

ANNE-MARIE MADORE

**DISTRIBUTION GÉOGRAPHIQUE DE CERTAINS
POLYMORPHISMES ASSOCIÉS À DES MALADIES
INFLAMMATOIRES CHRONIQUES DANS LES
RÉGIONS ADMINISTRATIVES
DU QUÉBEC**

Mémoire présenté
à la Faculté des études supérieures de l'Université Laval
comme exigence partielle du programme de maîtrise en médecine
expérimentale – génétique des populations humaines
offert à l'Université du Québec à Chicoutimi
en vertu d'un protocole d'entente avec l'Université Laval
pour l'obtention du grade de maître ès sciences (M.Sc.)

FACULTÉ DE MÉDECINE
UNIVERSITÉ LAVAL
QUÉBEC

et

DÉPARTEMENT DES SCIENCES HUMAINES
UNIVERSITÉ DU QUÉBEC À CHICOUTIMI
SAGUENAY

JUIN 2005



Mise en garde/Advice

Afin de rendre accessible au plus grand nombre le résultat des travaux de recherche menés par ses étudiants gradués et dans l'esprit des règles qui régissent le dépôt et la diffusion des mémoires et thèses produits dans cette Institution, **l'Université du Québec à Chicoutimi (UQAC)** est fière de rendre accessible une version complète et gratuite de cette œuvre.

Motivated by a desire to make the results of its graduate students' research accessible to all, and in accordance with the rules governing the acceptance and diffusion of dissertations and theses in this Institution, the **Université du Québec à Chicoutimi (UQAC)** is proud to make a complete version of this work available at no cost to the reader.

L'auteur conserve néanmoins la propriété du droit d'auteur qui protège ce mémoire ou cette thèse. Ni le mémoire ou la thèse ni des extraits substantiels de ceux-ci ne peuvent être imprimés ou autrement reproduits sans son autorisation.

The author retains ownership of the copyright of this dissertation or thesis. Neither the dissertation or thesis, nor substantial extracts from it, may be printed or otherwise reproduced without the author's permission.

RÉSUMÉ

Les études démographiques et génétiques ont souligné le caractère hétérogène de la population québécoise. Cette étude a pour but de vérifier la présence d'une variabilité génétique dans cette population pour deux traits complexes fréquents (asthme et maladies cardiovasculaires). Parmi les 15 régions administratives du Québec, 1680 échantillons ont été génotypés pour 11 SNPs inclus dans 8 gènes associés à ces traits. L'analyse de classification hiérarchique utilisée pour regrouper les régions et les variants selon les fréquences alléliques observées a révélé un regroupement de variants ainsi que trois groupes de régions par affection. Il n'est pas possible d'expliquer tous ces regroupements grâce à nos connaissances sur le peuplement du Québec. Toutefois, ces résultats suggèrent qu'une étude de plus grande envergure pourrait permettre de caractériser l'hétérogénéité de la population afin de créer des outils adaptés à la structure génétique de celle-ci qui pourraient être utilisés par le système de santé public québécois.

AVANT-PROPOS

Je voudrais tout d'abord remercier Catherine Laprise, ma directrice de maîtrise. Déjà à travers sa façon de transmettre ses connaissances lors de ses cours théoriques et ensuite lors de mon apprentissage dans son laboratoire, elle a su me communiquer son goût pour la génétique et sa passion pour la recherche dans ce domaine. Merci beaucoup Catherine pour ton esprit critique, ton intégrité et ton intelligence dans tes décisions en tant que chercheur. Merci beaucoup de m'avoir intégrée dans ton équipe de recherche, d'avoir cru en mes capacités, d'avoir toujours été disponible ainsi que de m'avoir guidée avec honnêteté. Finalement, un gros merci d'avoir été pour moi un modèle de ténacité, de courage, de rigueur ainsi que d'ambition et d'équilibre tout à la fois!

Merci également à Louis Houde pour son aide précieuse. Je veux te remercier pour ta patience lors des nombreuses fois où tu m'as expliqué les diverses analyses qui ont été explorées pour ce travail, pour l'intérêt que tu portais à ce projet ainsi que pour ta bonne humeur constante même lorsque je te harcelais à la fin de presque tous les cours de biostatistique! J'ai beaucoup apprécié ta capacité à simplifier la matière et à la vulgariser.

L'équipe parmi laquelle on évolue tout au long de notre maîtrise joue pour beaucoup dans l'impression que l'on conserve de cette période de notre vie. Ce que j'ai énormément apprécié c'est l'ambiance d'entraide qui règne entre tous les étudiants de ce laboratoire. Lors de mon arrivée j'ai eu la chance d'être accueillie par deux étudiantes, Nancy et Karine. Merci beaucoup Nancy pour la passion que tu as pour tout ce que tu fais car cette passion est contagieuse! J'ai beaucoup apprécié travailler à tes côtés et tu as contribué par ta bonne humeur, ton savoir-faire, ton entraide et ton amitié à rendre cette période de ma vie très agréable. Un gros merci Karine pour ton sens critique et ton franc-

parler et parce que il n'y en a pas deux comme toi pour partir des discussions sur tous les sujets inimaginables! Merci à toi aussi Josianne, à la fois pour ton dynamisme, ton intelligence ainsi que pour ton sens pratique et logique. Merci beaucoup Tarek pour ton sens de la répartie et également pour ta capacité à toujours nous mettre sous le nez qu'il y en a qui vivent des choses pires que nous! Finalement, merci beaucoup Annie. J'ai beaucoup apprécié cheminer à tes côtés tout au long de ma maîtrise. Merci pour ta disponibilité, ta complicité, ta générosité, ton amitié et en plus pour avoir été mon aide-mémoire en permanence!

Je voudrais également remercier Nadia Mior et Denise Morin, qui sont pour moi beaucoup plus que deux employées, merci pour votre disponibilité et votre aide. Mais surtout merci pour votre générosité et votre bonne humeur!

Certaines personnes ont cru en moi depuis des années et n'ont cessé de me soutenir dans toutes mes entreprises. Je veux remercier ma mère et mon père pour leur soutien inconditionnel, pour m'avoir épaulée dans mes décisions, pour leurs encouragements, pour m'avoir un peu fouettée au besoin et pour m'avoir comprise et écoutée dans mes moments de découragement. C'est grâce à vous si j'ai réussi à faire ma maîtrise! Je voudrais également remercier ma grande sœur et mon grand frère qui m'ont soutenue et motivée à leur propre façon. Ces remerciements sont également pour « petite Alice », ma filleule, la plus merveilleuse et la plus enjouée des fillettes de 2 ans que je connaisse (mais qui a un sacré caractère!). Je veux aussi remercier mes ami(e)s qui ont cru en moi et qui m'ont appuyée dans cette expérience par leurs encouragements. Merci pour votre présence, votre amitié et votre générosité.

TABLE DES MATIÈRES

RÉSUMÉ	ii
AVANT-PROPOS.....	iii
TABLE DES MATIÈRES.....	v
LISTE DES TABLEAUX.....	viii
LISTE DES FIGURES	ix
INTRODUCTION.....	1
CHAPITRE I: Le peuplement du Québec.....	5
1.1 Origine des fondateurs	5
1.2 Occupation du territoire	8
CHAPITRE II: Génétique des populations et traits complexes	15
2.1 Génétique des populations	15
2.1.1 <i>Les isolats et les dèmes</i>	16
2.1.2 <i>L'effet fondateur</i>	16
2.1.3 <i>La dérive génétique</i>	17
2.2 Les gènes et variants étudiés.....	18
2.2.1 <i>La différenciation des monocytes</i>	19
2.2.1.1 <i>Enzyme de conversion de l'angiotensine</i>	21
2.2.1.2 « <i>Peroxisome proliferator-activated receptor gamma 2</i> ».....	23
2.2.1.3 « <i>Peroxisome proliferator-activated receptor alpha</i> »	24
2.2.1.4 <i>Lipase lipoprotéique</i>	25
2.2.1.5 <i>Récepteur des lipoprotéines à faible densité</i>	27

2.2.2 <i>La présentation de l'antigène</i>	28
2.2.2.1 <i>Interleukine 4</i>	28
2.2.2.2 <i>Récepteur de haute affinité pour les IgE</i>	29
2.2.2.3 <i>Enzyme de conversion de l'angiotensine</i>	30
2.2.2.4 <i>Apolipoprotéine E</i>	31
2.3 La prévalence de l'asthme	33
2.4 La prévalence des maladies cardiovasculaires	33
CHAPITRE III: Distribution géographique de polymorphismes associés à l'asthme ou aux maladies cardiovasculaires dans une cohorte de la population du Québec.....	35
3.1 Avant-propos.....	36
3.2 Résumé.....	39
3.3 Abstract.....	40
3.4 Introduction.....	40
3.5 Methods.....	43
3.5.1 <i>Sample</i>	43
3.5.2 <i>DNA genotyping</i>	44
3.5.3 <i>Data analysis</i>	44
3.6 Results.....	45
3.7 Discussion	46
3.8 Conclusions.....	49
3.9 Acknowledgments.....	50
3.10 References.....	51
3.11 Figure titles and legends	58
CHAPITRE IV: Résultats et discussion	62
4.1 Modèle de division géographique	62
4.2 Fréquences alléliques	63
4.2.1 <i>Choix des gènes</i>	63
4.2.2 <i>Méthode de génotypage</i>	64
4.3 Équilibre de Hardy-Weinberg.....	64

4.4 Analyse de classification descendante hiérarchique	66
4.4.1 <i>Regroupement des régions selon les variants impliqués dans l'asthme et les maladies cardiovasculaires</i>	68
4.4.3 <i>Regroupement des variants impliqués dans l'asthme et les maladies cardiovasculaires selon leur patron de distribution dans les régions du Québec</i>	71
CONCLUSION	75
PERSPECTIVES	77
BIBLIOGRAPHIE.....	79
ANNEXE A: Résultats du calcul de l'équilibre d'Hardy-Weinberg pour chaque variant dans chaque échantillon des régions administratives du Québec séparément ainsi que pour la cohorte provinciale	95
ANNEXE B: Fréquence allélique (%) des 11 variants génotypés pour chaque échantillon des régions administratives de la province de Québec	98

LISTE DES TABLEAUX

Tableau 1: Proportion d'immigrants (%) en 1765 selon leur lieu d'origine.....	6
Tableau 2: Proportion d'immigrants (%) en 1765 selon le pays d'origine.....	7
Tableau 3: Quantité d'immigrants en nombre absolu et en pourcentage pour chacune des régions administratives du Québec.....	12
Tableau 4: Classement des régions administratives du Québec en fonction de la proportion (%) d'individus parlant français, anglais ou une autre langue à la maison.....	13
Tableau 5: Position et codification du <i>National Center for Biotechnology Information</i> pour les gènes et variants ciblés pour cette étude	20
Table 6: References and allelic frequencies for all SNPs tested.....	59

LISTE DES FIGURES

Figure 1: Carte des régions administratives du Québec selon la stratification actuelle	10
Figure 2: Intégration des 8 gènes ciblés pour cette étude à l'intérieur de la cascade inflammatoire reliée à l'asthme et aux maladies cardiovasculaires (A et B).	22
Figure 3: Administrative regions of the province of Quebec	60
Figure 4: Groups of the Quebec regions according to the allelic frequencies of SNPs involved in a) asthma and b) cardiovascular diseases and clusters of SNPs involved in c) asthma and d) cardiovascular diseases according to their distribution into the regions of Quebec	61

INTRODUCTION

La population du Québec a longtemps été considérée par les sociologues, anthropologues, ethnographes et historiens comme un tout homogène (Bouchard, 1990), l'homogénéité se définissant comme la « propriété d'un ou de plusieurs échantillons ou populations qui implique qu'ils ou elles sont identiques quant à l'ensemble ou à quelques-uns seulement de leurs paramètres », ceci faisant opposition à l'hétérogénéité qui est la « propriété d'un ou de plusieurs échantillons, ou populations, qui implique qu'ils (ou elles) ne sont pas identiques quant à l'ensemble ou à quelques-uns de leurs paramètres » (Office québécois de la langue française, 2004). L'hétérogénéité de la population québécoise, sur le plan démographique, a été discutée pour la première fois en 1962 par G. Fortin (Fortin, 1962). Des études historiques et démographiques, les travaux de recherche entourant la création du Programme de recherche en démographie historique de l'Université de Montréal (PRDH) et celle d'un fichier de population, le fichier BALSAC, ont par la suite appuyé cette hypothèse d'hétérogénéité de la population québécoise (Bouchard, 1990; Bouchard, 2004; PRDH, 2004). Du point de vue de la génétique, des études concernant les polymorphismes de la phénylalanine hydroxylase (PAH), des loci CFTR et HLA, ainsi que des études de distribution géographique de variants présents dans les gènes de la lipase lipoprotéique (LPL) et du récepteur des lipoprotéines à faible densité (LDLR) sont également arrivées aux mêmes conclusions quant à la présence d'une variabilité dans la population québécoise (Bergeron et al., 1992; Normand et al., 1992; Davignon et Roy, 1993; Dionne et al., 1993; Vohl et al., 1997; Scriver, 2001).

L'étude de cette diversité génétique ainsi que celle de certaines maladies monogéniques ont amené l'évidence que certaines régions se distinguent du reste du Québec (Vézina, 1996). Les différences que l'on observe, par exemple, dans les régions de

Charlevoix et du Saguenay–Lac-Saint-Jean quant à la prévalence de certains gènes, sont déjà assez bien documentées (De Braekeleer, 1995; Scriver, 2001). Il a été démontré que ces deux régions présentent des prévalences de certaines maladies monogéniques plus élevées qu'ailleurs au Québec (telles que l'acidose lactique, l'ataxie récessive spastique et la fibrose kystique), mais que l'inverse est également vrai (CORAMH). Ainsi, d'autres maladies comme la thalassémie, ont une prévalence beaucoup plus basse qu'ailleurs, voire presque inexistante (CORAMH). Les raisons de la différenciation de ces deux régions se trouvent dans la succession de plusieurs effets fondateurs (dont trois sont documentés), d'effets d'isolement et d'un taux élevé de natalité (CORAMH, ; Bouchard et De Braekeleer, 1990; Scriver, 2001). De tels phénomènes peuvent être survenus à des degrés divers dans différentes régions du Québec, accentuant la diversité de la population québécoise déjà présente dans le petit effectif des premiers fondateurs (De Braekeleer, 1990).

Cette diversité, ajoutée au jeune âge de cette population, donne un avantage certain pour l'élaboration d'études d'épidémiologie génétique. Jusqu'à ce jour, des études épidémiogénétiques ont été réalisées dans la population québécoise, soit concernant des maladies monogéniques, soit sur des cohortes de personnes atteintes d'une maladie spécifique, aucune étude n'ayant été réalisée pour des traits complexes sur une cohorte représentant la population générale (Bergeron et al., 1992; Normand et al., 1992; Davignon et Roy, 1993; Dionne et al., 1993; Vohl et al., 1997). Dans la présente étude, les affections choisies sont donc des traits complexes, c'est-à-dire des traits qui résultent de l'interaction entre des facteurs environnementaux et des facteurs génétiques (Cardon et Abecasis, 2003). Ce sont plus particulièrement l'asthme et les maladies cardiovasculaires. Elles ont été choisies pour leur forte prévalence dans la population québécoise et également pour les coûts et les taux de mortalité/morbidité importants qu'elles occasionnent. Il y a plus de 300 millions d'individus sur la planète qui souffrent d'asthme (Masoli et al., 2004) et il y a 16,7 millions de victimes des maladies cardiovasculaires chaque année (WHO, 2003). Ces maladies se rejoignent pathologiquement par la présence d'une composante inflammatoire.

Il y a une grande accumulation de données génétiques concernant ces deux types de maladies. Des bases de données disponibles sur Internet comme *Asthma & Allergy Gene Database* (AAGD, 2004) et *GenCards* ainsi que plusieurs articles récapitulatifs illustrent bien la quantité de connaissances acquises (Blumenthal, 2004; Lusis et al., 2004; GeneCards, 2005). C'est donc à partir de ces informations que 8 gènes et 11 variants déjà documentés dans la population québécoise, à l'exception du gène IL-4, ont été ciblés. Chacun de ces variants est caractérisé par un rôle direct ou indirect dans la cascade inflammatoire de l'un ou des deux groupes de maladies. La prévalence de ces deux types de maladie dans notre société, additionnée à la grande quantité d'informations disponibles, laisse croire qu'il serait intéressant de pouvoir définir l'ampleur des facteurs de risque génétiques inhérents à ces traits complexes afin de les intégrer au système de santé public. L'objectif visé par cette étude est donc d'utiliser une cohorte représentant la population générale québécoise à un moment précis (ici 1990), sans biais pour les traits ciblés, afin de modéliser et caractériser la diversité génétique interrégionale du Québec à l'aide de la distribution géographique de variants ciblés pour leur association documentée et leur rôle possible dans des maladies à composante inflammatoire fréquentes dans la population québécoise. Les sous-objectifs de cette étude sont:

- a) d'établir la fréquence allélique de chacun des variants ciblés pour toutes les régions administratives du Québec et ainsi augmenter les connaissances sur leur prévalence dans la population générale;
- b) d'effectuer des analyses statistiques pour regrouper premièrement les régions selon leur composition allélique et en second lieu les variants selon leur patron de distribution respectif;

En plus d'observer de quelle façon se distribuent les différents polymorphismes (SNPs) ciblés dans les régions du Québec, il est intéressant d'essayer de comprendre la raison de cette variabilité. C'est pourquoi le chapitre 1 a comme mission de faire un survol historique du peuplement du Québec. L'histoire des migrations interrégionales du Québec servira également à comparer les résultats statistiques des regroupements de régions et de

variants à des fins de validation. Le chapitre 2 est une revue de littérature en trois volets, d'abord une explication de divers phénomènes de génétique des populations, suivie des précisions sur les critères de sélection des gènes ciblés ainsi que leur intégration dans la cascade inflammatoire, et finalement une illustration de l'importance de l'asthme et des maladies cardiovasculaires dans le monde et au Canada. Le chapitre 3 est composé de l'article issu de ce projet de recherche et fait un bref rappel sur le contexte du projet dans son introduction mais surtout décrit la méthode utilisée pour le génotypage et les analyses statistiques ainsi que les résultats obtenus de même qu'une brève discussion de ces résultats. Cet article est en préparation et devrait être soumis d'ici quelques semaines à la revue *Genetic epidemiology*. Une section résultats et discussion où les résultats sont discutés de façon plus élaborée et détaillée que dans l'article du chapitre 3 est présentée dans le chapitre 4. Cette section commente les méthodes statistiques utilisées et reprend chacun des résultats, tout en faisant des rapprochements avec l'histoire du peuplement du Québec. Finalement, les deux dernières sections concernent les conclusions qui découlent des résultats obtenus jusqu'à maintenant dans ce projet et les perspectives que nous laissent envisager cette étude.

CHAPITRE I

Le peuplement du Québec

1.1 Origine des fondateurs

La presque totalité des premiers immigrants au Québec étant d'origine française, on a supposé longtemps que la population de la province de Québec formait plus ou moins un tout homogène (Bouchard, 1990). Toutefois, certaines données génétiques ainsi que des informations sur l'origine des premiers immigrants apportent des informations pertinentes sur la présence d'une certaine diversité génétique dans la population québécoise ((Bouchard, 1990; Vézina, 1996; Scriver, 2001)).

La première ville de la Nouvelle-France, soit la ville de Québec, fut fondée en 1608 (Charbonneau et al., 2000). Les débuts de la colonie sont plutôt difficiles et c'est seulement à partir de 1632, soit après une brève occupation britannique qui dura de 1629 à 1632, que le peuplement se fait de façon plus efficace (Charbonneau et al., 2000). Pendant le règne du roi Louis XIII (1610-1643) l'immigration demeure peu encouragée mais dès le sacre de Louis XIV on observe une hausse de l'immigration grâce à la nomination de l'intendant Jean Talon, d'abord de 1665 à 1668, puis de 1670 à 1672 (Salone, 1970). Pour l'ensemble de la période couvrant le début du 17^{ème} siècle jusqu'en 1763, année de la conquête anglaise, on estime le nombre d'immigrants à 25 000 mais on sait que seulement environ 8 000 à 10 000 individus, dont environ 1 600 femmes, se seraient installés de façon permanente sur les rives du Saint-Laurent pour fonder une famille (voir tableau 1) (Boleda, 1984; Charbonneau et al., 2000).

Tableau 1: Proportion d'immigrants (%) en 1765 selon leur lieu d'origine

Régions	Proportion d'immigrants (%)
Bretagne	4,8
Normandie	13,6
Paris	13,3
Loire	6,5
Nord	3,2
Est	7,2
Poitou-Charentes	21,3
Centre	3,3
Sud	11,6
Indéterminé	5,4
Total estimé	100

*Source: (Charbonneau et al., 2000)

Ces nouveaux immigrants proviennent de plusieurs provinces, particulièrement du centre et du nord-ouest de la France telles que Poitou, Charentes et Normandie (voir tableau 1) (Charbonneau et al., 2000). Ces différentes provinces d'origine apportent un aspect d'hétérogénéité à la population fondatrice puisque les régions de France étaient cloisonnées et culturellement et socialement très diversifiées (De Braekeleer, 1990). Les recherches antérieures ont aussi souligné la présence d'une immigration non française composée de Suisses, d'Allemands et de Portugais ainsi que d'une minorité d'Acadiens et d'Amérindiens (voir tableau 2) (Jetté, 1983; Harris et Matthews, 1987; Charbonneau, 1990). Finalement, il faut spécifier que l'on ne sait à quel point l'échantillon des fondateurs était diversifié sur le plan de l'apparentement. Guillemette et Légaré ont démontré que 30% des immigrants était apparentés et De Braekeleer souligne le fait que les émigrants provenaient de 38 provinces françaises et qu'ils migraient le plus souvent en solitaire (Charbonneau et al., 1987; Guillemette et Légaré, 1989; De Braekeleer, 1990).

Tableau 2: Proportion d'immigrants (%) en 1765 selon le pays d'origine

Pays de provenance	Proportion d'immigrants (%)
France	90,1
Acadie	3,4
Pays d'en haut	1,1
Colonies anglaises	1,1
Europe (sauf France)	3,1
Amérindiens	0,9
Autres	0,3
Total	100

*Source: (Charbonneau et al., 2000)

Jusqu'à la conquête anglaise, parmi les immigrants on retrouve plus de 90% d'individus d'origine française, c'est-à-dire provenant de France ou bien d'Acadie (voir tableau 2) (Charbonneau, 1990). Dès la fin du régime français, en 1763, on remarque une immigration anglophone. Les Britanniques (Loyalistes et immigrants provenant directement d'Angleterre) forment alors le groupe d'immigrants non français le plus important (Charbonneau, 1990). Ils se concentreront surtout vers le Haut-Canada, qui s'appellera plus tard l'Ontario. Néanmoins, certains vont s'établir dans le Bas-Canada (la future province de Québec), préférant les grandes villes comme Québec et Montréal ainsi que les Cantons de l'Est (Bouchard et Tremblay, 1995). On remarque également une immigration provenant d'Écosse et d'Irlande (Boileau, 1999b). En 1825, la population francophone demeurait cependant la plus importante avec 370 000 individus sur un total de 480 000 (77,1%) et environ 930 000 sur 1 112 000 (83,6%) habitants en 1861 (Bouchard et Tremblay, 1995).

Des travaux récents effectués par l'équipe d'Hélène Vézina et Marc Tremblay sur le peuplement du Québec et son patrimoine génétique viennent confirmer les informations recueillies antérieurement concernant les différents pays ayant participé prioritairement au

peuplement québécois selon les époques, et amènent de nouveaux renseignements concernant leur contribution génétique propre (Tremblay et al., 2004). Selon leurs résultats, la France a contribué pour 73,4% de l'immigration et ce, prioritairement dans la première période du peuplement. Les Îles britanniques et l'Allemagne, ayant contribué respectivement pour 4,7% et 1,7% des fondateurs, se sont quant à eux établis au Québec en majorité dans la période de 1700 à 1765. On observe également que la contribution au patrimoine génétique québécois n'est pas forcément dans les mêmes proportions. En effet, la France a contribué pour 86,8% du patrimoine génétique québécois alors que les Îles britanniques et l'Allemagne ont participé pour 2,35% et 0,36% respectivement. Cette contribution génétique accentuée pour les représentants de la France par rapport aux autres pays peut être expliquée en partie par la date de leur établissement respectif (Desjardins, 1990). En effet, les immigrants français sont les premiers à s'être installés au Québec alors que les immigrants d'origine anglophone ou allemande ont accusé un retard pour leurs dates d'établissement (Desjardins, 1990).

1.2 Occupation du territoire

Pour entrer dans le territoire de la Nouvelle-France, la voie la plus pratique était le fleuve Saint-Laurent. Les premiers peuplements se sont donc établis sur les rives de ce fleuve et de ces principaux affluents dans les alentours de ce qui sera plus tard la ville de Québec (Laberge et Mathieu, 1996). Dès 1636 on pouvait observer trois villes: Québec, Trois-Rivières et Montréal (Laberge et Mathieu, 1996). Il faut attendre jusqu'en 1672 pour pouvoir observer le début de l'expansion du territoire lorsque Jean Talon procède à une distribution massive de nouveaux fiefs (Laberge et Mathieu, 1996). Le territoire s'étend alors des seigneuries d'Argenteuil, de Vaudreuil, de Soulages et de Chateauguay à l'ouest de l'île de Montréal jusqu'à celle des Éboulements sur la rive nord et de Trois-Pistoles sur la rive sud en aval de Québec. De plus, on retrouve à cette époque une vingtaine de seigneuries distribuées autour de la péninsule gaspésienne et de la Côte-Nord (Laberge et Mathieu, 1996). Le mode d'expansion du territoire est influencé entre autres par le besoin qu'ont les immigrants d'être protégés contre les attaques des Amérindiens, par le besoin

d'être à proximité des principales rivières et par les caractéristiques physiques du territoire (Laberge et Mathieu, 1996).

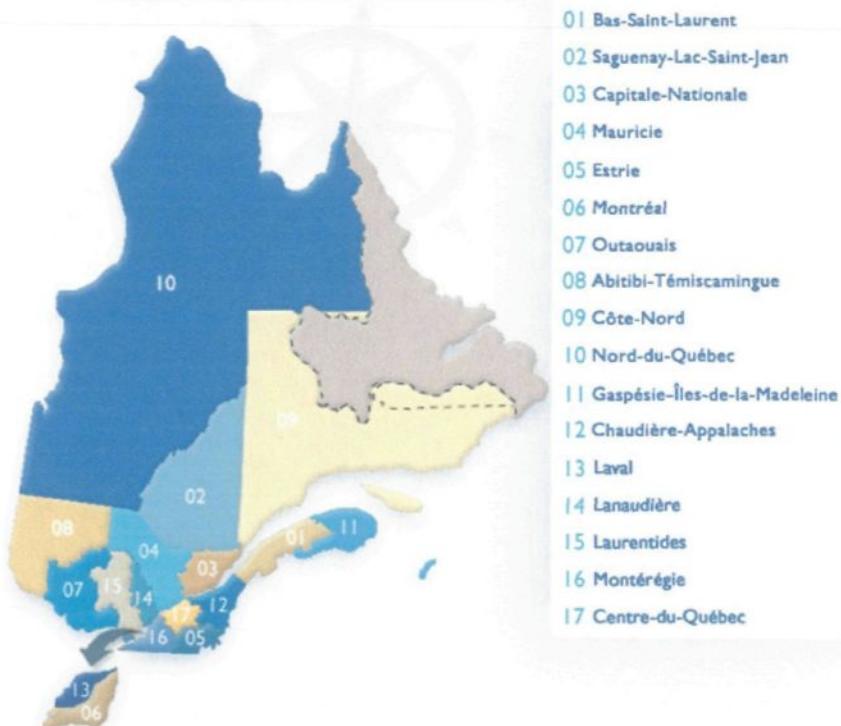
Vers 1720-1730 le territoire strictement laurentien est pratiquement constitué (Laberge et Mathieu, 1996). On observe par la suite un avancement dans la profondeur des terres qui va se poursuivre jusqu'aux premières décennies du 19^{ème} siècle (Laberge et Mathieu, 1996). En général, l'expansion du territoire se faisait beaucoup plus sur la rive sud que sur la rive nord du Saint-Laurent. Les mouvements de population le long du fleuve se faisaient du nord vers le sud-ouest, c'est-à-dire que le surplus de gens de la région de Québec et des environs avait tendance à migrer vers Montréal et les alentours, tandis que l'inverse ne se faisait aucunement ou très peu car les Montréalais s'installaient plutôt dans la plaine environnante (Bouchard et Tremblay, 1995). Peu après le début de l'expansion du territoire, c'est-à-dire de 1755 et 1763, a lieu la plus grande immigration acadienne qui est causée par leur déportation. De 1 500 à 4 000 Acadiens viennent s'installer dans la vallée du Saint-Laurent (Bouchard et Tremblay, 1995).

Entre 1835 et 1840, le surplus de population de la région de Charlevoix (qui avait été fondée majoritairement par des individus venant de Québec vers la fin du 17^{ème} siècle), entreprend la colonisation du Saguenay en s'installant à Chicoutimi (Bouchard et De Braekeleer, 1992). Vers 1850, le Bas-Saint-Laurent, Chaudière-Appalaches et le Centre-du-Québec sont explorés et ces territoires sont habités partiellement par de nouveaux individus (pour une carte des régions administratives du Québec voir figure 1). L'Outaouais compte lui aussi de nouveaux villages et l'exploration du territoire s'amorce lentement (Salone, 1970; Laberge et Mathieu, 1996; Boileau, 1999b). L'Estrie, qui avait premièrement été colonisée par les anglophones, reçoit à cette époque une vague importante d'immigration francophone (Boileau, 1999b; Bilodeau, 2002; Gobeil, 2002; Grégory, 2002).

Finalement, on remarque aussi dans la deuxième moitié du 19^{ème} siècle une nouvelle expansion du territoire le long des principaux cours d'eau telles les rivières Saguenay, Richelieu, Chaudière et Outaouais (Bouchard et De Braekeleer, 1992; Bouchard et al., 1996). Les régions des Laurentides et de l'Outaouais ont été peuplées principalement

par des Montréalais. Le Saguenay–Lac-Saint-Jean, quant à lui, a été essentiellement colonisé par des individus de Charlevoix. Finalement, ce sont majoritairement des individus des régions de Charlevoix et de la Gaspésie–Îles-de-la-Madeleine qui ont immigré dans la région de la Côte-Nord (Bouchard et De Braekeleer, 1992). Vers le début du 20^{ème} siècle se rajoute la région de l’Abitibi–Témiscamingue. Son peuplement est particulièrement stimulé par l’apparition du chemin de fer, qui traverse cette région pour la première fois en 1911 et qui favorise ainsi une diversification dans la région d’origine des fondateurs (Bouchard et De Braekeleer, 1992; Boileau, 1999a; Boileau, 2001).

Figure 1: Carte des régions administratives du Québec selon la stratification actuelle La cohorte de cette étude provient de l’enquête sur la santé cardiovasculaire de la



population québécoise effectuée en 1990. La stratification régionale utilisée alors était quelque peu différente, les régions 13 et 06 ainsi que les régions 16 et 05 y étaient regroupées. Source: (Portail Québec, 2005)

Le peuplement de plusieurs de ces régions, tels l’Abitibi–Témiscamingue, la Gaspésie et la Côte-Nord, à titre d’exemple, a été caractérisé par l’arrivée d’un nombre

restreint de premiers fondateurs (Bouchard et De Braekeleer, 1990). On sait que plusieurs autres régions ont subi un isolement plus ou moins marqué dans les premières décennies de leur colonisation. En effet, l'élaboration de routes a été beaucoup ralentie pour des causes géographiques autant sur la rive sud que la rive nord du Saint-Laurent à la hauteur de la Mauricie et des Bois-Francs (Bouchard, 1990). On sait aussi que les échanges entre la région du Témiscamingue et le reste du Québec étaient pour ainsi dire inexistant avant la construction du chemin de fer en 1911 (Boileau, 1999a; Boileau, 2001). Tous ces éléments laissent supposer que la diversité interrégionale du pool génique du Québec pourrait ne pas être qu'observable au Saguenay–Lac-Saint-Jean mais aussi dans les autres régions du Québec. Un autre aspect démographique qui influence encore aujourd'hui la diversité génétique de la population et dont il faut tenir compte est l'immigration multiethnique.

En 1996, au Canada on dénombrait 4 971 070 immigrants, dont 664 495 au Québec (Statistique Canada, 2001a). Ce nombre équivalait alors à 9,4% de la population québécoise totale (Institut de la statistique du Québec, 1996). Bien que les chiffres détaillés résultant du recensement de 2001 au Québec ne soient pas tous disponibles, on sait que la proportion d'immigrants continue d'y augmenter, passant à 9,9% en 2001 (Statistique Canada, 2001b). Les immigrants ne s'installent pas de façon aléatoire sur le territoire mais ont tendance à choisir des régions particulières (voir tableau 3). Les régions les plus touchées par l'immigration sont, en ordre d'importance, Montréal/Laval, Estrie/Montérégie, l'Outaouais et les Laurentides. Ensuite on descend sous la barre du 3% pour chuter jusqu'à 0,6% au Saguenay–Lac-Saint-Jean et dans le Bas–Saint-Laurent (Institut de la statistique du Québec, 1996). Cette immigration différentielle selon les régions peut être en partie liée aux langues parlées dans chacune de ces régions, comme l'illustre le tableau 4 (Institut de la statistique du Québec, 1996). En effet, on remarque dans les tableaux 3 et 4 que les régions qui attirent une plus grande quantité d'immigrants sont également celles qui possèdent le plus haut pourcentage d'individus utilisant l'anglais comme langue principale à la maison, c'est-à-dire les régions de Montréal/Laval, de l'Outaouais et de l'Estrie/Montérégie. Les grandes villes comme Montréal possèdent aussi des communautés ethniques utilisant leur langue maternelle respective ce qui est également un élément important comme pôle d'attraction. Les informations sur l'immigration sont importantes à

prendre en compte puisque l'arrivée d'un nombre important d'immigrants dans des régions spécifiques modifie la composition allélique de ces sous-populations en augmentant la diversité génétique présente dans les régions en question (Laberge, 1990).

Tableau 3: Quantité d'immigrants en nombre absolu et en pourcentage pour chacune des régions administratives du Québec

Régions	Quantité d'immigrants				Rang [†]
	An 1996 (nombre)	An 1996 (%)	Bilan de l'an 1986 à 1996 (nombre)		
Bas-Saint-Laurent	1 235	0,6	15		14
Saguenay–Lac-Saint-Jean	1 610	0,6	-75		14
Québec	16 490	2,6	3 600		7
Mauricie	3 045	1,2	350		10
Estrie / Montérégie	9 360/75 535	3,4/6,1	1 255/14 180		5/3
Lanaudière	7 245	2,0	2 170		8
Outaouais	14 955	4,9	5 080		4
Abitibi–Témiscamingue	1 695	1,1	-480		11
Côte-Nord	905	0,9	-195		12
Nord-du-Québec	285	0,9	285		12
Gaspésie–Îles-de-la-Madeleine	665	0,7	55		13
Laurentides	13 410	3,1	3 240		6
Chaudière-Appalaches	3 540	0,9	185		12
Centre-du-Québec	3 700	1,8	1 050		9
Montréal / Laval	462 905/47 825	26,5/14,6	89 085/17 830		1/2
Province de Québec	664 495	9,4	137 360		---

*Source: (Institut de la statistique du Québec, 1996)

[†]Le rang détermine le classement des régions en fonction de la quantité d'immigrants.

Tableau 4: Classement des régions administratives du Québec en fonction de la proportion (%) d'individus parlant français, anglais ou une autre langue à la maison

Régions	Langue maternelle parlée et rang associé					
	Français	Rang [†]	Anglais	Rang [†]	allophones	Rang [‡]
Bas-Saint-Laurent	99,4	1	0,3	15	0,1	13
Saguenay–Lac-Saint-Jean	99,1	2	0,4	14	0,1	13
Québec	97,2	5	1,2	11	0,9	8
Mauricie	97,7	4	0,8	12	1,1	6
Estrie / Montérégie	89,6 / 86,6	10 / 12	8,1 / 9,2	4	1,0 / 2,4	7 / 5
Lanaudière	96,9	6	1,6	10	0,8	9
Outaouais	78,6	13	16,7	2	2,4	5
Abitibi-Témiscamingue	95,0	7	3,4	9	0,6	10
Côte-Nord	87,7	11	5,0	7	6,6	4
Nord-du-Québec	49,0	16	4,4	8	44,3	1
Gaspésie–Îles-de-la-Madeleine	89,7	9	9,0	5	0,5	11
Laurentides	91,6	8	6,2	6	0,9	8
Chaudière-Appalaches	99,1	2	0,5	13	0,1	13
Centre-du-Québec	98,4	3	0,8	12	0,4	12
Montréal / Laval	53,8 / 76,9	15 / 14	24,1 / 10,1	1 / 3	17,1 / 9,3	2 / 3
Province de Québec	81,9	---	10,1	---	5,8	---

*Source: (Institut de la statistique du Québec, 1996)

[†]Le rang détermine le classement des régions en fonction de la proportion d'individus ayant comme langue maternelle le français, l'anglais ou tout autre langue que le français ou l'anglais.

Toutes les données disponibles sur la provenance des premiers fondateurs, sur leurs déplacements principaux lors des différents peuplements du territoire québécois ainsi que les données sur l'immigration au Québec sont des outils pouvant servir à expliquer ou peut-être même à anticiper les similarités et différences que l'on pense observer entre les régions du Québec. C'est donc en partie sur la base de ces informations du peuplement et des mouvements de populations que les résultats des regroupements régionaux en fonction de leur ressemblance génétique seront comparés.

CHAPITRE II

Génétique des populations et traits complexes

2.1 Génétique des populations

Chaque population possède son fardeau génétique. Ce fardeau est représenté par une prévalence plus élevée de certaines maladies génétiques dans une population, en réponse à son développement et à son évolution (Laberge, 1990). Le hasard de la méiose et de la présence des polymorphismes amène un risque pour chaque nouveau zygote, ou chaque nouvel individu, de posséder un ou deux gènes délétères à un locus quelconque de son génome. Dans une population, on appelle locus polymorphique d'un gène un locus où l'allèle le plus commun a une fréquence de moins de 95%, considérant qu'un allèle est une des formes possibles d'un gène donné qui diffère dans sa séquence d'ADN (King et Stansfield, 1997). Actuellement, on estime que la présence de gènes délétères amène un risque d'environ 10% de mourir d'une maladie génétique (Laberge, 1990). Chaque population possède son jeu propre de gènes délétères et leur prévalence fluctue selon les hasards de l'évolution et les mouvements de population (immigration et émigration), puisque les gènes « bougent » avec les individus. Pour expliquer l'évolution sur le plan génétique on peut considérer deux écoles de pensées différentes. Les darwinistes classiques (gradualistes) expliquent l'évolution par des changements génétiques qui se font de manière lente et graduelle alors que les darwinistes modernes (saltiques) considèrent que parfois ces changements peuvent survenir de manière plus soudaine (Laberge, 1990). Ce type de changements survient habituellement dans les populations isolées, par exemple de façon géographique ou culturelle. Certains facteurs tels la dérive génétique, les migrations et la consanguinité les influencent (Laberge, 1990). Comme le Québec n'a pas connu des taux de consanguinité plus élevé que la moyenne lors de son peuplement (De Braekeleer et al.,

1990; Scriver, 2001), ce sont surtout les phénomènes d'isolation, de migration et de dérive génétique qui seront approfondis dans les paragraphes suivants.

2.1.1 Les isolats et les dèmes

La création d'isolat, c'est-à-dire la création d'une nouvelle population ne bénéficiant d'aucune immigration ou échange avec les populations voisines, est fréquente lors de la colonisation de nouvelles terres (Laberge, 1990). Dans le cas du Québec, lors du peuplement de plusieurs régions, par exemple le Saguenay–Lac-Saint-Jean, il y a eu formation d'isolats. L'isolation du Saguenay–Lac-Saint-Jean a été causée particulièrement par sa situation géographique, étant entourée de montagnes avec, au départ de la colonie, aucune route praticable en dehors de la rivière Saguenay (Laberge et Mathieu, 1996). Encore de nos jours, comme le démontrent les statistiques actuelles, les régions périphériques comme le Saguenay–Lac-Saint-Jean (à l'exception de l'Outaouais et de l'Estrie) sont plus isolées tel que démontré par un faible pourcentage d'individus parlant l'anglais ou d'autres langues, européennes ou asiatiques par exemple, ce qui, comme le soulignent les tableaux 3 et 4, attire peu les nouveaux immigrants parlant une autre langue que le français (Institut de la statistique du Québec, 1996). Ces facteurs expliquent qu'il n'y ait pas eu beaucoup d'immigration interprovinciale et internationale. Le secteur de Charlevoix (situé dans la région de Québec) est un bon exemple de dème. La création de dèmes survient lorsque la fondation d'un nouveau territoire se fait par une quantité restreinte d'ancêtres (Laberge, 1990). Effectivement, on sait que les premiers fondateurs de Charlevoix provenaient en très grande partie de la ville de Québec et des environs et qu'ils formaient un petit effectif (Bouchard et De Braekeleer, 1990). Tant et aussi longtemps que cet effectif reste en dessous de la capacité socioéconomique et culturelle de la région, elle est considérée comme un dème (Laberge, 1990).

2.1.2 L'effet fondateur

L'effet fondateur est un autre phénomène qui survient parfois lors de la colonisation d'un nouveau territoire. Celui-ci est provoqué lorsqu'une petite partie de la population

souche quitte celle-ci pour aller fonder une nouvelle population (Laberge, 1990). Cette seconde population sera donc formée d'un nouvel assortiment aléatoire d'allèles provenant de la population souche. Habituellement, les allèles compris dans la sous-population seront ceux qui étaient représentés en abondance dans la population mère. Plusieurs allèles seront donc absents de cette nouvelle population. Par contre, il peut aussi arriver que des allèles rares soient surreprésentés dans celle-ci. La nouvelle population peut donc être très différente de la population souche. C'est ce qui est arrivé lors des premiers peuplements de la Nouvelle-France, puis lors de la fondation de la région de Charlevoix et du Saguenay–Lac-Saint-Jean. Les effets fondateurs sont souvent accompagnés d'un phénomène de dérive génétique, ce qui augmente la différenciation de la nouvelle population par rapport à la population mère.

2.1.3 La dérive génétique

La dérive génétique est le phénomène par lequel certains allèles seront fixés alors que d'autres disparaîtront selon la fécondité différentielle des porteurs (Laberge, 1990). Les fréquences alléliques d'une génération sont toujours différentes de celles de la génération précédente parce qu'elles sont le fruit des méioses faites au hasard. Plus la population est petite, plus la dérive génétique sera importante puisqu'il y aura moins de diversité possible et que l'absence de descendance de certains individus ne sera pas compensée dans le pool génique global de la population. Dans une grande population, le fait qu'un adulte ne puisse transmettre son allèle est contrebalancé par la présence de plusieurs autres adultes qui ont le même allèle et qui vont le transmettre aux générations suivantes. C'est cet effet de moyenne qui se produit beaucoup moins dans une petite population. La dérive génétique est donc un effet statistique, découlant entièrement du hasard (Laberge, 1990). Certains facteurs, comme l'isolation, ont comme effet d'augmenter la dérive génétique. En effet, lorsqu'il y a échange du bagage génétique entre les populations cela vient contrer la dérive génétique par un apport en diversité (Laberge, 1990). Le Saguenay–Lac-Saint-Jean est un bon exemple de dérive génétique subie dans une petite population fondatrice et augmentée par un phénomène d'isolation.

Les aspects démogénétiques d'isolat, de dème, d'effet fondateur et de dérive génétique, intimement liés aux mouvements et migrations des populations, ont donc un effet considérable sur la composition génétique d'une population. Étant donné que le peuplement du Québec a été un terrain propice à la survenue de ce type d'événements, il est donc possible que l'on observe une hétérogénéité génétique entre les différentes régions du Québec. Ce phénomène s'observe d'ailleurs par la variation de la prévalence de traits monogéniques d'une région à l'autre du Québec. Par exemple il y a entre autres sept maladies beaucoup plus fréquentes au Saguenay–Lac-Saint-Jean dont 5 sont récessives (l'acidose lactique, l'ataxie de Charlevoix–Saguenay, la tyrosinémie de type I, la fibrose kystique et la neuropathie sensitivo-motrice) et deux dominantes (la dystrophie myotonique et l'hypercholestérolémie familiale) (CORAMH). On retrouve également dans l'est du Québec des maladies particulières comme la maladie de Lafora et le syndrome de Morquio (Chan et al., 2003; Rioux et al., 2003). C'est pourquoi les objectifs de cette étude sont de définir des patrons de distribution de polymorphismes impliqués dans des maladies complexes fréquentes et d'explorer des techniques de modélisation de cette nouvelle information pour éventuellement développer des outils pour des fins de santé publique.

2.2 Les gènes et variants étudiés

Le présent projet porte à la fois sur l'asthme et sur les maladies cardiovasculaires parce que ces affections sont fréquentes dans la population et comportent un aspect commun de leur physiopathologie, soit la cascade inflammatoire. Tous les gènes ciblés par cette étude ont donc un rôle direct ou indirect documenté dans la cascade inflammatoire de l'un ou de ces deux types de maladies, ils seront donc traités dans ce mémoire en fonction de leur rôle biologique dans cette voie inflammatoire. Ils sont tous (sauf IL4) documentés dans la population canadienne-française du Québec. Les bases de données *Asthma and Allergy Genes Database* (AAGD, 2004) et *GenBack* (OGMD, 2004) ont été très utiles lors du choix des gènes puisqu'elles renferment des données sur les gènes impliqués respectivement dans les maladies de type respiratoire ou cardiovasculaire. Par exemple, on y retrouve les variants documentés pour chacun des gènes répertoriés, les études d'associations et les mutations. Le tableau 5 dresse la liste de tous les gènes et variants

ciblés par ce projet de maîtrise ainsi que leur codification dans le *National Center of Biotechnology Information* (NCBI) en regard de leur code d'accès (NM) ainsi que pour les variants génétiques choisis (rs).

La cascade inflammatoire est un processus biologique complexe caractérisé par la mise en scène des cellules du système immunitaire (tels les lymphocytes T et B, les macrophages, les mastocytes et les éosinophiles) et de plusieurs molécules de signalisation immunes (par exemple les cytokines et les interleukines). On peut regrouper les huit gènes ciblés pour cette étude dans deux « voies » différentes de la cascade inflammatoire (voir figure 2). La première est caractérisée par la différenciation des monocytes en macrophages puis en cellules spumeuses (figure 2a). La deuxième est définie par le processus débutant par la présentation de l'antigène aux lymphocytes T auxiliaires, précisément T_{H1} et T_{H2} , jusqu'à l'activation des cellules B et des mastocytes (figure 2b). La fonction de chacun des gènes sélectionnés, expliquant leur intégration dans la première ou la seconde voie, est décrite dans les paragraphes suivants.

2.2.1 *La différenciation des monocytes*

Cinq des gènes ciblés sont impliqués dans le processus de différenciation des monocytes (figure 2a). Les monocytes, qui font partie intégrante du système phagocytaire dit mononucléé (un seul noyau), sont des cellules circulantes, c'est-à-dire qu'elles se retrouvent dans le système sanguin (Goldsby et al., 2001a). Elles y demeurent pour environ 8 heures, pour ensuite se diriger vers les tissus et se différencier en macrophages fixes ou mobiles, spécifiques au tissu en question (Goldsby et al., 2001a). Ces derniers font également partie du système phagocytaire mononucléé et servent de première ligne de défense du corps humain grâce à leur capacité pour la phagocytose (Goldsby et al., 2001a). Les macrophages présents dans les vaisseaux sanguins accumulent parfois en grande quantité du cholestérol dérivé des lipoprotéines, devenant ce que l'on appelle des cellules spumeuses. Leur présence a comme effet d'augmenter les probabilités de développer de l'athérosclérose (Libby, 2002). Les paragraphes suivants, en plus de définir chacun des 5

Tableau 5: Position et codification du *National Center for Biotechnology Information* pour les gènes et variants ciblés pour cette étude

Associations	Symbole du gène	Position génétique	Codification		
			dans la base NCBI (NM)	Variant(s)	rs
à l'asthme	IL4	5q31.1	NM_000589	C-589T*	rs_2243250
	MS4A2 (FCER1B)	11q13	NM_000139	E237G*	rs_569108
aux maladies cardiovasculaires	PPARA	22q12-	NM_005036	L162V	rs_1800206
		11q13.1			
	LPL	8p22	NM_000237	D9N	rs_1801177
				G188E	ND [†]
	APOE	19q13.2	NM_000041	P207L	ND [†]
				N291S	rs_268
				Isoforme E2	ND [‡]
aux deux types d'affection	ACE	17q23	NM_000789	I/D	ND [‡]
				Isoforme E3	ND [‡]
	PPARG2	3p25.2	NM_015869	P12A	rs_1801282
				Isoforme E4	ND [‡]

*Dans l'exemple du variant C-589T, on retrouve le nucléotide de départ (T), la position négative de la substitution (-589), donc avant le premier codon du gène, et le nouveau nucléotide (C). Pour les autres variants tel E237G, tous situés à l'intérieur même du gène, on peut observer l'acide aminé de départ (E), la position du codon contenant le polymorphisme (237) et l'acide aminé résultant (G).

[†]Ces polymorphismes d'une seule paire de base n'ont pas encore été répertoriés et n'ont donc pas de numéro rs (NCBI, 2004).

[‡]La nomenclature ne s'applique donc pas dans ces cas-ci.

gènes ciblés impliqués dans la différenciation des monocytes, indiqueront également par quel mécanisme ceux-ci sont impliqués dans les traits étudiés et quelles sont les associations documentées pour chacun d'entre eux avec l'asthme ou les maladies cardiovasculaires.

2.2.1.1 Enzyme de conversion de l'angiotensine

L'enzyme de conversion de l'angiotensine (ACE) a été localisé sur le chromosome 17 du génome humain en 1989, plus particulièrement au locus 17q23 (Mattei et al., 1989). Cet enzyme a comme fonction principale d'effectuer la conversion de l'angiotensine I en angiotensine II dans le système rénine-angiotensine, système de régulation de la pression corporelle (Erdos, 1977). On observe la présence de ACE à la fois dans les sites d'inflammation de l'asthme et dans ceux des maladies cardiovasculaires.

Ce gène a un rôle documenté à la fois dans la différenciation des monocytes et dans la présentation de l'antigène (voir figure 2a et 2b). En effet, des recherches ont démontré la présence de ACE dans la section de haute vascularisation des plaques athérosclérotiques (*shoulder region*), à l'endroit où on retrouve un regroupement de lymphocytes T et de macrophages. Cet enzyme participerait probablement à la croissance de la plaque (Diet et al., 1996). Son action dans la présentation de l'antigène sera expliquée dans la section suivante. Pour ce qui est de son rôle dans la différenciation des monocytes dans les plaques athérosclérotiques, par le biais de l'angiotensine II, ACE induirait l'expression des régulateurs de l'interleukine 6 (IL-6) dans les macrophages et les cellules musculaires lisses *in vitro* (voir figure 2a) (Schieffer et al., 1996). Il aurait également comme effet de stimuler la synthèse d'enzymes de dégradation matricielle et de récepteurs des LDL chez les macrophages, stimulant donc l'absorption de LDL dans ceux-ci (voir figure 2a) (Hamanaka et al., 1992).

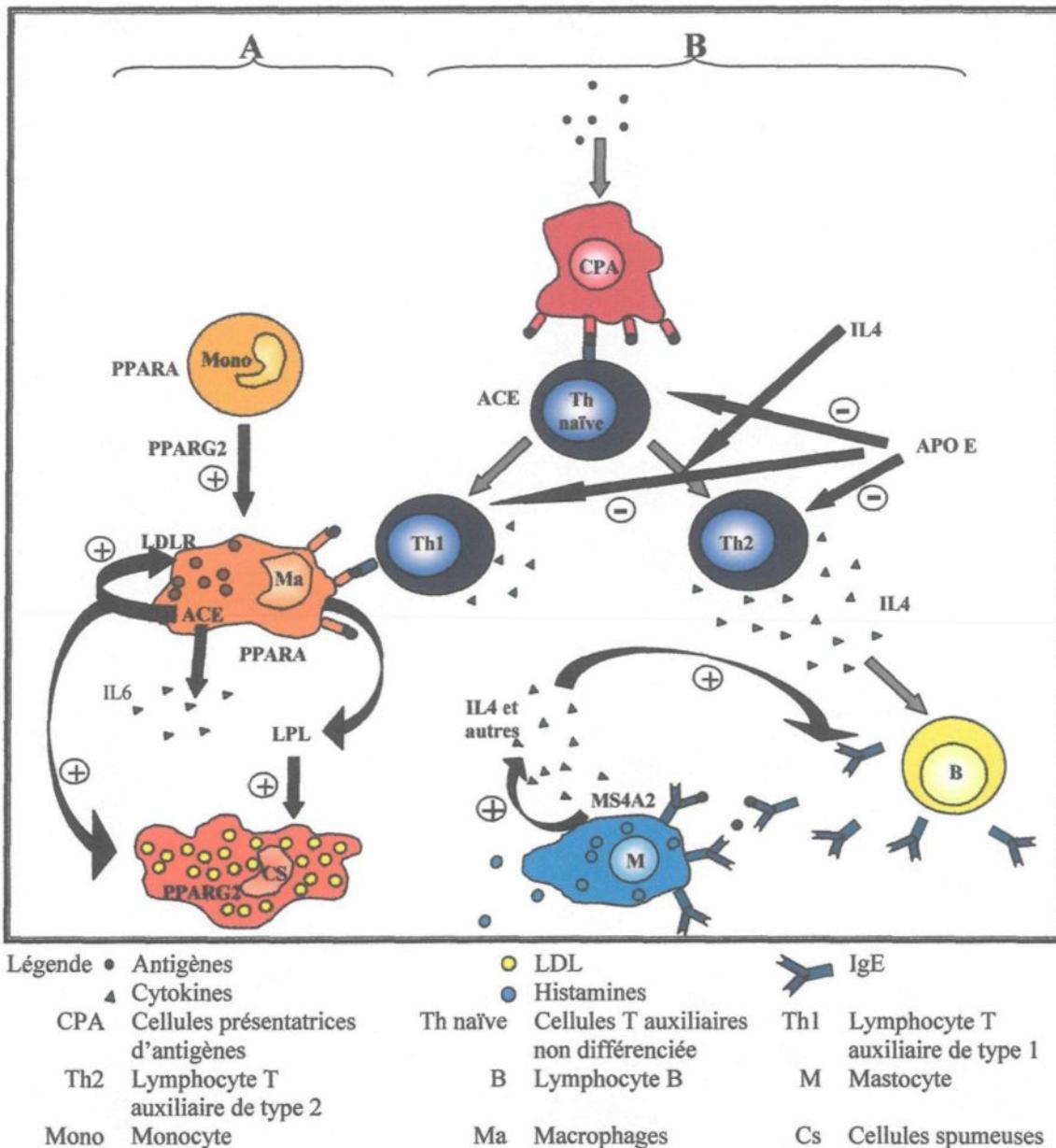


Figure 2: Intégration des 8 gènes ciblés pour cette étude à l'intérieur de la cascade inflammatoire reliée à l'asthme et aux maladies cardiovasculaires (A et B). La figure illustre deux voies possibles de la cascade inflammatoire. La voie A, dans laquelle s'insère cinq des gènes sélectionnés pour cette étude, correspond à la différenciation des monocytes en macrophages une fois que ceux-ci ont atteint leur tissu cible (Goldsby et al. 2001a). Ces macrophages servent de défense de première ligne pour l'organisme (Goldsby et al. 2001a). La seconde étape illustrée est la transformation des macrophages en cellules spumeuses suite à l'intégration d'une grande quantité de cholestérol (Libby 2002). Quatre gènes sont intégrés dans la voie B illustrant une des possibilités de cascade survenant suite à la présentation de l'antigène par une cellule présentatrice d'antigène à un lymphocyte T naïf. Ceci active les lymphocytes T et déclenche un processus de prolifération clonale (Goldsby et al. 2001b). La sécrétion de cytokines (IL4) active par la suite les lymphocytes B qui, via la sécrétion d'immunoglobulines (IgE), activent d'autres cellules du système immunitaire tels les mastocytes (M) (Goldsby et al. 2001b). Ces derniers, retrouvés entre autres dans les tractus gastriques et respiratoires, jouent un rôle important dans le développement de l'allergie (Goldsby et al. 2001a).

Le variant ciblé pour le gène ACE n'est pas un polymorphisme d'une seule paire de bases (SNP) mais bien une délétion de 287 paires de bases située dans l'intron 16 du gène (Rigat et al., 1990). Les génotypes D/D, c'est-à-dire que l'individu possède les deux copies du gène avec délétion, ou I/D, l'individu a une copie avec l'insertion et une seconde avec la délétion, ont été définis comme un facteur de risque important de l'athérosclérose (Arbustini et al., 1995).

2.2.1.2 « *Peroxisome proliferator-activated receptor gamma 2* »

Le *peroxysome proliferator-activated receptor gamma 2* (PPARG2) a été localisé sur le chromosome 3 du génome humain, au locus 3p25 (Greene et al., 1995). Les *peroxysome proliferator-activated receptors* (PPARs) font partie de la sous-famille des récepteurs nucléaires qui, une fois en hétérodimère avec les *retinoid X receptors*, régularisent la transcription de plusieurs gènes (NCBI, 2004). On retrouve dans cette sous-famille les PPAR-alpha, -gamma et -delta. Les PPARs gamma existent sous deux isoformes différents, c'est-à-dire deux formes du gène qui ont une séquence d'acides aminés qui diffère légèrement mais qui codent néanmoins pour deux protéines effectuant la même fonction avec plus ou moins la même efficacité, soit l'isoforme 1 (PPARG1) et l'isoforme 2 (PPARG2) (Tontonoz et al., 1994; King et Stansfield, 1997). Ces deux isoformes sont identiques à l'exception d'un ajout de 84 nucléotides à l'extrémité 5' de l'isoforme 2 (Elbrecht et al., 1996).

Plusieurs rôles dans la cascade inflammatoire ont été démontrés pour PPARG2 au niveau de la différenciation des monocytes (voir figure 2a). L'équipe de Welch lui suppose un rôle physiologique dans la régulation de la réponse immunitaire innée et acquise (Welch et al., 2003). Cette molécule induirait la différenciation des monocytes en macrophages (voir figure 2a). Il a également été démontré que PPARG2 est exprimé dans les cellules spumeuses dérivées des macrophages se trouvant dans les plaques athérosclérotiques (Marx et al., 1998; Ricote et al., 1998; Tontonoz et al., 1998). Il serait aussi un régulateur clef dans l'expression des gènes des cellules spumeuses (voir figure 2a) (Nagy et al., 1998). L'article de Clark (Clark, 2002) dresse une liste intéressante de ses rôles comme agent

inflammatoire et anti-inflammatoire. En ce qui concerne particulièrement l'asthme, une augmentation de l'expression du gène PPARG2 a été observée dans les voies respiratoires de personnes asthmatiques, plus particulièrement dans l'épithélium ainsi que dans les muscles lisses (Benayoun et al., 2001).

Un variant a été ciblé pour ce gène, soit le variant P12A, qui se situe uniquement dans l'isoforme 2 (NCBI, 2004). Ce variant est le résultat de la substitution dans le gène d'une cytosine (C) par une guanine (G) qui entraîne le changement d'une proline (P) pour une alanine (A) au codon 12 de la protéine (Yen et al., 1997). Bien que pour l'instant il n'y ait pas eu d'étude d'association entre P12A et la composante inflammatoire des maladies cardiovasculaires de façon directe, il a été démontré dans plusieurs cohortes que l'allèle le plus commun, Pro12, augmentait les risques de développer du diabète de type II alors que l'allèle Ala12 avait plutôt tendance à être protecteur (Altshuler et al., 2000; Lohmueller et al., 2003).

2.2.1.3 « *Peroxisome proliferator-activated receptor alpha* »

Le *peroxysome proliferator-activated receptor alpha* (PPARA) est situé sur le chromosome 22 au locus 22q12-q13.1 (Sher et al., 1993). PPARA est un autre membre de la sous-famille des récepteurs nucléaires ayant donc comme fonction la régularisation de la transcription de plusieurs gènes suite à sa liaison avec le *retinoid X receptor* (NCBI, 2004). La première fonction supposée de PPARA a été la modulation de l'inflammation. Certaines études démontrent que PPARA aurait des propriétés anti-inflammatoires (Marx et al., 1999; Aljada et al., 2001). En effet, il a été démontré que l'activation de PPARA inhibe l'activité inflammatoire des éicosanoides comme LTB4 en augmentant l'expression d'enzymes hépatiques (Devchand et al., 1996). Il aurait également des effets anti-athérosclérotiques via son action sur l'inhibition des cellules d'adhésion vasculaires ainsi que sur la suppression de la migration des monocytes/macrophages (voir figure 2a) (Marx et al., 1999).

Le polymorphisme L162V, variant documenté du gène PPARA, est causé par la substitution d'une leucine (L) par une valine (V) au 162^{ème} acide aminé de la protéine. En

fait, ce changement est dû au remplacement d'une cytosine (C) par une guanine (G) (Vohl et al., 2000). L'allèle V162 augmenterait les risques de développer de l'athérosclérose via une augmentation plasmatique de lipoprotéines contenant de l'apolipoprotéine B (Vohl et al., 2000). Par contre, l'équipe de recherche Gouni-Berthold émet l'hypothèse que le variant PPARA-L162V apporterait une protection contre l'athérosclérose et les maladies coronariennes pour les individus atteints du diabète de type II (Gouni-Berthold et al., 2004).

2.2.1.4 Lipase lipoprotéique

Le gène de la lipase lipoprotéique (LPL) a été localisé et séquencé pour la première fois en 1987 (Sparkes et al., 1987; Wion et al., 1987). Dans le génome humain on le retrouve au locus 8p22 du chromosome 8. La LPL, située dans l'endothélium des capillaires sanguins, est responsable du catabolisme des chylomicrons et des lipoprotéines de très faible densité (VLDL). De cette activité résulte des acides gras libres et des résidus riches en cholestérol (Nilsson-Ehle et al., 1980). Des études antérieures laissent supposer que la LPL pourrait avoir un rôle dans le processus de l'athérosclérose (Zilversmit, 1973; Steinberg, 1983). Une certaine proportion de macrophages synthétiseraient des LPL, ce qui favoriserait leur développement en cellules spumeuses (voir figure 2a). Ceci ayant été observé à la fois chez les rats, les lapins et chez l'humain (Khoo et al., 1981; Chait et al., 1982). Il a été aussi démontré que plusieurs types de cellules retrouvées dans les plaques athérosclérotiques répondaient aux anti-LPL (Jonasson et al., 1987). Dans la présente étude, quatre variants ont été ciblés pour la lipase lipoprotéique.

Premièrement, D9N, SNPs bien documenté, qui consiste en la substitution d'une guanine (G) pour une alanine (A), provoquant le changement d'un acide aspartique (D) pour une asparagine (N) au 9^{ème} codon de la protéine (Elbein et al., 1994; Mailly et al., 1995; Rouis et al., 1996). Ce variant est associé à l'hypertriglycéridémie et à une réduction de la concentration plasmatique de lipoprotéines à haute densité en cholestérol (de Bruin et al., 1996). La présence de ce variant a pour effet d'augmenter le taux de triglycérides plasmatiques de 20% et de diminuer celui des lipoprotéines à haute densité (HDL) de

0,08 mmol/L, ce qui a récemment été associé à une augmentation des risques de développer de l'athérosclérose (Wittrup et al., 1999).

À la position 188 du gène de la LPL on observe également le changement d'une glycine (G) par un acide glutamique (E). Ce SNP est causé par la substitution d'une cytosine (C) par une alanine (A) (Emi et al., 1990; Monsalve et al., 1990; Bergeron et al., 1992; Henderson et al., 1992). LPL-G188E a comme effet d'augmenter de 78% le taux plasmatique de triglycérides ainsi que de diminuer le taux d'HDL de 0,25 mmol/L, supposant ainsi une association avec une augmentation des risques de développer de l'athérosclérose (Wittrup et al., 1999).

Le variant P207L est retrouvé en majorité dans la population canadienne-française (Ma et al., 1991a). Il consiste en une substitution d'une cytosine (C) pour une thymine (T), résultat du changement d'une proline (P) par une leucine (L) à la position 207 de la protéine (Ma et al., 1991b; Normand et al., 1992; Levesque et al., 1994). Les variants G188E et P207L ne sont distants que de 19 acides aminés, ce qui laisse croire qu'ils seraient dans le même bloc haplotypique. Leurs effets sur le métabolisme des lipides seraient liés, et par conséquent les risques sous-jacents de développer de l'athérosclérose le seraient également (Normand et al., 1992).

Finalement, le dernier variant ciblé pour le gène LPL dans cette étude est le N291S qui est le résultat de la substitution d'une alanine (A) pour une guanine (G), provoquant le changement d'une asparagine (N) pour une sérine (S) au 291^{ème} acide aminé de la protéine (Ma et al., 1994; Reymer et al., 1995). Ce SNP, d'abord associé à l'hyperlipidémie combinée familiale (Reymer et al., 1995), jouerait aussi un rôle dans l'augmentation des risques de développer de l'athérosclérose car sa présence a comme effet d'augmenter de 31% le taux de triglycérides plasmatiques et de diminuer de 0,12 mmol/L celui des HDL plasmatiques (Wittrup et al., 1999).

2.2.1.5 Récepteur des lipoprotéines à faible densité

Le gène codant pour le récepteur des lipoprotéines de faible densité, soit LDLR, est situé sur le chromosome 19, plus précisément au locus 19p13.1-13.3 (Lindgren et al., 1985). Le récepteur des LDL joue un rôle important dans le métabolisme du cholestérol puisqu'il participe activement à son endocytose (Brown et Goldstein, 1979). D'ailleurs, la quantité de LDLR à la surface membranaire des macrophages joue un rôle dans leur transformation en cellules spumeuses et donc dans le processus athérosclérotique (voir figure 2a) (Hamanaka et al., 1992; Linton et al., 1999). Le LDLR a été associé à l'athérosclérose également parce que la présence de variants dans ce gène a comme effet d'augmenter la concentration plasmatique de cholestérol-LDL (c-LDL) et de cholestérol total, augmentant ainsi les risques de développer des maladies cardiovasculaires artérielles (Brown et Goldstein, 1986; Gaudet et al., 1999).

Le variant ciblé est le W66G. Il est caractérisé par la substitution d'une thymine (T) par une guanine (G) qui a comme effet de modifier le 66^{ème} acide aminé de la protéine, donnant une glycine (G) en lieu et place du tryptophane (W) habituel (Leitersdorf et al., 1990). Dans la littérature, ce variant est associé particulièrement avec l'hypercholestérolémie familiale (Leitersdorf et al., 1990). Ce SNP, qui apparaît dans le domaine ligand de la protéine, c'est-à-dire qui est situé exactement à l'emplacement où le récepteur se lie avec les LDL, empêche complètement toute liaison avec les molécules de LDL, laissant augmenter la concentration de cholestérol total et de c-LDL plasmatique. Ceci a pour effet d'augmenter les risques athérosclérotiques (Leitersdorf et al., 1990). Chez les hétérozygotes, on observe la moitié moins de cholestérol total et de c-LDL plasmatique, l'allèle normal du gène étant apte à maintenir son niveau d'activité (Moorjani et al., 1993).

Tous ces gènes ont donc un effet direct ou indirect sur l'activité des monocytes et macrophages dans la cascade inflammatoire (c'est-à-dire la migration des monocytes et des macrophages, la production d'interleukine et la régularisation de la transcription) ou encore sur la transformation des macrophages en cellules spumeuses. Les 4 autres gènes ciblés par cette étude, y compris le gène ACE qui est impliqué dans les deux « voies » de la cascade

inflammatoire, ont une fonction s'insérant dans la voie caractérisée par la présentation de l'antigène.

2.2.2 La présentation de l'antigène

Quatre des gènes ciblés sont impliqués dans le processus de présentation de l'antigène et d'activation du système immunitaire. La section B de la figure 2 représente une des possibilités de cascade biologique survenant suite à la présentation de l'antigène par une cellule présentatrice d'antigène (CPA) à un lymphocyte T auxiliaire (T_H) naïf. Lorsque les récepteurs à antigène des cellules T_H entrent en contact avec le complexe du CMH de classe II de la CPA, ces cellules T_H sont activées et entrent dans une phase de prolifération clonale (Goldsby et al., 2001b). Ensuite, via la sécrétion de cytokines telle IL4, elles participent à l'activation des lymphocytes B (Goldsby et al., 2001b). Dans notre exemple, cela sera suivi d'une réponse humorale au cours de laquelle les cellules B vont se différencier en deux populations distinctes, une qui va sécréter les anticorps et l'autre qui sera composée de cellules mémoires (Goldsby et al., 2001b). Une fois que les anticorps sont sécrétés, par exemple les IgE, ceux-ci vont entrer en contact avec les antigènes et vont se lier à des cellules du système immunitaire comme les mastocytes. Les mastocytes sont des cellules retrouvées dans de nombreux tissus comme la peau, les tissus conjonctifs et les muqueuses des tractus respiratoires et digestifs (Goldsby et al., 2001a). Elles possèdent des granules cytoplasmiques contenant entre autres de l'histamine et des cytokines et détiennent un rôle important dans le développement de l'allergie (Goldsby et al., 2001a). Les paragraphes suivants, en plus de définir chacun des 4 gènes ciblés impliqués dans la présentation de l'antigène, indiqueront également à quel niveau ceux-ci sont impliqués et quelles sont les associations documentées pour chacun d'entre eux avec l'asthme ou les maladies cardiovasculaires.

2.2.2.1 Interleukine 4

C'est en 1986 que le gène codant pour l'interleukine 4 (IL4) a été isolé chez l'humain (Yokota et al., 1986). Il est situé sur le chromosome 5 au locus 5q31.1

(Sutherland et al., 1988; Le Beau et al., 1993). Il existe deux isoformes différents de la cytokine IL4. La différence entre les deux isoformes réside dans la présence d'un exon en surplus chez l'isoforme 1 en comparaison avec l'isoforme 2 (Center for Biomolecular Science & Engineering, 2004; NCBI, 2004). L'interleukine 4 est une cytokine de première importance dans les étapes de la cascade inflammatoire. Elle est exprimée par les lymphocytes T auxiliaires de type 2 ($T_{H}2$) (Hamalainen et al., 2001). Elle agit comme un facteur de croissance et de différenciation auprès des cellules B (voir figure 2b) et incite la différenciation des monocytes en cellules dendritiques (Howard et al., 1982; Isakson et al., 1982). L'équipe de Swain a aussi démontré son important rôle trophique via son rôle dans la différenciation des cellules $T_{H}2$ (voir figure 2b) (Swain et al., 1990). Suite à ces nombreuses découvertes, une association, suscitée par la présence de polymorphismes (SNPs) dans le gène IL4, a été proposée avec l'allergie et l'asthme (Kelly-Welch et al., 2003).

Un des polymorphismes bien documentés du gène IL4 est le C-589T. Comme son nom l'indique, c'est le changement de la base azotée cytosine (C) à l'emplacement -589 dans le promoteur du gène, pour la base azotée thymine (T) (Rosenwasser et al., 1995). Ce SNP est présent chez l'isoforme 1 et 2 (Center for Biomolecular Science & Engineering, 2004). Bien que l'association n'ait jamais été retrouvée lors de subsequentes expérimentations, le SNP C-589T a été associé à une augmentation de la concentration totale sérique d'IgE *in vivo* (Rosenwasser et al., 1995). De plus, ce variant a été associé à plusieurs reprises avec l'asthme et l'atopie dans des populations caucasiennes et asiatiques (Rosenwasser et al., 1995; Walley et Cookson, 1996; Noguchi et al., 1998).

2.2.2.2 Récepteur de haute affinité pour les IgE

Le gène codant pour le récepteur de haute affinité pour les IgE, soit MS4A2 (FCERIB), est beaucoup moins documenté que l'IL4, n'ayant été localisé sur le locus 11q13 du chromosome 11 chez l'humain qu'en 1993 (Sandford et al., 1993). MS4A2 est impliqué dans l'initiation de la réaction allergique de type 1 chez les basophiles et les mastocytes (Cookson et al., 1992; Sandford et al., 1993). On le retrouve à la surface des

mastocytes (Stevenson et al., 1996) et son activation serait à l'origine de la libération par ceux-ci d'IL4 et d'autres cytokines (voir figure 2b), ceci ayant possiblement comme effet l'augmentation de la synthèse des IgE et de surcroît du processus inflammatoire (Plaut et al., 1989). Ce gène a été associé à l'atopie par l'équipe de Sandford en 1993 (Sandford et al., 1993).

Le variant ciblé pour cette étude, E237G, consiste en la substitution d'une alanine (A) pour une guanine (G), provoquant le changement d'un acide glutamique pour une glycine au 237^{ème} acide aminé de la protéine (Hill et Cookson, 1996). Dans les recherches antérieures, ce SNP a été associé à l'atopie ou bien à l'asthme avec atopie (asthme allergique) successivement dans une population australienne, une population japonaise et une population canadienne-française (Hill et Cookson, 1996; Shirakawa et al., 1996; Laprise et al., 2000).

2.2.2.3 Enzyme de conversion de l'angiotensine

L'enzyme de conversion de l'angiotensine (ACE), comme mentionné antérieurement, est également impliqué dans la voie caractérisée par la présentation de l'antigène. Cet enzyme a comme fonction principale d'effectuer la conversion de l'angiotensine I en angiotensine II dans le système rénine-angiotensine (Erdos, 1977), système de régulation de la pression corporelle. L'hypothèse serait que ACE aurait un effet chémotactique sur les lymphocytes T via l'angiotensine II (voir figure 2b) (Weinstock et al., 1987). Pour ce qui est des maladies cardiovasculaires, selon l'équipe de Diet (Diet et al., 1996), ACE serait présent dans les plaques athérosclérotiques, plus particulièrement dans la section de haute vascularisation (*shoulder region*) à l'endroit où on retrouve un regroupement de lymphocytes T et de macrophages, et participerait probablement à la croissance de la plaque (Diet et al., 1996). Sur le plan de l'asthme, on a observé que l'utilisation des inhibiteurs de ACE augmentait l'hyperréactivité bronchique chez les patients atteints d'asthme via sa capacité d'inactivation des bradykinines et des tachykinines (Semple et Herd, 1986; Kaufman et al., 1992; Gao et al., 2000).

En plus des associations mentionnées précédemment pour le polymorphisme ACE-I/D avec l'athérosclérose, il a aussi été démontré que le génotype D/D était plus fréquent dans une cohorte d'asthmatiques (Benessiano et al., 1997; Gao et al., 2000). Par contre, lors d'expérimentations dans une seconde cohorte, cette dernière affirmation n'a pas pu être corroborée (Tomita et al., 1998).

2.2.2.4 Apolipoprotéine E

Le gène codant pour l'apolipoprotéine E (APOE), se situe sur le chromosome 19 du génome humain, plus précisément au locus 19p13.2 (Das et al., 1985; Lusis et al., 1986). Cette molécule joue un rôle important dans le catabolisme des résidus de chylomicrons et de VLDLs (Biserte et Fruchart, 1973; Shepherd, 1994). Elle est intégrée à ces molécules lipidiques au cours du métabolisme des lipides et est reconnue par le récepteur des lipoprotéines à faible densité lors de la prise de ces molécules lipidiques par les cellules hépatiques (Brown et Goldstein, 1986). Le gène APOE a été associé au développement de l'athérosclérose suite à plusieurs observations. Entre autres, des études sur des souris ont démontré que l'APOE a des effets bénéfiques sur l'athérosclérose parce qu'elle inhibe: a) la prolifération et l'activation des lymphocytes (Kelly et al., 1994; Mistry et al., 1995; Ishigami et al., 1998); b) l'agrégation plaquettaire et la production d'oxyde nitrique par les plaquettes (Riddell et al., 1997) et; c) parce qu'elle a des effets antioxydants sur la peroxydation lipidique (Huang et al., 1997; Tangirala et al., 2001). Dans le cas de l'APOE, les variants choisis sont ses trois isoformes les plus documentées, soit E2, E3 et E4, chacun ayant une combinaison différente d'acides aminés aux positions 112 et 158 de la protéine résultante (Weisgraber et al., 1981; Rall et al., 1982).

L'isoforme E2, se distingue par la présence d'une cystéine aux positions 112 et 158 de la protéine (Weisgraber et al., 1981; Rall et al., 1982). L'allèle E2 est associé en général à un taux plus bas de c-LDL et de cholestérol que l'allèle E3, ce dernier étant considéré comme référence puisqu'il est le plus fréquent (Utermann et al., 1979; Eichner et al., 2002). On attribue aussi à l'isoforme E2 une augmentation des triglycérides plasmatiques (Dallongeville et al., 1992). Cette dernière affirmation entre par contre en contradiction

avec certains autres résultats de recherche témoignant d'une baisse des triglycérides plasmatiques chez les individus en santé (Xhignesse et al., 1991). Cette baisse du cholestérol total accompagnée d'une hausse des triglycérides plasmatiques a des influences sur les risques de développer de l'athérosclérose. D'ailleurs, il a été démontré que le génotype APO E3/E2 est associé à une élévation des risques de développer des troubles athérosclérotiques des carotides (de Andrade et al., 1995).

L'isoforme E3 est considéré comme celui qui confère une fonction normale à la protéine résultante, avec une efficacité optimale. C'est l'isoforme le plus fréquent dans toutes les populations humaines (Corbo et Scacchi, 1999). Il est composé d'une cystéine au 112^{ème} acide aminé de la protéine et d'une arginine à la position 158 (Weisgraber et al., 1981; Rall et al., 1982).

Finalement, l'allèle E4 codant pour l'isoforme E4 serait l'allèle ancestral, celui à l'origine de tous les autres (Davignon, 2001). À la position 112 de sa protéine résultante il possède une arginine ainsi qu'à son 158^{ème} acide aminé (Weisgraber et al., 1981; Rall et al., 1982). Cet allèle a été associé à une grande concentration de cholestérol sanguin et à une diminution du taux plasmatique de c-LDL (Weisgraber et al., 1981; Das et al., 1985; Paik et al., 1985; Eichner et al., 2002) mais cette diminution est de 2 à 3 fois moins importante que celle causée par l'allèle E2 (Eichner et al., 2002). Cette association entre par contre en contradiction avec les résultats de certains autres groupes de recherche (Bouthillier et al., 1983; Sing et Davignon, 1985). L'allèle E4 a également été associé à une augmentation du taux de triglycérides plasmatiques (Dallongeville et al., 1992). De façon plus spécifique, il est considéré comme un facteur de risque important pour le développement des maladies athérosclérotiques mais son effet diminuerait avec le vieillissement (Ilveskoski et al., 1999).

Les recherches antérieures ont démontré pour chacun de ces 4 derniers gènes un rôle dans la cascade inflammatoire, plus précisément dans l'activation et la progression du processus immunitaire acquis soit au niveau de la différenciation des lymphocytes T_H non différenciés, soit au niveau des lymphocytes T_{H1}, T_{H2} et B ou bien au niveau de

l'activation des mastocytes. Au total, les 8 gènes ciblés par cette étude ont donc un rôle dans le processus inflammatoire (soit au niveau de la différenciation des monocytes et de la formation des cellules spumeuses, soit au niveau de la réponse immunitaire suite à la présentation d'un antigène) en plus d'être associés de façon spécifique au développement de l'asthme ou des maladies cardiovasculaires qui sont très fréquentes dans la population tant mondiale que canadienne.

2.3 La prévalence de l'asthme

L'asthme est une maladie respiratoire très répandue dans le monde et affecte toutes les couches de la société. Elle représente un taux de morbidité/mortalité très élevé et est aussi à l'origine de coûts importants, mondialement plus lourds que ceux du VIH et de la tuberculose réunis (OMS, 2000). On compte plus de 300 millions de personnes souffrant d'asthme dans le monde (Masoli et al., 2004). Au Canada en 2004, la prévalence de l'asthme était estimée à 14,1% et le taux de mortalité 1,6% (Masoli et al., 2004).

2.4 La prévalence des maladies cardiovasculaires

Les maladies cardiovasculaires sont une cause importante de mortalité et de morbidité au niveau mondial. Dans le monde on estime à 16,7 millions le nombre de décès par année causés par ces maladies (WHO, 2003). De plus, c'est un type de maladie qui touche toutes les couches de la société. Dans la population canadienne des 35 ans et plus, 4% de la population en 1997 ont déjà été diagnostiqués pour une maladie cardiovasculaire et on estime que ces affections sont responsables de 36% de la mortalité globale (Fondation des maladies du cœur du Canada, 2003).

La prévalence de ces deux types de maladies dans la population québécoise les désigne comme des cibles intéressantes pour des recherches sur les possibilités de caractérisation des facteurs de risque génétiques impliquées dans des traits complexes fréquents dans la population générale. Une meilleure définition du rôle de ces facteurs

génétiques dans ces maladies pourrait donc améliorer les services de prévention et de traitements disponibles et par le fait même la qualité de vie d'une grande quantité de personnes.

CHAPITRE III

Distribution géographique de polymorphismes associés à l'asthme ou aux maladies cardiovasculaires dans une cohorte de la population du Québec

Article en préparation.

3.1 Avant-propos

Ce chapitre souligne les principaux résultats obtenus lors de cette étude de maîtrise visant à caractériser une portion de la variabilité génétique pour deux traits complexes fréquents dans la province de Québec qui sont l'asthme et les maladies cardiovasculaires, en utilisant 11 variants génétiques documentés pour leur rôle dans l'un de ces types de maladie (ou les deux).

Les différentes analyses réalisées lors de ce projet ont démontré pour la première fois la présence d'une variabilité génétique observable pour les maladies complexes fréquentes dans la population générale du Québec. Des possibilités de modélisation de cette information sous forme de regroupement des régions possédant le même profil génétique pour les polymorphismes ciblés ainsi que de regroupements des variants ayant une distribution géographique semblable ont également été évaluées.

Ma contribution dans cet article débute avec l'accomplissement de toutes les étapes de la manipulation (hormis les deux variants génotypés par les équipes de Marie-Claude Vohl et Louis Pérusse). J'ai également participé aux analyses statistiques, sous la direction de Louis Houde. Par la suite j'ai rédigé le manuscrit en tant que premier auteur. Toutes ces étapes, de la manipulation à l'écriture, ont été réalisées sous la direction de Catherine Laprise.

Cet article scientifique est présentement en cours de correction par les coauteurs et a nécessité la collaboration de plusieurs chercheurs appartenant à des domaines variés. Premièrement, Louis Houde, statisticien à l'UQAC, a effectué les analyses de classification descendante hiérarchique et a participé à l'écriture des aspects statistiques de cet article. Hélène Vézina, chercheure au Groupe de recherche interdisciplinaire en démographie et épidémiologie génétique (GRIG), a corrigé les aspects concernant l'histoire et la démographie de la population québécoise, alors que Marie-Claude Vohl, chercheure à l'Unité de recherche sur les maladies lipidiques de l'Université Laval, en plus de génotyper le variant PPARA-L162V dans son laboratoire, a révisé les aspects concernant les maladies

cardiovasculaires. L'équipe de Louis Pérusse, chercheur à l'Unité de recherche sur les maladies cardiovasculaires athérosclérotiques de l'Université Laval, a génotypé le variant ACE-I/D. Les échantillons d'ADN provenant de l'enquête sur la santé cardiovasculaire de la population québécoise effectuée en 1990 ont été fournis par Philip W. Connelly, chercheur au département de médecine de l'Université de Toronto. Ce mémoire et cet article s'intègrent dans le volet III du projet multicentrique « ÉCOGÈNE-21; de l'ADN à la communauté », coordonné par Daniel Gaudet du Centre de médecine génique communautaire de l'Université de Montréal. De plus, ce projet est un pilote du projet CART@GENE, idée originale de Claude Laberge, chercheur à l'Université Laval au département de génétique humaine. Ces deux chercheurs figurent également parmi les coauteurs. L'ensemble des coauteurs a révisé le manuscrit.

Title: Modeling genetic information from geographic distribution of SNPs associated with chronic inflammatory diseases in the Quebec population

Authors: Madore A-M^{1,2}, Houde L¹, Vézina H¹, Vohl M-C^{3,4}, Pérusse L^{3,5}, Mior N², Connelly PW⁶, Laberge C², Gaudet D², Laprise C^{1,2}.

Affiliations: 1) Université du Québec à Chicoutimi, Saguenay, Canada 2) University of Montreal Community Genomic Medicine Center, Chicoutimi University Hospital, Saguenay, Canada 3) Lipid Research Center and Department of Food Science and Nutrition, Laval University, Québec city, Canada 5) Department of Social and Preventive Medicine, division of Kinesiology, Laval University, Québec city, Canada 6) Departments of Medicine and Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Canada.

Address for correspondence:

UQAC and University of Montreal Community Genomic Center
Chicoutimi University Hospital
305 Saint-Vallier,
Chicoutimi, Québec, G7H 5H6
(418) 541-1077
claprise@saglac.qc.ca

Key words: Genetic risk factors, hierarchical clustering, asthma, cardiovascular diseases, French-Canadian population.

3.2 Résumé

Le but de cette recherche est de vérifier la présence d'une hétérogénéité génétique dans la population québécoise en utilisant des variants associés à deux traits complexes et d'explorer des moyens de la caractériser afin de créer des outils utilisant ce paramètre de santé publique. Pour ce faire, 1680 échantillons provenant de toutes les provinces du Québec ont été génotypés pour 11 SNPs distribués dans 8 gènes impliqués dans l'asthme et/ou les maladies cardiovasculaires. Avec les fréquences alléliques, une analyse de classification hiérarchique a été faite pour regrouper les régions puis les variants géniques. Trois groupes de régions et un groupe de SNPs (explicables en partie par les données démographiques), ont ainsi été formés pour chaque type de maladie. Ce projet pilote laisse penser qu'une étude utilisant un plus grand nombre de participants et de SNPs permettrait d'organiser les informations génétiques de façon utile pour le système de santé québécois.

3.3 Abstract

Objectives: The purpose of this project was to evaluate the potential of the downward hierarchical clustering analysis (DHCA) for studying genetic heterogeneity in 15 sociosanitary regions of the Province of Quebec (Canada). **Methods:** The study relied on an anonymized sample of 1680 individuals from 15 different sociosanitary regions who had participated in the Quebec Hearth Health Survey in 1990-91 (QHHS). Genotyping of 11 variants in 8 candidate genes known to be involved in chronic inflammatory diseases, namely asthma and cardiovascular diseases (CVD) was performed using the ARMS and RFLP techniques. From this set of 11 variants, only those showing an allelic frequency $> 2\%$ in the QHHS study ($N=8$) were selected. DHCA techniques were then applied to model the geographical distribution of these 8 gene variants in 15 Quebec sociosanitary regions and to study genetic heterogeneity. **Results:** The DHCA allowed to group sociosanitary regions and gene variants on the basis of genetic variability. For both asthma and CVD, three significant clusters of sociosanitary regions and one cluster of gene variants were identified. **Discussion:** This study illustrates that DHCA might be useful in studying genetic heterogeneity at the population level. It suggests that this analysis may constitute an additional tool for the integration of genetic information in population studies or public health activities.

3.4 Introduction

The demographic history of the province of Quebec (Canada) is well documented. At the present time, the Quebec population numbers more than 7 million people. Approximately 55% of the population lives in the southwestern part of the province, which is more heterogeneous (19 languages spoken in the vicinity of Montreal) than the rest of the Quebec population, composed mainly of French Canadian descendants [Institut de la statistique du Québec 2004]. It has been estimated that 25,000 settlers came from various provinces of France, especially from the northwestern part, between the beginning of the 17th century and the British conquest of 1763 [Charbonneau, et al. 2000; Harris and

Matthews 1987]. The population size increased steadily after the institution of a settlement plan in 1660, and reached over 70,000 inhabitants by the time of the British conquest. The subsequent growth of the French Canadian population is notable due to the high reproduction rate of the following generations [Charbonneau, et al. 2000]. Immigration waves of people from the British Isles as well as American loyalists also contributed to the population increase [Harris and Matthews 1987]. In recent decades, however, immigration has occurred from all continents, explaining the actual increasing diversity of the Quebec population.

Some research about the introduction and dissemination of monogenic traits in the French Canadian population of Quebec have underlined a genetic variability at the regional level [Vézina 1996]. Moreover, the study of the worldwide distribution of polymorphisms (SNPs) associated with phenylalanine hydroxylase (PAH) and alleles at the CFTR and HLA loci point toward a genetic differentiation between the regional populations of the province of Quebec [Scriver 2001]. This genetic diversity is also observed for variants in lipoprotein lipase (LPL) and the low-density lipoprotein-receptor (LDLR) genes that cause hypertriglyceridemia and familial hypercholesterolemia, respectively [Bergeron, et al. 1992; Davignon and Roy 1993; Dionne, et al. 1993; Normand, et al. 1992; Vohl, et al. 1997].

Even though the studies described above demonstrate a certain genetic diversity at the regional level, little is known about the distribution of SNPs involved in complex traits or biological processes throughout regional populations. Moreover these studies were limited to the French Canadian component of the Quebec population. In the pilot research presented herein, we wanted to verify whether it is possible to identify sociosanitary regions at higher risk for common chronic inflammatory diseases such as allergic asthma and cardiovascular diseases (CVD), two pro-inflammatory phenotypes [Libby 2002]. We selected these two common diseases because they have an important morbidity/mortality rate and represent a large proportion of health expenses in our society. Throughout the world, approximately 300 million persons are asthmatic [Masoli, et al. 2004]. In Canada for

the year 2004, the prevalence was estimated at 14.1% and the fatality rate at 1.6% [Masoli, et al. 2004]. CVD are also widespread around the world, with 16.7 million deaths each year [WHO 2003]. CVD account for 36% of the global mortality in Canada [Heart and Stroke Foundation of Canada 2003]. Thus, asthma and CVD are common worldwide and increasingly contribute to the public health burden of diseases. It has been recently proposed that a significant portion of common diseases such as asthma and CVD are the result of common genetic variants, and that susceptibility mutations that are at a relatively high frequency are responsible for much of the clinical and economic burden of diseases seen today [Botstein, 2003 #246]. While there is some controversy as to the magnitude of this burden, there are good reasons to expect that at least some of the genetic risks of asthma, CVD and other chronic inflammatory diseases are due to common variants, based on evolutionary arguments, and the fact that most of the human genetic variation is common.

The list of common variants that cause common diseases, like inflammatory diseases, is growing steadily. Among the significant ones, known to be associated with asthma or CVD are: the interleukine-4 (IL4) and the IgE Fc receptor subunit beta (MS4A2), which were selected for their involvement in the Th2 pathway of allergic respiratory diseases [Hamalainen, et al. 2001; Sandford, et al. 1993]; the peroxysome proliferators-activated receptor alpha (PPARA), the lipoprotein lipase (LPL) and the low-density lipoprotein receptor (LDLR) genes, which were selected for their association with cardiovascular diseases [Chait, et al. 1982; Marx, et al. 1999] and the apolipoprotein E (APOE) gene for its pleiotropic influence on cardiovascular diseases and other inflammatory diseases, among which its ability to inhibit lymphocyte proliferation and activation [Kelly, et al. 1994]. Finally, the peroxysome proliferator-activated receptor gamma 2 (PPARG2) and the angiotensin 1 converting enzyme (ACE) genes have been reported to be involved in the inflammatory pathway of both allergic respiratory and cardiovascular diseases, among others [Hamanaka, et al. 1992; Kaufman, et al. 1992; Welch, et al. 2003]. Not all of these variants are directly causing inflammation. However, they are all associated with biological processes leading directly or indirectly to chronic inflammation. Their association with inflammatory diseases and their suspected high allelic

frequency (based on several case-control studies) explain their selection for the purpose of the present study. Although the selected gene variants all contribute to chronic inflammatory disease relative risk (RR), their contribution to the population attributable risk (PAR) is not well documented and no population-based samples have been used yet for estimation. By virtue of its random sampling design, the Quebec Heart Health Study (QHHS) cohort contains a mix of individuals from different ethnic background, with and without disease. It has the power to estimate the prevalence of variants with allelic frequency (AF) of $\geq 5\%$ in the general population. From the list of genes listed above, we have selected 11 common variants which have been tested in the present study (Table 1).

Obviously our selection of gene variants is not exhaustive and much genetic information has already been collected about asthma and cardiovascular diseases [Blumenthal 2004; Lusis, et al. 2004]. This study is a pilot designed to explore ways of describing genetic heterogeneity and geographical clusters in a manner that would be helpful and meaningful for population studies and public health management. Within the limits of our sample, we were able to observe the genetic heterogeneity of the Quebec population at the sociogeographical regional level, using the selected variants. Regardless of the bias of the study, we tried to find out if it was possible a) to group geographical regions on the basis of their allelic structure and b) to make clusters of 11 selected susceptibility SNPs for chronic inflammatory diseases, particularly asthma and cardiovascular diseases according to their geographical distribution patterns. The primary goal was to evaluate the possibility to model genetic and sociogeographic information at the community and population levels.

3.5 Methods

3.5.1 Sample

The sample used in this study is composed of anonymized DNA drawn from 1680 individuals having participated in the QHHS conducted in 1990 [Institut de la statistique du

Québec 1990] as part of a pan-canadian survey (Figure 1). The data collection was anonymous and performed with appropriate stratification for age (between 18 and 74 years old) and sex and without selection bias for cardiovascular diseases. The QHHS sampling was representative of the Quebec population and covered the 15 Quebec sociosanitary regions (Figure 1): Bas-Saint-Laurent (01), Saguenay–Lac-Saint-Jean (02), Québec (03), Mauricie (04), Estrie (05), Montréal (06), Outaouais (07), Abitibi-Témiscamingue (08), Côte-Nord (09), Nord-du-Québec (10), Gaspésie-Îles-de-la-Madeleine (11), Chaudière-Appalaches (12), Centre-du-Québec (13), Lanaudière (14) and Laurentides (15).

3.5.2 DNA Genotyping

DNA was extracted from blood lymphocytes using the guanidine hydrochloride-proteinase K method [Jeanpierre 1987]. Genotyping was done using the standard polymerase chain reaction (PCR) with the amplification refractory mutation system (ARMS) [Newton, et al. 1989] for C-589T-IL4, E237G-MS4A2 variants or the restriction fragment length polymorphism technique (RFLP) [Wyman and White 1980] for P12A-PPARG2, D9N-LPL, G188E-LPL, N291S-LPL, P207L-LPL, I/D-ACE, E2/E3/E4-APOE, W66G-LDLR, L162V-PPARA variants. Denaturation, annealing, and extension temperature of standard PCR and quantities for ARMS or RFLP mix are subject to little changes according to the SNPs tested (data not shown).

3.5.3 Data analysis

For the 11 selected SNPs, AF was calculated at the provincial level and for each sociosanitary region separately. We used the AF to verify whether the study sample was in Hardy-Weinberg equilibrium in the different sociosanitary regions. Three of the 11 SNPs (G188E-LPL, P207L-LPL and W66G-LDLR) had an allelic frequency lower than 2% in this sample and were not used in subsequent analyses. The N291S-LPL variant, had an AF of 1.25%, and was included in the analysis because its AF reaches 2% in many regions.

A contingency table was formed with the AF obtained for each SNP in each sociosanitary region to use downward hierarchical clustering analysis (DHCA) in order to characterize the distribution patterns of the variants [Kaufman and Rousseeuw 1990]. The principle is to separate the initial group into two subgroups on the chi-square distance basis and to recursively apply the procedure downward until every element forms its own group. At each step, the best grouping or cluster is obtained by maximizing the distance over every possible grouping [Kaufman and Rousseeuw 1990]. The p-value is evaluated as an indicator of the relevance of splitting the groups into two different subgroups at this step. However, p-value is influenced by the statistical power of this analysis. During the procedure, it is impossible to reconsider previous steps of division. The contingency table is formed by the geographical regions and the SNPs AF so it can be used to group regions or cluster variants. Analyses were also performed separately for variants known to be specifically involved in asthma or CVD.

3.6 Results

Eight SNPs chosen for their biological relevance with one or both of the complex diseases selected were conserved for statistical analyses. They all had an allelic frequency ranging between 1.3% and 55.2% (Table 1). Hardy-Weinberg equilibrium was tested for conserved SNPs in all regions ($8 \times 15 = 120$ combination). Only 11 combinations were in disequilibrium.

Downward hierarchical clustering analysis have then been performed and allowed the identification of three clusters significantly associated with asthma SNPs (Figure 2a): 1) Nord-du-Québec, Gaspésie–Îles-de-la-Madeleine, Québec, Mauricie and Centre-du-Québec; 2) Estrie, Lanaudière and Laurentides; 3) Outaouais, Saguenay–Lac-Saint-Jean, Abitibi-Témiscamingue, Bas-Saint-Laurent, Côte-Nord and Chaudière-Appalaches. Montréal was not included in any group of regions by the DHCA. Applying the same analysis to CVD SNPs, 3 different clusters were identified (Figure 2b): 1) Nord-du-Québec, Montréal, Chaudière-Appalaches, Bas-Saint-Laurent and Laurentides; 2) Saguenay–Lac-

Saint-Jean and Outaouais; 3) Estrie, Gaspésie–Îles-de-la-Madeleine, Côte-Nord, Centre-du-Québec, Mauricie, Québec and Abitibi-Témiscamingue. The DHCA did not include Lanaudière in any of these groups.

Then, using DHCA, we clustered together variants that had a similar distribution pattern throughout the regions. This led to the identification of one genetic cluster for asthma (I/D-ACE and P12A-PPARG2 SNPs) (Figure 2c) and another for CVD (I/D-ACE, P12A-PPARG2, E4-APOE and N291S-LPL variants) (Figure 2d).

3.7 Discussion

Our results illustrate the genetic heterogeneity of the Quebec population at the regional level for common variants associated with chronic inflammatory diseases. This study also suggests that DHCA might be useful for grouping regions and variants on the basis of genetic variability at the population level.

To underline the genetic variability of the population, the sample was divided according to administrative areas following the 1990 cardiovascular survey division [Santé Québec 1990]. To group regions according to their allelic frequency composition and variants on the basis of their distribution pattern, we used the DHCA. With the use of the p-values given by the chi-square distances calculated for each separation step, this analysis is appropriate to obtain groups of regions or variants that are similar. The downward classification is an exploratory analysis. Therefore, p-values should be interpreted with caution, and the size of the cohort of each group should be considered.

Indeed, although DHCA provided interesting and promising results, its potential for population studies and the validity of these results, as applied to the QHHS sample, have limitations due to potential *selection bias, regression-dilution bias, population admixture and multiple testing*. The majority of participants in the QHHS represent a population-based sample. Likely, those who agreed to be screened represent a selected group of

“interested volunteers”, and if they are more health conscious than those who did not agree to be screened, our estimates of relative risks may “underestimate” the true population AF of inflammatory disease-associated variants. *Selection bias* applies to the criteria used to select the gene variants. The list of selected variants is not exhaustive, and the panel only represents a subset of genome variations potentially associated with asthma, CVD or other chronic inflammatory diseases. Indeed, the use of other SNPs would produce different groupings. *Regression dilution* is associated with an underestimation of the true association between a given factor and its outcome overtime when classification of individuals is based on only one measurement taken at baseline. Since our analysis is solely based on cross sectional data, we cannot statistically adjust the AF coefficients over time. While the population is always in movement, modifying at the same time the genetic pool of regions, the results obtained in this pilot with the QHHS samples cannot be applied to the present population. *Population stratification* could also occur. This phenomenon is caused by undetected population substructure that can potentially produce false estimation of allelic frequencies. The way the sample was divided could also produce the same kind of error. In this pilot the divisions were made according to the sociosanitary regions of the 1990 cardiovascular survey, but other kinds of division, for example based on the principal known movements of population, would probably give other groupings. Another obstacle in the analysis of AF in different regions of a given population are the *multiple testing* problems. Multiple testing in the present study relies on the number of geographical sub-regions (n=15), variants (N=8) and on the criteria used to define them.

However, Hardy-Weinberg equilibrium tests demonstrated that only 11 combinations out of 120 combinations of frequency distribution of a SNP in a particular region are in disequilibrium with significant results. This is expected with the level used for the tests and the size of the cohort within each region separately.

As with historical studies that have underlined the existence of a demographic heterogeneity of founders and epidemiogenetic studies of the contemporary population of Quebec, the results obtained with the DHCA reveal the presence of a heterogeneity

between certain regions or groups of regions [Bouchard 1990; Bouchard and De Braekeleer 1990; Charbonneau, et al. 2000; Gagnon and Heyer 2001; Scriver 2001; Vézina 1996]. We used this analysis to group regions according to allele frequencies of susceptibility variants for asthma and CVD separately and to reveal modeling possibilities.

Comparisons with major events of Quebec settlement history or geographic proximity can partly explain our findings, but in each of our generated groups, the presence of some regions cannot be explained in that fashion. A possible explanation is the presence of smaller migration movements between regions. Heyer hypothesized that, considering all demogenetic events that occurred in the Saguenay–Lac-Saint-Jean region (for example differential fecundity and a high level of birthrate), if only 2% of the first founders coming from Charlevoix (68 individuals) were carriers of the same deleterious gene, this was sufficient to account for the allelic frequencies of the common monogenic diseases in today's population [Heyer 1995]. Moreover, with more variants to represent each type of disease, we could probably obtain a better overview of the possible groups of regions. Another explanation could be the size of the cohort. After the distribution of the samples, some regions were represented by a very small cohort (only 70 alleles for the Nord-du-Québec region). A small cohort could induce allele frequencies that do not really represent the population of the region and provoke wrong grouping of regions or SNPs. It also decreases the power of this analysis, increasing chances that classification analysis can not detect differences between regions.

The DHCA was also performed for clustering genetic variants. It gave us one cluster for each type of disease. Regional profile of AF obtained for each SNP did not validate all p-values for each cluster. The lack of power can be a factor because of the small size of the cohort and because the number of low frequencies (under 5%) has a high influence on the chi-square distance. Like the grouping of regions, the small size of our cohort per region could affect the way in which clusters of variants were made.

The genes selected in our investigation were all associated with an increased risk for asthma and/or CVD [Berger and Moller 2002; Clark 2002; Diet, et al. 1996; Gaudet, et al. 1999; Sandford, et al. 1993; Semple and Herd 1986; Swain, et al. 1990; Tiret, et al. 1994; Zilversmit 1973]. Consequently, developing a genetic map of the regional populations of Quebec and grouping similar regions could provide a very useful tool to target zones at higher risk in the province and to calculate relative risks, taking into account both environmental and genetic factors. Clustering selected variants with similar distribution patterns is also a way to better describe a relative risk. Moreover, it could make it possible to design a screening procedure that includes susceptibility variants and clustering information.

3.8 Conclusions

We found that it is possible to organize a great quantity of data in a comprehensive and practical way based on statistical tools such as the DHCA. This project merely demonstrated that it could be possible to model genetic information. Optimal use of genetic information in health care requires an understanding of 1) genetic risk alleles, 2) the distribution of these risk factors in a population, and 3) the environmental factors that interact with genetic variants. In the last decade, substantial progress was made to understand highly penetrant monogenic diseases. Conversely, our understanding of the genetics of complex disorders that depends on a contribution of multiple genes and environmental determinants, such as inflammatory diseases remains limited. The discovery of common risk alleles is consistent with the common disease-common variant hypothesis for complex diseases, which proposes that common disorders are primarily due to the segregation of hundreds or thousands of common genetics variants [Cargill, et al. 1999]. Common and rare diseases that have an underlying genetic etiology can systematically be studied with tools afforded by the Human Genome Project and large-scale epidemiological approaches. The latter require an understanding of genetic diversity at the population level. Research initiatives are needed to translate the knowledge of genetic diversity into individual genetic predispositions and collectively into determinants of public health.

Furthermore, they will allow the integration of genomics into medical services, health promotion and prevention of diseases.

3.9 Acknowledgments

A.-M. Madore was supported by the Quebec Respiratory Health Training Program of the Quebec Respiratory Network of the Fonds de recherche en santé du Québec (FRSQ) (supported by the Canadian Institutes of Health Research (CIHR)) and by the FRSQ. M.-C. Vohl is a research scholar from the FRSQ. D. Gaudet is the chairholder of the Canada Research Chair in preventive genetics and community genomics. C. Laprise is the chairholder of the Canada Research Chair for genetic determinants in asthma (www.chairs.gc.ca). This work was supported by the CIHR, as part of the ECOGENE-21: from DNA to the community project (CAR43283).

3.10 References

- Berger J, Moller DE. 2002. The mechanism of action of PPARs. *Annu. Rev. Med.* 2002. 53:409-35 53:409-35.
- Bergeron J, Normand T, Bharucha A, Ven Murthy MR, Julien P, Gagne C, Dionne C, De Braekeleer M, Brun D, Hayden MR and others. 1992. Prevalence, geographical distribution and genealogical investigations of mutation 188 of lipoprotein lipase gene in the French Canadian population of Quebec. *Clin Genet.* 41(4):206-210.
- Blumenthal M. 2004. The immunopathology and genetics of asthma. *Minn Med* 87(5):53-6.
- Bouchard G. 1990. Reproduction familiale et "effets multiplicateurs". *Histoire d'un génome.* Sillery: Presses de l'Université du Québec. p 213-252.
- Bouchard G, De Braekeleer M. 1990. Mouvements migratoires, effets fondateurs et homogénéisation génétique. *Histoire d'un génome.* Sillery: Presses de l'Université du Québec. p 281-322.
- Brown MS, Goldstein JL. 1986. A receptor-mediated pathway for cholesterol homeostasis. *Science* 232(4746):34-47.
- Institut de la statistique du Québec. 1990. Survey on the cardiovascular health of the Québec population. http://www.stat.gouv.qc.ca/cadrisq/enquete_an.htm. Consulted in February 2005.
- Cargill M, Altshuler D, Ireland J, Sklar P, Ardlie K, Patil N, Shaw N, Lane CR, Lim EP, Kalyanaraman N and others. 1999. Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nat Genet* 22(3):231-8.

Chait A, Iverius PH, Brunzell JD. 1982. Lipoprotein lipase secretion by human monocyte-derived macrophages. *J Clin Invest* 69(2):490-3.

Charbonneau H, Desjardins B, Légaré J, Denis H. 2000. The Population of the St. Lawrence Valley, 1608-1760. In: Michael Haines et Richard Steckel, Editor. *A Population History of North America*. Cambridge: Cambridge University Press. p 99-142.

Clark RB. 2002. The role of PPARs in inflammation and immunity. *Journal of Leukocyte Biology* 71:388-400.

Cookson WO, Young RP, Sandford AJ, Moffatt MF, Shirakawa T, Sharp PA, Faux JA, Julier C, Nakamura Y, Nakamura Y and others. 1992. Maternal inheritance of atopic IgE responsiveness on chromosome 11q. *Lancet* 340(8816):381-4.

Davignon J, Roy M. 1993. Familial hypercholesterolemia in French-Canadians: taking advantage of the presence of a "founder effect". *Am J Cardiol*. 72(10):6D-10D.

Diet F, Pratt RE, Berry GJ, Momose N, Gibbons GH, Dzau VJ. 1996. Increased accumulation of tissue ACE in human atherosclerotic coronary artery disease. *Circulation* 94(11):2756-67.

Dionne C, Gagne C, Julien P, Murthy MR, Roederer G, Davignon J, Lambert M, Chitayat D, Ma R, Henderson H and others. 1993. Genealogy and regional distribution of lipoprotein lipase deficiency in French-Canadians of Quebec. *Hum Biol*. 65(1):29-39.

Gagnon A, Heyer E. 2001. Fragmentation of the Quebec population genetic pool (Canada): evidence from the genetic contribution of founders per region in the 17th and 18th centuries. *Am J Phys Anthropol* 114(1):30-41.

Gaudet D, Vohl MC, Couture P, Moorjani S, Tremblay G, Perron P, Gagne C, Despres JP. 1999. Contribution of receptor negative versus receptor defective mutations in the LDL-receptor gene to angiographically assessed coronary artery disease among young (25-49 years) versus middle-aged (50-64 years) men. *Atherosclerosis* 143(1):153-161.

Hamalainen H, Zhou H, Chou W, Hashizume H, Heller R, Lahesmaa R. 2001. Distinct gene expression profiles of human type 1 and type 2 T helper cells. *Genome Biol* 2(7).

Hamanaka R, Kohno K, Seguchi T, Okamura K, Morimoto A, Ono M, Ogata J, Kuwano M. 1992. Induction of low density lipoprotein receptor and a transcription factor SP-1 by tumor necrosis factor in human microvascular endothelial cells. *J Biol Chem* 267(19):13160-5.

Harris RC, Mathews G. 1987. *Historical Atlas of Canada. I:From the beginning to 1800*. University of Toronto Press. Toronto.

Heart and Stroke Foundation of Canada. 2003. *The Growing Burden of Heart Disease and Stroke in Canada 2003*. Ottawa: Health Canada and the Canadian Cardiovascular Society. http://www.cvdinfobase.ca/cvdbook/CVD_Fr03.pdf. Consulted in February 2005.

Heyer E. 1995. Genetic consequences of differential demographic behaviour in the Saguenay region, Quebec. *Am J Phys Anthropol* 98(1):1-11.

Hill MR, Cookson WO. 1996. A new variant of the beta subunit of the high-affinity receptor for immunoglobulin E (Fc epsilon RI-beta E237G): associations with measures of atopy and bronchial hyper-responsiveness. *Human Molecular Genetics* 5(7):959-962.

Institut de la statistique du Québec. 2004. La situation démographique au Québec, bilan 2004.

Jeanpierre M. 1987. A rapid method for the purification of DNA from blood. *Nucleic Acids Res* 15(22):9611.

Kaufman J, Schmitt S, Barnard J, Busse W. 1992. Angiotensin-converting enzyme inhibitors in patients with bronchial responsiveness and asthma. *Chest* 101(4):922-5.

Kaufman L, Rousseeuw PJ. 1990. Finding groups in data; an introduction to cluster analysis. Wiley J. New York.

Kelly ME, Clay MA, Mistry MJ, Hsieh-Li HM, Harmony JA. 1994. Apolipoprotein E inhibition of proliferation of mitogen-activated T lymphocytes: production of interleukin 2 with reduced biological activity. *Cell Immunol* 159(2):124-39.

Leitersdorf E, Tobin EJ, Davignon J, Hobbs HH. 1990. Common low-density lipoprotein receptor mutations in the French Canadian population. *J Clin Invest* 85(4): 1014-23.

Libby P. 2002. Inflammation in atherosclerosis. *Nature* 420(6917):868-74.

Lusis AJ, Mar R, Pajukanta P. 2004. Genetics of atherosclerosis. *Annu Rev Genomics Hum Genet* 5:189-218.

Marx N, Sukhova GK, Collins T, Libby P, Plutzky J. 1999. PPARalpha activators inhibit cytokine-induced vascular cell adhesion molecule-1 expression in human endothelial cells. *Circulation* 99(24):3125-31.

Masoli M, Fabian D, Holt S, Beasley R. 2004. Global burden of asthma. Developed for the Global Initiative for Asthma. Medical Research Institute of New-Zealand, Wellington, New Zealand and University of Southampton, Southampton, United Kingdom. 122 p. <http://www.ginasthma.com/ReportItem.asp?l1=2&l2=2&intId=94>. Consulted in February 2005.

Murthy V, Julien P, Gagne C. 1996. Molecular pathobiology of the human lipoprotein lipase gene. *pharmacol. ther.* 70:101-135.

Newton CR, Graham A, Heptinstall LE, Powell SJ, Summers C, Kalsheker N, Smith JC, Markham AF. 1989. Analysis of any oint mutation in DNA. The amplification refractory mutation system (ARMS). *Nucleic Acids Res* 11(17):2503-2516.

Normand T, Bergeron J, Fernandez-Margallo T, Bharucha A, Ven Murthy MR, Julien P, Gagne C, Dionne C, De Braekeleer M, Ma R and others. 1992. Geographic distribution and genealogy of mutation 207 of the lipoprotein lipase gene in the French Canadian population of Quebec. *Hum Genet* 89(6):671-675.

Ramsay SG, Dagg KD, McKay IC, Lipworth BJ, McSharry C, Thomson NC. 1997. Investigations on the renin-angiotensin system in acute severe asthma. *Eur Respir J* 10(12):2766-71.

Rigat B, Hubert C, Alhenc-Gelas F, Cambien F, Corvol P, Soubrier F. 1990. An insertion/deletion polymorphism in the angiotensin I-converting enzyme gene accounting for half the variance of serum enzyme levels. *J Clin Invest* 86(4):1343-6.

Rosenwasser LJ, Klemm DJ, Dresback JK, Inamura H, Mascali JJ, Klinnert M, Borish L. 1995. Promoter polymorphisms in the chromosome 5 gene cluster in asthma and atopy. *Clin Exp Allergy* 25 Suppl 2:74-8; discussion 95-6.

Sandford AJ, Shirakawa T, Moffatt MF, Daniels SE, Ra C, Faux JA, Young RP, Nakamura Y, Lathrop GM, Cookson WO and others. 1993. Localisation of atopy and beta subunit of high-affinity IgE receptor (Fc epsilon RI) on chromosome 11q. *Lancet* 341(8841):332-4.

Santé Québec. 1990. Enquête sur la santé cardiovasculaire de la population québécoise. Institut de la statistique du Québec.

Scriver CR. 2001. Human genetics: Lessons from Quebec populations. *Annu Rev Genomics Hum* 2:69-101.

Seppala PF, Herd GW. 1986. Cough and wheeze caused by inhibitors of angiotensin-converting enzyme. *N Engl J Med* 314(1):61.

Smith KA. 2002. Medical immunology: a new journal for a new subspecialty. *Med Immunol* 1(1):1.

Swain SL, Weinberg AD, English M, Huston G. 1990. IL-4 directs the development of Th2-like helper effectors. *J Immunol* 145(11):3796-806.

Tiret L, de Knijff P, Menzel HJ, Ehnholm C, Nicaud V, Havekes LM. 1994. ApoE polymorphism and predisposition to coronary heart disease in youths of different European populations. The EARS Study. European Atherosclerosis Research Study. *Arterioscler Thromb* 14(10):1617-24.

Utermann G, Hees M, Steinmetz A. 1977. Polymorphism of apolipoprotein E and occurrence of dysbetalipoproteinaemia in man. *Nature* 269(5629):604-7.

Vézina H. 1996. Démographie génétique et maladies héréditaires au Québec: l'état des recherches. *Cah Que Demogr* 25(2):293-322.

Vohl MC, Lepage P, Gaudet D, Brewer CG, Betard C, Perron P, Houde G, Cellier C, Faith JM, Despres JP and others. 2000. Molecular scanning of the human PPAR α gene: association of the L162v mutation with hyperapobetalipoproteinemia. *Journal of Lipid Research* 41:945-952.

Vohl MC, Moorjani S, Roy M, Gaudet D, Torres AL, Minich A, Gagné C, Tremblay G, Lambert M, Bergeron J and others. 1997. Geographic distribution of French-Canadian low-density lipoprotein receptor gene mutations in the Province of Quebec. *Clinical genetics*. 52:1-6.

Welch JS, Ricote M, Akiyama TE, Gonzalez FJ, Glass CK. 2003. PPARgamma and PPARdelta negatively regulate specific subsets of lipopolysaccharide and IFN-gamma target genes in macrophages. *Proc Natl Acad Sci U S A* 100(11):6712-7.

WHO. 2003. Cardiovascular disease: prevention and control. World Health Organization. <http://www.who.int/dietphysicalactivity/publications/facts/cvd/en/>. Consulted in February 2005.

Wyman AR, White R. 1980. A highly polymorphic locus in human DNA. *Proc Natl Acad Sci U S A* 77(11):6754-8.

Yen CJ, Beamer BA, Negri C, Silver K, Brown KA, Yarnall DP, Burns DK, Roth J, Shuldiner AR. 1997. Molecular scanning of the human peroxisome proliferator activated receptor gamma (hPPAR-gamma) gene in diabetic Caucasians: identification of a pro12ala PPAR-gamma-2 missense mutation. *Biochem. Biophys. Res. Commun.* 241:270-274.

Zilversmit DB. 1973. A proposal linking atherogenesis to the interaction of endothelial lipoprotein lipase with triglyceride-rich lipoproteins. *Circ Res* 33(6):633-8.

3.11 Figure titles and legends

Figure 3: Administrative regions of the province of Quebec. This figure shows the administrative regions of the province of Quebec according to divisions of 1990, year in which our samples were collected and classified. For each region we can see the size of the cohort in number of alleles.

Legend: 01: n*=342; 02: n=616; 03: n=340; 04: n=258; 05: n=204, 06: n=248; 07: n=144; 08: n=158; 09: n=112; 10: n=70; 11: n=104; 12: n=112; 13: n=156; 14: n=492; 15: n=130.

*n: number of alleles.

Figure 4: Groups of the Quebec regions according to the allelic frequencies of SNPs involved in a) asthma and b) cardiovascular diseases and clusters of SNPs involved in c) asthma and d) cardiovascular diseases according to their distribution into the regions of Quebec. The three groups of regions obtained by DHCA are not composed by the same regions for asthma (4a) and cardiovascular (4b) diseases. The presence of most of the regions in each cluster cannot be explained by historical or geographical information. The presence of Mauricie, Centre-du-Québec and Québec in the first cluster, that of Lanaudière and Laurentides in the second one and that of Bas-Saint-Laurent and Chaudière-Appalaches and also Saguenay–Lac-Saint-Jean and Abitibi-Témiscamingue in the third one can be explained by geographic proximity. This figure also illustrates one cluster of variants for asthma (4c) and cardiovascular diseases (4d) that have a similar distribution pattern. A part of the cluster of variants for asthma diseases (I/D-ACE and P12A-PPARG2) is found again in the cardiovascular one. All groups of regions and clusters of SNPs are separated by a p-value under 0.05. The grouping index indicates the order of division of all groups of regions or variants from the higher number to zero. The small size of the cohort and, for some SNPs (D9N-LPL, N291S-LPL and E237G-MS4A2), the small allelic frequencies, induce a decrease of power for this analysis.

Table 6: References and allelic frequencies for all SNPs tested

Genes	Variants	Allelic frequency (%)
IL4 ¹	C-589T ²	11.4
MS4A2 ^{3,4}	E237G ⁵	2.6
ACE ^{6,7}	I/D ⁸	55.2
PPARG2 ⁹	P12A ¹⁰	12.2
PPARA ¹¹	L162V ¹²	7.9
LPL ^{13,14}	D9N ¹⁵	2.3
	G188E ¹⁵	0*
	P207L ¹⁵	0.1*
APOE ¹⁶	N291S ¹⁵	1.3 [†]
	E2/E3/E4 ¹⁶	9.2/75.7/15.1
LDLR ^{17,18}	W66G ¹⁹	0*

*Those variants were eliminated because of their allelic frequencies lower then 2%.

[†]This variant was conserved because in some regions its allelic frequency reaches 2%.

Legend: 1: [Smith 2002]; 2:[Rosenwasser, et al. 1995]; 3: [Cookson, et al. 1992]; 4: [Sandford, et al. 1993]; 5: [Hill and Cookson 1996]; 6: [Diet, et al. 1996]; 7: [Ramsay, et al. 1997]; 8: [Rigat, et al. 1990]; 9: [Clark 2002]; 10: [Yen, et al. 1997]; 11: [Berger and Moller 2002]; 12: [Vohl, et al. 2000]; 13: [Zilversmit 1973]; 14: [Chait, et al. 1982]; 15: [Murthy, et al. 1996]; 16: [Utermann, et al. 1977]; 17: [Brown and Goldstein 1986]; 18: [Gaudet, et al. 1999]; 19: [Leitersdorf, et al. 1990].

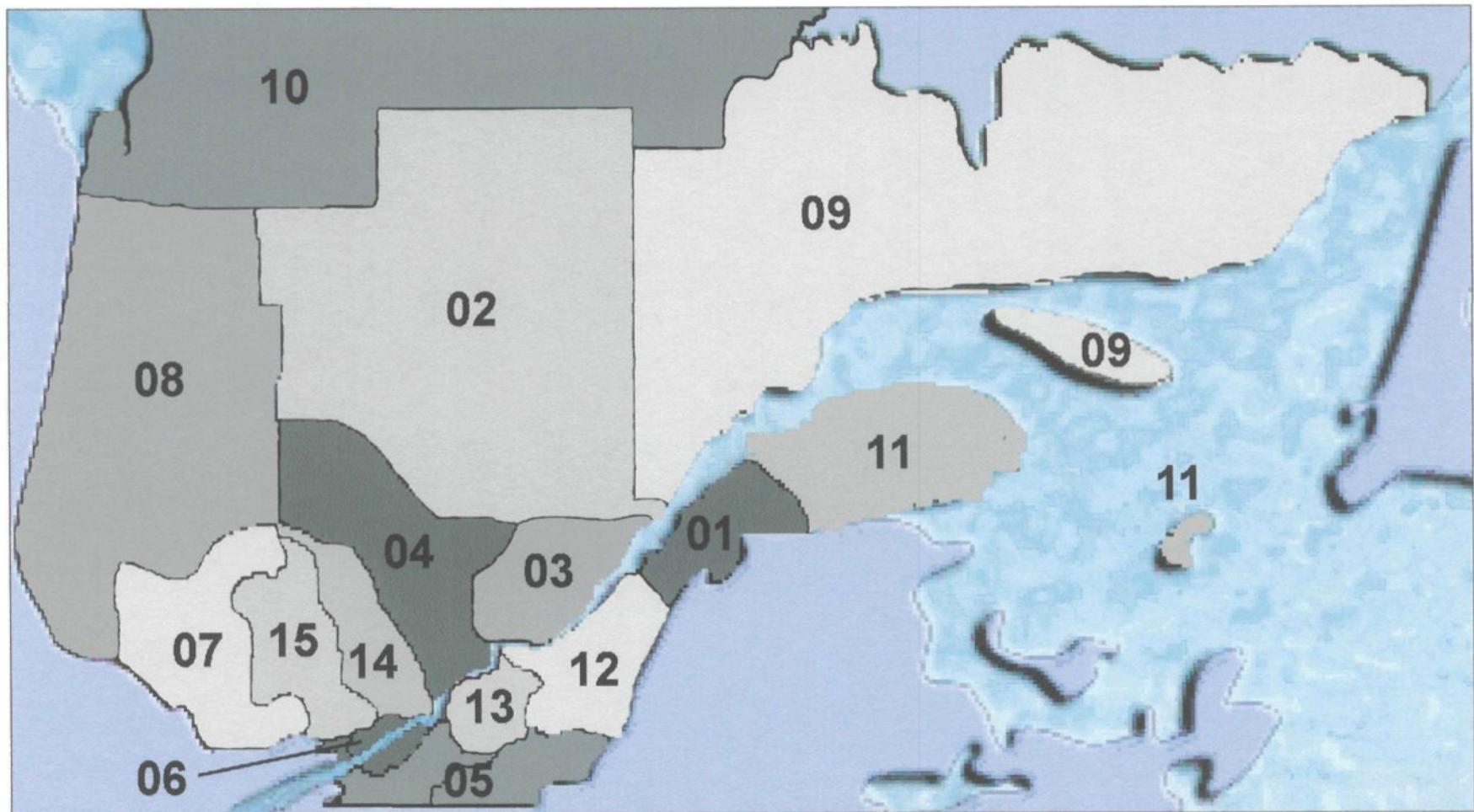


Figure 3: Administrative regions of the province of Quebec

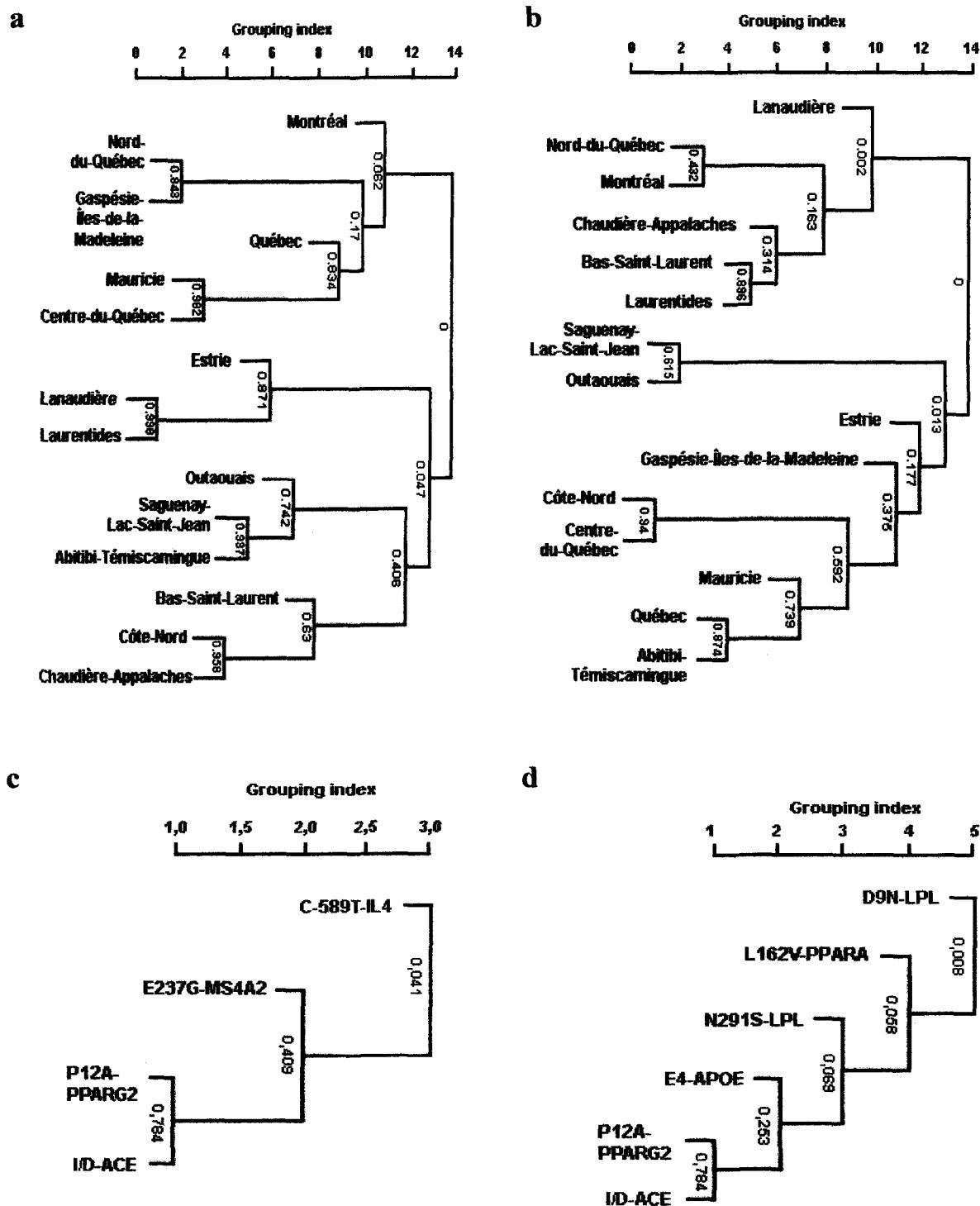


Figure 4: Groups of the Quebec regions according to the allelic frequencies of SNPs involved in a) asthma and b) cardiovascular diseases and clusters of SNPs involved in c) asthma and d) cardiovascular diseases according to their distribution into the regions of Quebec

CHAPITRE IV

Résultats et discussion

4.1 Modèle de division géographique

Dans le cadre de ce projet pilote, les régions administratives (selon les divisions de 1990) ont été utilisées pour séparer le territoire québécois. En effet, les échantillons utilisés pour l'expérimentation, qui proviennent de l'enquête de santé cardiovasculaire de 1990, avaient été préalablement répartis selon les régions administratives de l'époque (Santé cardiovasculaire, 1990). De plus, le gouvernement québécois, lorsqu'il décrit le risque inhérent à des facteurs environnementaux pour des maladies données, fonctionne selon les 17 régions administratives (en 1990 elles étaient au nombre de 15, la Montérégie et l'Estrie ainsi que Montréal et Laval étaient regroupées) ou selon les 18 régions sociosanitaires (la 18^{ème} représentant les Terres-Cries-de-la-Baie-James). Ces deux types de division du territoire québécois sont similaires. Donc, si le but est de pouvoir regrouper les données génétiques avec les données de type environnementales, il est plus facile de le faire si les résultats obtenus sont divisés de manière à ce qu'ils soient comparables avec les données environnementales déjà disponibles.

Toutefois, on remarque que les divisions des régions administratives du Québec ne reflètent pas nécessairement les mouvements historiques de peuplement du territoire (voir chapitre 1, section *organisation sur le territoire*). Étant donné que les mouvements de population importants laissent présager une ressemblance entre certains secteurs de régions différentes, un regroupement basé sur celles-ci aurait peut-être permis de mieux différencier des groupes de régions par rapport aux autres. Par exemple, il aurait peut-être fallu regrouper le secteur de Charlevoix avec la région du Saguenay–Lac-Saint-Jean et non avec

la région urbaine de Québec puisque l'histoire indique de façon précise que ce sont des individus de Charlevoix et non de la région de Québec qui ont en grande majorité peuplé le Saguenay–Lac-Saint-Jean à ses débuts (Bouchard et De Braekeleer, 1992). Par contre, toujours en référence avec l'idée d'intégration des données génétiques aux données environnementales, cela augmenterait probablement les difficultés lors de la caractérisation de facteurs de risques intégrant ces deux aspects.

4.2 Fréquences alléliques

Pour estimer l'importance d'un facteur de risque génétique, on peut se baser sur la prévalence d'un ou de plusieurs allèle(s) de susceptibilité pour une maladie donnée dans une population cible. Dans ce projet de maîtrise, nous avons donc calculé la fréquence allélique de plusieurs polymorphismes documentés pour leur rôle dans le développement de l'asthme ou des maladies cardiovasculaires dans les régions administratives du Québec.

4.2.1 *Choix des gènes*

Le choix des gènes reposait sur une association documentée dans au moins un des deux types de maladie à l'étude en tant que facteur augmentant la probabilité de développement ou la sévérité. Pour ce qui est des variants sélectionnés dans ces gènes, les critères de sélection étaient une fréquence allélique d'au moins 2,5% documentée dans une population contrôle ou générale, donc formée d'individus qui ne sont atteints d'aucune maladie particulière (voir tableau 5). Par contre, certains polymorphismes, pour lesquels aucune fréquence allélique obtenue d'une population contrôle n'était disponible, ont tout de même été choisis pour leur association connue dans l'asthme ou dans les maladies cardiovasculaires (voir les variants IL4C-589T et LDLR-W66G dans le tableau 5). Pour ces polymorphismes il est impossible de prévoir quelle sera leur fréquence allélique dans la population générale du Québec. En effet, la fréquence allélique sera forcément plus basse dans la population générale que dans la cohorte d'individus atteints déjà documentée. Par la suite, certains autres polymorphismes ayant une fréquence allélique plus basse ont été ajoutés parce que les gènes en question avaient une association pertinente avec les maladies

ciblées (MS4A2-E237G , LPL-D9N et LPL-G188E) ou parce que des études de distribution géographique effectuées dans la province de Québec avaient déjà proposé la présence d'une variabilité génétique interrégionale dans leur distribution et en ce sens servaient de contrôle (LPL-G188E, LPL-P207L et LDLR-W66G) (Zilversmit, 1973; Steinberg, 1983; Plaut et al., 1989; Bergeron et al., 1992; Cookson et al., 1992; Normand et al., 1992; Davignon et Roy, 1993; Dionne et al., 1993; Sandford et al., 1993; Vohl et al., 1997).

4.2.2 Méthode de génotypage

Afin de pouvoir calculer leur fréquence allélique, les variants ciblés par cette étude ont été génotypés soit par la méthode ARMS, soit par la méthode RFLP (Wyman et White, 1980; Newton et al., 1989). Pour certains échantillons, l'utilisation de ces techniques n'a pas donné de résultats concluants. Dans le cas de notre projet, il y a eu de 50 (PPARG2-P12A) à 283 (MS4A2-E237G) échantillons sur 1680 d'éliminés faute de résultats tout dépendant des variants. Il se peut que ce problème découle d'erreurs de manipulation, c'est pourquoi les échantillons ne donnant pas de résultats ont été génotypés une seconde fois. Ceci peut également être imputable à la durée et aux conditions de conservation des échantillons. Les sanguins totaux ont été congelés pendant 10 ans avant l'extraction, ce qui a pu causer la formation de cristaux qui auraient pu endommager l'ADN. La technique d'extraction d'ADN utilisée lors de l'étude québécoise sur la santé cardiovasculaire en 1990 (*guanidine hydrochloride-proteinase K method* (Jeanpierre, 1987)) peut également être responsable de la perte de certains échantillons d'ADN.

4.3 Équilibre de Hardy-Weinberg

Le test utilisé pour vérifier si la répartition des proportions d'homozygotes mutés, d'hétérozygotes et d'homozygotes normaux est équilibrée est le test d'EHW (Hardy, 1908; Weinberg, 1908). L'EHW repose sur le principe de panmixie et considère chaque fécondation comme deux tirages aléatoires. Ainsi, pour des fréquences alléliques p et q dans une population donnée on observe les fréquences génotypiques p^2 , $2pq$ et q^2 (Hartl et

Clark, 1997). Sachant cela, il est possible de comparer à l'aide d'un chi-deux les fréquences génotypiques observées lors de l'expérimentation avec les fréquences génotypiques attendues calculées grâce aux fréquences alléliques retrouvées dans les populations ciblées. Suite à ce calcul, il a été observé que certaines des 15 régions du Québec ne sont pas en EHW pour tous les polymorphismes étudiés. Dans 11 cas (polymorphisme/région) sur 165 (donc dans 6,7% des cas), l'EHW n'est pas respecté (voir annexe A).

L'EHW devrait s'observer pour tous les variants si l'échantillon était vraiment représentatif de la population générale. Cet écart de l'EHW pour ces 11 cas s'explique en partie par le niveau du test utilisé qui était de 5% mais également par la petite taille de la cohorte pour certaines régions, amenant une mauvaise représentation de la population de celles-ci. Ainsi, comme l'énonce J.L. Serre dans son livre sur la génétique des populations, « la fréquence d'un événement [ici d'un génotype] est égale à sa probabilité si le nombre de tirages est très grand » (Serre, 1997). L'échantillon le plus petit est celui représentant le Nord-du-Québec qui ne comporte que 35 individus (70 allèles) qui, de plus, ne sont pas répartis selon la densité de population des différentes ethnies. En effet, aucune communauté autochtone n'est représentée dans les échantillons des différentes régions du Québec alors que l'on sait, par exemple, que les autochtones représentent 48,9% de la population du Nord-du-Québec (Institut de la statistique du Québec, 1996). Observer un déséquilibre dans le test de Hardy-Weinberg dans une population générale peut aussi s'expliquer par la présence d'erreurs de génotypage ou d'erreurs dans la compilation des résultats. Dans le cas du génotypage, pour éviter le plus possible les erreurs lors des manipulations, chaque résultat incertain était soit génotypé une seconde fois, soit éliminé. De plus, pour ce qui est de la prise de donnée, la table de compilation des résultats a été vérifiée entièrement révélant un taux d'erreur lors de l'entrée des données d'environ 2% (incluant le rejet de nouveaux résultats considérés comme trop incertains ainsi que les erreurs lors de l'entrée des génotypes dans la table de données observées), ce qui est considéré comme plus qu'acceptable. Malgré un déséquilibre de l'écart de l'EHW pour certaines régions, elles ont toutes été conservées pour les étapes subséquentes visant à les regrouper ou à regrouper les SNPs sur la base de la variabilité génétique interrégionale grâce à l'analyse de classification hiérarchique.

4.4 Analyse de classification descendante hiérarchique

Dans cette étude les données à analyser sont sous forme de fréquences alléliques et celles-ci sont utilisées pour vérifier les possibilités de regroupements entre les régions ou entre les SNPs. En tenant compte de ces informations, le test le plus adéquat à utiliser est le test du chi-deux, qui est adapté à la comparaison de fréquences (Scherer, 1984). L'utilisation uniquement de ce test pour ces données engendre une comparaison de dépendance entre les variants et les régions plutôt que de souligner une similitude ou une différence possible entre les régions par rapport à leur pool génique respectif. C'est pourquoi l'analyse de classification hiérarchique a été ciblée comme test statistique. Cette analyse peut être utilisée avec une table de contingence renfermant les fréquences observées pour chaque variant pour chacune des régions administratives et peut utiliser un test de chi-deux pour les comparaisons de distances entre les éléments. L'emploi de la table de contingence dans la classification hiérarchique permet de vérifier les similitudes existant entre les régions en fonction de leur profil génétique, c'est-à-dire en fonction de tous les variants à l'étude, ou bien de comparer les profils de distribution des variants à travers les régions du Québec. Parmi les analyses de classification hiérarchique, on compte l'analyse ascendante, qui considère au départ chaque élément séparément pour les regrouper un à un jusqu'à l'obtention d'une seule classe, et l'analyse descendante, qui considère au départ tous les élément comme similaires, pour ensuite procéder à des subdivisions jusqu'à ce que chaque élément forme sa propre classe.

Dans cette étude, l'analyse de classification descendante hiérarchique a été choisie principalement pour son hypothèse de départ puisque l'on considère toutes les régions du Québec comme similaires pour ensuite vérifier si certaines d'entre elles se diffèrentient des autres et si des regroupements pourraient être formés. De plus, étant donné le type de résultats devant être comparé, cette analyse a l'avantage de conférer une plus grande puissance statistique que l'analyse de classification ascendante. Pour ce projet, lors du processus de classification, à chacune des étapes de division une distance du chi-deux a été réévaluée. Ceci alourdit de beaucoup l'algorithme mathématique utilisé mais la petite quantité d'éléments à considérer pour ce projet (15 régions) a permis ce procédé. En plus

d'apporter une précision supplémentaire, l'obtention d'un chi-deux réévalué pour chacun des nouveaux groupes formés lors des divisions permet l'obtention d'une valeur p. Celle-ci permet de mieux interpréter le dendrogramme qui résulte de la classification hiérarchique. Parmi les fréquences alléliques utilisées pour classer les régions ou bien les variants en classes distinctes, certaines d'entre elles sont de petites valeurs (<5%). Ces petites fréquences ont comme effet de diminuer la puissance du test, ce qui occasionne certaines erreurs de regroupement. C'est-à-dire que parfois certaines régions sont considérées similaires tout simplement parce que le test n'est pas apte à voir leurs différences par manque de puissance.

L'analyse de classification descendante hiérarchique a été utilisée premièrement afin de regrouper les régions administratives du Québec selon la variabilité génétique pour certains polymorphismes prédéterminés. Cet exercice a été tenté en considérant les variants impliqués dans l'asthme et les maladies cardiovasculaires de façon séparée. Une telle classification pourrait mener à la détection de zones à risque pour ces maladies et/ou à la caractérisation génotypique propre à chaque région afin de faciliter l'élaboration des outils de prévention, de sensibilisation et de traitement dans chaque région selon leurs besoins propres.

Elle a également servi à regrouper ces mêmes variants sur la base de leur patron de distribution interrégional. Encore une fois, ces polymorphismes ont été traités de façon séparée selon leur(s) association(s) respective(s). La connaissance du patron de distribution des variants à risque élevé est une autre méthode de détection des zones à risque. Le regroupement des variants ayant des patrons de distribution similaires apporte également des informations pertinentes lors du processus de création d'outils de prévention et de traitement pour les maladies inflammatoires.

Tel que mentionné dans la section discussion du chapitre 3, cette analyse, autant pour le regroupement des régions que pour celui des SNPs, comporte quelques limitations, soit la possibilité; a) d'un biais de sélection lors du recrutement des individus et également d'un biais de sélection lors du choix des quelques gènes testés; b) d'un biais de régression-

dilution du fait que cette analyse ne permet pas un suivi des données dans le temps alors que la population est en constant mouvement, amenant des changements plus ou moins important du pool génique de chaque région; c) d'un biais de stratification de la population causé par la présence non détectée de sous-populations dans l'échantillon ou bien causé par la façon dont l'échantillon est divisé et; d) d'un dernier biais du aux multiples tests effectués entre les 15 régions administratives et les 8 variants testés.

4.4.1 Regroupement des régions selon les variants impliqués dans l'asthme et les maladies cardiovasculaires

Une analyse de classification descendante hiérarchique a été effectuée pour regrouper les régions sur la base de leur composition allélique pour des variants impliqués dans le processus inflammatoire de l'asthme, soit les variants IL4-C-589T, MS4A2-E237G, ACE-I/D et PPARG2-P12A, et pour des variants impliqués dans les maladies cardiovasculaires, soit ACE-I/D, PPARG2-P12A, PPARA-L162V, LPL-D9N, LPL-N291S et APOE-E4 (Utermann et al., 1977; Rigat et al., 1990; Rosenwasser et al., 1995; Hill et Cookson, 1996; Murthy et al., 1996; Yen et al., 1997; Vohl et al., 2000). Pour les maladies cardiovasculaires, 3 polymorphismes, soit LPL-G188E, LPL-P207L et LDLR-W66G, n'ont pas été pris en compte puisqu'ils n'ont pas été observés avec une fréquence allélique assez élevée dans la cohorte génotypée (voir tableau 5) (Leitersdorf et al., 1990; Murthy et al., 1996). De plus, le variant APOE-E2 n'a pas été sélectionné comme variant à risque puisque dans certaines études il semble qu'il puisse être protecteur (Utermann et al., 1977; Xhignesse et al., 1991). Si on considère les résultats obtenus (figures 4a et 4b), on observe trois regroupements de régions pour chacun des deux types de maladies ainsi qu'une région non regroupée dans chacun d'eux. Les regroupements ne sont pas formés des mêmes régions pour les deux types de maladies. La formation de tels groupes souligne la présence d'une hétérogénéité génétique interrégionale dans la population québécoise.

Chacun des regroupements a été comparé avec les connaissances acquises sur le peuplement du Québec et les mouvements de population. Un parallèle a également été fait avec les positions géographiques des régions rassemblées dans une même classe. On

remarque que chacun des regroupements qui ont été formés, autant pour l'asthme que pour les maladies cardiovasculaires, ne sont explicables qu'en partie par les connaissances démographiques et historiques du peuplement québécois. Par contre, on peut valider chacun de ces regroupements à l'aide des données de fréquences alléliques (voir annexe B), toutes les régions d'une même classe ayant des fréquences comparables pour la majorité des variants. La présence de différences dans les fréquences alléliques des variants MS4A2-E237G, LPL-D9N et LPL-N291S n'influence pas beaucoup les calculs de l'analyse de classification étant donné que ceux-ci ont des fréquences en dessous de 5%. Un choix de variants avec une fréquence allélique plus importante dans la population générale serait donc peut-être préférable pour tenter de regrouper ensemble les régions génétiquement similaires.

La présence de regroupements qui ne peuvent être validés par les données historiques et géographiques, peut être expliquée entre autres par la survenue de mouvements de populations à plus petite échelle entre les régions qui auraient pu advenir soit au début du peuplement ou bien plus récemment, quelques générations étant suffisantes pour modifier les fréquences alléliques d'une région. Une autre explication est la quantité d'échantillons disponibles pour cette étude. En effet, la cohorte comprend 1 680 échantillons, ce qui équivaut à 3 360 allèles. Par contre, une fois divisée dans les 15 régions administratives du Québec, on obtient des cohortes allant de seulement 70 allèles pour le Nord-du-Québec, à 616 allèles pour le Saguenay–Lac-Saint-Jean. Une trop petite cohorte peut induire de mauvais regroupements dus à son manque de représentativité de la population régionale en question. De petits effectifs diminuent la puissance du test, augmentant ainsi les chances que certaines différences entre régions ne soient pas reconnues.

La petite quantité des polymorphismes ciblés a pu également provoquer des erreurs de regroupement. En effet, après élimination des variants ayant de trop basses fréquences alléliques (LPL-G188E, LPL-P207L et LDLR-W66G), seulement 6 SNPs représentaient les maladies cardiovasculaires et 4 SNPs représentaient l'asthme. Une plus grande quantité de variants ciblés pour leur association aux deux types de maladies sélectionnés aurait dressé

un portrait plus juste de chacune des régions et aurait possiblement permis de rassembler les régions administratives québécoises de façon à être plus en accord avec nos connaissances historiques et démographiques. Effectivement, le fait que les variants ciblés aient été peu nombreux a peut-être entraîné un manque de représentativité de la population générale de chaque région et fait ressortir des évènements mineurs des mouvements interrégionaux. Une plus grande quantité de SNPs aurait peut-être mieux fait ressortir les phénomènes de génétique des populations comme la dérive génétique, les effets fondateurs ou la formation d'isolats et de dèmes qui ont peut-être fait partie intégrante du peuplement de plusieurs régions.

Dans chacune des classes formées pour les deux types de maladies à l'étude, on observe par contre certaines régions dont le regroupement peut être expliqué historiquement ou géographiquement. Dans le cas de l'asthme, les régions de la Mauricie, de Québec et du Centre-du-Québec dans le premier groupe, dans le deuxième groupe celles de Lanaudière et des Laurentides et dans le troisième groupe celles du Saguenay–Lac-Saint-Jean et de l'Abitibi-Témiscamingue ainsi que celles du Bas-Saint-Laurent et de Chaudière-Appalaches s'expliquent toutes par une proximité géographique (figure 4a). Une telle proximité permet un échange entre les populations et favorise une certaine homogénéisation sur le plan génétique en diminuant entre autres l'effet de certains phénomènes tels la dérive génétique (Laberge, 1990). De plus, les similitudes entre les régions de Lanaudière et des Laurentides peuvent également être expliquées historiquement. En effet, il est admis que lors de l'occupation du territoire par les immigrants, les Montréalais se sont beaucoup établis dans les plaines environnantes, entre autres vers les régions des Laurentides et de Lanaudière (Bouchard et Tremblay, 1995). Finalement, l'exclusion de la région de Montréal par rapport aux 3 groupes formés s'explique démographiquement. En effet, Montréal n'a pas été intégrée dans une des classes malgré une valeur p de 0,062 en raison de la forte multiethnicité caractéristique de cette région (Institut de la statistique du Québec, 1996). En effet, les recherches de l'équipe de Cao et al. en 2001 concernant la distribution de 8 525 SNPs chez 84 individus provenant de quatre populations américaines différentes ainsi que l'étude de l'équipe de Shriver et al. sur la diversité des loci HLA-A, B et C dans 5 populations américaines représentées par

1 296 individus, comprenant des autochtones, ont toutes deux démontré la présence d'une différence concernant la diversité génétique entre les différentes ethnies étudiées (Cao et al., 2001; Shriver et al., 2004).

Dans le cas des maladies cardiovasculaires on retrouve dans le premier groupe la présence de la région de Chaudière-Appalaches associée à celle du Bas-Saint-Laurent et dans le troisième groupe celles de l'Abitibi-Témiscamingue et de la Mauricie qui peut être expliquée par une proximité géographique favorisant un échange entre les populations (figure 4b). Toujours dans le troisième groupe, la similarité entre les pools géniques des régions de la Gaspésie-Îles-de-la-Madeleine et de la Côte-Nord s'explique historiquement puisque la Côte-Nord a été en bonne partie colonisée par des individus provenant des régions de Charlevoix et de la Gaspésie-Îles-de-la-Madeleine (Bouchard et De Braekeleer, 1992).

Puisque la classification des régions en fonction de leur pool génique respectif a donné des regroupements qui n'ont pu être que partiellement validés par les connaissances acquises concernant le peuplement du Québec et les principaux mouvements de population ainsi que par la proximité géographique des régions d'une même classe, le regroupement des polymorphismes impliqués dans chacun des deux types de maladies en fonction de leur patron de distribution interrégional a également été réalisé. Cet autre type de classification apporte également des informations sur la caractérisation du pool génique québécois et sur l'identification de zones à risque.

4.4.3 Regroupement des variants impliqués dans l'asthme et les maladies cardiovasculaires selon leur patron de distribution dans les régions du Québec

Après avoir tenté de regrouper les régions sur la base d'une variabilité génétique interrégionale, le même exercice a été fait pour regrouper les variants entre eux sur la base de cette même variabilité. L'exercice de regroupement des variants impliqués dans l'asthme, ainsi que celui des variants impliqués dans les maladies cardiovasculaires, témoigne de la possibilité que certains de ces polymorphismes puissent avoir un patron de

distribution très similaire à travers les régions. En effet, un regroupement a été obtenu pour chaque type de maladie.

Pour effectuer les analyses de classification descendantes hiérarchiques afin de classer les variants, la même table de contingence que celle utilisée pour classer les régions en fonction de leur pool génique respectif a été employée. Donc, pour chacun des polymorphismes ciblés, l'échantillon total comporte 3 360 allèles divisés à travers les 15 régions administratives du Québec formant ainsi des cohortes contenant de 70 à 616 allèles selon les régions. Étant donné que les cohortes régionales sont petites, il se peut que celles-ci ne soient pas réellement représentatives de chacun des pools géniques régionaux. De plus, comme stipulé dans la section précédente, l'emploi d'une petite cohorte diminue la puissance statistique du test et peut empêcher que certaines différences entre régions soient reconnues par le test utilisé. Malgré la petite taille de la cohorte, la présence de presque tous les variants d'un même groupe, si on se fie aux fréquences alléliques de l'annexe B, peut être validée par les fréquences observées dans les cohortes régionales. Par contre, pour 3 de ces SNPs (MS4A2-E237G, LPL-D9N et LPL-N291S) les fréquences alléliques sont très petites, soit en dessous de 5%, ce qui pourrait biaiser les regroupements. Pour obtenir des résultats de regroupement représentatifs des données de fréquences observées, il faut que, pour un même polymorphisme, les fréquences alléliques soient pour la grande majorité plus élevées que 5%.

Dans le regroupement formé pour les variants impliqués dans l'asthme, 2 des 3 variants semblent avoir une distribution similaire à travers les régions du Québec (ACE-I/D et PPARG2-P12A), le troisième ayant des fréquences alléliques trop petites pour pouvoir valider son regroupement avec les deux autres SNPs (voir figure 4c). En effet, ceux-ci possèdent des fréquences alléliques constantes ou similaires d'une région à l'autre. Le variant IL4-C-589T, par contre, témoigne d'une baisse marquée des fréquences alléliques observées tout particulièrement dans les régions du Nord-du-Québec et de la Gaspésie–Îles-de-la-Madeleine. Son exclusion du groupe est donc justifiée par cette distribution non uniforme à travers les régions du Québec par rapport à celle des trois autres polymorphismes.

Un regroupement de 4 des SNPs sélectionnés pour leur rôle dans les maladies cardiovasculaires soit ACE-I/D, PPARG2-P12A, APOE-E4 et PPARA-L162V a été formé, le variant LPL-N291S ayant des fréquences alléliques trop petites pour être pris en compte (voir figure 4d). Dans ce regroupement, les variants ACE-I/D et PPARG2-P12A semblent, comme dans le cas de la figure 4c, avoir un patron de distribution géographique très similaire entre les régions du Québec. L'isoforme E4 du gène APOE a un patron de distribution interrégional un peu moins constant. En effet, il y a entre autres deux fréquences alléliques plus basses pour les régions de la Gaspésie–Îles-de-la-Madeleine et Chaudière-Appalaches (voir annexe B). Le dernier variant faisant partie de ce groupe, PPARA-L162V, a une valeur *p* près du seuil de 0,05 avec 0,058, ce qui témoigne d'une distribution moins constante d'une région à l'autre en opposition aux polymorphismes ACE-I/D et PPARG2-P12A. La petite valeur *p* de ce SNP est expliquée par des fréquences alléliques beaucoup plus basses dans les régions du Bas-Saint-Laurent, du Nord-du-Québec, de la Gaspésie–Îles-de-la-Madeleine et des Laurentides (voir annexe B). Finalement, pour le variant LPL-D9N, en plus d'avoir des fréquences alléliques en-dessous de 5%, on observe pour les régions de l'Estrie et de l'Outaouais des fréquences relativement élevées alors que pour tous les autres polymorphismes les fréquences alléliques dans ces régions sont toujours constantes. Même type d'observation pour les régions de la Côte-Nord et du Centre-du-Québec où cette fois-ci les fréquences du variant LPL-D9N sont très basses (voir annexe B). Il est donc logique qu'il n'ait pas été regroupé avec les autres polymorphismes.

On remarque grâce aux figures 4c, 4d et à l'annexe B que certains variants ont un patron de distribution similaire et que les regroupements formés lors de cette analyse comprennent les polymorphismes ayant les distributions interrégionales les plus constantes. Parmi les variants ciblés, aucun SNP ayant un patron de distribution inégal n'a été groupé avec un autre SNP. Les SNPs provoquant une variabilité interrégionale quant à l'augmentation ou la diminution d'un risque relatif associé aux deux types de maladie à l'étude seraient donc les SNPs n'ayant pas été groupés avec les autres. Les polymorphismes utilisés lors des analyses de classification descendantes hiérarchiques dans ce projet sont tous situés sur des chromosomes différents. Aucun biais dû à des déséquilibres de liaison

ou bien à la présence de blocs haplotypiques parmi les variants n'interfère donc dans les résultats de regroupement. On peut en conséquence poser l'hypothèse que les regroupements formés pourraient illustrer, par exemple, des migrations et des mouvements de populations plus ou moins importants qui seraient survenus entre les régions.

CONCLUSION

Dans la littérature on retrouve des informations sur la répartition géographique dans la province de Québec de certains variants ciblés pour leur rôle dans les maladies cardiovasculaires tels les variants LPL-P207L, LPL-G188E ainsi que certains variants du gène LDLR dont LDLR-W66G (Bergeron et al., 1992; Normand et al., 1992; Davignon et Roy, 1993; Dionne et al., 1993; Vohl et al., 1997). Cette étude a d'original l'utilisation d'une cohorte aléatoire représentant la population générale du Québec, plutôt que d'une cohorte d'individus atteints d'une maladie ciblée, afin de caractériser la variabilité génétique interrégionale du Québec. Dans le cadre de cette maîtrise, les polymorphismes impliqués dans le développement de traits complexes fréquents, soit l'asthme et les maladies cardiovasculaires, ont été ciblés dans le but de décrire leur répartition géographique mais également et surtout dans le but de chercher des méthodes pour essayer de regrouper ces informations de façon utilisable, ce qui n'avait jamais été tenté auparavant.

Les résultats obtenus suite à cette étude ont démontré premièrement qu'il existe une variabilité génétique interrégionale au Québec pour les traits complexes fréquents tels les maladies inflammatoires. Grâce à cette variabilité génétique, il a été possible de regrouper les régions ainsi que les variants sur la base du patron de distribution de chaque polymorphisme. Par contre, la comparaison des regroupements de régions avec les données historiques et démographiques a démontré qu'il était souvent impossible de valider nos résultats de cette façon. Ceci a été expliqué entre autres par la petite taille de la cohorte une fois celle-ci répartie dans chaque région et par la méthode de sélection des variants ainsi que la quantité de variants sélectionnée. De plus ces résultats soulignent une adéquation qui n'est pas toujours très bonne entre les régions historiques et les régions administratives contemporaines.

Cette étude a également souligné que l'analyse de classification hiérarchique s'est avérée un moyen intéressant dans le processus de modélisation des informations génétiques. La classification descendante hiérarchique semble être un outil adéquat à utiliser pour ce type de modélisation puisqu'elle suppose comme hypothèse de départ que toutes les régions ou tous les variants sont similaires, pour ensuite tenter de les subdiviser en groupes et qu'elle indique l'ordre dans lequel les éléments se regroupent. Toutefois, cette analyse étant une analyse de type exploratoire, elle doit être considérée avec réserve car elle ne donne qu'un indice de la façon dont on peut faciliter la caractérisation de nos données. De plus, cette analyse ne permet pas de faire un suivi dans le temps et donne les regroupements possibles de régions ou de variants seulement pour l'année pour laquelle les données ont été recueillies.

Finalement, les groupements effectués sur la base de l'analyse de classification hiérarchique ne peuvent pas tous être validés mais cette étude à tout de même démontré qu'il existe une variabilité dans les fréquences alléliques pour les gènes candidats à l'asthme et aux maladies cardiovasculaires au Québec. Cet élément souligne une avenue intéressante dans la gestion du système de santé publique. En effet, il serait pertinent et même important de tenir compte de cette variabilité génétique dans la caractérisation des facteurs de risque propre à chaque région administrative dans le processus de création d'outils pour la prévention, la sensibilisation et le traitement des maladies inflammatoires fréquentes au Québec.

PERSPECTIVES

Ce projet a fait ressortir plusieurs éléments qui seraient à améliorer pour de futures analyses de la caractérisation de la variabilité génétique dans la population générale québécoise. L'observation de cette variabilité et des possibilités de regroupement des régions et des variants laisse également supposer plusieurs alternatives quant aux avenues futures à explorer.

Parmi les aspects à améliorer dans d'éventuelles études, il faudrait premièrement disposer d'une cohorte plus importante et réfléchir à la quantité voulue d'échantillons par région. L'objectif visé par le projet CART@GENE est de 50 000 échantillons répartis dans toutes les régions du Québec. Est-ce qu'il faut une cohorte dont la quantité d'échantillons est proportionnelle à la densité de population de chaque région administrative ou bien une quantité égale d'échantillons peu importe la densité de population? Les informations concernant l'origine ethnique devraient également être disponibles pour s'assurer d'une bonne représentativité de la population québécoise qui est de plus en plus multiethnique. De plus, il serait pertinent d'obtenir les informations phénotypiques concernant chaque individu afin de pouvoir procéder éventuellement à des études d'association et également afin de recouper ces deux types d'information.

Le deuxième aspect important à améliorer est la quantité de polymorphismes ciblés pour chacune des maladies sélectionnées afin d'avoir une meilleure idée des régions qui présentent un facteur de risque génétique similaire. Ce projet pilote a bien démontré qu'une certaine proportion des variants choisis est éliminée faute de fréquences alléliques supérieures à 2,5% dans une population générale et également qu'en fonction de l'analyse utilisée, une certaine fréquence allélique minimale est requise (environ 5%). De plus, il n'est pas toujours possible d'avoir des données théoriques concernant tous les variants dans

une population générale ou contrôle. Il faut donc avoir sélectionné une grande quantité de polymorphismes au commencement de l'étude. Un autre aspect important serait de pouvoir leur associer un risque relatif en rapport avec leur impact sur la prévalence et le développement des maladies ciblées.

Donc, l'utilisation d'une plus grande cohorte mieux définie et d'une plus grande quantité de polymorphismes auxquels serait associé un risque relatif connu permettrait d'atteindre les objectifs ciblés. Parmi les avenues possibles, les données recueillies pour cette étude pourraient servir à effectuer des tests de puissance afin de déterminer la quantité requise d'échantillons ainsi que la quantité de variants qui devrait être utilisée pour l'étude de la variabilité génétique interrégionale. Pour une utilisation optimale des informations génétiques en santé il faudrait avoir une meilleure connaissance des allèles à risque ainsi que de leur distribution et des facteurs de risque environnementaux. Pour y arriver, les étapes subséquentes à cette étude pourraient être premièrement, à l'aide de modèles statistiques pouvant par exemple découler d'analyses de classification hiérarchique, de calculer un risque relatif total pour chaque région en se basant sur les risques relatifs associés à chacun des variants sélectionnés et définir quels sont les polymorphismes les plus importants dans chaque région pour chaque type de maladie. Cet exercice permettrait par la suite de cerner les zones à risque et de les caractériser. De plus, chaque facteur de risque génétique obtenu pour chaque région pourrait être intégré dans un facteur de risque global où seraient tenu en compte à la fois les facteurs génétiques et les facteurs environnementaux calculés lors des études de population gouvernementales. Finalement, il serait intéressant, afin de mieux valider les résultats de regroupements régionaux, d'obtenir les données nécessaires à la construction de généalogies à l'aide du fichier de population BALSAC (Bouchard, 2004) afin de faire une étude conjointe entre les données moléculaires et démographiques. Toutes ces étapes pourraient éventuellement permettre aux gestionnaires du système de santé québécois d'incorporer concrètement les risques génétiques à l'élaboration des outils de prévention et de traitement des maladies complexes fréquentes.

BIBLIOGRAPHIE

AAGD (2004). *Asthma and Allergy Gene Database*.
<http://cooke.gsf.de/asthmagene/main.cfm>. Consulté en mai 2005.

Aljada, A, H Ghanim, J Friedman, R Garg, P Mohanty et P Dandona (2001). *Troglitazone reduces the expression of PPARgamma while stimulating that of PPARalpha in mononuclear cells in obese subjects*. J Clin Endocrinol Metab 86(7): p.3130-3.

Altshuler, D, JN Hirschhorn, M Klannemark, CM Lindgren, MC Vohl, J Nemesh, CR Lane, SF Schaffner, S Bolk, C Brewer, T Tuomi, D Gaudet, TJ Hudson, M Daly, L Groop et ES Lander (2000). *The common PPARgamma Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes*. Nat Genet. 26(1): p.76-80.

Arbustini, E, M Grasso, R Fasani, C Klersy, M Diegoli, E Porcu, N Banchieri, P Fortina, C Danesino et G Specchia (1995). *Angiotensin converting enzyme gene deletion allele is independently and strongly associated with coronary atherosclerosis and myocardial infarction*. Br Heart J 74(6): p.584-91.

Benayoun, L, S Letuve, A Druilhe, J Boczkowski, MC Dombret, P Mechighel, J Megret, G Leseche, M Aubier et M Pretolani (2001). *Regulation of peroxisome proliferator-activated receptor gamma expression in human asthmatic airways: relationship with proliferation, apoptosis, and airway remodeling*. Am J Respir Crit Care Med 164(8): p.1487-1494.

Benessiano, J, B Crestani, F Mestari, W Klouche, F Neukirch, S Hacein-Bey, G Durand et M Aubier (1997). *High frequency of a deletion polymorphism of the angiotensin-converting enzyme gene in asthma*. J Allergy Clin Immunol 99(1 Pt 1): p.53-7.

Bergeron, J, T Normand, A Bharucha, MR Ven Murthy, P Julien, C Gagne, C Dionne, M De Braekeleer, D Brun, MR Hayden et al. (1992). *Prevalence, geographical distribution and genealogical investigations of mutation 188 of lipoprotein lipase gene in the French Canadian population of Quebec*. Clin Genet. 41(4): p.206-210.

Bilodeau, M (2002). *La présence écossaise dans les Cantons de l'Est. Histoire d'une grande région*. Histoire Québec 8(1).

- Biserte, G et JC Fruchart (1973). *Le métabolisme des lipoprotéines*. Ann Biol Clin (Paris) 31(2): p.119-22.
- Blumenthal, M (2004). *The immunopathology and genetics of asthma*. Minn Med 87(5): p.53-6.
- Boileau, G (1999a). *L'annexion du territoire de l'Abitibi au Québec. Il y a 100 ans*. Histoire Québec 4(2bis).
- Boileau, G (1999b). *Réflexion sur les villages du Québec*. Histoire Québec 4(2bis).
- Boileau, G (2001). *Le Témiscamingue, entre le peuplement volontaire et la colonisation organisée. Depuis le frère Moffet et le père Gendreau*. Histoire Québec 6(3).
- Boleda, M (1984). *Les migrations au Canada sous le régime français (1608-1760)*. Cah Que Demogr 13(1): p.23-40.
- Bouchard, G (1990). *Représentations de la population et de la société québécoise: apprentissage de la diversité*. Cah Que Demogr 19(1): p.7-28.
- Bouchard, G (2004). *Projet BALSAC-rapport annuel 2003-2004*. 55 p.
- Bouchard, G et M De Braekeleer (1990). *Mouvements migratoires, effets fondateurs et homogénéisation génétique*. IN Histoire d'un génome. Sillery, Presses de l'Université du Québec. p. 281-322.
- Bouchard, G et M De Braekeleer (1992). *Pourquoi les maladies héréditaires?* Sillery, 184 p.
- Bouchard, G, D Gauvreau et M Saint-Hilaire (1996). *Vers les plateaux et vers les villes. La formation des espaces régionaux*. IN Atlas historique du Québec: population et territoire. Les Presses de l'Université Laval. Sainte-Foy. p. 65-80.
- Bouchard, G et M Tremblay (1995). *Le peuplement francophone au Canada: survol historique et géographique (17e-20e siècle)*. IN Français de France et français du Canada: les parlers de l'Ouest de la France, du Québec et de l'Acadie (Sous la direction de Pierre Gauthier et Thomas Lavoie). J. M. Université Lyon 3. Paris, Centre d'études linguistiques Jacques Goudet. p. 109-343.
- Bouthillier, D, CF Sing et J Davignon (1983). *Apolipoprotein E phenotyping with a single gel method: application to the study of informative matings*. J Lipid Res 24(8): p.1060-9.

- Brown, MS et JL Goldstein (1979). *Receptor-mediated endocytosis: insights from the lipoprotein receptor system*. Proc Natl Acad Sci U S A 76(7): p.3330-7.
- Brown, MS et JL Goldstein (1986). *A receptor-mediated pathway for cholesterol homeostasis*. Science 232(4746): p.34-47.
- Cao, K, J Hollenbach, X Shi, W Shi, M Chopek et MA Fernandez-Vina (2001). *Analysis of the frequencies of HLA-A, B, and C alleles and haplotypes in the five major ethnic groups of the United States reveals high levels of diversity in these loci and contrasting distribution patterns in these populations*. Hum Immunol 62(9): p.1009-30.
- Cardon, LR et GR Abecasis (2003). *Using haplotype blocks to map human complex trait loci*. Trends Genet 19(3): p.135-40.
- Center for Biomolecular Science & Engineering (2004). *UCSC Genome Bioinformatics Group*. Baskin School of Engineering, University of California, Santa Cruz. <http://genome.ucsc.edu/>. Consulté en
- Chait, A, PH Iverius et JD Brunzell (1982). *Lipoprotein lipase secretion by human monocyte-derived macrophages*. J Clin Invest 69(2): p.490-3.
- Chan, EM, DE Bulman, AD Paterson, J Turnbull, E Andermann, F Andermann, GA Rouleau, AV Delgado-Escueta, SW Scherer et BA Minassian (2003). *Genetic mapping of a new Lafora progressive myoclonus epilepsy locus (EPM2B) on 6p22*. J Med Genet 40(9): p.671-5.
- Charbonneau, H (1990). *Le caractère français des pionniers de la vallée laurentienne*. Cah Que Demogr 19(1): p.49-62.
- Charbonneau, H, B Desjardins, A Guillemette, Y Landry, J Légaré et F Nault (1987). *Naissance d'une population. Les Français établis au Canada au XVIIe siècle*. Presses universitaires de France et Les Presses de l'Université de Montréal. Paris et Montréal, 232 p.
- Charbonneau, H, B Desjardins, J Légaré et H Denis (2000). *The Population of the St. Lawrence Valley, 1608-1760*. IN *A Population History of North America*. Michael Haines et Richard Steckel. Cambridge, Cambridge University Press. p. 99-142.
- Clark, RB (2002). *The role of PPARs in inflammation and immunity*. Journal of Leukocyte Biology 71: p.388-400.

Cookson, WO, RP Young, AJ Sandford, MF Moffatt, T Shirakawa, PA Sharp, JA Faux, C Julier, Y Nakumuura, Y Nakumura et al. (1992). *Maternal inheritance of atopic IgE responsiveness on chromosome 11q*. Lancet 340(8816): p.381-4.

CORAMH *Corporation de recherche et d'action sur les maladies héréditaires*. (2004). <http://www.coramh.org/>. Consulté en février 2005.

Corbo, RM et R Scacchi (1999). *Apolipoprotein E (APOE) allele distribution in the world. Is APOE*4 a 'thrifty' allele?* Ann Hum Genet 63 (Pt 4): p.301-10.

Dallongeville, J, S Lussier-Cacan et J Davignon (1992). *Modulation of plasma triglyceride levels by apoE phenotype: a meta-analysis*. J Lipid Res 33(4): p.447-54.

Das, HK, J McPherson, GA Bruns, SK Karathanasis et JL Breslow (1985). *Isolation, characterization, and mapping to chromosome 19 of the human apolipoprotein E gene*. J Biol Chem 260(10): p.6240-7.

Davignon, J (2001). *Apolipoprotéine E, une molécule polymorphe et pléiotrope: rôle dans l'athérosclérose et au-delà*. Groupe de recherche sur les hyperlipidémies et l'athérosclérose; Institut de recherches cliniques de Montréal.

Davignon, J et M Roy (1993). *Familial hypercholesterolemia in French-Canadians: taking advantage of the presence of a "founder effect"*. Am J Cardiol. 72(10): p.6D-10D.

de Andrade, M, I Thandi, S Brown, A Gotto, Jr., W Patsch et E Boerwinkle (1995). *Relationship of the apolipoprotein E polymorphism with carotid artery atherosclerosis*. Am J Hum Genet 56(6): p.1379-90.

De Braekeleer, M (1990). *Homogénéité génétique des Canadiens français du Québec: mythe ou réalité?* Cah Que Demogr 19(1): p.29-48.

De Braekeleer, M (1995). *Geographic distribution of 18 autosomal recessive disorders in the French Canadian population of Saguenay-Lac-Saint-Jean, Quebec*. Ann Hum Biol 22(2): p.111-22.

De Braekeleer, M, G Bouchard et M Gradie (1990). *Consanguinité et parenté au Saguenay*. IN *Histoire d'un génoïme*. Sillery, Presses de l'Université du Québec. p. 323-342.

de Bruin, TW, F Mailly, HH van Barlingen, R Fisher, M Castro Cabezas, P Talmud, GM Dallinga-Thie et SE Humphries (1996). *Lipoprotein lipase gene mutations D9N and N291S in four pedigrees with familial combined hyperlipidaemia*. Eur J Clin Invest 26(8): p.631-9.

- Desjardins, B (1990). *Homogénéité ethnique de la population québécoise sous le Régime français*. Cah Que Demogr 19(1): p.63-76.
- Devchand, PR, H Keller, JM Peters, M Vazquez, FJ Gonzalez et W Wahli (1996). *The PPARalpha-leukotriene B4 pathway to inflammation control*. Nature 384(6604): p.39-43.
- Diet, F, RE Pratt, GJ Berry, N Momose, GH Gibbons et VJ Dzau (1996). *Increased accumulation of tissue ACE in human atherosclerotic coronary artery disease*. Circulation 94(11): p.2756-67.
- Dionne, C, C Gagne, P Julien, MR Murthy, G Roederer, J Davignon, M Lambert, D Chitayat, R Ma, H Henderson, PJ Lupien, MR Hayden et M de Braekeleer (1993). *Genealogy and regional distribution of lipoprotein lipase deficiency in French-Canadians of Quebec*. Hum Biol. 65(1): p.29-39.
- Eichner, JE, ST Dunn, G Perveen, DM Thompson, KE Stewart et BC Stroehla (2002). *Apolipoprotein E polymorphism and cardiovascular disease: a HuGE review*. Am J Epidemiol 155(6): p.487-495.
- Elbein, SC, C Yeager, LK Kwong, A Lingam, I Inoue, JM Lalouel et DE Wilson (1994). *Molecular screening of the lipoprotein lipase gene in hypertriglyceridemic members of familial noninsulin-dependent diabetes mellitus families*. J Clin Endocrinol Metab 79(5): p.1450-6.
- Elbrecht, A, Y Chen, CA Cullinan, N Hayes, M Leibowitz, DE Moller et J Berger (1996). *Molecular cloning, expression and characterization of human peroxisome proliferator activated receptors gamma 1 and gamma 2*. Biochem Biophys Res Commun 224(2): p.431-7.
- Emi, M, DE Wilson, PH Iverius, L Wu, A Hata, R Hegele, RR Williams et JM Lalouel (1990). *Missense mutation (Gly----Glu188) of human lipoprotein lipase imparting functional deficiency*. J Biol Chem 265(10): p.5910-6.
- Erdos, EG (1977). *The angiotensin I converting enzyme*. Fed Proc 36(5): p.1760-5.
- Fondation des maladies du coeur du Canada (2003). *Le fardeau croissant des maladies cardiovasculaires et des accidents vasculaires cérébraux au Canada, 2003*. Statistique Canada, Institut canadien d'information sur la santé. 79 p.
- Fortin, G (1962). *L'étude du milieu rural*. Dumont et Martin: p.105-116.

- Gao, J, Y Lin, Y Xiao, K Xu, W Xu, Y Zhu, Y Ma et Y Bai (2000). *Polymorphism of angiotensin-converting enzyme gene and genetic susceptibility to asthma with familial aggregation*. Chin Med Sci J 15(1): p.24-8.
- Gaudet, D, MC Vohl, P Couture, S Moorjani, G Tremblay, P Perron, C Gagne et JP Despres (1999). *Contribution of receptor negative versus receptor defective mutations in the LDL-receptor gene to angiographically assessed coronary artery disease among young (25-49 years) versus middle-aged (50-64 years) men*. Atherosclerosis 143(1): p.153-161.
- GeneCards (2005). Weizmann Institute of Science in association with Xennex. <http://bioinfo.weizmann.ac.il/cards/index.shtml>. Consulté en avril 2004.
- Gobeil, A (2002). *Cookshire: essor d'un village au 19e siècle. Histoire d'une grande région*. Histoire Québec 8(1).
- Goldsby, RA, TJ Kindt et BA Osborne (2001a). *Cellules et organes du système immunitaire*. IN Immunologie. Le cours de Janis Kuby. Paris, Dunod, Paris, 2001. p. 27-60.
- Goldsby, RA, TJ Kindt et BA Osborne (2001b). *Vue d'ensemble du système immunitaire*. IN Immunologie. Le cours de Janis Kuby. Paris, Dunod, Paris, 2001. p. 3-26.
- Gouni-Berthold, I, E Giannakidou, D Muller-Wieland, M Faust, J Kotzka, HK Berthold et W Krone (2004). *Association between the PPARalpha L162V polymorphism, plasma lipoprotein levels, and atherosclerotic disease in patients with diabetes mellitus type 2 and in nondiabetic controls*. Am Heart J 147(6): p.1117-24.
- Greene, ME, B Blumberg, OW McBride, HF Yi, K Kronquist, K Kwan, L Hsieh, G Greene et SD Nimer (1995). *Isolation of the human peroxisome proliferator activated receptor gamma cDNA: expression in hematopoietic cells and chromosomal mapping*. Gene Expr 4(4-5): p.281-99.
- Grégory, P (2002). *Quand Sherbrooke s'appelait Hyatt's Mill*. Histoire Québec 8(1).
- Guillemette, A et J Légaré (1989). *The influence of kinship on seventeenth-century immigration to Canada*. Continuity and Change 4(1): p.79-102.
- Hamalainen, H, H Zhou, W Chou, H Hashizume, R Heller et R Lahesmaa (2001). *Distinct gene expression profiles of human type 1 and type 2 T helper cells*. Genome Biol 2(7).

- Hamanaka, R, K Kohno, T Seguchi, K Okamura, A Morimoto, M Ono, J Ogata et M Kuwano (1992). *Induction of low density lipoprotein receptor and a transcription factor SP-1 by tumor necrosis factor in human microvascular endothelial cells*. J Biol Chem 267(19): p.13160-5.
- Hardy, GH (1908). *Mendelian proportions in a mixed population*. Science 28: p.49-50.
- Harris, RC et G Matthews (1987). *Historical Atlas of Canada. I:From the beginning to 1800*. University of Toronto Press. Toronto, p.
- Hartl, DL et AG Clark (1997). *Principles of population genetics*. Sunderland, Sinauer, p.
- Henderson, HE, F Hassan, GM Berger et MR Hayden (1992). *The lipoprotein lipase Gly188----Glu mutation in South Africans of Indian descent: evidence suggesting common origins and an increased frequency*. J Med Genet 29(2): p.119-22.
- Hill, MR et WO Cookson (1996). *A new variant of the beta subunit of the high-affinity receptor for immunoglobulin E (Fc epsilon RI-beta E237G): associations with measures of atopy and bronchial hyper-responsiveness*. Human Molecular Genetics 5(7): p.959-962.
- Howard, M, J Farrar, M Hilfiker, B Johnson, K Takatsu, T Hamaoka et WE Paul (1982). *Identification of a T cell-derived b cell growth factor distinct from interleukin 2*. J Exp Med 155(3): p.914-23.
- Huang, Y, SC Rall, Jr. et RW Mahley (1997). *Genetic factors precipitating type III hyperlipoproteinemia in hypolipidemic transgenic mice expressing human apolipoprotein E2*. Arterioscler Thromb Vasc Biol 17(11): p.2817-24.
- Ilveskoski, E, M Perola, T Lehtimaki, P Laippala, V Savolainen, J Pajarinne, A Penttila, KH Lalu, A Mannikko, KK Liesto, T Koivula et PJ Karhunen (1999). *Age-dependent association of apolipoprotein E genotype with coronary and aortic atherosclerosis in middle-aged men: an autopsy study*. Circulation 100(6): p.608-13.
- Institut de la statistique du Québec (1996). *Recensement de la population. Immigration, langue et origine ethnique*. Gouvernement du Québec. (2004). <http://www.stat.gouv.qc.ca/publications/regional/theme3.htm>. Consulté en mars 2005.
- Isakson, PC, E Pure, ES Vitetta et PH Krammer (1982). *T cell-derived B cell differentiation factor(s). Effect on the isotype switch of murine B cells*. J Exp Med 155(3): p.734-48.

- Ishigami, M, DK Swertfeger, NA Granholm et DY Hui (1998). *Apolipoprotein E inhibits platelet-derived growth factor-induced vascular smooth muscle cell migration and proliferation by suppressing signal transduction and preventing cell entry to G1 phase.* J Biol Chem 273(32): p.20156-61.
- Jeanpierre, M (1987). *A rapid method for the purification of DNA from blood.* Nucleic Acids Res 15(22): p.9611.
- Jetté, R (1983). *Dictionnaire généalogique des familles québécoises. Des origines à 1730.* Presses de l'Université de Montréal. Montréal, 1176 p.
- Jonasson, L, G Bondjers et GK Hansson (1987). *Lipoprotein lipase in atherosclerosis: its presence in smooth muscle cells and absence from macrophages.* J Lipid Res 28(4): p.437-45.
- Kaufman, J, S Schmitt, J Barnard et W Busse (1992). *Angiotensin-converting enzyme inhibitors in patients with bronchial responsiveness and asthma.* Chest 101(4): p.922-5.
- Kelly, ME, MA Clay, MJ Mistry, HM Hsieh-Li et JA Harmony (1994). *Apolipoprotein E inhibition of proliferation of mitogen-activated T lymphocytes: production of interleukin 2 with reduced biological activity.* Cell Immunol 159(2): p.124-39.
- Kelly-Welch, AE, EM Hanson, MR Boothby et AD Keegan (2003). *Interleukin-4 and interleukin-13 signaling connections maps.* Science 300(5625): p.1527-8.
- Khoo, JC, EM Mahoney et JL Witztum (1981). *Secretion of lipoprotein lipase by macrophages in culture.* J Biol Chem 256(14): p.7105-8.
- King, RC et WD Stansfield (1997). *A dictionary of genetics. Fifth edition.* Oxford University Press. New York, 439 p.
- Laberge, C (1990). *De la génétique des populations à l'épidémiologie génétique. IN Histoire d'un génome.* Sillery, Presses de l'Université du Québec. p. 477-500.
- Laberge, C et J Mathieu (1996). *Le peuplement des basses terres. L'expansion de l'écoumène. IN Atlas historique du Québec: population et territoire.* Les Presses de l'Université Laval. Sainte-Foy. p. 55-62.
- Laprise, C, LP Boulet, J Morissette, E Winstall et V Raymond (2000). *Evidence for association and linkage between atopy, airway hyper-responsiveness, and the beta subunit Glu237Gly variant of the high-affinity receptor for immunoglobulin E in the French-Canadian population.* Immunogenetics 51: p.695-702.

- Le Beau, MM, R Espinosa, 3rd, WL Neuman, W Stock, D Roulston, RA Larson, M Keinanen et CA Westbrook (1993). *Cytogenetic and molecular delineation of the smallest commonly deleted region of chromosome 5 in malignant myeloid diseases.* Proc Natl Acad Sci U S A 90(12): p.5484-8.
- Leitersdorf, E, EJ Tobin, J Davignon et HH Hobbs (1990). *Common low-density lipoprotein receptor mutations in the French Canadian population.* J Clin Invest 85(4): p.1014-23.
- Levesque, G, P Julien, Y Deshaies, PJ Lupien et MR Ven Murthy (1994). *A rapid restriction site screening method for the Pro207-->Leu mutation in the lipoprotein lipase gene.* Hum Mutat 3(4): p.416-7.
- Libby, P (2002). *Inflammation in atherosclerosis.* Nature 420(6917): p.868-74.
- Lindgren, V, KL Luskey, DW Russell et U Francke (1985). *Human genes involved in cholesterol metabolism: chromosomal mapping of the loci for the low density lipoprotein receptor and 3-hydroxy-3-methylglutaryl-coenzyme A reductase with cDNA probes.* Proc Natl Acad Sci U S A 82(24): p.8567-71.
- Linton, MF, VR Babaev, LA Gleaves et S Fazio (1999). *A direct role for the macrophage low density lipoprotein receptor in atherosclerotic lesion formation.* J Biol Chem 274(27): p.19204-10.
- Lohmueller, KE, CL Pearce, M Pike, ES Lander et JN Hirschhorn (2003). *Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease.* Nat Genet 33(2): p.177-82.
- Lusis, AJ, C Heinzmann, RS Sparkes, J Scott, TJ Knott, R Geller, MC Sparkes et T Mohandas (1986). *Regional mapping of human chromosome 19: organization of genes for plasma lipid transport (APOC1, -C2, and -E and LDLR) and the genes C3, PEPD, and GPI.* Proc Natl Acad Sci U S A 83(11): p.3929-33.
- Lusis, AJ, R Mar et P Pajukanta (2004). *Genetics of atherosclerosis.* Annu Rev Genomics Hum Genet 5: p.189-218.
- Ma, Y, HE Henderson, P Julien, G Roederer, J Brunzell et MR Hayden (1991a). *A missense mutation (pro-to-leu207) in the human lipoprotein lipase gene is the major cause of type I hyperlipoproteinemia in French Canadians. (Series) Miami Short Reports.* Advances in Gene Technology: The Molecular Biology of Human Genetic Disease. New York: IRL Press (pub.) 1.: p.34.

- Ma, Y, HE Henderson, V Murthy, G Roederer, MV Monsalve, LA Clarke, T Normand, P Julien, C Gagne, M Lambert et et al. (1991b). *A mutation in the human lipoprotein lipase gene as the most common cause of familial chylomicronemia in French Canadians.* N Engl J Med 324(25): p.1761-6.
- Ma, Y, TC Ooi, MS Liu, H Zhang, R McPherson, AL Edwards, IJ Forsythe, J Frohlich, JD Brunzell et MR Hayden (1994). *High frequency of mutations in the human lipoprotein lipase gene in pregnancy-induced chylomicronemia: possible association with apolipoprotein E2 isoform.* J Lipid Res 35(6): p.1066-75.
- Mailly, F, Y Tugrul, PW Reymer, T Bruin, M Seed, BF Groenemeyer, A Asplund-Carlson, D Vallance, AF Winder, GJ Miller et et al. (1995). *A common variant in the gene for lipoprotein lipase (Asp9-->Asn). Functional implications and prevalence in normal and hyperlipidemic subjects.* Arterioscler Thromb Vasc Biol 15(4): p.468-78.
- Marx, N, G Sukhova, C Murphy, P Libby et J Plutzky (1998). *Macrophages in human atheroma contain PPARgamma: differentiation-dependent peroxisomal proliferator-activated receptor gamma (PPARgamma) expression and reduction of MMP-9 activity through PPARgamma activation in mononuclear phagocytes in vitro.* Am J Pathol 153(1): p.17-23.
- Marx, N, GK Sukhova, T Collins, P Libby et J Plutzky (1999). *PPARalpha activators inhibit cytokine-induced vascular cell adhesion molecule-1 expression in human endothelial cells.* Circulation 99(24): p.3125-31.
- Masoli, M, D Fabian, S Holt et R Beasley (2004). *Global burden of asthma. Developped for the Global Initiative for Asthma.* Medical Research Institute of New-Zealand, Wellington, New Zealand and University of Southampton, Southampton, United Kingdom. 122 p.
- Mattei, MG, C Hubert, F Alhenc-Gelas, N Roeckel, P Corvol et F Soubrier (1989). *Angiotensin-I converting enzyme gene is on chromosome 17.* Cytogenet. Cell Genet. 51: p.1041.
- Mistry, MJ, MA Clay, ME Kelly, MA Steiner et JA Harmony (1995). *Apolipoprotein E restricts interleukin-dependent T lymphocyte proliferation at the G1A/G1B boundary.* Cell Immunol 160(1): p.14-23.
- Monsalve, MV, H Henderson, G Roederer, P Julien, S Deeb, JJ Kastelein, L Peritz, R Devlin, T Bruin, MR Murthy et et al. (1990). *A missense mutation at codon 188 of the human lipoprotein lipase gene is a frequent cause of lipoprotein lipase deficiency in persons of different ancestries.* J Clin Invest 86(3): p.728-34.

- Moorjani, S, M Roy, A Torres, C Betard, C Gagne, M Lambert, D Brun, J Davignon et P Lupien (1993). *Mutations of low-density-lipoprotein-receptor gene, variation in plasma cholesterol, and expression of coronary heart disease in homozygous familial hypercholesterolaemia*. Lancet 341(8856): p.1303-6.
- Murthy, V, P Julien et C Gagne (1996). *Molecular pathobiology of the human lipoprotein lipase gene*. pharmacol. ther. 70: p.101-135.
- Nagy, L, P Tontonoz, JG Alvarez, H Chen et RM Evans (1998). *Oxidized LDL regulates macrophage gene expression through ligand activation of PPARgamma*. Cell 93(2): p.229-40.
- NCBI (2004). *National Center of Biotechnology Information*. U.S. national library of medecine. <http://www.ncbi.nlm.nih.gov/>. Consulté en mai 2005.
- Newton, CR, A Graham, LE Heptinstall, SJ Powell, C Summers, N Kalsheker, JC Smith et AF Markham (1989). *Analysis of any oint mutation in DNA. The amplification refractory mutation system (ARMS)*. Nucleic Acids Res 11(17): p.2503-2516.
- Nilsson-Ehle, P, AS Garfinkel et MC Schotz (1980). *Lipolytic enzymes and plasma lipoprotein metabolism*. Annu Rev Biochem 49: p.667-93.
- Noguchi, E, M Shibasaki, T Arinami, K Takeda, Y Yokouchi, T Kawashima, H Yanagi, A Matsui et H Hamaguchi (1998). *Association of asthma and the interleukin-4 promoter gene in Japanese*. Clin Exp Allergy 28(4): p.449-53.
- Normand, T, J Bergeron, T Fernandez-Margallo, A Bharucha, MR Ven Murthy, P Julien, C Gagne, C Dionne, M De Braekeleer, R Ma et al. (1992). *Geographic distribution and genealogy of mutation 207 of the lipoprotein lipase gene in the French Canadian population of Quebec*. Hum Genet 89(6): p.671-675.
- Office québécois de la langue française (2004). *Le grand dictionnaire terminologique*. http://w3.granddictionnaire.com/btml/fra/r_motclef/index1024_1.asp. Consulté en avril 2005.
- OGMD (2004). *Obesity Gene Map Database*. <http://obesitygene.pbrc.edu/>. Consulté en décembre 2004.
- OMS (2000). *Asthme bronchitique*. Organisation mondiale de la Santé. (2004). <http://www.who.int/mediacentre/factsheets/fs206/fr/index.html>. Consulté en juin 2005.

- Paik, YK, DJ Chang, CA Reardon, GE Davies, RW Mahley et JM Taylor (1985). *Nucleotide sequence and structure of the human apolipoprotein E gene*. Proc Natl Acad Sci U S A 82(10): p.3445-9.
- Plaut, M, JH Pierce, CJ Watson, J Hanley-Hyde, RP Nordan et WE Paul (1989). *Mast cell lines produce lymphokines in response to cross-linkage of Fc epsilon RI or to calcium ionophores*. Nature 339(6219): p.64-7.
- Portail Québec (2005). *Portails régionaux*. Gouvernement du Québec. (2005). <http://www.gouv.qc.ca/wps/portal/pgs/commun/portailsregionaux?lang=fr>. Consulté en mars 2005.
- PRDH (2004). *Le Programme de recherche en démographie historique*. <http://www.genealogie.umontreal.ca/fr/>. Consulté en février 2005.
- Rall, SC, Jr., KH Weisgraber, TL Innerarity et RW Mahley (1982). *Structural basis for receptor binding heterogeneity of apolipoprotein E from type III hyperlipoproteinemic subjects*. Proc Natl Acad Sci U S A 79(15): p.4696-700.
- Reymer, PW, E Gagne, BE Groenemeyer, H Zhang, I Forsyth, H Jansen, JC Seidell, D Kromhout, KE Lie, J Kastelein et al. (1995). *A lipoprotein lipase mutation (Asn291Ser) is associated with reduced HDL cholesterol levels in premature atherosclerosis*. Nat Genet 10(1): p.28-34.
- Ricote, M, J Huang, L Fajas, A Li, J Welch, J Najib, JL Witztum, J Auwerx, W Palinski et CK Glass (1998). *Expression of the peroxisome proliferator-activated receptor gamma (PPARgamma) in human atherosclerosis and regulation in macrophages by colony stimulating factors and oxidized low density lipoprotein*. Proc Natl Acad Sci U S A 95(13): p.7614-9.
- Riddell, DR, A Graham et JS Owen (1997). *Apolipoprotein E inhibits platelet aggregation through the L-arginine:nitric oxide pathway. Implications for vascular disease*. J Biol Chem 272(1): p.89-95.
- Rigat, B, C Hubert, F Alhenc-Gelas, F Cambien, P Corvol et F Soubrier (1990). *An insertion/deletion polymorphism in the angiotensin I-converting enzyme gene accounting for half the variance of serum enzyme levels*. J Clin Invest 86(4): p.1343-6.
- Rioux, P, C Truchon et C Drouin (2003). *Étude généalogique de deux maladies héréditaires rares dans l'Est du Québec : la maladie de Lafora et le syndrome de Morquio*. 71ième congrès de l'ACFAS; session: C-108 La génétique humaine au Québec : de Clouston à Cartagène (colloque).

- Rosenwasser, LJ, DJ Klemm, JK Dresback, H Inamura, JJ Mascali, M Klinnert et L Borish (1995). *Promoter polymorphisms in the chromosome 5 gene cluster in asthma and atopy*. Clin Exp Allergy 25 Suppl 2: p.74-8; discussion 95-6.
- Rouis, M, P Lohse, KA Dugi, OU Beg, R Ronan, GD Talley, JD Brunzell et S Santamarina-Fojo (1996). *Homozygosity for two point mutations in the lipoprotein lipase (LPL) gene in a patient with familial LPL deficiency: LPL(Asp9-->Asn, Tyr262-->His)*. J Lipid Res 37(3): p.651-61.
- Salone, E (1970). *La colonisation de la Nouvelle-France: étude sur les origines de la nation canadienne française*. E. Guilmoto, [1905 Trois-Rivières:Éditions Le Boréal Express, c1970]. Paris, p.
- Sandford, AJ, T Shirakawa, MF Moffatt, SE Daniels, C Ra, JA Faux, RP Young, Y Nakamura, GM Lathrop, WO Cookson et al. (1993). *Localisation of atopy and beta subunit of high-affinity IgE receptor (Fc epsilon RI) on chromosome 11q*. Lancet 341(8841): p.332-4.
- Santé cardiovasculaire (1990). *Enquête sur la santé cardiovasculaire de la population québécoise*. Institut de la Statistique du Québec.
- Scherer, B (1984). *Biostatistique*. Gaëtan Morin. Montréal, Paris, Casablanca, 850 p.
- Schieffer, B, WG Paxton, MB Marrero et KE Bernstein (1996). *Importance of tyrosine phosphorylation in angiotensin II type 1 receptor signaling*. Hypertension 27(3 Pt 2): p.476-80.
- Scriver, CR (2001). *Human genetics: Lessons from Quebec populations*. Annu Rev Genomics Hum 2: p.69-101.
- Semple, PF et GW Herd (1986). *Cough and wheeze caused by inhibitors of angiotensin-converting enzyme*. N Engl J Med 314(1): p.61.
- Serre, JL (1997). *Génétique des populations. Modèles de base et applications*. Nathan. 250 p.
- Shepherd, J (1994). *Lipoprotein metabolism. An overview*. Drugs 47 Suppl 2: p.1-10.
- Sher, T, HF Yi, OW McBride et FJ Gonzalez (1993). *cDNA cloning, chromosomal mapping, and functional characterization of the human peroxisome proliferator activated receptor*. Biochemistry 32(21): p.5598-604.

- Shirakawa, T, XQ Mao, S Sasaki, T Enomoto, M Kawai, K Morimoto et J Hopkin (1996). *Association between atopic asthma and a coding variant of Fc epsilon RI beta in a Japanese population.* Hum Mol Genet 5(8): p.1129-30.
- Shriver, MD, GC Kennedy, EJ Parra, HA Lawson, V Sonpar, J Huang, JM Akey et KW Jones (2004). *The genomic distribution of population substructure in four populations using 8,525 autosomal SNPs.* Hum Genomics 1(4): p.274-86.
- Sing, CF et J Davignon (1985). *Role of the apolipoprotein E polymorphism in determining normal plasma lipid and lipoprotein variation.* Am J Hum Genet 37(2): p.268-85.
- Sparkes, RS, S Zollman, I Klisak, TG Kirchgessner, MC Komaromy, T Mohandas, MC Schotz et AJ Lusis (1987). *Human genes involved in lipolysis of plasma lipoproteins: mapping of loci for lipoprotein lipase to 8p22 and hepatic lipase to 15q21.* Genomics 1(2): p.138-44.
- Statistique Canada (2001a). *Population immigrante selon le lieu de naissance par provinces et territoires (Recensement de 2001).* Gouvernement du Canada. (2004). http://www.statcan.ca/francais/Pgdb/demo34a_f.htm. Consulté en avril 2005.
- Statistique Canada (2001b). *Proportion des personnes nées à l'étranger, par provinces et territoires (Recensement 2001, 1996, 1991).* Gouvernement du Canada. http://www.statcan.ca/francais/Pgdb/demo46a_f.htm. Consulté en mars 2005.
- Steinberg, D (1983). *Lipoproteins and atherosclerosis. A look back and a look ahead.* Arteriosclerosis 3(4): p.283-301.
- Stevenson, FK, RE Snow, CJ Chapman, AJ Frew et ST Holgate (1996). *Genetic analysis of IgE.* Thorax 51(5): p.458-60.
- Sutherland, GR, E Baker, DF Callen, VJ Hyland, G Wong, S Clark, SS Jones, LK Eglinton, MF Shannon, AF Lopez et al. (1988). *Interleukin 4 is at 5q31 and interleukin 6 is at 7p15.* Hum Genet 79(4): p.335-7.
- Swain, SL, AD Weinberg, M English et G Huston (1990). *IL-4 directs the development of Th2-like helper effectors.* J Immunol 145(11): p.3796-806.
- Tangirala, RK, D Pratico, GA FitzGerald, S Chun, K Tsukamoto, C Maugeais, DC Usher, E Pure et DJ Rader (2001). *Reduction of isoprostanes and regression of advanced atherosclerosis by apolipoprotein E.* J Biol Chem 276(1): p.261-6.

- Tomita, H, S Sato, R Matsuda, N Ogisu, T Mori, T Niimi et S Shimizu (1998). *Genetic polymorphism of the angiotensin-converting enzyme (ACE) in asthmatic patients.* Respir Med 92(12): p.1305-10.
- Tontonoz, P, E Hu, RA Graves, AI Budavari et BM Spiegelman (1994). *mPPAR gamma 2: tissue-specific regulator of an adipocyte enhancer.* Genes Dev 8(10): p.1224-34.
- Tontonoz, P, L Nagy, JG Alvarez, VA Thomazy et RM Evans (1998). *PPARgamma promotes monocyte/macrophage differentiation and uptake of oxidized LDL.* Cell. 98 ou93: p.241-252.
- Tremblay, M, H Vézina, B Desjardins et L Houde (2004). *Distant kinship in the Quebec population: an analysis of the founder effects using extended genealogies,* Paris, 1-2 octobre.
- Utermann, G, M Hees et A Steinmetz (1977). *Polymorphism of apolipoprotein E and occurrence of dysbetalipoproteinaemia in man.* Nature 269(5629): p.604-7.
- Utermann, G, N Pruin et A Steinmetz (1979). *Polymorphism of apolipoprotein E. III. Effect of a single polymorphic gene locus on plasma lipid levels in man.* Clin Genet 15(1): p.63-72.
- Vézina, H (1996). *Démographie génétique et maladies héréditaires au Québec: l'état des recherches.* Cah Que Demogr 25(2): p.293-322.
- Vohl, MC, P Lepage, D Gaudet, CG Brewer, C Betard, P Perron, G Houde, C Cellier, JM Faith, JP Despres, K Morgan et TJ Hudson (2000). *Molecular scanning of the human PPAR α gene: association of the L162v mutation with hyperapobetalipoproteinemia.* Journal of Lipid Research 41: p.945-952.
- Vohl, MC, S Moorjani, M Roy, D Gaudet, AL Torres, A Minich, C Gagné, G Tremblay, M Lambert, J Bergeron, P Couture, P Perron, S Blaichman, LD Brun, J Davignon, PJ Lupien et JP Després (1997). *Geographic distribution of French-Canadian low-density lipoprotein receptor gene mutations in the Province of Quebec.* Clinical genetics. 52: p.1-6.
- Walley, AJ et WO Cookson (1996). *Investigation of an interleukin-4 promoter polymorphism for associations with asthma and atopy.* J Med Genet 33(8): p.689-92.
- Weinberg, W (1908). *Über den Nachweis der Verebung beim Menschen.* Jahresh. Verein f. Vaterl. 64: p.368-82.

- Weinstock, JV, AM Blum et JT Kassab (1987). *Angiotensin II is chemotactic for a T-cell subset which can express migration inhibition factor activity in murine schistosomiasis mansoni*. Cell Immunol 107(1): p.180-7.
- Weisgraber, KH, SC Rall, Jr. et RW Mahley (1981). *Human E apoprotein heterogeneity. Cysteine-arginine interchanges in the amino acid sequence of the apo-E isoforms*. J Biol Chem 256(17): p.9077-83.
- Welch, JS, M Ricote, TE Akiyama, FJ Gonzalez et CK Glass (2003). *PPARgamma and PPARdelta negatively regulate specific subsets of lipopolysaccharide and IFN-gamma target genes in macrophages*. Proc Natl Acad Sci U S A 100(11): p.6712-7.
- WHO (2003). *Cardiovascular disease: prevention and control*. World Health Organization. <http://www.who.int/dietphysicalactivity/publications/facts/cvd/en/>. Consulté en février 2005.
- Wion, KL, TG Kirchgessner, AJ Lusis, MC Schotz et RM Lawn (1987). *Human lipoprotein lipase complementary DNA sequence*. Science 235(4796): p.1638-41.
- Wittrup, HH, A Tybjaerg-Hansen et BG Nordestgaard (1999). *Lipoprotein lipase mutations, plasma lipids and lipoproteins, and risk of ischemic heart disease. A meta-analysis*. Circulation 99(22): p.2901-7.
- Wyman, AR et R White (1980). *A highly polymorphic locus in human DNA*. Proc Natl Acad Sci U S A 77(11): p.6754-8.
- Xhignesse, M, S Lussier-Cacan, CF Sing, AM Kessling et J Davignon (1991). *Influences of common variants of apolipoprotein E on measures of lipid metabolism in a sample selected for health*. Arterioscler Thromb 11(4): p.1100-10.
- Yen, CJ, BA Beamer, C Negri, K Silver, KA Brown, DP Yarnall, DK Burns, J Roth et AR Shuldiner (1997). *Molecular scanning of the human peroxisome proliferator activated receptor gamma (hPPAR-gamma) gene in diabetic Caucasians: identification of a pro12ala PPAR-gamma-2 missense mutation*. Biochem. Biophys. Res. Commun. 241: p.270-274.
- Yokota, T, T Otsuka, T Mosmann, J Banchereau, T DeFrance, D Blanchard, JE De Vries, F Lee et K Arai (1986). *Isolation and characterization of a human interleukin cDNA clone, homologous to mouse B-cell stimulatory factor 1, that expresses B-cell- and T-cell-stimulating activities*. Proc. Nat. Acad. Sci. 83: p.5894-5898.
- Zilversmit, DB (1973). *A proposal linking atherogenesis to the interaction of endothelial lipoprotein lipase with triglyceride-rich lipoproteins*. Circ Res 33(6): p.633-8.

ANNEXE A

Résultats du calcul de l'équilibre d'Hardy-Weinberg pour chaque variant dans chaque échantillon des régions administratives du Québec séparément ainsi que pour la cohorte provinciale

Régions		Bas-Saint-Laurent	Saguenay-Lac-Saint-Jean	Québec	Mauricie	Estrie	Lanaudière	Outaouais	Abitibi-Témiscamingue
SNPs	Valeurs								
IL4-C-589T	χ^2 P*	0,0008 1,00	0,1003 0,80	6,2043 0,02	7,2709 0,01	0,6145 0,45	0,2549 0,68	0,00002 1,00	0,1319 0,75
MS4A2-E237G	χ^2 P	0,0537 0,90	0,1508 0,75	0,0921 0,85	0,0603 0,90	0,1393 0,75	0,2174 0,68	0,0061 1,00	0,0345 1,00
ACE I/D	χ^2 P	0,0262 0,95	1,8368 0,20	0,6774 0,56	1,7827 0,20	0,0408 0,90	0,4512 0,55	0,0060 1,00	2,8442 0,10
PPARG2-P12A	χ^2 P	0,0140 1,00	0,0015 1,00	0,3257 0,60	0,1845 0,72	5,8770 0,02	4,6890 0,04	0,5086 0,50	0,8698 0,38
PPARA-L162V	χ^2 P	0,5570 0,50	0,9287 0,35	1,1363 0,32	2,2933 0,15	0,0862 0,80	0,3899 0,55	0,5002 0,50	0,5420 0,50
LPL-D9N	χ^2 P	0,0799 0,85	0,2016 0,70	0,0734 0,85	0,0561 0,90	0,2272 0,68	— —	0,1471 0,75	0,0312 1,00
LPL-G188E	χ^2 P	— —	— —	— —	— —	— —	— —	— —	— —
LPL-P207L	χ^2 P	— —	0,0012 1,00	— —	— —	— —	— —	— —	— —
LPL-N291S	χ^2 P	0,1067 0,80	0,0036 1,00	0,0024 1,00	0,0086 1,00	0,0104 1,00	0,1193 0,78	— 1,00	0,0167 1,00
APOE E2/E3/E4	χ^2 P	2,8902 0,09	15,9870 0,00	0,4124 0,48	2,8082 0,12	1,5905 0,23	2,2836 0,15	2,0330 0,16	0,6235 0,42
LDLR-W66G	χ^2 P	— —	— —	— —	— —	— —	— —	— —	— —

* La valeur P limite est de 0,05. Donc toutes les populations ayant une probabilité en dessous de 0,05 ne sont pas considérées comme étant en équilibre de Hardy-Weinberg.

Régions		Côte-Nord	Nord-du-Québec	Gaspésie-Îles-de-la-Madeleine	Laurentides	Chaudière-Appalaches	Centre-du-Québec	Montréal	Province de Québec
SNPs	Valeurs								
IL4-C-589T	χ^2 P*	0,6486 0,42	0,0747 0,90	0,2082 0,70	0,0101 1,00	0,0907 0,80	0,4619 0,52	0,6481 0,42	1,3805 0,27
MS4A2-E237G	χ^2 P	0,0070 1,00	— 0,50	0,5110 0,85	0,0700 1,00	0,0050 0,90	0,0316 0,15	2,1822 0,15	0,0029 1,00
ACE I/D	χ^2 P	0,0091 1,00	5,7647 0,02	0,0306 0,95	0,0061 1,00	0,0963 0,78	1,5778 0,22	3,0104 0,09	0,2885 0,65
PPARG2-P12A	χ^2 P	0,3549 0,60	1,2658 0,28	3,4124 0,07	5,9265 0,02	3,1490 0,08	0,1615 0,75	2,0317 0,17	1,6162 0,25
PPARA-L162V	χ^2 P	— —	— —	0,1949 0,70	33,7277 0,00	0,2543 0,68	0,7589 0,40	0,7147 0,40	0,0004 1,00
LPL-D9N	χ^2 P	— 1,00	0,0075 0,83	0,0830 0,83	0,0715 0,83	0,1792 0,72	0,0037 1,00	0,0775 0,85	0,8588 0,38
LPL-G188E	χ^2 P	— —	— —	— —	— —	— —	— —	— —	— —
LPL-P207L	χ^2 P	— —	— —	— —	— —	— —	— —	— —	0,0007 1
LPL-N291S	χ^2 P	— —	— —	— —	0,0387 1,00	0,0424 1,00	0,0033 1,00	0,0189 1,00	0,2393 0,68
APOE E2/E3/E4	χ^2 P	10,1186 0,00	2,3920 0,13	1,7053 0,22	1,6103 0,23	0,9613 0,32	14,4392 0,00	1,8422 0,20	10,8861 0,00
LDLR-W66G	χ^2 P	— —	— —	— —	— —	— —	— —	— —	— —

* La valeur P limite est de 0,05. Donc toutes les populations ayant une probabilité en dessous de 0,05 ne sont pas considérées comme étant en équilibre de Hardy-Weinberg.

ANNEXE B

**Fréquence allélique (%) des 11 variants génotypés pour chaque
échantillon des régions administratives de la province de Québec**

Régions	Gènes	IL4	MS4A2	ACE	PPARG2	PPARA	LPL			APOE		LDLR	
	SNPs	C-589T	E237G	I/D	P12A	L162V	D9N	G188E	P207L	N291S	E2	E4	W66G
Bas-St-Laurent		15,57	2,07	56,75	10,61	5,52	2,19	0,00	0,00	2,53	8,86	14,87	0,00
Saguenay–Lac-Saint-Jean		12,25	2,60	55,26	10,03	7,68	2,55	0,00	0,20	0,34	10,62	18,86	0,00
Québec		11,65	2,88	57,94	14,62	9,35	2,78	0,00	0,00	0,48	5,34	14,56	0,00
Mauricie		10,16	2,29	61,07	14,40	13,54	2,16	0,00	0,00	0,84	12,62	12,62	0,00
Estrie		13,86	3,68	50,98	12,00	11,27	4,59	0,00	0,00	1,02	9,28	13,92	0,00
Lanaudière		11,43	2,93	52,49	12,24	8,77	0,00	0,00	0,00	2,27	8,80	16,52	0,00
Outaouais		11,80	1,19	53,62	7,86	7,75	4,48	0,00	0,00	0,00	4,35	12,32	0,00
Abitibi-Témiscamingue		9,62	2,21	50,63	9,49	7,69	2,00	0,00	0,00	1,61	13,01	11,64	0,00
Côte-Nord		10,42	1,35	55,13	11,90	6,94	0,00	0,00	0,00	0,00	10,00	12,86	0,00
Nord-du-Québec		4,55	0,00	50,00	16,18	0,00	1,47	0,00	0,00	0,00	4,55	16,67	0,00
Gaspésie-Îles-de-la-Madeleine		5,94	1,02	57,69	15,69	5,77	3,85	0,00	0,00	0,00	7,00	10,00	0,00
Laurentides		12,30	3,28	55,00	12,50	3,28	3,33	0,00	0,00	2,46	9,02	13,93	0,00
Chaudière-Appalaches		15,18	0,98	57,55	15,18	6,36	5,36	0,00	0,00	2,68	8,04	8,93	0,00
Centre-du-Québec		8,33	2,03	59,21	13,46	8,97	0,74	0,00	0,00	0,65	12,14	12,86	0,00
Montréal		7,32	4,91	54,51	15,45	6,15	2,46	0,00	0,00	1,23	9,43	18,44	0,00
Province de Québec		11,42	2,61	55,21	12,24	7,87	2,27	0,00	0,08	1,25	9,24	15,10	0,00
Écart type		3,07	1,20	3,14	2,43	3,08	1,55	0,00	0,05	0,97	2,62	2,75	0,00