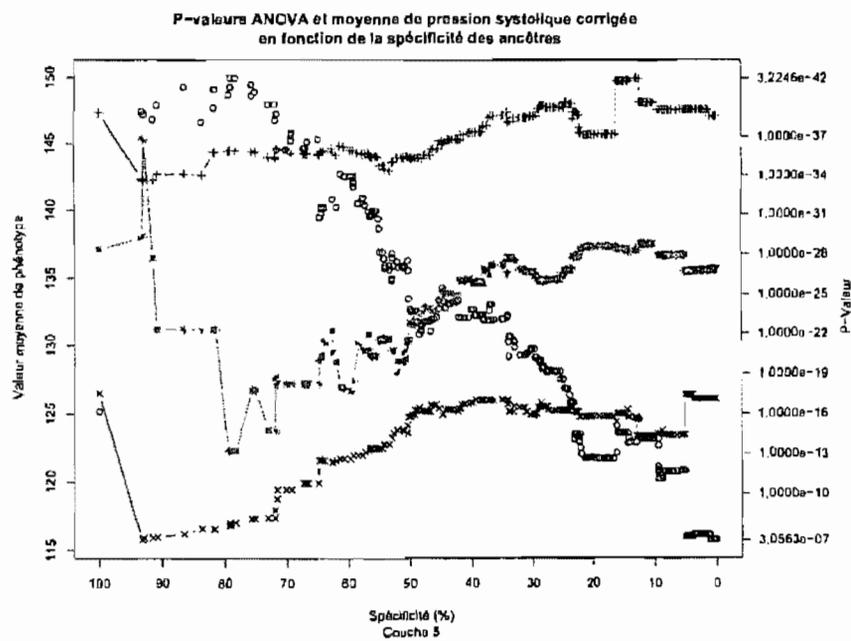
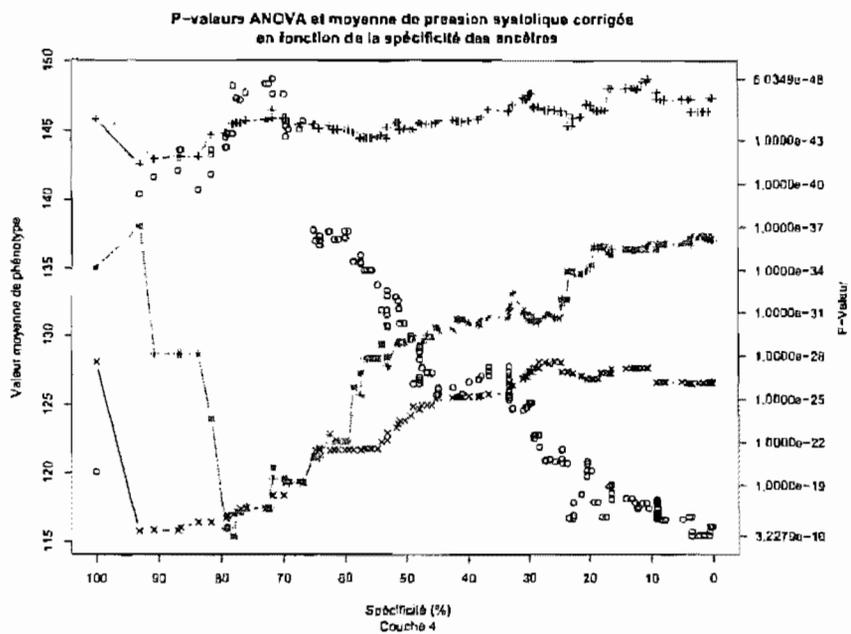
Annexe II

Graphiques d'analyses de descendance supplémentaires pour la pression systolique corrigée

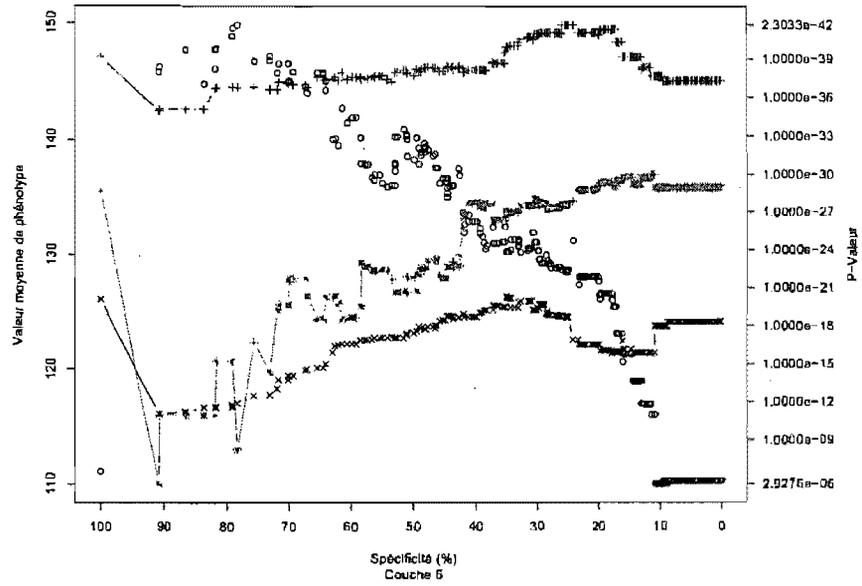
Analyses ANOVA de l'indice de la pression systolique corrigée des descendance en fonction de la spécificité des ancêtres pour toutes les couches généalogiques. Voir la section 10.1 pour une explication détaillée des graphiques.



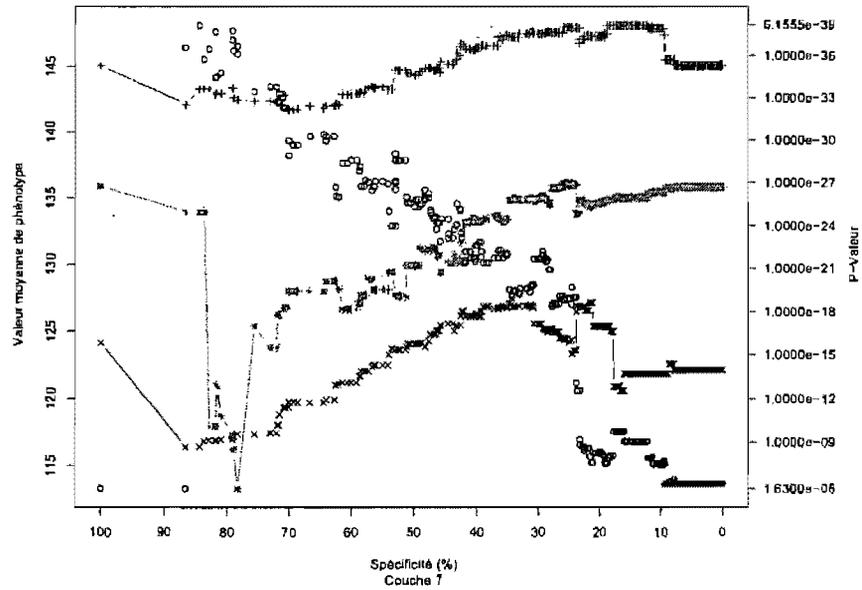


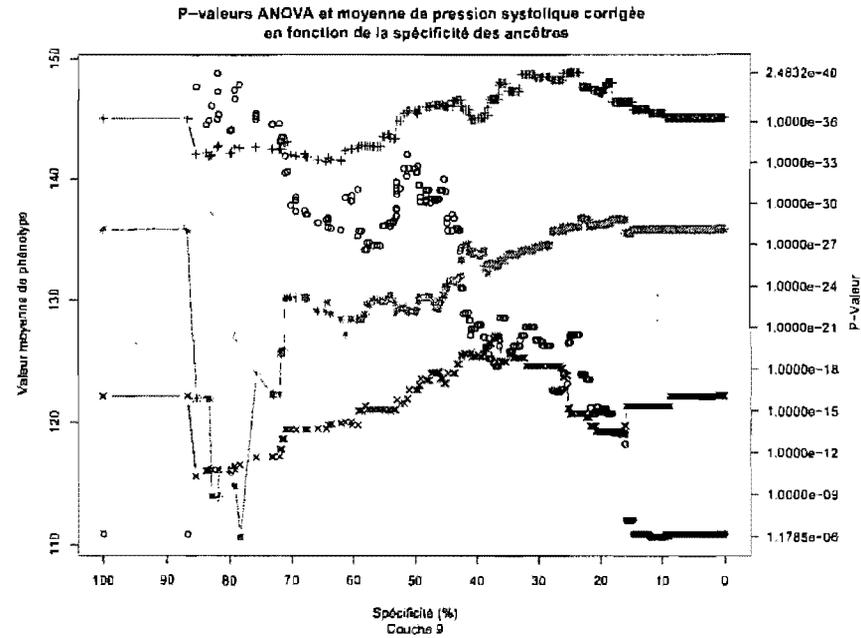
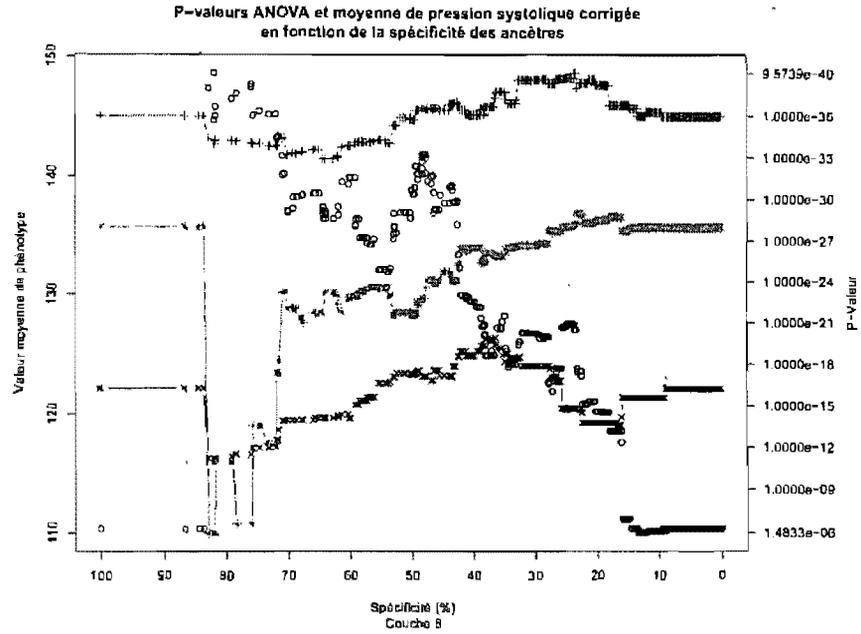


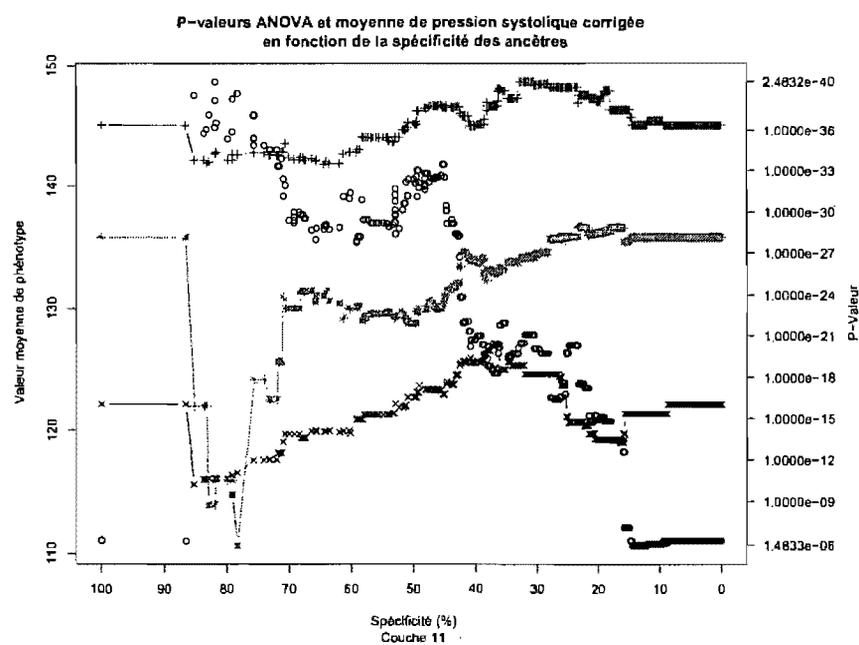
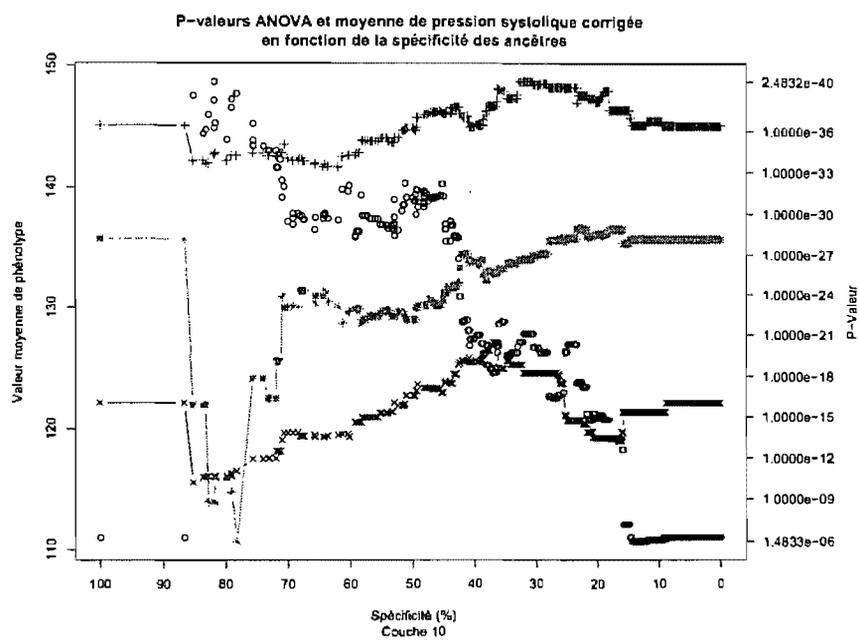
P-valeurs ANOVA et moyenne de pression systolique corrigée en fonction de la spécificité des ancêtres

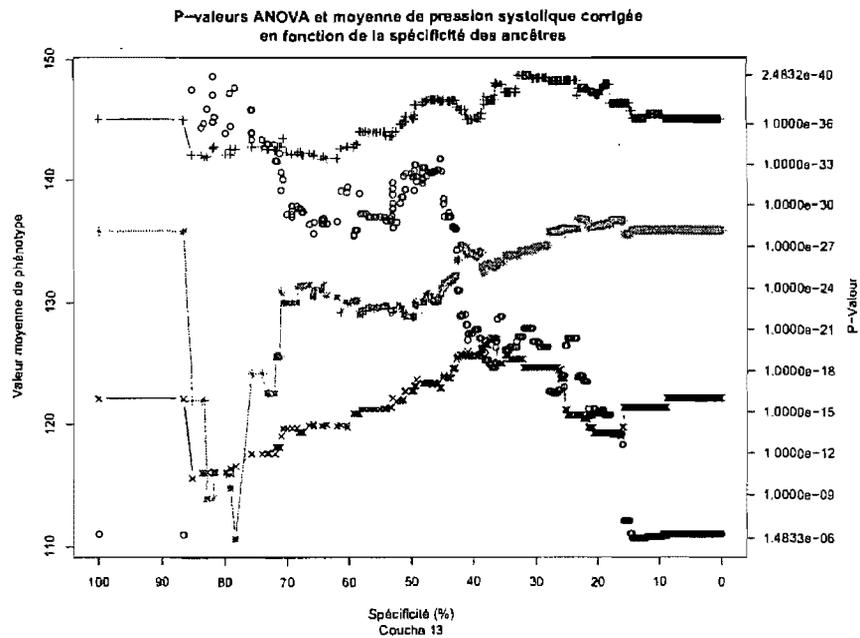
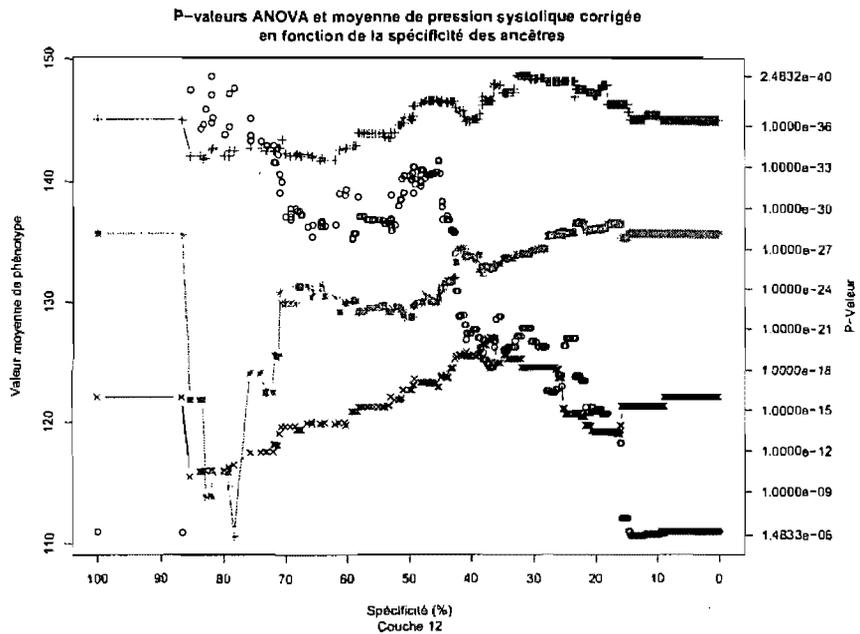


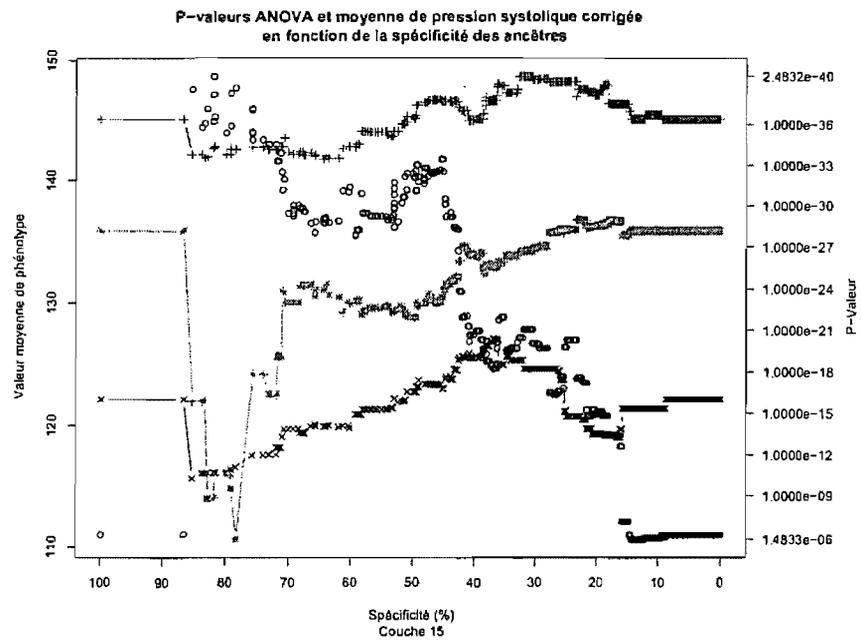
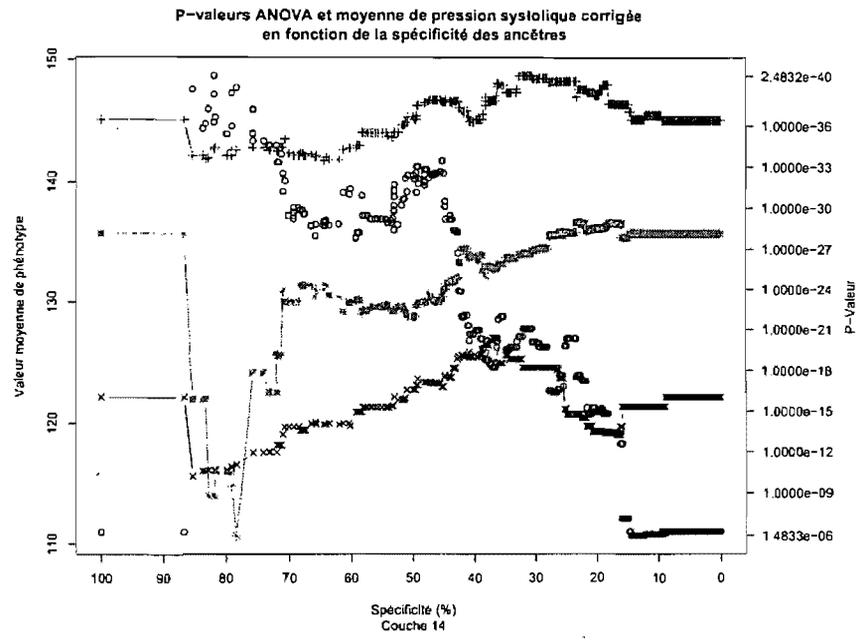
P-valeurs ANOVA et moyenne de pression systolique corrigée en fonction de la spécificité des ancêtres







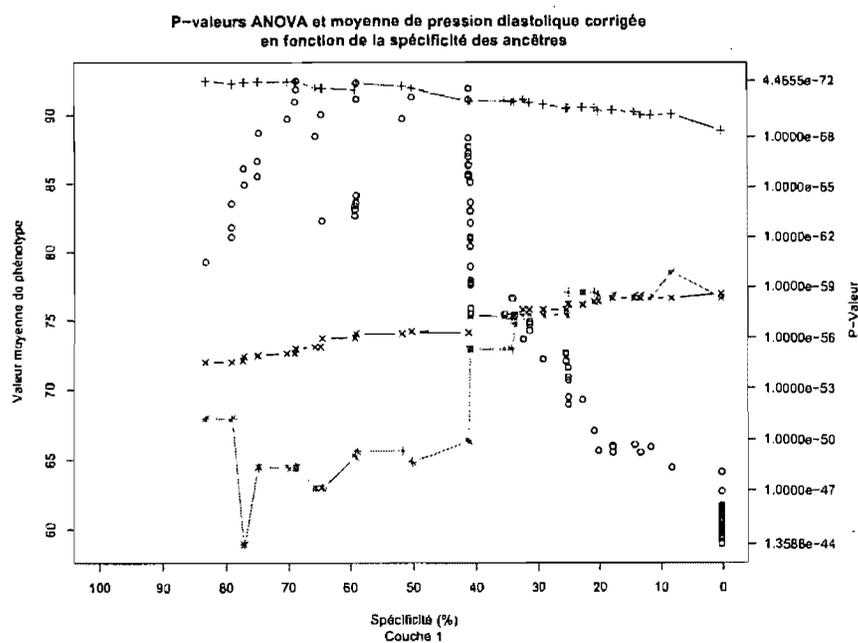




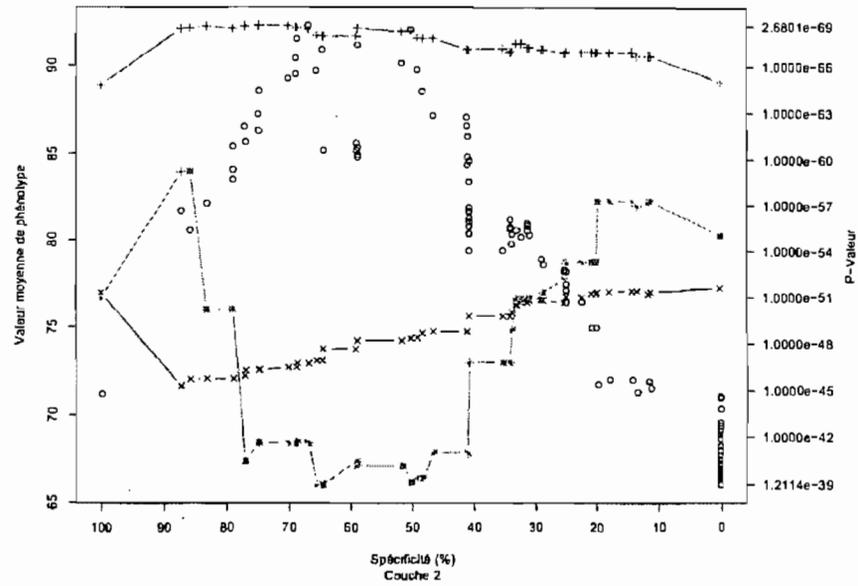
Annexe III

Graphiques d'analyses de descendance supplémentaires pour la pression diastolique corrigée

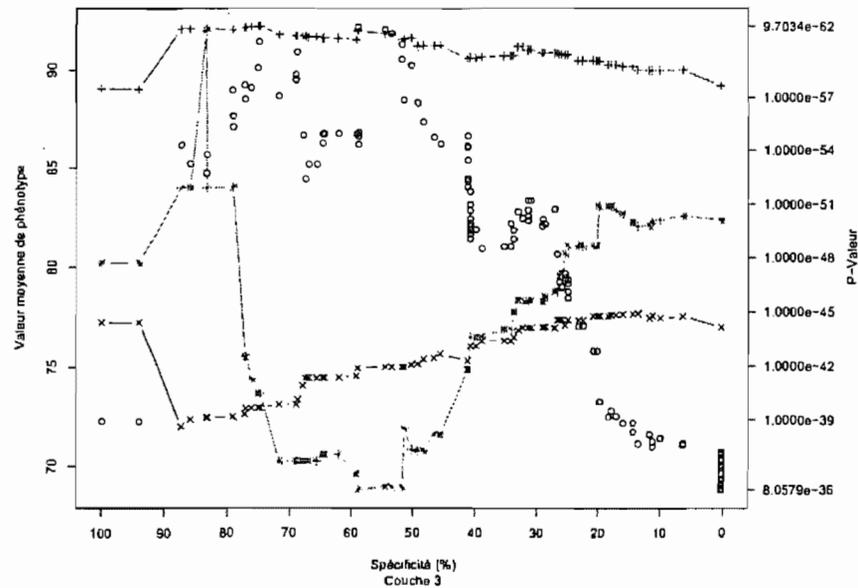
Analyses ANOVA de la pression diastolique corrigée des descendance en fonction de la spécificité des ancêtres pour toutes les couches généalogiques. Voir la section 10.1 pour une explication détaillée des graphiques.



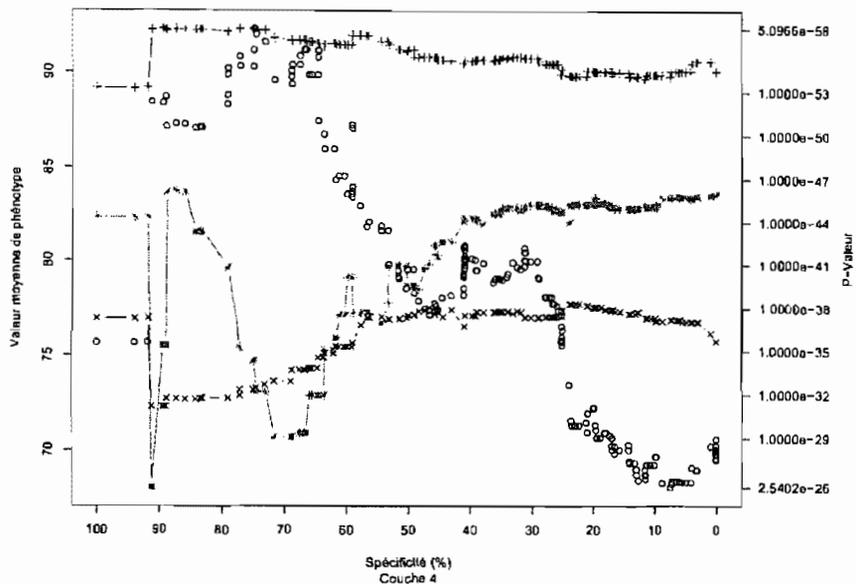
P-valeurs ANOVA et moyenne de pression diastolique corrigée en fonction de la spécificité des ancêtres



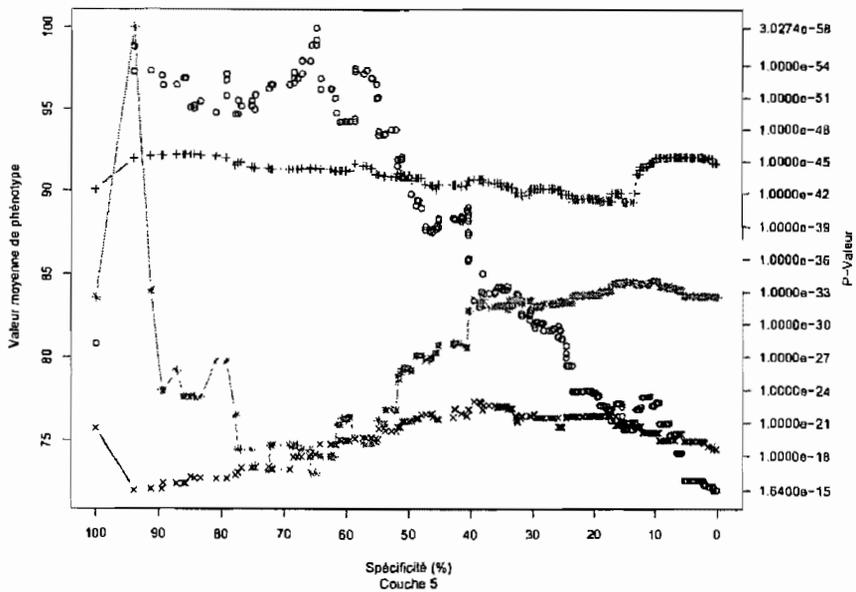
P-valeurs ANOVA et moyenne de pression diastolique corrigée en fonction de la spécificité des ancêtres



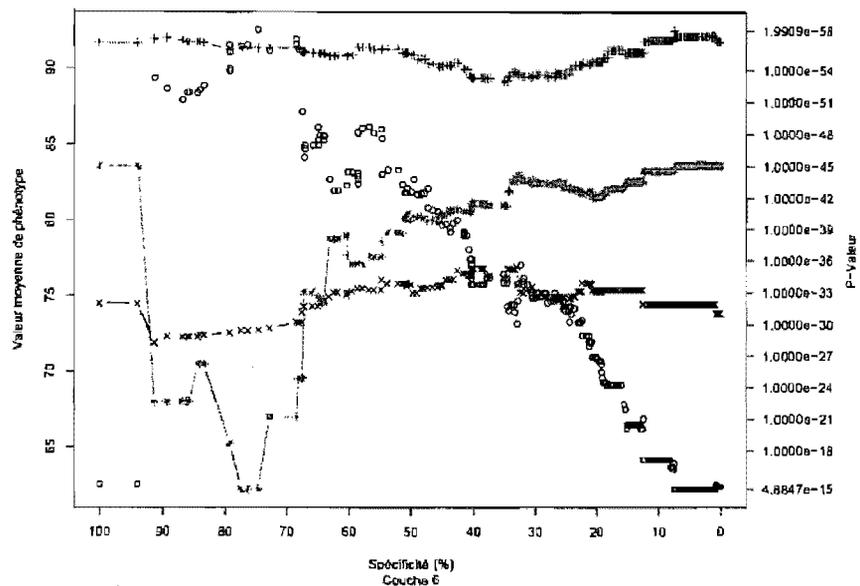
P-valeurs ANOVA et moyenne de pression diastolique corrigée en fonction de la spécificité des ancêtres



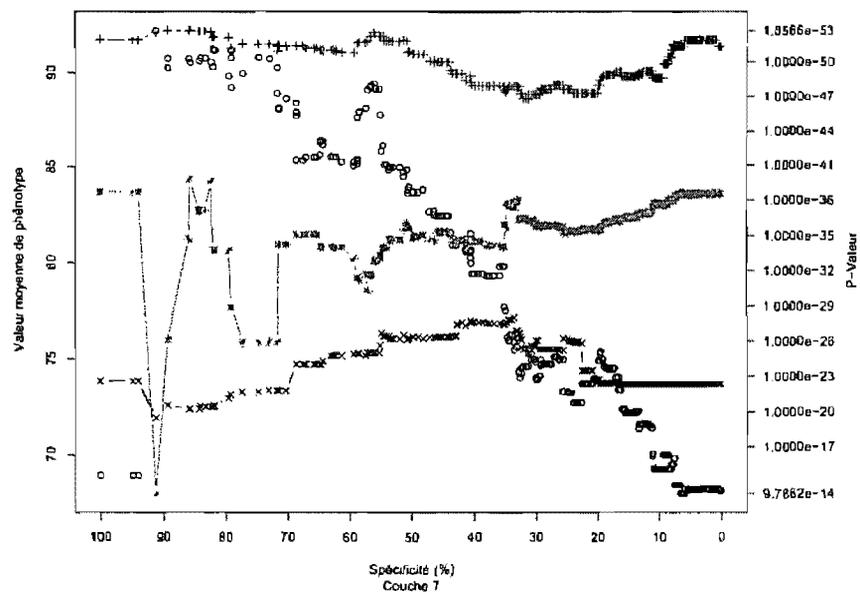
P-valeurs ANOVA et moyenne de pression diastolique corrigée en fonction de la spécificité des ancêtres



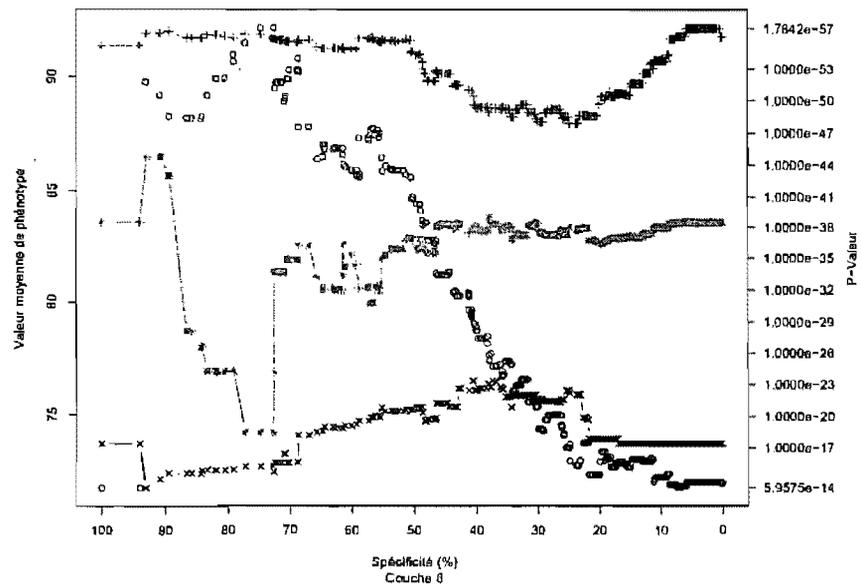
P-valeurs ANOVA et moyenne de pression diastolique corrigée en fonction de la spécificité des ancêtres



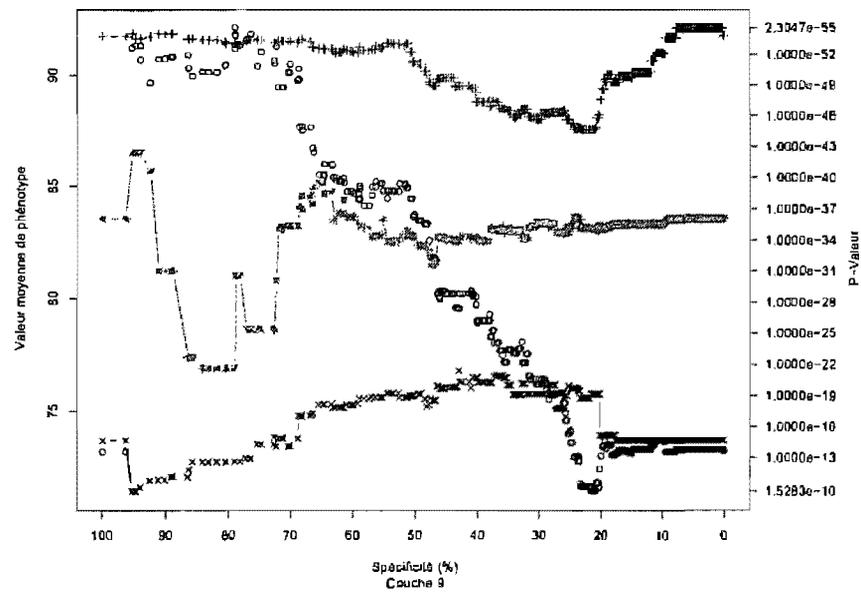
P-valeurs ANOVA et moyenne de pression diastolique corrigée en fonction de la spécificité des ancêtres

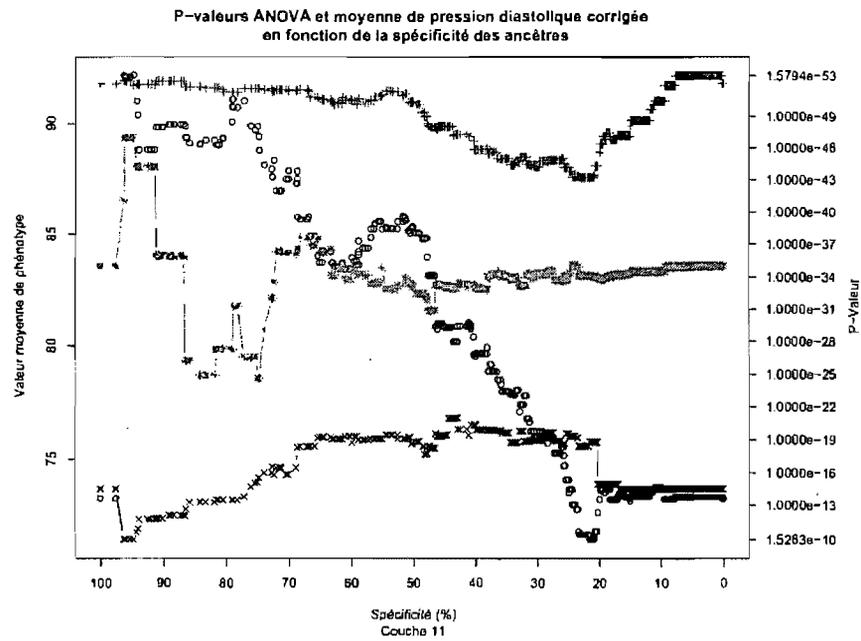
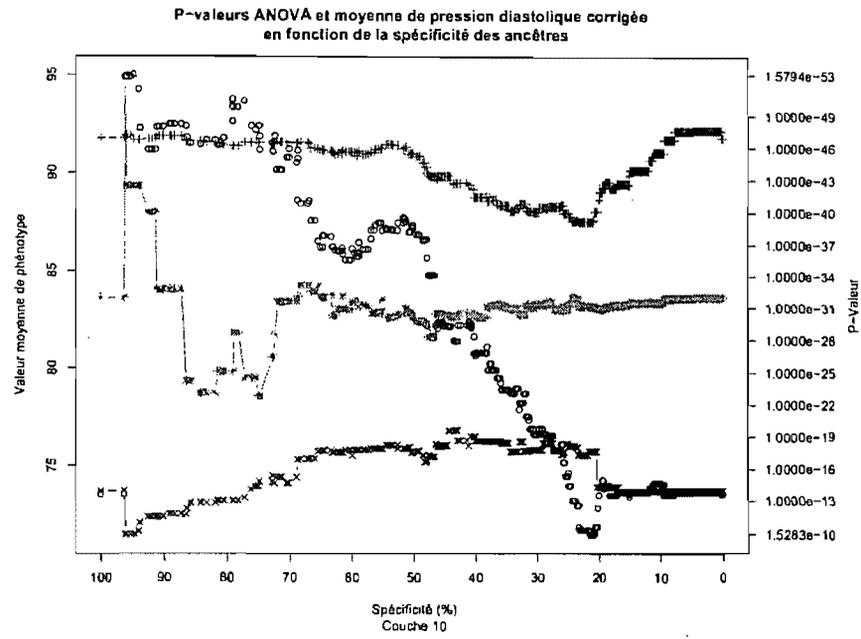


P-valeurs ANOVA et moyenne de pression diastolique corrigée en fonction de la spécificité des ancêtres

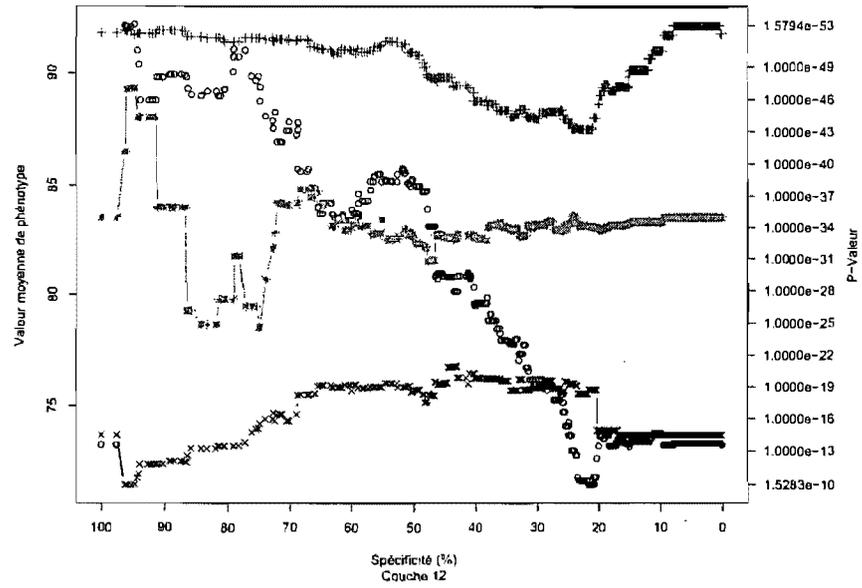


P-valeurs ANOVA et moyenne de pression diastolique corrigée en fonction de la spécificité des ancêtres

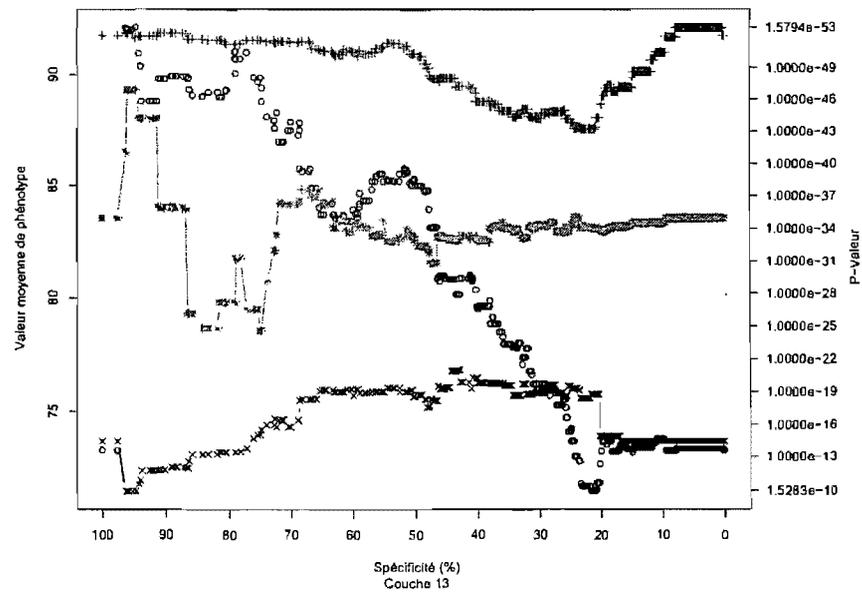




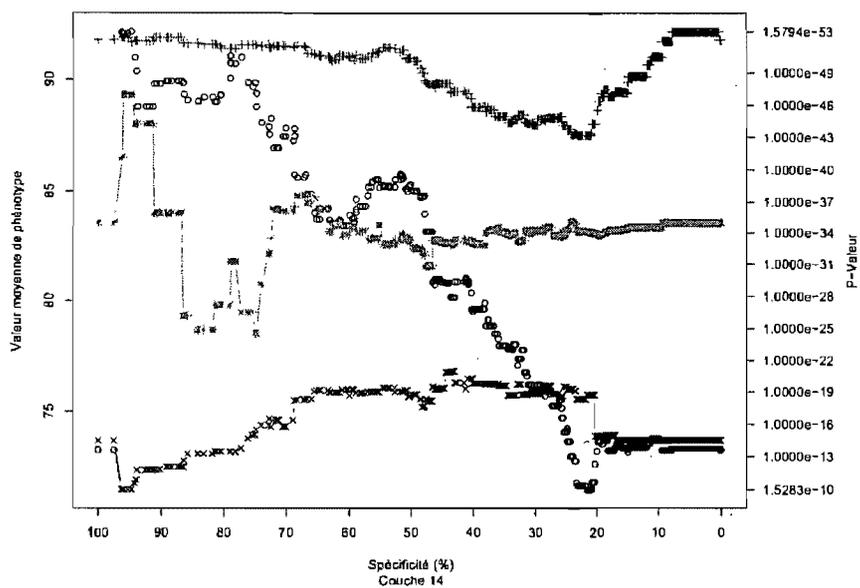
P-valeurs ANOVA et moyenne de pression diastolique corrigée en fonction de la spécificité des ancêtres



P-valeurs ANOVA et moyenne de pression diastolique corrigée en fonction de la spécificité des ancêtres



P-valeurs ANOVA et moyenne de pression diastolique corrigée en fonction de la spécificité des ancêtres



P-valeurs ANOVA et moyenne de pression diastolique corrigée en fonction de la spécificité des ancêtres

