

Université de Montréal

**Étude pangénomique de la variabilité dans le nombre de
copies liée à l'hypertension artérielle et ses anomalies
métaboliques associées**

par

Mahiné Ivanga

Sciences Biomédicales

Médecine

Thèse présentée à la Faculté de Médecine
en vue de l'obtention du grade de Docteur
en Sciences Biomédicales

Mars, 2014

© Mahiné Ivanga, 2014

Université de Montréal
Faculté des études supérieures et postdoctorales

Cette thèse intitulée

**Étude pangénomique de la variabilité dans le nombre de
copies liée à l'hypertension artérielle et ses anomalies
métaboliques associées**

Présentée par :
Mahiné Ivanga

a été évaluée par un jury composé des personnes suivantes :

Sylvie Mader, président-rapporteur
Pavel Hamet, directeur de recherche
Daniel Sinnett, membre du jury
Claude Laberge, examinateur externe
Pierrette Gaudreau, représentante du doyen de la FES

Résumé

L'hypertension artérielle essentielle (HTA) est une pathologie complexe, multifactorielle et à forte composante génétique. L'impact de la variabilité dans le nombre de copies sur l'HTA est encore peu connu. Nous envisageons que des variants dans le nombre de copies (CNVs) communs pourraient augmenter ou diminuer le risque pour l'HTA. Nous avons exploré cette hypothèse en réalisant des associations pangénomiques de CNVs avec l'HTA et avec l'HTA et le diabète de type 2 (DT2), chez 21 familles du Saguenay-Lac-St-Jean (SLSJ) caractérisées par un développement précoce de l'HTA et de la dyslipidémie. Pour la réplique, nous disposons, d'une part, de 3349 sujets diabétiques de la cohorte ADVANCE sélectionnés pour des complications vasculaires. D'autre part, de 187 sujets de la cohorte Tchèque Post-MONICA (CTPM), choisis selon la présence/absence d'albuminurie et/ou de syndrome métabolique. Finalement, 134 sujets de la cohorte CARTaGENE ont été analysés pour la validation fonctionnelle.

Nous avons détecté deux nouveaux loci, régions de CNVs (CNVRs) à effets quantitatifs sur 17q21.31, associés à l'hypertension et au DT2 chez les sujets SLSJ et associés à l'hypertension chez les diabétiques ADVANCE. Un modèle statistique incluant les deux variants a permis de souligner le rôle essentiel du locus CNVR1 sur l'insulino-résistance, la précocité et la durée du diabète, ainsi que sur le risque cardiovasculaire. CNVR1 régule l'expression du pseudogène LOC644172 dont le dosage est associé à la prévalence de l'HTA, du DT2 et plus particulièrement au risque cardiovasculaire et à l'âge vasculaire ($P < 2 \times 10^{-16}$). Nos résultats suggèrent que les porteurs de la duplication au locus CNVR1 développent précocement une anomalie de la fonction bêta pancréatique et de l'insulino-

résistance, dues à un dosage élevé de *LOC644172* qui perturberait, en retour, la régulation du gène paralogue fonctionnel, *MAPK8IP1*.

Nous avons également identifié six CNVRs hautement hérités et associés à l'HTA chez les sujets SLSJ. Le score des effets combinés de ces CNVRs est apparu positivement et étroitement relié à la prévalence de l'HTA ($P=2\times 10^{-10}$) et à l'âge de diagnostic de l'HTA. Dans la population SLSJ, le score des effets combinés présente une statistique C, pour l'HTA, de 0.71 et apparaît aussi performant que le score de risque Framingham pour la prédiction de l'HTA chez les moins de 25 ans. Un seul nouveau locus de CNVR sur 19q13.12, où la délétion est associée à un risque pour l'HTA, a été confirmé chez les Caucasiens CTPM. Ce CNVR englobe le gène *FFAR3*. Chez la souris, il a été démontré que l'action hypotensive du propionate est en partie médiée par *Ffar3*, à travers une interférence entre la flore intestinale et les systèmes cardiovasculaire et rénal.

Les CNVRs identifiées dans cette étude, affectent des gènes ou sont localisées dans des QTLs reliés majoritairement aux réponses inflammatoires et immunitaires, au système rénal ainsi qu'aux lésions/réparations rénales ou à la spéciation. Cette étude suggère que l'étiologie de l'HTA ou de l'HTA associée au DT2 est affectée par des effets additifs ou interactifs de CNVRs.

Mots-clés : Associations pangénomiques, Hypertension (HTA), Diabète de type 2 (DT2), Région de variabilité dans le nombre de copies (CNVR), Variant dans le nombre de copies (CNV), Loci de caractères quantitatifs (QTL)

Abstract

Essential hypertension (HT) is a multifactorial complex disease with a strong genetic component. However, little is known about the effects of copy number variance on HT. We hypothesized common Copy Number Variants (CNVs) could increase or decrease the risk for HT. We performed GWAS of CNVs with HT and, with HT and Type 2 Diabetes (T2D), in 21 families of the Saguenay-Lac-St-Jean region of Quebec (FC) affected by early-onset hypertension and dyslipidemia. Replication was tested in a cohort of 3349 unrelated diabetic subjects of Caucasian origin from the ADVANCE trial. Replication was also tested in 187 individuals from the Czech Post-Monica (CPM) cross-sectional survey, ascertained by the presence/absence of albuminuria and/or metabolic syndrome. We performed locus-specific transcriptional analyses in 134 subjects from the CARTaGENE population cohort.

We identified two CNV Regions (CNVRs), at 17q21.31, associated with HT and T2D in FC and associated with hypertension in ADVANCE diabetics. A statistical model of association including both CNVRs underlined the main effect size of CNVR1 on insulin resistance, T2D early onset and duration, and risk for cardiovascular diseases (CVD). CNVR1 appeared to influence *LOC644172* expression, whose transcript abundance was associated with the prevalence of HT and T2D, and strongly with the risk of CVD and vascular age ($P < 2 \times 10^{-16}$). Our results suggest carriers of copy-number gain at these 17q21.31 loci, principally at the CNVR1 locus, undergo premature β -cell functional deregulation and insulin resistance, due to increase dosage of the *LOC644172* pseudogene,

which might in turn affect the regulation of expression of its functional paralog, *MAPK8IP1*.

We also report six different CNVR loci, highly heritable and contributing to the risk of hypertension, in French Canadians. The combined CNV risk score appeared robustly related to prevalence of hypertension ($p=2\times 10^{-10}$) and age at diagnosis of hypertension. In FC, this combined CNV risk score model showed a C-statistic of 0.71 for HT and appeared as powerful as Framingham HT risk score in predicting hypertension in individuals aged less than 25. We validated the association of a new locus, 19q13.12 deletion-CNVR, with hypertension, in CPM. *FFAR3* surrounds this 19q13.12 deletion-CNVR. It has been demonstrated that in mice, a portion of propionate hypotensive effect is mediated by *Ffar3*, and involves a cross-talk between the gut microbiota and the renal-cardiovascular system.

The identified CNVRs appear to influence genes and QTLs mainly related to immune and inflammatory responses and renal damaged and repair. Some CNVRs are exclusive to primates. This study suggests that additive and interactive actions of multiple copy-number variants are involved in the etiology of hypertension or of hypertension associated with T2D.

Keywords : Genome wide association study (GWAS), Hypertension (HT), Type 2 Diabetes (T2D), Copy Number Variant Region (CNVR), Copy Number Variant (CNV), Quantitative Trait Loci (QTL)

Table des matières

INTRODUCTION	1
I-L'hypertension artérielle	1
1.1- Physiopathologie de l'hypertension artérielle	1
1.1.1- Des déterminants biologiques de base	1
1.1.2- Des facteurs de risque	3
1.2- L'hypertension et les pathologies cardiovasculaires	6
1.2.1- Épidémiologie de l'hypertension artérielle	6
1.2.2- L'hypertension artérielle : un facteur de risque clef pour les maladies cardiovasculaires	7
1.3- L'association de l'hypertension avec des anomalies métaboliques renforce le risque cardiovasculaire	8
1.3.1- La résistance à l'insuline	9
1.3.2- Hypertension et dyslipidémies	9
1.3.3- Hypertension et obésité	10
1.3.4- Hypertension et diabète de type 2	10
1.3.5- Hypertension et syndrome métabolique	11
1.4- Génétique de l'hypertension artérielle	13
1.4.1- Évidences d'un déterminisme génétique de l'hypertension artérielle ..	13
1.4.2- Les formes monogéniques d'hypertension artérielle	14
1.4.3- Recherche pangénomique de loci reliés à l'hypertension artérielle essentielle	15
II-La variabilité dans le nombre de copies	19
2.1- Les Polymorphismes génétiques	19
2.2- La variation dans le nombre de copies	20
2.2.1- Définition	20
2.2.2- Mécanismes de formation et différentes classes de CNV	21
2.2.3- CNVs vs. SNPs	23
2.3- Techniques de détection des CNVs	24

2.4-	CNVs, impact fonctionnel, adaptation et évolution	25
2.4.1-	CNV et adaptation	26
2.4.2-	Duplication et spéciation	27
2.5-	CNVs, héritabilité et pathologies humaines.....	28
2.5.1-	Héritabilité des CNVs	29
2.5.2-	Les CNVs sont impliqués dans diverses pathologies humaines	30
	III-Conclusion	32
	HYPOTHÈSE DE TRAVAIL ET OBJECTIFS	33
	CHAPITRE I: Copy Number Variants on chromosome 17 Associated with Type 2 Diabetes and Hypertension	35
1.1	Abstract	36
1.2	Introduction.....	37
1.3	Research Design and Methods	38
1.3.1-	Study cohorts.....	38
1.3.2-	Phenotypes	39
1.3.3-	CNV detection and selection.....	41
1.3.4-	CNV analysis using TaqMan	43
1.3.5-	Analysis of expression levels	44
1.3.6-	Statistical analyses	46
1.4	Results.....	47
1.5	Discussion	50
1.6	References.....	57
1.7	Tables and Figures	60
	CHAPITRE II : GWAS Identifies CNVRs Contributing to Hypertension in French Canadian Families.....	70
2.1-	Abstract	71
2.2-	Introduction.....	72
2.3-	Methods.....	73
2.3.1-	Study cohorts.....	73
2.3.2-	Phenotypes	73

2.3.3-	The sibling risk ratio (λ_s) in French Canadians	76
2.3.4-	CNV detection and selection.....	76
2.4-	Statistical analyses	76
2.4.1-	Construction of a weighted CNV score.	77
2.4.2-	Transmission analysis	78
2.5-	Results.....	78
2.6-	Discussion	80
2.7-	References.....	87
	DISCUSSION	96
	CONCLUSION ET PERSPECTIVES.....	105
	BIBLIOGRAPHIE	107
	SUPPLEMENTAL APPENDIX.....	i

Liste des tableaux

Introduction

Tableau 1- Quelques formes monogéniques d'hypertension artérielle, P15

Tableau 2- Variants génétiques associés à l'hypertension artérielle et à des phénotypes de pression artérielle dans des méta-analyses pangénomiques, P17

Chapitre I

Table 1- CNVR1 showed directionally consistent association with hypertension and T2D in French Canadians and with hypertension in diabetic ADVANCE subjects, P60

Table 2-CNVR1 is associated with quantitative traits related to T2D and its complications, P61

Table S1- GWAS for hypertension and T2D in French Canadians, P66

Chapitre II

Table1- GWAS identified six CNV regions nominally associated with hypertension in French Canadians, P91

Table 2- Percentage of Inheritance of the identified CNVRs, P92

Table 3- Associations with quantitative traits related to blood pressure, P94

Liste des figures

Introduction

Figure 1. Physiopathologie de l'hypertension artérielle, P12

Figure 2. Mécanismes de formation des CNVs, P21

Figure 3. Quelques formes de Variation dans le nombre de Copies, P23

Chapitre I

Figure 1. CNVR1 is intergenic and highly enriched in H3K27Ac histone mark, P62

Figure 2. CNVR1 influences LOC644172 expression, P63

Figure 3. Hypertension and Diabetes prevalence increases with LOC644172 transcript level, P64

Figure 4. Risk of CVD and vascular aging according to LOC644172 transcript level, P65

Figure S1. TaqMan Validation, P67

Figure S2. Frequency distribution of copy-number at the CNVR1 locus, P68

Figure S3. LOC644172 relative expression is modulated by copy-number change at the LOC644172 pseudogene locus, P69

Chapitre II

Figure 1. ROC plots for models containing each CNVR effect and the resulting CNV Risk Score, P93

Figure 2. The CNV risk score was as powerful as Framingham hypertension risk score in predicting hypertension in individuals less than 25 years of age, P95

Supplemental Appendix

Figure S1. The sibling risk ratio (λ_s) of hypertension per age group in French Canadians, Pv

Figure S2. Summary of study design, Pvi

Figure S3. Prevalence of Hypertension and Duplication-CNVRs identified in FC, Pvii

Figure S4. Prevalence of Hypertension and Deletion-CNVRs identified in FC, Pviii

Figure S5. Frequency of selected CNVRs, Pix

Figure S6. Most selected CNVRs map to human QTLs, Px

Figure S7. The CNV risk score in pedigrees, Pxi

Figure S8. Replication in Czech Post-MONICA (CPM) and ADVANCE, P_{xii}

Liste des abréviations

HTA : hypertension artérielle

PA : pression artérielle

MCV : maladies cardiovasculaires

SM : syndrome métabolique

SRAA : système rénine angiotensine

SNS : système nerveux sympathique

c-HDL : lipoprotéines de haute densité

c-HLD : lipoprotéines de basse densité

TG : triglycérides

ACV : accidents cérébrovasculaires

IMC : indice de masse corporelle

CNV : variant ou variation dans le nombre de copies

CNVR : région de variabilité dans le nombre de copies

QTL : loci de caractères quantitatifs

GWAS : genome wide association study, étude pangénomique

*À Mamie Popo,
À l'énigme du renoncement...*

Avant-propos

Chapitre I : J'ai effectué les tests d'associations et suis également responsable de l'interprétation des résultats, de la rédaction du manuscrit et de l'élaboration des figures.

Chapitre II : J'ai effectué les tests d'associations dans la population du Saguenay-Lac-St-Jean et dans la population Tchèque Post-MONICA. J'ai construit le score de risque pondéré, les courbes ROC, calculé le pourcentage de transmission, interprété les résultats et rédigé le manuscrit. J'ai également réalisé toutes les figures et tableaux, à l'exception de la figure S1 (Supplemental Appendix).

Remerciements

Merci à mon directeur de thèse, Dr Pavel Hamet et aussi à Dre Johanne Tremblay. Merci également à Dr Ondřej Šeda qui m'a accompagné au début de cette aventure et à Lucie Šedova. Merci au Dre Anne-Marie Antchouey Ambourouet. Un énorme merci à tous ces gens formidables que j'ai côtoyés durant toutes ces années et particulièrement à Pierre Dumas et à l'efficace Marie-Noël Nadeau. Merci à ceux dont le passage a été court mais marquant, Carole Mekoudjou et mon stagiaire inoubliable, Éric Migeon. Merci à tous ceux dont le travail a rendu ce projet possible et particulièrement à François Harvey, John Raelson, F-C Marois-Blanchet, Carole Long, Gilles Corbeil, Mounsiif Haloui, Youssef Idaghdour et Gilles Godefroid. Merci à Marie-Pierre Sylvestre. Merci à Suzanne Cossette, Alexandru Gurau, Pierre-Luc Brunelle, Majid Nikpay et Rana El Bikaï... Aux anciens de Biogenix et aux anciens et nouveaux de Prognomix, au aki du midi ... Un merci tout particulier à Audrey Noël et Johanna Sandoval dont les départs ont laissés un gros vide.

Merci à mes parents qui m'ont permis d'arriver jusque-là. Merci à Sozè et Magali.

Merci à Annie, Nabbie et Sandra dont la présence a été cruciale dans les moments difficiles. Merci à Paola et Adeline qui ont été présentes et apporté leur aide, naturellement, avant même que je les sollicite. Merci à Neiss. Merci à Lino, à son humour, c'est bientôt ton tour. Merci à Abess, Al et Marie, si loin mais si proches depuis toujours. Merci à votre soutien sans faille. Merci à LD mon Superman.

INTRODUCTION

I-L'hypertension artérielle

L'hypertension artérielle est une condition chronique survenant lorsque la pression sanguine demeure haute. Elle est communément définie comme une pression artérielle systolique supérieure ou égale à 140 mm Hg et/ou une pression artérielle diastolique supérieure ou égale à 90 mm Hg [1]. L'hypertension essentielle est une pathologie idiopathique complexe, multifactorielle et héréditaire. Ainsi, de nombreux facteurs génétiques et environnementaux sont impliqués dans l'étiologie de l'hypertension essentielle qui représente environ 90% des cas d'hypertension chez l'adulte. L'hypertension secondaire représente environ 10% des cas d'hypertension artérielle et paraît découler de conditions clairement identifiées. L'incidence de l'hypertension artérielle augmente dans le monde entier où cette pathologie constitue un important problème de santé publique. En effet, l'hypertension artérielle est un facteur de risque majeur pour les maladies cardiovasculaires. De plus, son association fréquente avec des anomalies métaboliques accroît considérablement le développement des pathologies cardiovasculaires.

1.1- Physiopathologie de l'hypertension artérielle

1.1.1- Des déterminants biologiques de base

La pression sanguine ou tension artérielle représente la pression qu'exerce le sang sur la paroi des artères. Elle dépend du débit cardiaque, du volume sanguin et de la contractilité des petites artères et des artérioles [2]. Elle est régulée par de nombreux facteurs d'ordre

hémodynamique, neuroendocrine, cellulaire, moléculaire ou relatifs à la structure vasculaire. Des anomalies dans ces différents facteurs entraînent l'apparition de l'hypertension artérielle (HTA) essentielle. D'un point de vue hémodynamique, l'HTA peut résulter d'une augmentation du débit cardiaque ou de résistances périphériques dues à des agents vasoconstricteurs. De plus, l'hyperplasie des cellules du muscle lisse, constituerait l'une des causes principales de la résistance vasculaire caractéristique de nombreux cas d'HTA [3-5]. En effet, la rigidité artérielle augmente avec l'hypertension et chez le patient athéromateux [6].

En ce qui à trait au rein, il joue un rôle central dans la relation pression artérielle-natriurèse. L'apparition de l'hypertension est associée à une dérégulation de ce système et à un déficit de l'excrétion sodée. D'une part, l'activation du système rénine-angiotensine-aldostérone (SRAA) entraîne l'augmentation de la pression artérielle via la rétention sodée et la vasoconstriction des artères rénales [7]. Ainsi, chez l'humain, une sténose de l'artère rénale induit une HTA rénovasculaire, s'accompagnant d'une hypersécrétion de rénine. L'hyperaldostéronisme primaire (syndrome de Conn) et secondaire (associée à la rétention sodée) s'accompagnent également d'hypertension artérielle. La formation d'angiotensine II est assurée par l'Enzyme de Conversion de l'Angiotensine (ECA), dont les inhibiteurs (IEC) constituent les médicaments les plus utilisés pour contrôler l'hypertension. De plus, les récepteurs AT1, principaux médiateurs de l'action de l'angiotensine II sur la pression artérielle, sont les cibles de médicaments antihypertenseurs, les antagonistes des récepteurs de l'angiotensine II (ARA2).

D'autre part, le système nerveux sympathique (SNS) agit sur les vaisseaux directement à travers les récepteurs alpha et indirectement via le système rénine-angiotensine-aldostérone

[8]. Les prostaglandines, produites par le tissu-cible et les récepteurs alpha et bêta-adrénergiques pré-sympathiques, modulent la libération des catécholamines (noradrénaline et adrénaline) stockées dans la médullo-surrénale et les terminaisons synaptiques. L'hypertension artérielle peut résulter d'une hyperactivité du SNS et/ou d'une hypersécrétion des catécholamines (ex. phéochromocytome). En effet, les bêtabloquants sont des médicaments hypotenseurs qui obstruent ou inhibent l'action des médiateurs du système adrénergique tels que l'adrénaline. De plus, les antihypertenseurs centraux diminuent le tonus sympathique vasoconstricteur.

L'ensemble des facteurs décrits plus haut sont inter-reliés et l'hypertension artérielle hyperkinétique du jeune, caractérisée par une élévation du débit cardiaque, constitue une illustration de l'interrelation système nerveux sympathique - système rénine-angiotensine-aldostérone. Ce syndrome s'accompagne également de résistances vasculaires systémiques.

1.1.2- Des facteurs de risque

Le sexe, l'ethnicité, la sensibilité au sodium, la sédentarité, la consommation d'alcool et de tabac, ainsi que l'âge influencent aussi la pression artérielle [1].

1.1.2a- Sexe et ethnicité

Ainsi, chez les Caucasiens, les femmes pré-ménopausées sont communément moins hypertendues que les hommes d'âge similaire [9]. Ce qui suggère un rôle protecteur des œstrogènes ou témoigne d'une exposition moindre aux facteurs de risque. Toutefois, les femmes de descendance africaine semblent développer l'hypertension artérielle plus précocement et ont tendance à être plus hypertendues que les hommes [10][11]. En général,

les individus de descendance africaine sont plus hypertendus que la plupart des autres groupes ethniques. Par exemple, aux États-Unis, la prévalence de l'hypertension chez les Africains-Américains varie entre 43-44%, alors qu'elle est de 30% chez les Européens-Américains et de 26% chez les Mexicains-Américains [12]. Dans la population de descendance africaine, la prévalence accrue se caractérise principalement par des degrés plus importants d'hypertrophie du ventricule gauche [13], et par une tendance plus réduite de la pression à décroître durant le sommeil [14]. Dans cette population, les dommages rénaux sont plus importants (haute prévalence d'insuffisance rénale au stade final [15]).

1.1.2b- Sensibilité au sodium et sédentarité

Le sodium induit une augmentation de la résistance vasculaire et de l'hypertension artérielle (HTA). La sensibilité au sodium résulterait d'un déséquilibre entre les mécanismes hormonaux [15], neuronaux [16], hémodynamiques [17] et/ou génétiques [18] changeant l'équilibre sodé à travers des modifications dans la filtration glomérulaire ou dans la réabsorption tubulaire. Elle apparaît plus importante chez les hypertendus Africains-Américains (72%) que chez les hypertendus Européens-Américains (56%) [19]. D'autre part, la sédentarité ou l'inactivité physique favorise la prise de poids, l'HTA et les troubles des métabolismes glucidique et lipidique. La pratique d'exercice physique régulier (30-45 minutes au moins 4 jours par semaine) protégerait contre le développement de pathologies cardiovasculaires [20]. En effet, l'activité physique induit la baisse de pression artérielle, améliore l'équilibre glycémique, favorise l'élévation du bon cholestérol (c-HDL) et diminue la résistance à l'insuline [21].

1.1.2c- Alcool et tabagisme

Une consommation modérée et régulière d'alcool paraît avoir un effet protecteur contre l'hypertension et ses complications cardiovasculaires associées [23], du en partie à une action anti-inflammatoire [24]. Cependant, une consommation excessive d'alcool (>30 ml/jour) [25] augmente la pression artérielle et les complications cardiovasculaires liées à l'hypertension. Les larges quantités d'alcool (plus de 2 portions par jour) induiraient une augmentation du débit cardiaque et du volume sanguin, possiblement à travers une stimulation du système nerveux sympathique [26]. De plus, l'alcool altérerait les membranes cellulaires, permettant l'entrée du calcium vraisemblablement via l'inhibition du transport de sodium [27]. D'autre part, la nicotine augmente la pression artérielle à travers une stimulation de la libération de norépinéphrine, au niveau des terminaisons nerveuses adrénérgiques. De plus, le tabac induit une réduction accrue de la compliance de l'artère radiale, indépendamment de l'augmentation de la pression artérielle [28]. Le tabac majore également le risque de thrombose, renforce l'ischémie myocardique et potentialise les troubles du risque cardiaque. De plus, Nikpay *et al.* [29] ont souligné le rôle central de la plasticité synaptique dans la convergence substantielle entre les déterminants génétiques de la consommation de substances psychoactives (tabac, alcool et café), du stress, de l'obésité et des traits hémodynamiques (pression artérielle et fréquence cardiaque).

1.1.2d- L'âge

La prévalence de l'HTA augmente avec l'âge et 60-70% des sujets âgés de plus de 70 ans sont atteints d'hypertension artérielle. L'âge pourrait refléter l'exposition prolongée aux autres facteurs de risque. En effet, Hamet *et al.* ont montré que l'hypertension accélère

la sénescence des cellules cardiovasculaires [30]. De plus, la rigidité artérielle, qui augmente avec l'âge, est considérée comme un marqueur du vieillissement vasculaire [31]. Plus précisément, la pression artérielle systolique augmente avec l'âge jusqu'à la 8^{ème} décennie de vie. Tandis que la pression diastolique augmente jusqu'à la 5^{ème} décennie, pour ensuite demeurer constante, avant de légèrement décroître.

1.2- L'hypertension et les pathologies cardiovasculaires

1.2.1- Épidémiologie de l'hypertension artérielle

L'incidence de l'hypertension artérielle (HTA) augmente aussi bien dans les populations occidentales, que dans celles des pays en transition économique, avec une prévalence variant de 20 à 50% de la population adulte [32]. On estime à 1 milliard, la population mondiale atteinte d'HTA [33]. Dans les pays en transition économique, l'augmentation de la prévalence d'HTA correspond essentiellement à l'urbanisation et à l'occidentalisation des modes de vie, notamment à des changements dans l'alimentation. Par exemple, en 1997, une étude menée au Cameroun a mesuré une prévalence de l'HTA de 15,4% en zone rurale et de 19,1% en zone urbaine [34]. Généralement, la prévalence de l'HTA coïncide avec l'industrialisation. En effet, si la prévalence de l'HTA au Japon est semblable à celle mesurée chez les Japonais-Américains (~45%) [35], en Chine, la prévalence est de 27%, alors que 39% des Chinois-Américains sont hypertendus [36]. En occident, l'augmentation de la prévalence de l'hypertension artérielle paraît liée au vieillissement de la population et principalement à une augmentation de la prévalence de l'obésité. Par exemple, un tiers de la population américaine est hypertendue, et cette

prévalence est 6 fois plus élevée qu'au début du 20^e siècle [37]. Ce phénomène correspond à une augmentation dramatique de la prévalence de l'obésité, survenue entre 1900 (3% à 5%) et nos jours (30%) [38].

1.2.2- L'hypertension artérielle : un facteur de risque clef pour les maladies cardiovasculaires

L'hypertension artérielle est un facteur de risque majeur pour les maladies cardiovasculaires (MCV) qui étaient responsables en 2010, selon l'Organisation Mondiale de la Santé (OMS), de 29% de la mortalité mondiale totale. Les pathologies cardiovasculaires constituent ainsi la 1^{ère} cause de décès dans le monde, principalement due aux infarctus du myocarde et aux accidents cérébrovasculaires (ACV). Au niveau mondial, près de 2% des décès par MCV sont liés au rhumatisme articulaire aigu, 34% sont liés aux maladies cérébrovasculaires et 42% aux cardiopathies ischémiques [39].

1.2.2a- L'étude Framingham

De larges études de cohorte mis en place après la seconde guerre mondiale ont permis de mettre en évidence les principaux facteurs de risque des maladies cardiovasculaires, l'étude la plus importante étant l'étude de Framingham [40]. Cette étude a permis d'identifier les facteurs de risque majeurs des maladies cardiovasculaires : une pression artérielle élevée, une cholestérolémie élevée, la consommation de tabac, l'obésité, le diabète et l'inactivité physique. Framingham a également permis d'évaluer les effets

d'autres facteurs tels que l'âge, le sexe, les aspects psychologiques, et les niveaux sanguins des triglycérides (TG) et des lipoprotéines de haute densité (c-HDL).

1.2.2b- le fardeau d'une pression artérielle élevée

Récemment, des études ont permis de quantifier la contribution de la tension artérielle dans le développement des maladies cardiovasculaires à travers le monde [41]. Ainsi, la mesure « Disability Adjusted Life Year » (DALY) [42] permet d'évaluer le fardeau d'une pression artérielle élevée en intégrant les années de vie perdues (mortalité précoce), les années de travail perdues et la qualité de vie (morbidité). Globalement, en 2001, la pression artérielle élevée était impliquée dans 13.5% de la mortalité mondiale et 6% des cas de mortalité précoce et de morbidité dans le monde, avec un effet majeur sur les pathologies cardiaques ischémiques et cérébrovasculaires. En Asie de l'Est, dans le Pacifique et en Afrique Subsaharienne la pression artérielle élevée a un effet majeur sur les accidents cérébrovasculaires (ACV). Alors que dans les autres régions du monde, la tension artérielle élevée a un effet plus important sur les pathologies cardiaques ischémiques. 80% du fardeau attribuable à la tension artérielle élevée se retrouve dans les pays à faible et moyen revenus, avec une proportion plus grande d'individus plus jeunes affectés comparativement aux pays à revenu élevé.

1.3- L'association de l'hypertension avec des anomalies métaboliques renforce le risque cardiovasculaire

L'hypertension artérielle (HTA) apparaît souvent associée à une ou plusieurs anomalies métaboliques incluant la dyslipidémie, l'obésité, le diabète de type 2 ou le

syndrome métabolique. L'association de l'hypertension avec ces conditions accroît considérablement le développement des maladies cardiovasculaires. Plus particulièrement, l'insulino-résistance serait mise en cause dans ces associations.

1.3.1- La résistance à l'insuline

Le tissu adipeux viscéral des sujets obèses libère une grande quantité d'acides gras libres (FFA) qui activent des mécanismes hépatiques, musculaires et pancréatiques menant à une augmentation de la glycémie. L'insulino-résistance s'accompagne d'un phénomène de compensation au cours duquel les cellules bêta-pancréatiques augmentent la sécrétion d'insuline, afin de maintenir une glycémie normale. Ce phénomène se traduit également par une augmentation de la masse des cellules bêta-pancréatiques [43]. Au fil du temps, l'hyperinsulinisme conduit à des déficiences fonctionnelles des cellules bêta-pancréatiques qui s'épuisent. À terme, l'hyperglycémie du diabète de type 2 résulte de l'insulino-déficience consécutive aux dysfonctions cellulaires bêta-pancréatiques. D'autre part, l'insulino-résistance serait pré-existante à l'hypertension artérielle. En effet, elle affecte la sensibilité des récepteurs à l'insuline, mais également des voies de signalisation qui influencent divers organes dont le rein [44].

1.3.2- Hypertension et dyslipidémies

Les dyslipidémies désignent l'ensemble des troubles du métabolisme des lipides, plus particulièrement une concentration anormalement élevée dans le sang du cholestérol (hypercholestérolémie) et/ou des triglycérides (hypertriglycéridémie). L'hypertriglycéridémie et l'hypercholestérolémie augmentent le développement des maladies cardiovasculaires [45,46]. Plus particulièrement, la prévalence de pathologies

cardiovasculaires est plus que doublée chez les hypertendus dyslipidémiques. On observe également une augmentation de la prévalence d'accidents cérébrovasculaires dans cette catégorie d'individus [47]. Les lipoprotéines de haute densité (c-HDL) qui transportent le cholestérol de la périphérie vers le foie, jouent un rôle protecteur contre les maladies cardiovasculaires. Alors que les lipoprotéines de basse densité (c-LDL) transportent le cholestérol du foie vers les cellules périphériques, et correspondent à la fraction athérogène du cholestérol qui est associée à une augmentation des pathologies cardiovasculaires.

1.3.3- Hypertension et obésité

Un gain dans le poids corporel est associé à une augmentation de la tension artérielle [48]. On distingue l'obésité (Indice de Masse Corporelle, $IMC \geq 30$) de type androïde avec une prédominance de la graisse dans la zone abdominale et l'obésité de type gynoïde caractérisée par une prédominance des graisses au niveau des hanches et des cuisses. C'est l'obésité androïde ou viscérale qui est plus particulièrement associée à l'hypertension artérielle (HTA) [49]. L'obésité est associée à une multitude de mécanismes pro-hypertensifs [50]. En effet, les mécanismes impliqués dans le développement de l'HTA d'une part, et de l'HTA associée à l'obésité d'autre part, seraient distincts [51].

1.3.4- Hypertension et diabète de type 2

Comme l'obésité, le diabète de type 2 est considéré comme l'une des épidémies du XXIème siècle. En effet, cette pathologie est en progression fulgurante à l'échelle planétaire. Le diabète se définit par une hyperglycémie, c.-à-d. glycémie à jeun $\geq 1,26$ g/l (7 mmol/l) à au moins deux reprises et/ou une glycémie ≥ 2 g/l (11,1 mmol/l), 2 heures après une charge orale de 75g de glucose. Le diabète multiplie les risques de claudication

intermittente et d'amputation, et constitue un problème majeur de santé publique [52]. Chez les patients diabétiques, l'hypertension réfère à une pression systolique supérieure à 130 mmHg et une pression diastolique supérieure à 80 mmHg (en dépit de la médication). La comorbidité de l'hypertension et du diabète de type 2 aggrave la rigidité artérielle (via des dysfonctions endothéliales) [53], et conséquemment les risques de maladies rénales et cardiovasculaires. En effet, l'hyperglycémie favorise l'athérogenèse et la thrombose [54]. En 2006/7, plus de 5,1% des Canadiens de plus de 20 ans étaient atteints à la fois de diabète de type 2 et d'hypertension [55].

1.3.5- Hypertension et syndrome métabolique

Dans sa définition, le syndrome métabolique (SM) regroupe plusieurs anomalies incluant l'hypertension artérielle (HTA), l'obésité abdominale, l'hypertriglycémie, un taux de c-HDL bas, une intolérance au glucose ou un diabète de type 2. Ce syndrome, dont la prévalence augmente avec l'âge, constitue un facteur de risque pour le diabète de type 2 et les maladies cardiovasculaires [56]. Aux États-Unis, le SM touche 24% des individus en générale et plus de 40% des individus de plus de 60 ans [57]. La création de cette entité visait le regroupement d'anomalies glucido-lipidiques associées à l'insulino-résistance, à l'HTA et à l'obésité abdominale. Si cette mesure a été reconnue par l'OMS en 1998, puis par les instances américaines en 2001, elle ne fait toujours pas l'unanimité dans la communauté scientifique. Ainsi, les Européens ne reconnaissent pas ce syndrome comme une entité pathologique. D'autre part, plusieurs définitions du syndrome métabolique ont été proposées telles que la définition de l'OMS [58] ou la définition de la fédération

internationale de diabète IDF [59] qui diffèrent par la présence ou l'absence de la résistance à l'insuline dans leurs définitions.

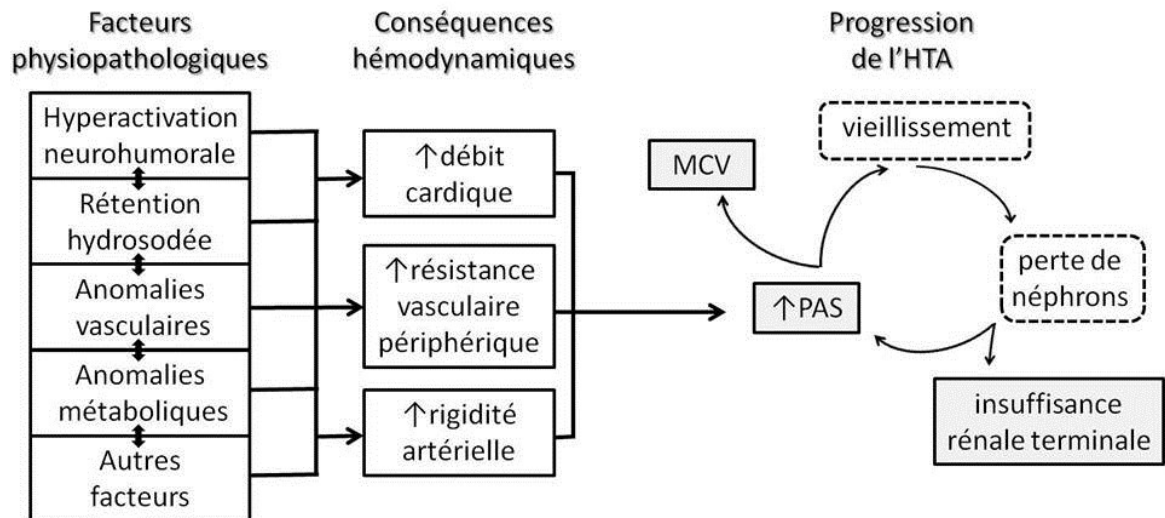


Figure 1 : Physiopathologie de l'hypertension artérielle. HTA : hypertension artérielle; PAS : pression artérielle systolique; MCV : maladies cardiovasculaires. Figure adaptée de Baudoin B et al., 2009 *Revue Francophone des Laboratoires* 409 : 67.

1.4- Génétique de l'hypertension artérielle

L'agrégation familiale et le taux de concordance élevé chez les jumeaux monozygotes témoignent de la composante génétique de l'hypertension artérielle (HTA). Dans la population, 30% à 70% de la variation de la pression artérielle (PA) serait attribuée à des facteurs génétiques [60]. Il existe des formes rares monogéniques de l'HTA. Toutefois, l'HTA apparaît principalement comme une pathologie polygénique complexe, dans laquelle des facteurs environnementaux sont impliqués. Récemment, de larges études, à l'échelle du génome entier, ont permis d'identifier plusieurs loci associés à la variation de la PA. Cependant, ces résultats apparaissent souvent difficilement reproductibles et l'impact de chacun des loci identifiés, sur la variance de la PA, demeure relativement modeste.

1.4.1- Évidences d'un déterminisme génétique de l'hypertension artérielle

Dans toutes les populations, on observe une agrégation familiale de l'hypertension artérielle où des enfants sont plus souvent hypertendus lorsque l'un ou les deux parents sont hypertendus. Par exemple, une étude menée sur 277 familles nucléaires a montré que 3% des enfants étaient hypertendus avec deux parents normotendus, 28% des enfants étaient hypertendus avec l'un des parents hypertendu et 45% des enfants étaient hypertendus avec les deux parents hypertendus [61]. D'autre part, la preuve et la quantification d'un effet génétique sur la pression artérielle (PA) ont été démontrées dans des études comparant des corrélations de PA entre des jumeaux monozygotes et des jumeaux dizygotes [62], ainsi qu'entre des enfants adoptés, naturels et leurs parents [63]. Ces différentes études ont

permis de souligner l'importance des facteurs génétiques et de leurs interactions avec des facteurs environnementaux dans l'étiologie de l'hypertension.

1.4.2- Les formes monogéniques d'hypertension artérielle

Des formes monogéniques rares d'hypertension artérielle (HTA) héréditaire ont été décrites (tableau 1). En effet, ces formes monogéniques d'HTA se caractérisent par des mutations à forte pénétrance qui affectent principalement des gènes impliqués dans la régulation de l'équilibre hydrosodé. Aussi, des mutations dans plusieurs gènes incluant *PCSK9*, *LDLR* et *APOB* sont à l'origine d'HTA secondaire à l'hypercholestérolémie [64-67]. L'étude de ces pathologies a permis une meilleure compréhension des mécanismes impliqués dans la régulation de la pression artérielle [68].

Hypertension artérielle	Mode de transmission	Gène	Mutations et conséquences fonctionnelles	Phénotype
Hyperaldostéronisme de type 1 (glucocorticoid remedial aldosteronism)	Autosomique dominante	Fusion de <i>CYP11b1</i> et <i>CYP11b2</i>	Gène chimérique sous contrôle d'ACTH	Hypertension, hypokaliémie, hyperaldostéronisme, PRA ↓, 18-hydroxycortisol ↑
Hyperaldostéronisme de type 2	Autosomique dominante	Locus chromosome 7p22	Surproduction d'aldostérone dans glandes surrénaliennes	Hypertension, hypokaliémie, hyperaldostéronisme, PRA ↓, hyperplasie/adénome glande surrénaliennes
Syndrome de Liddle	Autosomique dominante	<i>SCNN1B</i> <i>SCNN1G</i>	Activation constitutive d'ENaC (canal sodique épithéliale dans le tube distal/collecteur)	Hypertension, hypokaliémie, hypoaldostéronisme, PRA ↓
Hyperplasie congénitale surrénalienne	Autosomique récessive	<i>CYP11B1</i>	Déficit de l'enzyme 11β hydroxylase	Hypertension, hypokaliémie, hypoaldostéronisme, PRA ↓, déoxycortisone ↑
Déficit d'11 β-OH stéroïd déhydrogénase type 2 (apparent mineralocorticoid excess)	Autosomique récessive	<i>HSD11B1</i>	Déficit de désactivation de cortisol	Hypertension, hypokaliémie, hypoaldostéronisme, PRA ↓
Pseudohypoaldostéronisme type II (syndrome de Gordon)	Autosomique dominante	<i>WNK1</i> <i>WNK4</i>	Activation constitutive du cotransporteur Na/Cl dans le tube distale	Hypertension, hyperkaliémie, hypoaldostéronisme, PRA ↓, acidose métabolique
Mutations de récepteur PPAR-γ	Autosomique dominante	<i>PPARG</i>	Loss-of-function mutation du récepteur	Hypertension, résistance à l' diabète
Syndrome d'hypertension artérielle, hypercholestérolémie, hypomagnésémie	Mitochondriale	Non identifié	Transmission maternelle d'une mutation causant une substitution de cytidine dans les ARNt mitochondriaux	Hypertension, hypercholestérolémie, hypomagnésémie

Tableau 1. Quelques formes monogéniques d'hypertension artérielle. ACTH : hormone adrénocorticotrope ; PRA : activité de la rénine plasmatique ; ENaC : canal sodique épithélial. Tableau extrait de Rev Med Suisse 2009;5:1763-1770.

1.4.3- Recherche pangénomique de loci reliés à l'hypertension artérielle essentielle

L'hypertension artérielle (HTA) essentielle résulte de l'interaction complexe entre plusieurs variants génétiques et des facteurs environnementaux. Actuellement, deux approches sont essentiellement utilisées afin d'identifier les différents variants génétiques associés à l'HTA, à l'échelle du génome entier. Ces études pangénomiques ne nécessitent

pas d'hypothèse mécanistique *a priori* et facilitent l'identification de gènes insoupçonnés, au regard des connaissances sur la physiopathologie.

La liaison génétique désigne le fait que deux facteurs génétiques, situés à deux loci distincts, ont tendance à être transmis ensemble d'un individu à sa descendance. L'analyse de liaison étudie la fréquence avec laquelle des marqueurs génétiques de certaines régions du génome sont transmis ensemble et avec la maladie, à la prochaine génération [69]. Parmi les loci de caractères quantitatifs (QTLs) liés à la pression artérielle et à l'hypertension détectés, environ 34 ont présenté un score LOD¹, logarithme des probabilités (logarithm of odds), supérieur à 3. Pourtant, seul un locus (QTL en 2q34 lié à la pression artérielle diastolique) a pu être effectivement répliqué dans des familles indépendantes [60].

L'analyse d'association utilise des approches épidémiologiques classiques (études cas/contrôles, études transversales etc.) afin de déterminer si une maladie/trait est associé(e) à un facteur génétique, chez des familles ou des individus non apparentés. Toutefois, les premières études d'association pangénomique ont identifié des loci, associés à la pression artérielle ou à l'hypertension, qui n'atteignaient pas le seuil de significativité à l'échelle du génome entier (communément $P=5\times 10^{-8}$). Par exemple, dans l'étude menée par le Wellcome Trust Case Control Consortium (WTCCC) aucun locus en association avec l'HTA n'atteignait un $P<5\times 10^{-7}$ [70].

¹ À partir des probabilités de transmission des allèles de deux gènes observées dans la descendance, il est possible d'établir une valeur numérique qui quantifie le degré de liaison génétique, c'est le *LOD score*. Une liaison est déclarée significative lorsque le LOD score, ou le cumul des scores obtenus pour différentes familles, est supérieur ou égal à 3.

Région chromosomique	SNP principal	Position	Gène	Pertinence fonctionnelle	Phénotype associé	Consortium
1p36	rs17367504	11785365	MTHFR*	Intron	PAS,PAD, HTA,PAM	Global BPgen, AGEN,ICBP ^{2,3}
4q21	rs16998073	81403365	FGF5	Intergénique	PAS,PAD, HTA,PAM	Global BPgen, AGEN,ICBP ³
5p13	rs1173771	32850785	NPR3	Intergénique	PAS,PAD,HTA, PAM,PP	ICBP ^{2,3}
6p22	rs1799945	26199158	HFE*	Mutation faux-sens	PAM,PAS, PAD,HTA	ICBP ^{2,3}
6p21	rs805303	31724345	BAG6*	Intron	PAS,PAD, HTA	ICBP ²
7q36	rs3918226	150321109	NOS3	Promoteur à proximité	HTA	HYPERGENES
10p12	rs1813353	18747454	CACNB2*	Intron	PAM,PAS, PAD,HTA	ICBP ^{2,3}
10q21	rs4590817	63137559	C10orf107*	Intron	PAM,PAS, PAD,HTA	ICBP ^{2,3}
10q23	rs932764	95885930	PLCE1*	Intron	PAS,HTA	ICBP ^{2,3}
10q24	rs11191548	104836168	NT5C2	Intergénique	PP,PAM,PAS, PAD,HTA	Global BPgen, AGEN,ICBP ^{2,3}
11q22	rs633185	100098748	ARHGAP42*	Intron	PAM,PAS, PAD,HTA	ICBP ^{2,3}
12q21	rs2681472	88533090	ATP2B1*	Intron	PP,PAM,PAS, PAD,HTA	CHARGE, AGEN,ICBP ³
12q21	rs11105354	88550654	ATP2B1*	Intron	PAS,HTA	CHARGE
12q21	rs12579302	88574634	ATP2B1	Promoteur à proximité	PAS,HTA	CHARGE
12q21	rs11105364	88593407	ATP2B1	Intergénique	PAS,HTA	CHARGE
12q21	rs11105368	88598572	ATP2B1	Intergénique	PAS,HTA	CHARGE
12q21	rs11105378	88614872	ATP2B1	Intergénique	PAS,HTA	CHARGE
12q21	rs12230074	88614998	ATP2B1	Intergénique	PAS,HTA	CHARGE
15q24	rs1378942	72864420	CSK*	Intron	PAM,PAS, PAD,HTA	Global BPgen, AGEN,ICBP ^{2,3}
20q13	rs6015450	57184512	ZNF831	Intergénique	PAM,PAS, PAD,HTA	ICBP ^{2,3}

Tableau 2. Variants génétiques associés à l'hypertension artérielle et à des phénotypes de pression artérielle dans des méta-analyses pangénomiques. AGEN=Asian Genetic Epidemiology Network; CHARGE=Cohorts for Hearts and Aging Research in Genomic Epidemiology; Global BPgen=Global Blood Pressure Genetics; ICBP=International Consortium of Blood Pressure; HTA=hypertension artérielle; PAS=pression artérielle systolique ; PAD=pression artérielle diastolique; PAM=pression artérielle moyenne; PP=pression pulsée; *Variant localisé dans le gène ; ICBP², Ehret et al., 2011; ICBP³ Wain et al., 2011. Tableau adapté de Zhao et al., 2013.

Récemment, l'amélioration des techniques de génotypage et la formation de méga-consortiums ont permis l'identification effective de nouveaux loci associés à l'HTA et atteignant le seuil de significativité fixé, $P=5 \times 10^{-8}$ (Tableau 2) [60]. Ces consortiums comprenaient les échantillons de dizaines de milliers d'individus provenant du monde entier, ce qui a considérablement amélioré la puissance de détection de nouveaux loci.

Cependant, les variants génétiques communs identifiés ont un impact relativement modeste sur la variance de la pression artérielle ($<1\text{mmHg}$). Ainsi, l'ensemble de ces variants n'expliquerait que 0,9% de la variance de la pression artérielle dans la population générale [71,72], laissant une importante portion de la composante génétique de l'hypertension inexplicée. La prise en compte des interactions entre gènes ou des gènes avec des facteurs environnementaux pourrait faciliter la détection de variants génétiques avec des effets phénotypiques relativement plus importants. De plus, la plupart des études pangénomiques sont basées sur les polymorphismes nucléotidiques (SNPs), alors que d'autres variants génétiques, tels que la variabilité dans le nombre de copie (CNV), ont potentiellement un plus grand impact sur la variance de la pression artérielle.

II-La variabilité dans le nombre de copies

2.1- Les Polymorphismes génétiques

On a pu assister, ces 25 dernières années, à la découverte de nombreux polymorphismes génétiques ayant permis de définir les bases génétiques de nombreux traits ou affections mendéliennes humaines. Ainsi, le génome humain présente différents types de variations incluant les polymorphismes de taille des fragments de restriction [73], les minisatellites (ou les séquences en nombre variable répétées en tandem), les microsatellites (ou séquences courtes répétées en tandem, STRs) [74] qui ont été utilisés pour créer des cartes génétiques de l'ensemble des chromosomes humains et les polymorphismes nucléotidiques simples (« Single Nucleotide Polymorphisms » ou SNPs) [75,76]. Le SNP désigne la variation d'une seule paire de bases et constitue une source majeure de variabilité génétique et phénotypique interindividuelle. En effet, le SNP est la variation génétique la plus commune avec au moins 10 millions de SNPs recensés à travers le génome humain, les allèles² SNP les plus rares (mineurs) ayant une fréquence d'au moins 1%. En 2005, le projet HapMap a permis de génotyper des millions de SNPs à l'échelle du génome entier. Aujourd'hui, le SNP est la variation génétique la plus étudiée et essentiellement utilisée comme marqueur pour les pathologies.

² Dans la plupart des cas un SNP est présent sous deux formes : les allèles. Mais il se peut qu'un SNP existe sous plus de deux formes alléliques.

2.2- La variation dans le nombre de copies

2.2.1- Définition

Les premières larges variations génomiques structurales identifiées (duplications ou délétions) étaient associées à des complications cliniques majeures. En effet, les premiers cas de trisomie 21 [77], syndromes de Turner [78] et Klinefelter [79] ont été rapportés dès 1959. En 2004, les avancées technologiques et le décryptage du génome humain ont permis d'observer que les variations génomiques structurales sont largement répandues à travers le génome humain et qu'elles ne sont pas uniquement liées à des pathologies, mais constituent aussi une source de variabilité phénotypique considérable [80-82]. Ainsi, la variation dans le nombre de copie (CNV ; Copy Number Variant) a initialement été définie comme un segment d'au moins 1 kilobase d'ADN présent en un nombre variable de copies comparativement à un génome de référence [83]. Actuellement, l'amélioration de la résolution des méthodes de détection permet la détermination d'un seuil de 50 paires de base (pb) pour distinguer les indels des CNVs. En conséquence, un CNV peut également être défini comme une séquence d'ADN d'au moins 50pb présente en un nombre variable de copies dans le génome [84]. Plus particulièrement, les CNVs communs sont des CNVs que l'on retrouve dans au moins 1% de la population générale. Les CNVs communs, essentiellement ceux ayant une fréquence populationnelle d'au moins 5%, jouent un rôle important dans la variabilité génétique interindividuelle [85].

Les CNV sont catalogués dans des bases de données incluant la Database of Genomic Variants (DGV ; <http://dgvbeta.tcag.ca/dgv/app/>), la European Cytogeneticists Association Register of Unbalanced Chromosome Aberrations (ECARUCA ; www.ecaruca.net) ou la Database of Chromosomal Imbalance and Phenotype in Human using Ensembl Resources

(DECIPHER ; <https://decipher.sanger.ac.uk/>). En septembre 2013, la DGV contenait environ 109863 régions de CNV recensées à travers le génome humain.

On distingue différents types de CNVs : des duplications, délétions, insertions, inversions, ou réarrangements à des sites multiples [84,86]. Les CNVs varient également en taille et en complexité, ce qui suggère des différences dans les mécanismes d'origine [87].

2.2.2- Mécanismes de formation et différentes classes de CNV

Quatre mécanismes principaux génèrent des réarrangements dans le génome et sont très probablement à l'origine de la plupart des CNVs : la recombinaison homologue non-allélique (Non-Allelic Homologous Recombination, NAHR), la ligature d'extrémités non-homologues (Non-Homologous End-Joining, NHEJ), l'interruption de la fourche de réplication et commutation de cible (Fork Stalling and Template Switching, FoSTeS) et la rétrotransposition [88-90].

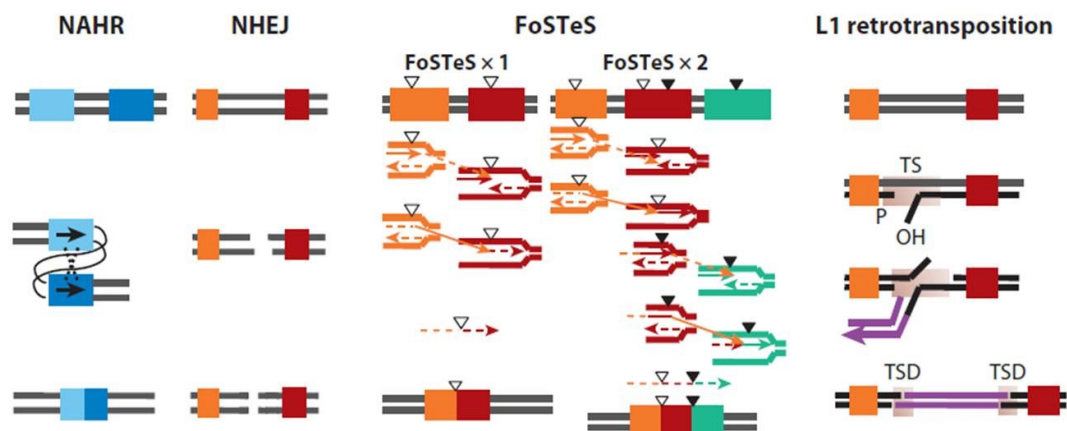


Figure 2. Mécanismes de formation des CNVs (tirée de Zhang F et al., 2009)

NAHR implique un échange (« crossing-over ») entre des séquences homologues non alléliques en orientation directe qui peut entraîner une délétion, duplication ou inversion du segment intermédiaire. La présence de séquences génétiques hautement similaires favoriserait ce type de recombinaison (ex. LCRs, SDs, séquences Alu, pseudogènes et rétrotransposons L1). NHEJ est le mécanisme principal de réparation des cassures doubles brins qui implique une soudure des extrémités d'ADN cassées et qui peut conduire à la formation de translocations chromosomiques, de délétions ou de duplications. FoSTeS désigne l'interruption d'une fourche de réplication initiale et le désengagement de ses brins qui vont migrer vers une autre fourche de réplication afin de poursuivre la synthèse d'ADN (via des microhomologies). FoSTeS résulte en des réarrangements plus ou moins complexes de l'ADN. La rétrotransposition désigne la mobilité des rétrotransposons actifs, particulièrement L1 (ou Long Interspersed Element-1, LINE1), qui induisent des mutations insertionnelles et des instabilités génétiques, et Kidd *et al.*, [91] estiment qu'elle serait à l'origine de 15% des variations structurelles. Les CNVs les plus récurrents sont simples et stables en taille (délimitations stables des segments sur le génome) et seraient essentiellement formés par NAHR. Par contre, les autres mécanismes seraient à l'origine de CNVs très diverses avec des structures complexes et des tailles variables (délimitations variables des segments sur le génome) [92] (Figure 2).

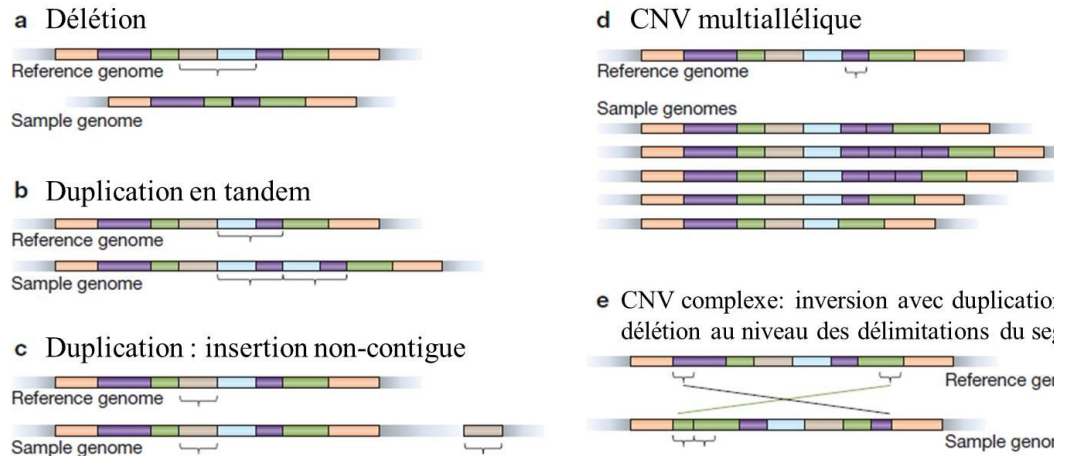


Figure 3. Quelques formes de Variation dans le nombre de Copies. Les variations sont décrites relativement à un génome de référence. Les couleurs représentent différents segments d'ADN de séquence identique (adaptée de Lee et Sherer, 2010).

Les CNVs présentent un nombre de copies qui diffère du nombre que l'on pensait « usuel », c.à.d. 2 (pour 1 copie du père et 1 copie de la mère). Dans les formes simples de CNVs on compte, des délétions (nombre de copies inférieur à 2) et des duplications (nombre de copies supérieur à 2) en tandem ou par insertion. Certains CNVs présentent des réarrangements plus complexes avec une combinaison de délétions et de duplications (Figure 3).

2.2.3- CNVs vs. SNPs

On peut envisager que les CNVs pourraient avoir plus d'impact que les SNPs sur les différences entre individus, du fait du plus grand nombre de paires de bases impliquées et de leur grande variabilité en nombre copies [82]. En effet, il a été récemment démontré que les CNVs sont plus fortement corrélés à l'expression des gènes que les SNPs [93].

2.3- Techniques de détection des CNVs

Le nombre de régions de CNVs identifiées était, il y a encore peu de temps, surestimé du fait de la variabilité de leurs délimitations sur le génome et/ou de la faible résolution des plateformes. Récemment, les techniques utilisées pour la détection des CNVs ont considérablement gagné en résolution, permettant une estimation plus précise de la taille des CNVs sur le génome, ainsi qu'une meilleure corrélation des résultats obtenus par les différentes études [94].

Ce sont principalement les biopuces d'hybridation génomique comparative (CGH) et les biopuces de SNP qui ont permis de découvrir l'abondance des CNVs de taille relativement petite dans le génome humain. Dans l'hybridation génomique comparative, la détection des CNVs est basée sur la cohybridation d'un ADN à tester avec un ADN de référence en utilisant un marquage différentiel [95]. Les tailles des segments d'ADN utilisés comme sondes sur les biopuces CGH ont grandement variées entre 80-200kb pour les BAC/PAC, 25-80pb pour les oligonucléotides et 100pb-1.5kb pour les produits de PCR génomique [96]. Si à l'origine les biopuces de SNPs n'avaient pas été élaborées pour la détection de variabilité dans le nombre de copies, elles ont été utilisées dans les premières détections de CNVs [97]. Les dernières générations de biopuces de SNPs contiennent plus de SNPs et des sondes spécifiques à la détection de CNVs et offrent ainsi une meilleure couverture du génome [98]. Depuis la complétion du génome humain, les méthodes de séquençage se sont grandement développées. On peut maintenant séquencer un génome entier en quelques analyses et détecter des CNVs par différentes méthodes incluant le séquençage bidirectionnel des extrémités (paired-end mapping, PEM) [91,99] ou le

séquençage à haut débit (ou next-generation sequencing, NGS) [100-102]. D'autre part, la détection ciblée de CNVs situés à des loci bien spécifiques peut s'effectuer en utilisant divers méthodes telles que le TaqMan [103], la Multiplex Ligation-dependent Probe Amplification (MLPA) [104,105], ou la PCR en multiplex [106]. Le TaqMan ou PCR quantitative en temps réel étant l'une des méthodes les plus fiables et les plus utilisées pour la validation de CNVs candidats.

Des efforts importants et en constante évolution sont fournis pour des développements technologiques et en ressources. Ainsi, des nouvelles ressources comme le « 1000 Genome Project » (<http://www.1000genomes.org/>), sont disponibles afin d'identifier l'ensemble du spectre de variabilité génétique.

2.4- CNVs, impact fonctionnel, adaptation et évolution

Un CNV peut être neutre ou influencer l'expression des gènes directement, via un effet de dosage, ou indirectement en altérant la régulation transcriptionnelle dans la région de variabilité et même jusqu'à une mégabase plus loin (effet cis-régulateur) [107-109]. Chez l'humain, des CNVs causeraient l'altération dans le dosage d'environ 1 gène sur 20 [94] et expliqueraient plus de 17.7% de la variation de l'expression des gènes [107], généralement positivement corrélée au nombre de copies du CNV [110]. Ainsi, l'impact fonctionnel des CNVs s'exercerait à travers une modulation subtile de l'expression de gènes impliqués dans diverses voies biologiques incluant principalement des sentiers de signalisation cellulaire, des sentiers impliqués dans des processus biologiques

extracellulaires [94] et des sentiers métaboliques [111]. Notamment, les petits CNVs (<1kb) et les gains auraient un impact fonctionnel plus important que les larges CNVs et les délétions, respectivement. En effet, Banerjee et al [112] ont montré que les petits CNVs sont plus susceptibles de réguler la transcription des gènes que les grands CNVs, alors que parmi les larges variants (>1kb) les gains ont un impact plus grand sur la transcription des gènes que les délétions. De la sorte, la variabilité dans le nombre de copie contribue de façon majeure à la diversité phénotypique animale. Ainsi, la couleur de la robe du cheval, du porc et du mouton serait en partie déterminée par des CNVs affectant des gènes impliqués dans la pigmentation. De plus, chez le poulet, la croissance et la forme des plumes sont régulées par des CNVs [113]. Finalement, chez l'humain, des CNVs seraient impliqués dans la régulation de la taille [114].

2.4.1- CNV et adaptation

Les CNVs sont enrichis en gènes impliqués dans l'adaptation aux réponses environnementales (ex. gènes liés à l'immunité, aux processus neurophysiologiques et de perceptions sensorielles) et appauvris en gènes liés aux processus cellulaires basiques [82,115], ce qui en fait une variation génétique particulièrement soumise à l'influence des pressions sélectives. Ces dernières auraient influencé la distribution de la variabilité dans le nombre de copies, chez les populations. Il existe quelques évidences de pressions sélectives positives sur les CNVs. Par exemple, la diète aurait influencé le nombre de copies du gène de l'amylase³ salivaire (*AMY1*) qui apparaît plus élevé chez les populations ayant une alimentation riche en amidon, chez lesquelles l'efficacité de la digestion de l'amidon est

³ L'amylase est une enzyme d'origine salivaire ou pancréatique qui catalyse l'hydrolyse de l'amidon en maltose et dextrines.

accrue [116]. Le virus de l'influenza aurait influencé le nombre de copies du gène DEFB103⁴ qui est particulièrement accru en Asie de l'est où il favoriserait les résistances à l'infection de la grippe [117]. Il existe également des cas d'équilibre sélectif. Par exemple, la plupart des individus possède 4 copies du gène de l'alpha-globine. La délétion de 2 copies de l'alpha-globine cause une thalassémie légère, celle de 3 copies une thalassémie sévère, alors que la délétion complète est létale. En Asie du sud-est, les 5% de la population hétérozygotes pour la cis-délétion (nombre de copies=2) de l'alpha-globine présentent une morbidité réduite pour le paludisme [118,119]. Généralement, la distribution populationnelle des hémoglobinopathies coïncide avec la prévalence du paludisme.

2.4.2- Duplication et spéciation

On dénombre, communément, plus de duplications que de délétions au sein des génomes des différentes espèces [120]. En effet, les duplications de gènes font partie des mécanismes génomiques primaires qui auraient favorisé la prolifération de nombreuses espèces. Par exemple, les primates sont apparus il y a environ 90 millions d'années et n'ont cessé depuis de subir une croissance importante. On compte aujourd'hui environ 300 espèces de primates différentes, chez lesquels un biais de sélection en faveur des duplications a été observé. Ainsi, il apparaît que l'augmentation du nombre de copies a un impact plus important que la réduction du nombre de copies, sur l'évolution des primates [94]. Aussi, les duplications permettraient d'estimer les différences évolutives entre les espèces. Par exemple, si *DUF1220*⁵ est inexistant chez les non mammifères, une seule

⁴ Le gène DEFB103 codant pour la défensine bêta 103B est un inhibiteur clef de la fusion du virus de la grippe à la membrane cellulaire.

⁵ Le gène DUF1220 code pour une protéine à domaines multiples, liée aux phénomènes cognitifs.

copie de *DUF1220* existe chez les mammifères non primates. Chez l'humain, le nombre de copies de *DUF1220* est particulièrement élevé comparativement à celui des grands singes d'Afrique et à celui des orangs-outans chez lesquels il est encore plus réduit [121]. Surtout, le dosage de *DUF1220* est positivement associé à la taille du cerveau et semble impliqué dans des cas de micro- et macroencéphalie [122].

D'autre part, l'augmentation du nombre de copies de l'aquaporine 7 (AQP7) apparaît également spécifique à la lignée humaine. AQP7 est impliquée dans le transport de l'eau à travers les membranes et jouerait un rôle clef dans le processus de sudation. En effet, chez l'humain, la marche et la course d'endurance ont été favorisées par la disparition du pelage, qui a contribué à la thermorégulation en favorisant la sudation et conséquemment la dissipation efficace de la chaleur générée [120]. AQP7 serait également impliquée dans les mécanismes de transport d'acides gras et de glycogène.

2.5- CNVs, héritabilité et pathologies humaines

Un CNV est hérité ou non, c.-à-d. *de novo*, et peut parfois avoir un impact sur la variabilité phénotypique humaine. Ce polymorphisme génétique n'est pas distribué de façon aléatoire sur le génome humain. On retrouve les CNVs principalement à proximité des centromères et des télomères, ainsi qu'au niveau des séquences simples répétées en tandem [123]. Chez l'humain, les régions péri-centromériques sont extrêmement dynamiques et présentent un taux important de mutations qui facilitent une évolution rapide du génome. Notamment, les gènes qui présentent un gain de copies spécifique à la lignée humaine sont essentiellement péri-centromériques [120]. Cependant, ce dynamisme induit

aussi une instabilité génomique qui favorise l'émergence de pathologies [124]. En effet, des CNVs ont déjà été associés à divers types de pathologies humaines et particulièrement à des susceptibilités pour des traits complexes.

2.5.1- Héritabilité des CNVs

Des études menées sur des jumeaux monozygotes ont montré que ces derniers arborent des CNVs différents. En effet, les CNVs peuvent survenir dans les cellules germinales ou somatiques [125]. Lorsque le CNV survient dans les cellules germinales, à moins qu'il soit létal, il est transmis à la descendance. La plupart des CNVs les plus communs, c.-à-d. présents dans au moins 5% de la population, seraient hérités [126] et suivraient les lois de transmission mendélienne [127]. Ces CNVs dériveraient d'un évènement mutationnel ancien et auraient par la suite été fixés dans la population, probablement du fait d'un avantage d'un point de vue adaptatif. Particulièrement, si les segments de CNVs varient grandement en taille, il a été observé que les délétions récurrentes de relativement petite taille (2-37kb) sont transmises de manière particulièrement stable à travers les générations [87]. La ségrégation des CNVs plus rares paraît plus spécifique à certains haplotypes [126].

Lorsque le CNV survient dans les cellules somatiques, il n'est pas transmis à la descendance. Un CNV *de novo* ou non hérité découle d'un évènement mutationnel récent. La contribution de ces CNVs rares *de novo* ou non hérités à la diversité génétique des individus est cependant cent fois inférieure à celle des CNVs hérités [83]. Les CNVs peuvent aussi être spécifiques aux tissus, ce qui suggère l'intervention de mécanismes épigénétiques dans l'émergence et la transmission des CNVs [125,128].

2.5.2- Les CNVs sont impliqués dans diverses pathologies humaines

Les premières descriptions de pathologies reliées à des CNVs concernaient de larges duplications et délétions, à très forte pénétrance, *de novo* pour la plupart, associées à des désordres intellectuelles et/ou des malformations congénitales (ex. la trisomie 21 [77], syndromes de Turner [78] ou Klinefelter [79]). Plus spécifiquement, les larges segments de CNVs (duplication ou délétion) d'une centaine de paires de kilobases sont plutôt rares dans la population et souvent impliqués dans des pathologies telles que les désordres neurologiques et neurocognitifs. De plus, la contribution fréquente des CNVs à la batterie de mutations menant au développement du cancer, est aujourd'hui largement reconnue. En effet, des études pangénomiques ont permis d'identifier des CNVs somatiques impliqués dans le développement de multiples cancers tels que l'adénocarcinome du poumon (57 CNVs) [129] ou le glioblastome [130].

D'autre part, des CNVs ont été associés à des pathologies mendéliennes autosomales dominantes telles que le syndrome de Smith-Magenis (délétion de *Rai1*) [131] ou la leucodystrophie de l'adulte (duplication de *LMNB1*) [132]. On retrouve également des CNVs liés à des maladies mendéliennes autosomales récessives comme la maladie de Gaucher (délétion de *GBA*) [133]. Des CNVs situés sur le chromosome X ont aussi été associés à des maladies humaines telles que le syndrome de Hunter (délétion-inversion de *IDS*) [134] et l'hémophilie A (inversion-délétion de *F8*) [135].

Plus récemment, des CNVs ont été impliqués dans l'étiologie de traits complexes. On distingue des maladies inflammatoires comme le psoriasis ou la maladie de Crohn. De

plus, les délétions complètes de *GSTT1* et *GSTMI* accompagnées du polymorphisme *GSTP1* Val/Val ont un rôle significatif dans la pathogenèse de l'asthme [136]. Un nombre diminué de copies du gène *FCGR3B* constitue un facteur de susceptibilité accru pour le développement de plusieurs maladies auto-immunes systémiques incluant la glomérulonéphrite et le lupus érythémateux systémique [137-139]. Plus particulièrement, un nombre accru de copies du gène *CCL3L1* a été associé à une susceptibilité diminuée pour le SIDA de type 1 [140].

Quelques études ont souligné l'effet de CNVs sur des pathologies cardiométaboliques. Par exemple, la duplication au locus du gène *LEPR* a été associée à la glycémie, à la cholestérolémie et à une élévation du risque pour le diabète de type 2 [141]. D'autre part, le nombre de copies au locus 16p11.2 a été associé à des indices de masse corporelle (IMC) extrêmes (c.à.d. obésité et maigreur), chez plusieurs cohortes européennes [142-144]. Plus particulièrement, une étude a montré qu'un CNV commun en amont de *NEGR1* (cis-régulation) est associé à l'IMC chez des Caucasiens [145]. En effet, les CNVs communs constitueraient de bons candidats pour l'estimation du risque pour des traits complexes [146]. Cependant aucune étude pangénomique n'a encore pu détecter d'associations significatives de CNVs communs avec l'hypertension artérielle [70].

III-Conclusion

L'hypertension artérielle essentielle (HTA) est une pathologie complexe, multifactorielle et polygénique. Récemment, des études pangénomiques ont identifié des SNPs impliqués dans la variation de la pression artérielle et la prévalence de l'HTA. Cependant, leur impact paraît modeste et une portion importante de la composante génétique de l'HTA demeure non élucidée. La variabilité dans le nombre de copies contribue de façon majeure à la diversité phénotypique, ainsi qu'à l'évolution et l'adaptation animale. Aussi, les CNVs apparaissent plus fortement corrélés à l'expression des gènes que les SNPs. En conséquence, les CNVs auraient plus d'impact sur la diversité interindividuelle que les SNPs. Des CNVs ont déjà été associés à des pathologies humaines et notamment à des traits complexes. Pourtant, les effets des CNVs communs sur l'HTA essentielle sont encore peu connus.

HYPOTHÈSE DE TRAVAIL ET OBJECTIFS

Le rationnel

Dans le monde entier, l'hypertension artérielle (HTA) constitue un important problème de santé publique. En effet, l'HTA est l'un des principaux facteurs de risque des maladies cardiovasculaires, responsables de la plupart des cas de mortalité et morbidité humaines. L'incidence de cette maladie polygénique et multifactorielle augmente aussi bien dans les populations occidentales que dans celles des pays en transition économique. Les facteurs génétiques jouent un rôle notable dans la pathogenèse de l'hypertension artérielle avec une contribution variant de 30% à 70% [60]. Cependant, bien que des études pangénomiques aient permis de relier certains loci à l'hypertension artérielle, la composante génétique de cette pathologie n'est pas entièrement élucidée. Récemment, le développement de technologies de criblage, telles que les biopuces de génotypage, a permis d'observer que la variabilité dans le nombre de copies est largement répandue à travers le génome des mammifères. Un CNV est une séquence d'ADN d'au moins 50pb, présente en un nombre variable de copies dans le génome [84]. Le CNV apparaît comme un contributeur majeur à la diversité phénotypique, ainsi qu'à l'adaptation et à l'évolution animale. Principalement, les CNVs communs, particulièrement ceux que l'on retrouve dans au moins 5% de la population générale, constitueraient de bons candidats pour l'estimation du risque pour des traits complexes [146]. En effet, si des CNVs sont clairement impliqués dans l'étiologie de la schizophrénie ou de l'autisme, leur impact sur l'HTA est encore peu connu.

L'hypothèse de travail

Nous envisageons que des CNVs pourraient accroître ou diminuer le risque pour l'hypertension. Les CNVs peuvent être hérités ou *de novo*.

Les objectifs spécifiques

Nous souhaitons, plus précisément, identifier des CNVs hérités et estimer leur impact sur le développement de l'hypertension artérielle. D'autre part, l'hypertension est une maladie très hétérogène et souvent associée à des anomalies métaboliques. L'étude des comorbidités devrait faciliter l'identification des voies/systèmes biologiques affectées par les CNVs d'intérêts. Plus particulièrement, l'hypertension est un facteur de risque important pour les maladies cardiovasculaires. L'identification de CNVs communs associés à l'hypertension devrait contribuer à caractériser des individus à haut risque et à permettre un diagnostic présymptomatique.

CHAPITRE I: Copy Number Variants on chromosome 17 Associated with Type 2 Diabetes and Hypertension

Copy Number Variants on chromosome 17 Associated with Type 2 Diabetes and Hypertension

Mahine Ivanga, Youssef Idaghdour, François Harvey, Jean-Philippe Goulet, John Raelson, François-Christophe Marois-Blanchet, Gilles Corbeil, John Chalmers, Stephen Harrap, Stephen McMahon, Michel Marre, Mark Woodward, Daniel Gaudet, Philip Awadalla, Johanne Tremblay, and Pavel Hamet

Manuscript submitted to Diabetes Journal

1.1 Abstract

Genome-wide association studies (GWAS) have identified single-nucleotide polymorphisms associated with hypertension and diabetes; however, little is known about the effects of copy-number variable regions (CNVRs) on these diseases. We report a GWAS of CNVRs in 165 French Canadians (FC) from 19 families, genotyped using Affymetrix Genome-Wide Human SNP Arrays 6.0. Replication was tested using statistical models accounting for interaction between selected CNVRs in 3,301 unrelated diabetic subjects from the ADVANCE trial. Locus-specific transcriptional analyses were performed in 134 subjects from the CARTaGENE population-cohort. 2 CNVRs on 17q21.31 were identified. CNVR1 showed directionally consistent association with hypertension and diabetes in FC, odds ratio (OR) per copy-number gain of 4.69 ($P=0.02$), and with hypertension in ADVANCE diabetic subjects, OR per copy-number gain of 3.76 ($P=5.87 \times 10^{-7}$). Increased copy-number at CNVR1 was also associated with more insulin resistance ($P=0.01$), diabetes younger-onset ($P=0.005$) and risk for cardiovascular diseases (CVD, $P<0.05$). CNVR1 influenced *LOC644172* expression ($P<0.04$), whose transcript abundance was higher in CARTaGENE copy-number gain carriers and strongly associated with risk for CVD ($P<6 \times 10^{-32}$). This study implicates common CNVRs in susceptibility to hypertension and diabetes, diabetes younger-onset and thus, in risk for CVD in Caucasian subjects. These effects appear to be modulated through expression of *LOC644172*.

1.2 Introduction

Hypertension and type 2 diabetes (T2D) are major worldwide health problems, known to increase the risk for cardiovascular diseases and kidney damage (1; 2). Population studies have shown the importance of genetic factors in the development of these diseases, with a variable degree of heritability (3-5). Genome-wide association studies (GWAS) have been successful in identifying common SNPs, but their overall contribution to T2D susceptibility or blood pressure variance is relatively modest, and many of the associations have not been replicated in subsequent studies (6; 7).

A copy-number variant (CNV), sequence greater than 50bp (8; 9) present in a variable number of copies in the genome, may represent a major source of both genomic and phenotypic variability (10). A few CNV GWAS studies have reported associations of common CNVs with type 1 diabetes or T2D in different ethnic groups (11-15); however, none of the CNV GWAS testing common CNVs that can be tagged by SNPs have provided evidence not previously proven through SNP studies of association with T2D and hypertension (15). Here we aimed through a systematic GWAS of CNVs followed by a large replication study to identify novel associations of diabetes and hypertension with common CNVs.

1.3 Research Design and Methods

1.3.1- Study cohorts

French Canadians

We selected 165 French Canadian (FC) individuals from 19 multigenerational families (16) from the Saguenay-Lac-St-Jean region of Quebec, ascertained by the presence of at least one sib pair affected by early-onset hypertension and dyslipidemia. The present study includes subjects aged from 19 to 92 years (mean age 50.6) with a mean BMI of 27.2 Kg.m⁻² and 48.5% males. The sample consists of 54.5% hypertensive, 9.7% diabetic and 7.3% hypertensive diabetic subjects.

ADVANCE

We also studied 3,301 Caucasian individuals from the Action in Diabetes and Vascular Disease: Peterax and Diamicron MR Controlled Evaluation (ADVANCE) study (17). All subjects were unrelated diabetic subjects aged from 48 to 87 (mean age 66.7) with a mean BMI of 30.1 Kg.m⁻² and 64.3% males. The sample includes 59.8% patients treated for hypertension at baseline.

CARTaGENE

CARTaGENE (18) is a population-based health survey and biobank of 20,000 subjects in Quebec, Canada. A sample of 44 individuals was selected based on Framingham risk scores. A second sample of 90 individuals was selected based on *LOC644172* transcript

levels. Expression levels were measured using Illumina's Human HT-12 BeadChips and pair-ends RNA sequencing in the 44 and 90 individuals, respectively.

1.3.2- Phenotypes

French Canadians

Cohort selection and phenotyping have been previously described (19; 20). The population living in the Chicoutimi/Saguenay-Lac-St-Jean region of Quebec is relatively isolated, displaying a founder effect for several conditions and documented with computerized genealogical records going back to 1608 (16; 21). The affected sib pair–inclusion criteria were essential hypertension (systolic blood pressure [SBP] >140 mmHg and/or diastolic blood pressure [DBP] >90 mmHg on two occasions or the use of antihypertensive medication), dyslipidemia (plasma cholesterol \geq 5.2 mmol/l and/or HDL cholesterol \leq 0.9 mmol/l or the use of lipid-lowering medication), BMI <35 kg/m², age 18–55 years, and of Catholic French Canadian origin. Exclusion criteria of affected sibpairs included secondary hypertension, DBP >110 mmHg in spite of therapy, diabetes mellitus, renal or liver dysfunction, malignancy, pregnancy, and substance abuse. Once the affected sib pairs were selected, all first- and second-degree relatives aged >18 years were invited to participate in the study, independent of health status. The total recruited population comprised 120 families (mean size 7.3 persons; median size 5 persons), comprising 897 subjects constituting 1,617 sib pairs. We used the Acute Insulin Response (AIRg, μ U/mL \times min), which represents the acute pancreatic β -cell response to glucose, as an insulin resistance trait. AIRg was defined as the secretion of insulin during the first 10 minutes following a

glucose bolus calculated using the Bergman minimal model. The study was approved by the Ethics Committees of Complexe hospitalier de la Sagamie (Chicoutimi, Quebec), Université du Québec à Chicoutimi, and the Centre hospitalier de l'Université de Montréal. All subjects gave their informed consent. All phenotyping was performed by trained personnel who followed standard operating procedures as described previously (19; 20). Particularly, antihypertensive drugs and lipid-lowering medications were withdrawn, respectively, for at least 1 week and 1 month, before the onset of phenotyping protocol.

ADVANCE

The ADVANCE (Action in Diabetes and Vascular Disease: Peterax and Diamicron MR Controlled Evaluation) study is a factorial, multicentre, randomised controlled clinical trial of 11,140 participants recruited from 215 centers in 20 countries (17). The ADVANCE genetic substudy was designed to identify the genomic signatures of complications in Type 2 diabetes mellitus (T2D) and more specifically to determine the pre-symptomatic predictors of the risk of myocardial infarction, stroke and nephropathy in patients with T2D. Hypertension status was defined by treatment with blood pressure-lowering drugs used at baseline. Risk for cardiovascular diseases was defined using a 4-year CVD risk score developed for patients with diabetes (22).

The CARTaGENE project (18) : the cohort consists of over 20,000 deeply endophenotyped individuals, aged 40-69 who underwent deep phenotyping with 190 physiological parameters including peripheral and central blood pressure, glycemic traits, anthropometric

measurements, medical history and socioeconomic status. CARTaGENE includes 2,000 T2D subjects and an equivalent number of individuals are in the pre-diabetes stage. It also includes 6,700 hypertensive and 2,300 hyperglycemic subjects. About 9,000 subjects have had their blood collected for transcriptomic analysis and 200 subjects have had their whole exomes sequenced. Risk for cardiovascular diseases (lipid-based) and vascular age (BMI-based) were estimated using the 10-year Framingham CVD risk scores (23).

1.3.3- CNV detection and selection

182 FC subjects were genotyped using Affymetrix Genome-Wide Human SNP Array 6.0, while 3629 ADVANCE subjects were genotyped with Affymetrix Genome-Wide Human SNP Arrays 5.0 (~30%) and 6.0, according to the manufacturer's protocol.

A principal component analysis (PCA) was performed using 147,088 SNPs to quantify the genetic divergence in ADVANCE subjects of European ancestry, which outputs each individual's coordinates along the principal axes of variation. The two first principal components (PC1 and PC2) corresponding to the Northwest and Southeast Europe axes of variation were obtained using smartpca (EIGENSOFT 3.0 (24)). These two PCs were used as covariates in subsequent association analysis.

The Birdsuite tool set (25) developed at the Broad Institute in combination with Plink (26) was used to evaluate CNV calls in FC and ADVANCE subjects. The Birdseye algorithm from Birdsuite implements a hidden Markov model (HMM) that integrates multiple sources of information, including Log R ratio (LRR; a measure of total probe signal intensity) and B allele frequency (BAF; a measure of the relative intensity ratio of the two allelic probes),

to infer copy-number from individual genotype samples; while a second Birdsuite algorithm, Canary, uses a one-dimensional Gaussian mixture model to detect common CNVs. The outputs from Birdseye and Canary were combined to produce a unified call of CNVs and these two algorithms are hereafter referred to as one algorithm. A quality control (QC) filtering step was applied at both the sample genotype and CNV level. Arrays with an overall call rate less than 86% were discarded from further analysis, leaving a data set of 180 FC and 3,323 genotyped individuals.

ParseCNV (27) was used in combination with PennCNV (28) to determine copy-number variant regions (CNVRs) in FC subjects. PennCNV joint-calling algorithm produces CNV calls in one single step for parents-offspring trio (29). We then further analyzed these calls with the ParseCNV algorithm. To infer copy-number from parents-offspring trio genotype samples, PennCNV also implements a HMM that integrates LRR and BAF followed by a QC filtering step. Duplicate samples and samples with bad contrast readings were removed from cell files. The remaining cell files were analyzed using the Affymetrix power tools and only files with a QC call-rate higher than 86% were retained. Individuals with sex-mismatch, ethnicity or relatedness issues were also removed, leaving a data set of 78 FC genotyped individuals. Raw CNV calls were further filtered to retain only those CNVs with a CNV quality score greater than 10 and with detection by a minimum of 6 probes and a minimum CNV call length of 500bp. ParseCNV creates probe-based statistics for the cleaned CNVs called by PennCNV, which are then used to determine CNVRs. These probe-based statistics are tested for case control association in Plink (26) using a Fisher exact test and probes without nominal significance ($P < 0.05$) are discarded from further

association testing. The output from the probe-based association tests is then used by ParseCNV to merge significant probes into CNVRs based on probe proximity (<1 MB) and comparable significance ($\pm 1 \log P$ -value) of neighbouring probes. CNVRs are defined in a dynamic manner, allowing flexible boundaries for complex CNVs and maintaining precise state-specific association region.

1.3.4- CNV analysis using TaqMan

TaqMan was used to validate FC microarrays data (Fig. S1, data supplement) and to genotype the 44 CARTaGENE subjects for the selected region. The copy-number was determined by a custom TaqMan® copy-number assay in a duplex real-time PCR (Applied Biosystems™ by life technologies™, Foster City, CA, USA) according to the manufacturer's protocol. PCR amplification was carried out using the 7900HT Fast Real-Time PCR System with a 20 µl reaction mixture containing 10 µl TaqMan® genotyping master mix (2X), 1.0 µl TaqMan® copy-number assay (20X), 1.0 µl TaqMan® copy-number reference assay (RNase P gene, 20X), 4 µl nuclease-free water and 4 µl of 5 ng/µl gDNA (20 ng). Custom assays were designed using Primer3Plus, Repeat Masker and NCBI/BLAST genome and SNP databases. The CNVR1 Custom assay maps to chromosome 17: 43657590-43657694, Hg19 genome Assembly; Forward: 5'-CTGTTTGAGGTAAATCTATGGTTCA-3', Reverse: 5'-TGGTGTTATCCTGGGACAGTT-3'; FAM Reporter: 5'-TCTGCCACTGATCTCCACTG-3'. ILMN_1772603 (Illumina Human HT-12 probe LOC644172) is included within the *LOC644172* TaqMan Custom assay, chromosome 17: 43678940-43679049 (Hg19 Assembly); Forward primer: 5'-

GAAGGACTCGGCACTGGAG-3', Reverse primer: 5'-
CCTGCCTCTCCAAGGACTC-3', FAM Reporter: 5'-CGGCCACTCATGAAGATGTT-
3'. Each sample was run in quadruplicates. Manual C_T threshold was set at 0.2 and
automatic baseline was activated. The relative copy-number in each gDNA sample was
obtained using Applied Biosystems software SDS 2.3 and CopyCaller™ software v2.0.

1.3.5- Analysis of expression levels

Illumina's Human HT-12 BeadChips

Blood samples were collected from CARTaGENE participants into PAXgene tubes (Qiagen) and total RNA extracted using PAXgene Blood RNA Kit (Qiagen). Quantification and quality control of total RNA was performed using a *2100 Bioanalyzer* instrument and an *RNA 6000 Nano LabChip* kit (Agilent Technologies). Samples with an RNA Integrity Number (RIN) > 8 were retained for further experiments. cDNA and cRNA synthesis were performed with the Illumina TotalPrep RNA Amplification kit (Ambion) following the manufacturer's instructions. Illumina's Human HT-12 BeadChips were used to generate expression profiles with 500 ng of labeled cRNA for each sample and following manufacturer's recommended protocols. A randomized design was used to minimize chip effects. Four individuals were replicated; these clustered adjacent to one another in hierarchical analysis and the expression intensities were averaged in the statistical analysis. Expression intensity measures were obtained from an average of 30 beads for each transcript. The BeadChips were imaged with an Illumina BeadArray Reader. The raw intensities were extracted with the Gene Expression Module in Illumina's BeadStudio software. Expression intensities were log₂ transformed and median-centered by subtracting

the mean value of each array from each intensity value. Intensities for the genes located in the 17q21.31 CNV region and for the *MAPK8IP1* loci were extracted to perform association testing with CNV variation.

Pair-ends RNA sequencing

Approximately 3 mL of blood was collected Tempus Blood RNA Tubes (Life Technologies). Total RNA was extracted by using a Tempus Spin RNA Isolation kit (Life Technologies) followed by globin mRNA depletion using a GLOBINclear-Human kit (Life Technologies). Quantification and quality control of globin mRNA-depleted total RNA was performed using a *2100 Bioanalyzer* instrument and an *RNA 6000 Nano LabChip* kit (Agilent Technologies). Only samples with an RNA Integrity Number (RIN) > 8 were retained for RNASeq library construction. RNAseq 100bp pair-ends indexed libraries were constructed using the TruSeq RNASeq library kit (Illumina) and sequencing was done on a HiSeq 2000 instrument (Illumina), multiplexing three samples per lane. After initial filtering based on sequencing read quality, paired-end reads were aligned using TopHat (V1.4.0) and PCR removal was performed using Picard (picard_tools/1.56). Raw gene-level count data was generated using HTseq 0.5.3p3. These counts were then normalized using EDASeq v1.4.0 and a procedure that adjust for GC-content as well as for distributional differences between and within sequencing lanes (30). Individual-level normalized gene expression levels of *LOC644172* were then extracted for further statistical analysis. Consistent normalized *LOC644172* expression levels ($r^2 > 0.91$) were obtained using two other methods: Cufflinks (31) and TMM (32).

1.3.6- Statistical analyses

Analyses of CNVRs were conducted using the R software (version 2.15.3). Results for the GWAS, in FC subjects, were estimated using a Generalized Estimating Equation (GEE) model with an exchangeable working correlation matrix to account for familial connection. The model included BMI, sex and age as fixed effects, hypertension & T2D and the Copy-Number (CN) change at a given CNVR locus.

Results for the replication analysis, in ADVANCE subjects, were estimated using a Generalized Linear Model (GLM). The combined model included PC1, PC2, genotyping chip and sex as fixed effects, treated hypertension, CN change at selected CNVR loci and their interaction.

Results for additional quantitative phenotypes were tested using GEE or GLM, as appropriate, in combined models where CN 0, 1, 2, 3 and 4 were represented by -2, -1, 0, 1, and 2, respectively. Explanatory variables were scaled in the model to obtain the beta coefficients (β) so that one SD increase in CN was associated with $\beta \times \text{SD}$ in quantitative trait unit. The threshold for statistical significance was fixed at 0.05.

rs413778 genotyping

The H2 haplotype (inverted orientation) is in nearly perfect linkage disequilibrium with SNP rs413778 minor allele ($r^2 > 0.99$) (15). SNP rs413778 genotype was assessed in 2,301 ADVANCE subjects to infer the inversion haplotype, H1 corresponding to the major allele and H2 to the minor allele. The Birdseed algorithm and the Plink program were used for genotyping calling and quality control procedures. The genotypes were imputed using

IMPUTE v2. SNP with low MAF (<5%) and deviation from Hardy-Weinberg were excluded.

1.4 Results

GWAS and CNVR Selection

From the 77 CNVRs determined by ParseCNV, 26 CNVRs (length \geq 1 Kb, frequency above 9% of our population sample) were selected for genome-wide screening of CNVs in 165 FC subjects (Table S1, data supplement). Only two CNVRs on chromosome 17 were found nominally significantly associated with hypertension and T2D (Table 1): CNVR1 at chr17: 43,656,380 - 43,658,471 (OR per CN increase is 4.69, 95% CI: 1.22-17.95, P=0.02) and CNVR2 at chr17: 44,394,412 - 44,648,367 (OR per CN increase is 0.38, 95% CI: 0.18-0.77, P=0.007), NCBI37/Hg19 Assembly.

Association between CNVR1 and CNVR2 with hypertension and diabetes was detected but not at genome-wide significance levels. This can be due to reduced power in our discovery cohort, complex interactions between CNVRs and potentially other contributing causal genetic variants. Stringer et al., showed that GWAS for dichotomous phenotypes could misestimate effect sizes of causal variants, as they don't consider multiple risk variants together (33). We therefore hypothesize that interactions between the two detected CNVRs could be implicated in modulating risk to hypertension and T2D, particularly given their location within the same chromosomal region. Consequently, we run statistical models (combined models) that account for interaction between the two CNVRs:

T2D and Hypertension (Table 1)

CNVR1 (frequency distribution in Fig. S2, data supplement) showed directionally consistent association with hypertension in the ADVANCE diabetic patients replication set, OR per CN increase of 3.76 (95% CI: 1.36-2.01, $P=5.87\times 10^{-7}$).

Additional phenotypes: traits related to Diabetes and its complications (Table 2)

In FC subjects, AIRg, the acute insulin response to intravenous glucose, increased by $259.08 \mu\text{U}\cdot\text{mL}^{-1}\times\text{min}^{-1}$ per CN gain at the CNVR1 locus ($P=0.01$). AIRg represents the incremental insulin area 0–10 min after the glucose injection and is used as a sensitive measure of beta-cell well-being, reflecting a combination of beta-cell mass and function (34). Duplication-carriers with CN=3 are more insulin-resistant, with $518.16 \mu\text{U}\cdot\text{mL}^{-1}\times\text{min}^{-1}$ further AIRg, than deletion-carriers with CN=1. In ADVANCE, a CN gain at CNVR1 was associated with a drop and a rise of ~ 1.1 year, in T2D age at diagnosis ($P=0.005$) and duration of diabetes ($P=0.0002$), respectively. Duplication-carriers with CN=4 are diagnosed for T2D ~ 3.3 years earlier than deletion-carriers with CN=1. At both CNVR1 and CNVR2 loci, a CN gain was associated with, $\sim 1/4$ ($P=2.84\times 10^{-5}$) and $\sim 1/5$ ($P=2.05\times 10^{-6}$), more blood pressure-lowering drug intake, respectively. The CVD risk score (22) increased of 0.48 per CN gain at the CNVR1 locus ($P<0.05$), in ADVANCE.

Content of CNVR1 and CNVR2

CNVR1 is an intergenic region (2.091 kb) highly enriched in H3K27Ac histone (acetylation of lysine 27 of the H3 histone protein) mark thought to enhance transcription of proximal gene by blocking the spread of the repressive histone mark H3K27Me3 (Fig. 1).

CNVR1 includes the BPTF duplicated promoter 5' upstream *LRRC37A4P* (35) (92% identities with isoform BPTF subunits (36)).

We therefore used the 44 CARTaGENE subjects to evaluate the impact of CNVR1 on neighbouring genes' expression (Fig. 2A). The total region covered by all the segments overlapping CNVR1 (118.058 kb in length, chr17: 43,585,756-43,703,814; NCBI37/Hg19 Assembly) harbours four genes encoding five transcripts. Of the four genes located within the total region of interest, only *LOC644172*'s relative expression was significantly higher in copy-number gain carriers than in non-gain carriers (raw $P=0.01$ and adjusted $P=0.05$). In addition, *LOC644172* relative expression appeared to be quantitatively modulated by CN change within the CNVR1 locus ($P<0.04$, Fig. 2B) and particularly by CN change within the *LOC644172* locus ($P<5.0\times 10^{-5}$, Fig. S3 data supplement). *LOC644172* is a transcribed pseudogene of *MAPK8IP1* that encodes the IB1 protein.

CNVR2 (253.955 kb) contains five referenced sequence genes. Particularly, the total region covered by all the segments overlapping CNVR2 includes an untranscribed pseudogene of *MAPK8IP1* (*MAPK8IPP*) and two members of the leucine rich repeat containing 37A gene family (*LRRC37A* and *LRRC37A2*). The total region covered by all the segments overlapping CNVR1 contains several DNA sequences including a transcribed pseudogene of the leucine rich repeat containing 37A gene family (*LRRC37A4P*) and a transcribed pseudogene of *MAPK8IP1*, *LOC644172*. Several studies have established a role of expressed pseudogenes in the regulation of expression of their functional paralog, namely coding parent gene (37).

LOC644172 Functional Analysis

Data from the 90 CARTaGENE subjects showed that *LOC644172* expression level was positively associated with the prevalence of hypertension ($P=0.02$) and T2D ($P=0.02$). Particularly, SBP (11.4 mmHg between lowest and highest tercile of expression) and cholesterol-HDL ratio presented a positive association with *LOC644172* group of expression, $P=3.9\times 10^{-5}$ and $P=7.75\times 10^{-6}$, respectively (*Fig. 3*). Accordingly, percentage risk of CVD and vascular age appeared strongly associated with *LOC644172* expression level, $P=6.0\times 10^{-32}$ and $P=2.4\times 10^{-17}$, respectively (*Fig. 4*). We specifically noted that the delta of vascular vs. chronological age, reflecting vascular aging, markedly increased with *LOC644172* expression level, $P=1.3\times 10^{-12}$. A Spearman rank test showed a positive correlation between *LOC644172* and *MAPK8IP1* expression level ($\rho=0.23$, $P=0.01$), suggesting that *LOC644172* accounts for 5% of *MAPK8IP1* transcript abundance.

H1 and H2 haplotypes

The inverted H2 haplotype is tagged by the minor allele of SNP rs413778. The minor allele of rs413778 is present in 857 ADVANCE subjects out of 2,150: i.e. ~40% of our ADVANCE subjects carry H2. However, a Pearson test indicated a weak correlation between the SNP rs413778 and CNVR1 ($r=-0.05$, $P=0.01$), in ADVANCE subjects.

1.5 Discussion

This study represents the first genome-wide screening of CNVs in Caucasians with diabetes and hypertension. We identified novel common susceptibility CNVR loci, showing dose effects, at 17q21.31 in two independent cohorts. The main effect size is observed for the

CNVR1 locus, with an odds ratio (OR) per one copy-number increase for hypertension varying from 2.23 to 6.32, with a mean risk of 3.76 in ADVANCE diabetic patients. The mean OR per copy-number gain for hypertension and diabetes is 4.69 in FC. The observed effects are substantially higher than usually reported in complex traits. No SNPs, at these loci, have previously been reported to be associated with these diseases in humans.

Chromosome 17 is particularly rich in segmental duplications in humans (38). Studies that provided linkage for essential hypertension to markers on chromosome 17 had diabetic hypertensives included in the cohort (39). Our study identified two 17q21.31 CNVRs which coincide with human blood-pressure quantitative trait loci (QTL) 34 and 35, suggesting their implication in the aetiology of hypertension.

This study specifically points to two highly complex DNA region within 17q21.31, where two major haplotypes were previously described (40), H1 (direct orientation) and H2 (inverted orientation). *LRRC37A* (partially overlaps CNVR2) and *LRRC37A4P* (3' downstream CNVR1) define the boundary for this common human inversion polymorphism (35). CNVR1 locus maps to a region proximal to the 5' end of *CRHR1* gene, where the sequences of haplotypes H2 and H1 have been found to be relatively similar (40). H2, the ancestral haplotype, which is present in ~40% of our ADVANCE subjects, is enriched in European. European and Mediterranean populations have a marked enrichment of duplicated haplotypes in this region, with duplications showing greater population stratification (40). Recently, we identified 8 SNPs in *CRHR1* (previously proposed as a

candidate for human hypertension (39)) in linkage with ANP levels in supine position in the same FC cohort (41), 51.2 Kb away from CNVR1 and 337.8 Kb away from CNVR2. CNVR1 lies next to the CNVR7113.6 locus, which gave weak evidence for association with two immune disorders, Type 1 diabetes and Crohn's disease (CD), in a previous GWAS (15). The CNVR7113.6R association with CD was then confirmed by Roberts et al. (42). CNVR1 and 2 map to two human chronic obstructive pulmonary disease QTLs (12 and 22) and the human rheumatoid arthritis QTL 28, further suggesting a role for these CNVR loci in immune and inflammatory responses. Interestingly, in a GWAS with CNVs, most of the candidate genes for T2D identified were related to transcriptional regulations and immune responses (11). Activated immunity and inflammation may represent the common trigger of T2D and hypertension pathogeneses (43).

Kudo et al. (12) have identified a CNV on chromosome 4 associated with early onset of T2D, which contains genes related to glucose-induced insulin secretion cascade of pancreatic β -cells, pancreatic development and differentiation, and adipose tissue functions. Here we show that increased CN within CNVR1 is associated with early onset of T2D in ADVANCE. In addition, in FC, carriers of copy-number gain at this locus present increased AIRg, an insulin resistance (IR) trait reflecting both β -cell mass and function. The T2D pathogenesis involves progressive development of IR and a relative deficiency in insulin secretion by β -cell that lead to hyperglycemia. Enhanced β -cell function and expansion of β -cell mass are involved in the development of IR (34). There is also strong evidence of a relationship between insulin resistance, insulin concentration and blood pressure (44). We

can hypothesize that gain in copy numbers at this locus may contribute to premature β -cell functional deregulation and IR.

IR is a critical factor in the development of T2D and in driving the expression of a phenotype that usually includes both atherogenic dyslipidemia and hypertension (45). Accordingly, we show that increased CN at 17q21.31 loci is associated with higher T2D and hypertension prevalence in ADVANCE and FC subjects. Earlier appearance of clustering of risk factors and longer the time of exposure, are associated with higher the risk of developing CVD. IR/T2D and hypertension exert a cumulative effect on the risk of CVD, which appears higher in ADVANCE carriers of copy-number gain at the CNVR1 locus. In ADVANCE, carriers of copy-number gain have an early onset of T2D and longer T2D duration.

We also show that *LOC644172* relative expression is modulated by the number of copies within CNVR1 in a dose-dependent manner. *LOC644172* is a transcribed pseudogene of *MAPK8IP1* that encodes the IB1 protein. Several studies have established a role of expressed pseudogenes in the regulation of expression of their functional paralog, namely coding parent gene (37). IB1 is expressed mainly in the brain and in Langerhans islets (46) and is considered a key regulator of the JNK pathway in neuronal and β -cells (47). Besides, IB1 level has been shown to be increased in white adipose tissue of obese mice, suggesting IB1 might trigger JNK activity and insulin resistance in obesity (48). IB1 is also a DNA-binding transactivator of the glucose transporter GLUT2 (encoded by *SLC2A2*) and, a

missense mutation in the coding region of *MAPK8IP1* segregated with T2D in a French family (49).

LOC644172 is among the genes most stratified by copy number in the human genome ($V_{sr}=0.45$) (50). Our data show that *LOC644172* expression is modulated by copy-number change at the CNVR1 locus and may account for ~5% of *MAPK8IP1* gene expression regulation. However, additional experiments are needed to characterize the mechanism by which *LOC644172* regulates *MAPK8IP1* gene expression. The present study demonstrates that augmented *LOC644172* expression is associated with greater prevalence of T2D and hypertension. Besides, increased *LOC644172* expression is associated with higher SBP (by 11.4 mmHg in the third vs. first tercile of expression) and cholesterol-HDL ratio. These traits are involved in the development of CVD, and our study implicates increased *LOC644172* expression in the molecular processes taking place during vascular aging and onset of CVD.

The main limitation of the study is that the impact of haplotypes on CNV effects has not been clearly explored, as well as the nature of CNVR1 and CNVR2 interaction. However, while studies showing the association of the CNVR7113.6 locus with T1D or CD (15; 42) in Caucasians have been performed using tag SNPs, the approach used here reveals a dose-dependent effect of the identified CNVR loci on hypertension and T2D. The major strengths of the present analysis are the inclusion of individuals from two different cohorts

and expression data from a third cohort. GWAS and functional analyses led to the identification of a novel candidate biomarker, *LOC644172*, for susceptibility to CVD.

In summary, carriers of copy-number gain at these 17q21.31 loci, particularly at CNVR1, have augmented insulin resistance, early-onset T2D, increased prevalence of hypertension and an elevated risk of CVD. Our transcriptional analysis shows that the number of copies at this locus significantly influences expression levels of *LOC644172*, which in turn is associated with endophenotypes promoting vascular aging and elevating risk for CVD. Taken together, these findings suggest a strong implication of *LOC644172* in the pathogenesis of T2D complications.

Acknowledgments

Authors thank participants who volunteered for the French Canadian, ADVANCE and CARTaGENE studies and Dr. Mounsif Haloui, CHUM Research Centre, for help with preparing the figures.

Funding

This work was supported by grants from Genome Quebec (“**Genomics research in human health - translational stream**”) and CIHR (ISO-106797 and MOP-93629), and by Prognomix Inc. and Les Laboratoires Servier.

Duality of interest

P.H. and J.T. are members of the Collège International de recherche Servier and consultants for Prognomix Inc.

J.C., M.W., S.H., M.M. received speaker fees from Les Laboratoires Servier.

Author Contributions

MI performed data and statistical analyses and wrote the manuscript. Y.I. researched the transcriptional data and reviewed/edited the manuscript. J.-P.G. researched the transcriptional data. F.H., F.-C. M.-B. J.R. and G.C. researched the genetic data. S.H., S. McM., M.M. and M.W. designed and conducted the ADVANCE clinical study and reviewed the manuscript. J.C. designed the ADVANCE clinical study and contributed to discussion. DG contributed to the design of the French Canadian familial study. P.A. designed the CARTaGENE transcriptional study and reviewed the manuscript. J.T. helped with the design of the genetic part of the study and reviewed/edited the manuscript. PH designed the clinical and genetic parts of the study and reviewed/edited the manuscript.

P.H. is the guarantor of this work and, as such, had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

1.6 References

1. Sarwar N, Gao P, Seshasai SR, et al.: Diabetes mellitus, fasting blood glucose concentration, and risk of vascular disease: a collaborative meta-analysis of 102 prospective studies. *Lancet* 2010;375:2215-2222
2. Seshasai SR, Kaptoge S, Thompson A, et al.: Diabetes mellitus, fasting glucose, and risk of cause-specific death. *N Engl J Med* 2011;364:829-841
3. El Shamieh S, Visvikis-Siest S: Genetic biomarkers of hypertension and future challenges integrating epigenomics. *Clin Chim Acta* 2012;
4. Vattikuti S, Guo J, Chow CC: Heritability and genetic correlations explained by common SNPs for metabolic syndrome traits. *PLoS Genet* 2012;8:e1002637
5. Tang ZH, Fang Z, Zhou L: Human genetics of diabetic vascular complications. *J Genet* 2013;92:677-694
6. Voight BF, Scott LJ, Steinthorsdottir V, et al.: Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nat Genet* 2010;42:579-589
7. Ehret GB, Munroe PB, Rice KM, et al.: Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature* 2011;478:103-109
8. Alkan C, Coe BP, Eichler EE: Genome structural variation discovery and genotyping. *Nat Rev Genet* 2011;12:363-376
9. Pang AW, MacDonald JR, Pinto D, et al.: Towards a comprehensive structural variation map of an individual human genome. *Genome Biol* 2010;11:R52
10. Valsesia A, Mace A, Jacquemont S, Beckmann JS, Kutalik Z: The Growing Importance of CNVs: New Insights for Detection and Clinical Interpretation. *Front Genet* 2013;4:92
11. Bae JS, Cheong HS, Kim JH, et al.: The genetic effect of copy number variations on the risk of type 2 diabetes in a Korean population. *PLoS One* 2011;6:e19091
12. Kudo H, Emi M, Ishigaki Y, et al.: Frequent loss of genome gap region in 4p16.3 subtelomere in early-onset type 2 diabetes mellitus. *Exp Diabetes Res* 2011;2011:498460
13. Irvin MR, Wineinger NE, Rice TK, et al.: Genome-wide detection of allele specific copy number variation associated with insulin resistance in African Americans from the HyperGEN study. *PLoS One* 2011;6:e24052
14. Grassi MA, Tikhomirov A, Ramalingam S, Below JE, Cox NJ, Nicolae DL: Genome-wide meta-analysis for severe diabetic retinopathy. *Hum Mol Genet* 2011;20:2472-2481
15. Craddock N, Hurles ME, Cardin N, et al.: Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls. *Nature* 2010;464:713-720
16. Hamet P, Merlo E, Seda O, et al.: Quantitative founder-effect analysis of French Canadian families identifies specific loci contributing to metabolic phenotypes of hypertension. *Am J Hum Genet* 2005;76:815-832
17. Patel A, MacMahon S, Chalmers J, et al.: Effects of a fixed combination of perindopril and indapamide on macrovascular and microvascular outcomes in patients with type 2 diabetes mellitus (the ADVANCE trial): a randomised controlled trial. *Lancet* 2007;370:829-840

18. Awadalla P, Boileau C, Payette Y, et al.: Cohort profile of the CARTaGENE study: Quebec's population-based biobank for public health and personalized genomics. *Int J Epidemiol* 2012;
19. Barrett JC, Fry B, Maller J, Daly MJ: Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005;21:263-265
20. Gabriel SB, Schaffner SF, Nguyen H, et al.: The structure of haplotype blocks in the human genome. *Science* 2002;296:2225-2229
21. Heyer E, Tremblay M: Variability of the genetic contribution of Quebec population founders associated to some deleterious genes. *Am J Hum Genet* 1995;56:970-978
22. Kengne AP, Patel A, Marre M, et al.: Contemporary model for cardiovascular risk prediction in people with type 2 diabetes. *Eur J Cardiovasc Prev Rehabil* 2011;18:393-398
23. D'Agostino RB, Sr., Vasan RS, Pencina MJ, et al.: General cardiovascular risk profile for use in primary care: the Framingham Heart Study. *Circulation* 2008;117:743-753
24. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D: Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 2006;38:904-909
25. Korn JM, Kuruvilla FG, McCarroll SA, et al.: Integrated genotype calling and association analysis of SNPs, common copy number polymorphisms and rare CNVs. *Nat Genet* 2008;40:1253-1260
26. Purcell S, Neale B, Todd-Brown K, et al.: PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007;81:559-575
27. Glessner JT, Li J, Hakonarson H: ParseCNV integrative copy number variation association software with quality tracking. *Nucleic Acids Res* 2013;41:e64
28. Wang K, Li M, Hadley D, et al.: PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Res* 2007;17:1665-1674
29. Wang K, Chen Z, Tadesse MG, et al.: Modeling genetic inheritance of copy number variations. *Nucleic Acids Res* 2008;36:e138
30. Rapaport F, Khanin R, Liang Y, et al.: Comprehensive evaluation of differential gene expression analysis methods for RNA-seq data. *Genome Biol* 2013;14:R95
31. Trapnell C, Roberts A, Goff L, et al.: Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc* 2012;7:562-578
32. Robinson MD, Oshlack A: A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol* 2010;11:R25
33. Stringer S, Wray NR, Kahn RS, Derks EM: Underestimated effect sizes in GWAS: fundamental limitations of single SNP analysis for dichotomous phenotypes. *PLoS One* 2011;6:e27964
34. Buchanan TA, Xiang AH, Peters RK, et al.: Preservation of pancreatic beta-cell function and prevention of type 2 diabetes by pharmacological treatment of insulin resistance in high-risk hispanic women. *Diabetes* 2002;51:2796-2803
35. Bekpen C, Tastekin I, Siswara P, Akdis CA, Eichler EE: Primate segmental duplication creates novel promoters for the LRRC37 gene family within the 17q21.31 inversion polymorphism region. *Genome Res* 2012;22:1050-1058

36. Ruthenburg AJ, Li H, Milne TA, et al.: Recognition of a mononucleosomal histone modification pattern by BPTF via multivalent interactions. *Cell* 2011;145:692-706
37. Li W, Yang W, Wang XJ: Pseudogenes: pseudo or real functional elements? *J Genet Genomics* 2013;40:171-177
38. Zody MC, Garber M, Adams DJ, et al.: DNA sequence of human chromosome 17 and analysis of rearrangement in the human lineage. *Nature* 2006;440:1045-1049
39. Knight J, Munroe PB, Pembroke JC, Caulfield MJ: Human chromosome 17 in essential hypertension. *Ann Hum Genet* 2003;67:193-206
40. Steinberg KM, Antonacci F, Sudmant PH, et al.: Structural diversity and African origin of the 17q21.31 inversion polymorphism. *Nat Genet* 2012;44:872-880
41. Arenas IA, Tremblay J, Deslauriers B, et al.: Dynamic Genetic Linkage of Intermediate Blood Pressure Phenotypes during Postural Adaptations in a Founder Population. *Physiol Genomics* 2012;
42. Roberts RL, Diaz-Gallo LM, Barclay ML, et al.: Independent replication of an association of CNVR7113.6 with Crohn's disease in Caucasians. *Inflamm Bowel Dis* 2012;18:305-311
43. Xiao L, Liu Y, Wang N: New paradigms in inflammatory signaling in vascular endothelial cells. *Am J Physiol Heart Circ Physiol* 2014;306:H317-325
44. Ferrannini E, Natali A, Capaldo B, Lehtovirta M, Jacob S, Yki-Jarvinen H: Insulin resistance, hyperinsulinemia, and blood pressure: role of age and obesity. European Group for the Study of Insulin Resistance (EGIR). *Hypertension* 1997;30:1144-1149
45. Semenkovich CF: Insulin resistance and atherosclerosis. *J Clin Invest* 2006;116:1813-1822
46. Mooser V, Maillard A, Bonny C, et al.: Genomic organization, fine-mapping, and expression of the human islet-brain 1 (IB1)/c-Jun-amino-terminal kinase interacting protein-1 (JIP-1) gene. *Genomics* 1999;55:202-208
47. Beeler N, Riederer BM, Waeber G, Abderrahmani A: Role of the JNK-interacting protein 1/islet brain 1 in cell degeneration in Alzheimer disease and diabetes. *Brain Res Bull* 2009;80:274-281
48. Brajkovic S, Marenzoni R, Favre D, et al. Evidence for tuning adipocytes ICER levels for obesity care. *Adipocyte* 2012;1(3):157-60
49. Waeber G, Delplanque J, Bonny C, et al.: The gene MAPK8IP1, encoding islet-brain-1, is a candidate for type 2 diabetes. *Nat Genet* 2000;24:291-295
50. Sudmant PH, Kitzman JO, Antonacci F, et al.: Diversity of human copy number variation and multicopy genes. *Science* 2010;330:641-646

1.7 Tables and Figures

Table 1- CNVR1 showed directionally consistent association with hypertension and T2D in French Canadians and with hypertension in diabetic ADVANCE subjects.

	<i>Chr</i>	<i>Start</i>	<i>End</i>	GWAS		Replication	
				<i>OR</i> (95%CI)	<i>P value</i>	<i>OR</i> (95%CI)	<i>P value</i>
				French Canadians N=165		ADVANCE N=3,301	
<i>CNVR1</i>	17	43 656 380	43 658 471	4.69 (1.22- 17.95)	0.0242	3.76 (2.23-6.32)	5.87×10⁻⁷
<i>CNVR2</i>	17	44 394 412	44 648 367	0.38 (0.18-0.77)	0.0072	3.38 (2.12-5.40)	3.37×10 ⁻⁷

Positions are given (in bp) based on Hg19 Assembly. Odds ratios (OR) are given per copy number gain. For the GWAS, results are adjusted for BMI, sex and age. For the replication, results are adjusted for PC1, PC2, genotyping chip and sex. Chr: chromosome. Directionally consistent results are shown in bold.

Table 2-CNVR1 is associated with quantitative traits related to T2D and its complications.

	CNVR1		CNVR2	
	β	<i>P value</i>	β	<i>P value</i>
French Canadians				
Acute Insulin Response	0.281	0.0142	0.189	0.0858
ADVANCE				
Duration of T2D	0.083	0.0002	0.012	0.5405
Age at T2D diagnosis	-0.062	0.0053	0.001	0.9619
Number of blood pressure-lowering drug	0.090	2.84×10^{-5}	0.089	2.05×10^{-6}
CVD risk score	0.045	0.0474	0.023	0.2494

The beta coefficients, β , are given per copy number gain. Results for Acute Insulin Response to glucose (AIRg) are adjusted for BMI, sex and age. Results for duration of T2D, age at diagnosis and number of blood pressure-lowering drug are adjusted for PC1, PC2, genotyping chip and sex. Results for cardiovascular risk (CVD score, 4 years) are adjusted for PC1, PC2 and genotyping chip. In the ADVANCE CVD model of risk prediction (22), CVD was defined as fatal or non-fatal myocardial infarction, stroke or cardiovascular death. Nominally significant results are shown in bold.

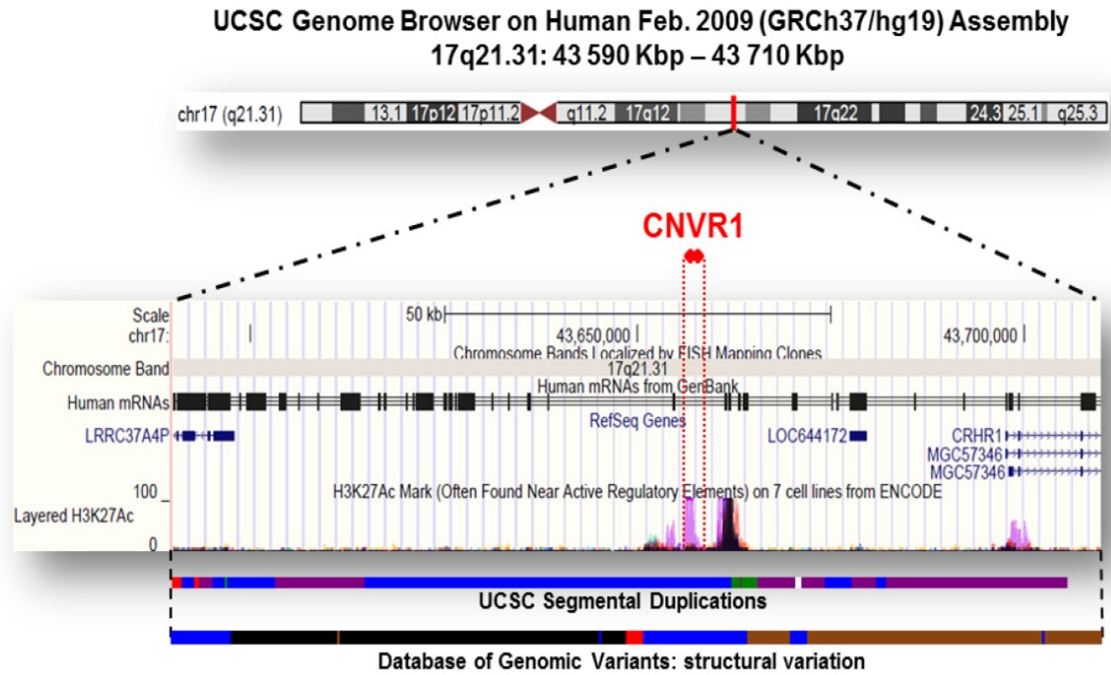


Figure 1. CNVR1 is intergenic and highly enriched in H3K27Ac histone mark. The total region covered by all the CNV segments overlapping CNVR1 harbours four genes encoding five transcripts. CNVR1 is intergenic and highly enriched in H3K27Ac histone mark thought to enhance transcription of proximal gene. Database of Genomic Variant: blue=gain, red=loss, brown=complex, and black=unkown; UCSC Segmental Duplications: level of similarity between duplications (inter/intrachromosomal), red=99-100%, purple=96-98.99%, green=93-95.99%, and blue=90-92.99%.

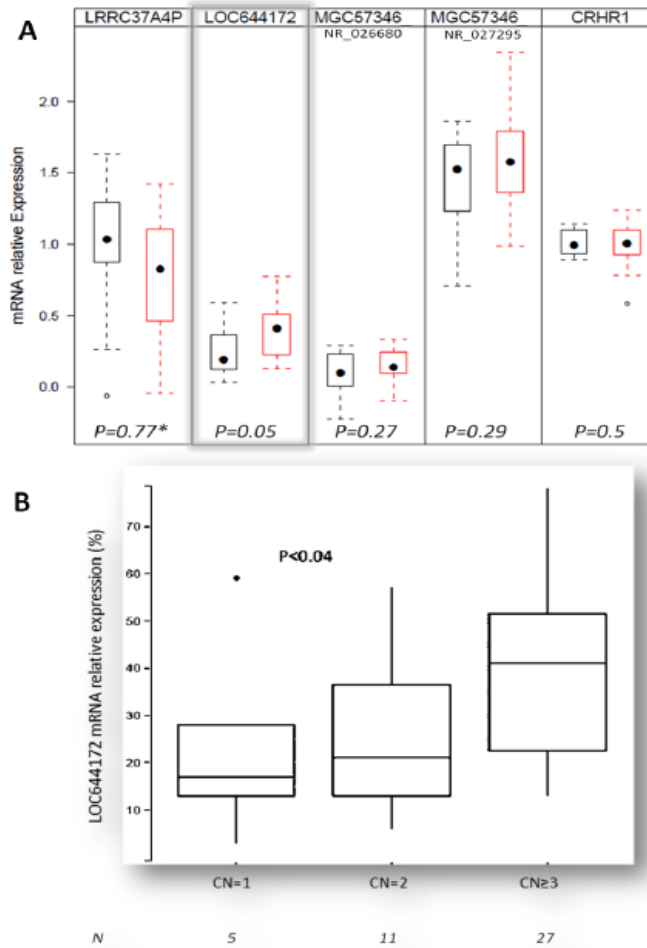


Figure 2. CNVR1 influences LOC644172 expression. The figures were drawn based on 43 CARTaGENE samples. **A:** LOC644172 mRNA relative expression was significantly higher in copy-number gain carriers (nominal $P=0.01$). P-values were estimated using a Wilcoxon rank sum testing expression levels higher in copy-number gain (red boxplots) carriers than in non-carriers (black boxplots). Adjusted P-values, with the adaptive Benjamini & Hochberg step-up FDR-controlling procedure, are indicated. The threshold for statistical significance was fixed at 0.05.*Adjusted P value when testing for expression level of transcripts of copy-number gain carriers lower than those of non-carriers: 0.19. **B:** LOC644172 relative expression is quantitatively modulated by copy-number change at the CNVR1 locus. P-value was estimated using a GLM. CN: copy-number, loss: $CN < 2$, gain: $CN > 2$.

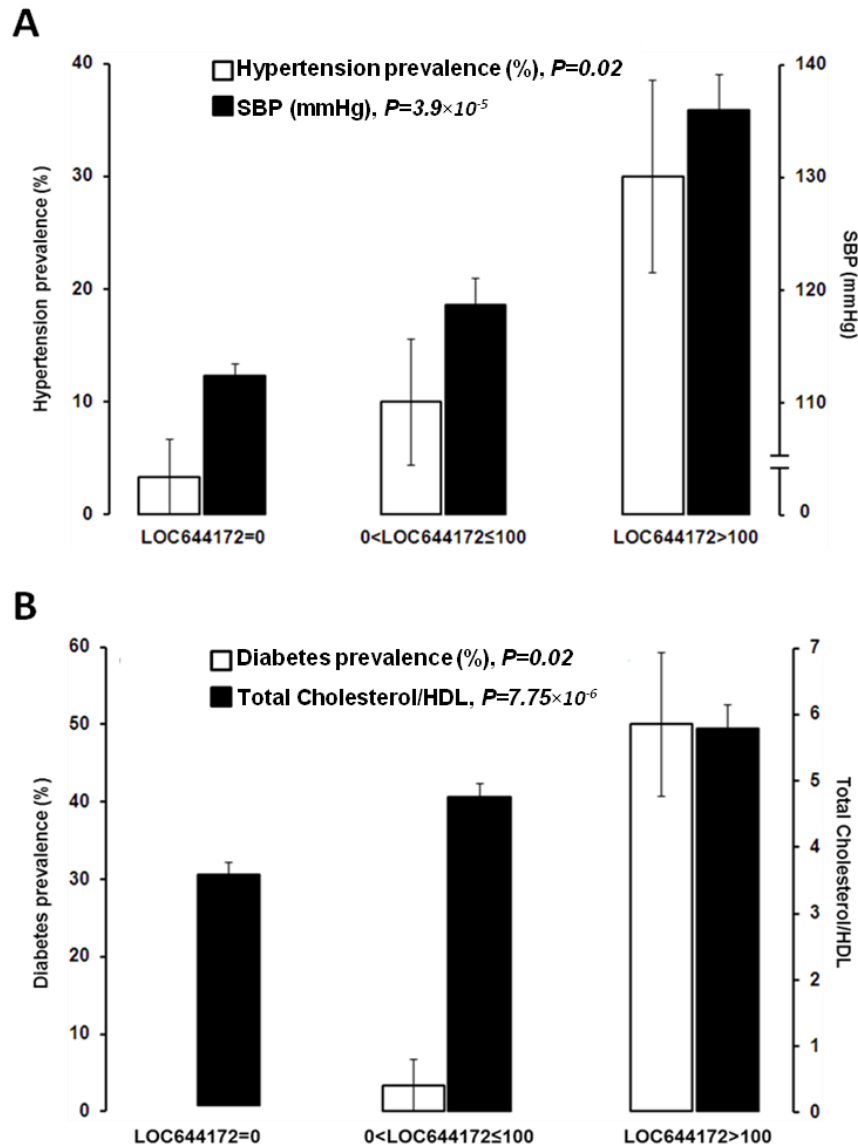


Figure 3. Hypertension and Diabetes prevalence increases with LOC644172 transcript level. **A:** Hypertension prevalence and SBP according to LOC644172 gene expression level group. **B:** Diabetes prevalence and Chol/HDL according to LOC644172 gene expression level group. Results were estimated (using glm) according to LOC644172 gene expression level, each group containing 30 subjects from the CARTaGENE cohort with as much males as females, and adjusted for BMI and age. I bars indicate standard errors. SBP: systolic blood pressure (mmHg); HDL: high density lipoprotein (mmol.L^{-1}); Chol: cholesterol (mmol.L^{-1}).

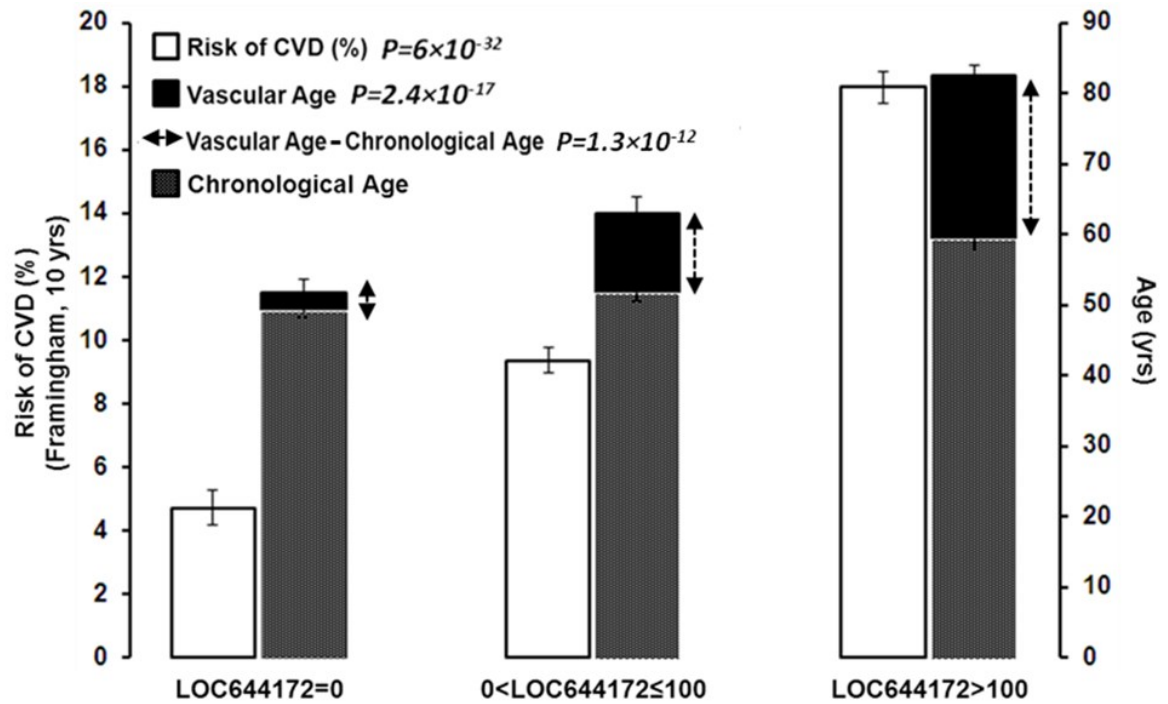


Figure 4. Risk of CVD and vascular aging according to LOC644172 transcript level. Results were estimated (using glm) according to LOC644172 gene expression level, each group containing 30 subjects from the CARTaGENE cohort. Percentage risk of CVD (Framingham, 10 yrs.) [152] strongly increased with expression level group. Similarly, vascular age (23), which reflects the age (yrs.) of the vascular system of the individuals, robustly increased with expression level group. The delta of vascular vs. chronological age augmented notably with expression level group, as well. I bars indicate standard errors.

Data Supplement

Ch	CNVR	Start (bp)	End (bp)	Size (bp)	Dup Freq	Del Freq	P value	OR
1	4	152,555,706	152,586,594	30,888	0.00	56.97	0.8332	0.85
1	5	169,233,974	169,242,401	8,427	0.00	44.85	0.8389	0.89
3	9	98,944,763	98,949,421	4,658	0.00	24.24	0.0758	0.24
3	10	162,137,600	162,142,705	5,105	0.00	9.09	0.7298	0.74
3	11	175,080,747	175,914,479	833,732	1.21	9.09	0.6595	0.49
4	12	9,461,230	9,479,345	18,115	0.00	49.09	0.8452	0.91
4	13	63,669,936	63,671,662	1,726	0.00	26.06	0.5830	1.66
4	14	69,244,434	69,288,142	43,708	0.00	61.82	0.3658	1.35
5	17	17,601,553	17,603,462	1,909	0.00	10.91	0.8332	0.85
6	20	33,938,743	33,942,705	3,962	0.00	9.09	0.8215	1.27
8	24	24,974,443	24,984,250	9,807	0.61	60.00	0.4572	1.36
8	25	39,235,603	39,386,965	151,362	0.00	60.61	0.4468	0.76
9	27	40,476,154	40,619,760	143,606	0.61	9.09	0.9632	1.04
10	32	58,846,314	59,210,991	364,677	0.00	40.61	0.5431	1.42
11	34	49,708,375	49,710,861	2,486	0.00	10.91	0.9217	1.09
11	35	55,447,237	55,453,009	5,772	0.00	39.39	0.1676	0.57
12	37	70,872,675	70,878,027	5,352	0.61	27.88	0.6925	1.23
14	40	20,210,997	20,213,055	2,058	40.00	0.00	0.5165	1.45
15	46	22,373,070	22,383,201	10,131	71.52	1.21	0.3933	0.59
16	50	78,372,440	78,379,675	7,235	0.00	73.33	0.5930	1.19
17	51	18,355,392	18,412,558	57,166	1.21	71.52	0.1768	0.33
17	1	43,656,380	43,658,471	2,091	13.94	5.45	0.0242	4.69
17	2	44,394,412	44,648,367	253,955	13.33	61.82	0.0072	0.38
18	54	51,207,209	51,209,577	2,368	0.00	16.36	0.8543	1.12
19	58	20,596,206	20,667,064	70,858	0.00	11.52	0.2677	0.47
20	59	29,420,352	29,448,955	28,603	0.00	25.45	0.2488	0.58

Table S1. GWAS for hypertension and T2D in French Canadians Positions are given (in base pair, bp) based on Hg19 Assembly. Odds ratios (OR) are given per copy-number gain. The 26 CNVRs were selected based on their length (≥ 1 Kb) and on their variation frequency (Dup or Del above 9% in our FC population sample). Results for the GWAS are adjusted for BMI, sex and age. Associations nominally significant are shown in bold. Ch: chromosome; CNVR: Copy Number Variant Region; bp: base pair; Dup: duplication (copy-number (CN) >2); Del: deletion (CN <2); Freq: frequency (%).

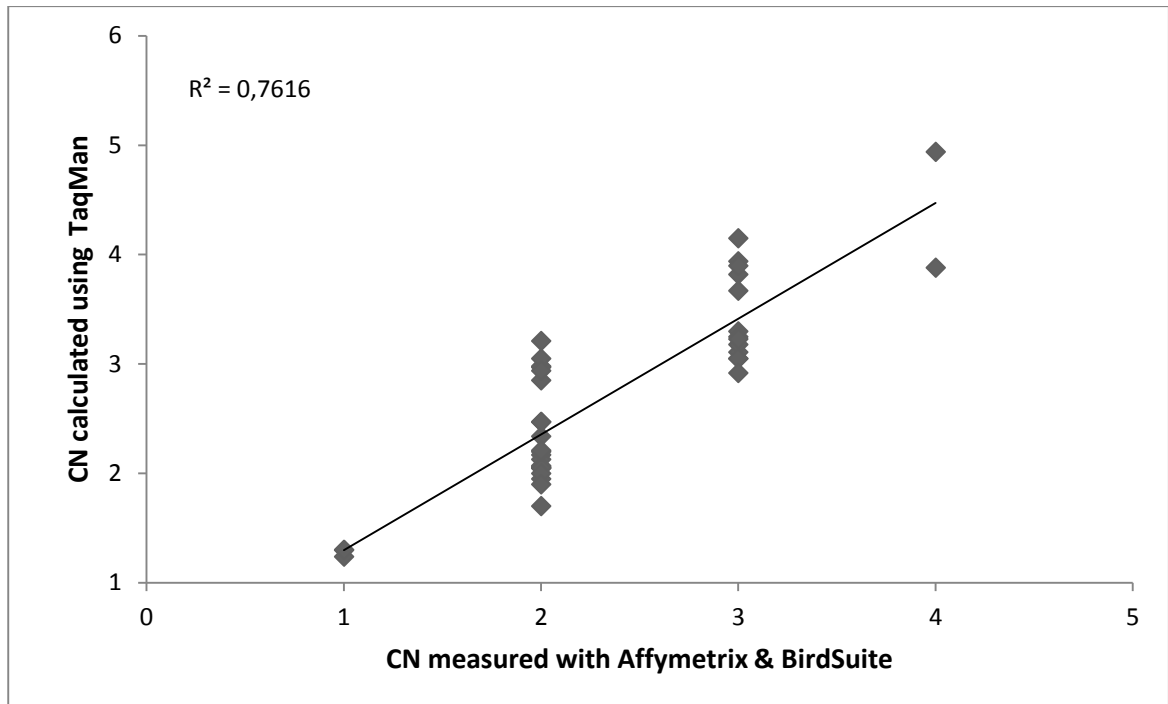


Figure S1. TaqMan Validation. Comparison of copy-number (CN) assessment at the CNVR1 locus using Affymetrix microarrays (&BirdSuite) vs. TaqMan, in 46 FC from the Saguenay-Lac-St-Jean Region of Quebec. The R-squared value is indicated.

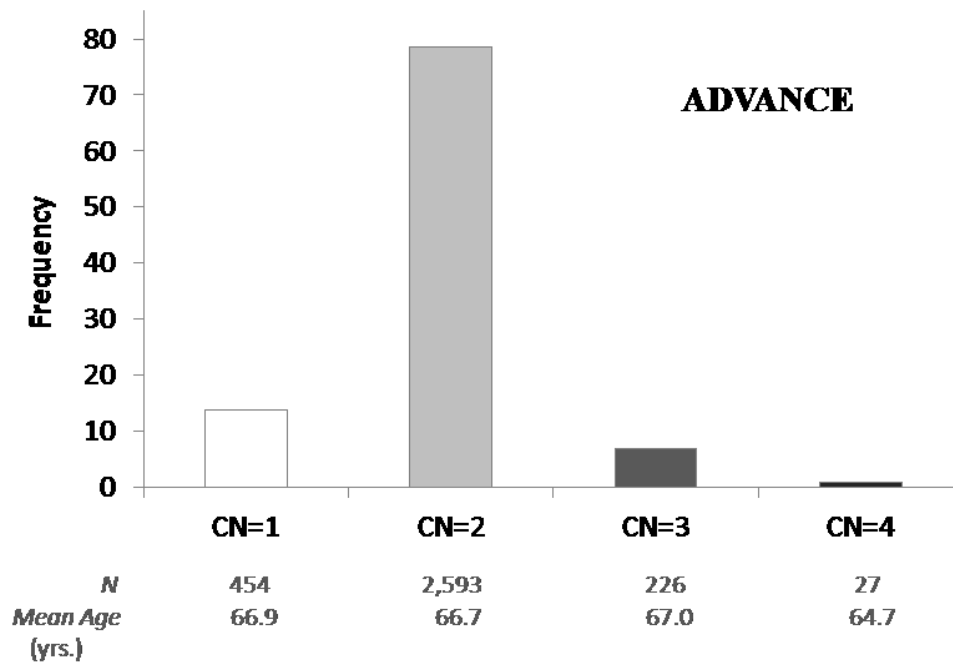
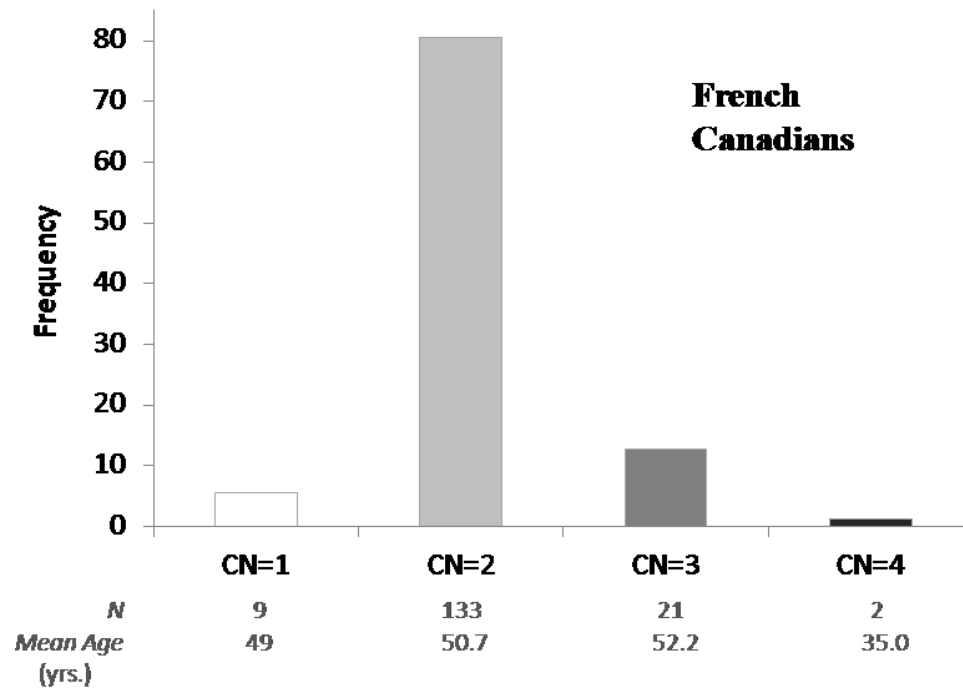


Figure S2. Frequency distribution of copy-number at the CNVR1 locus

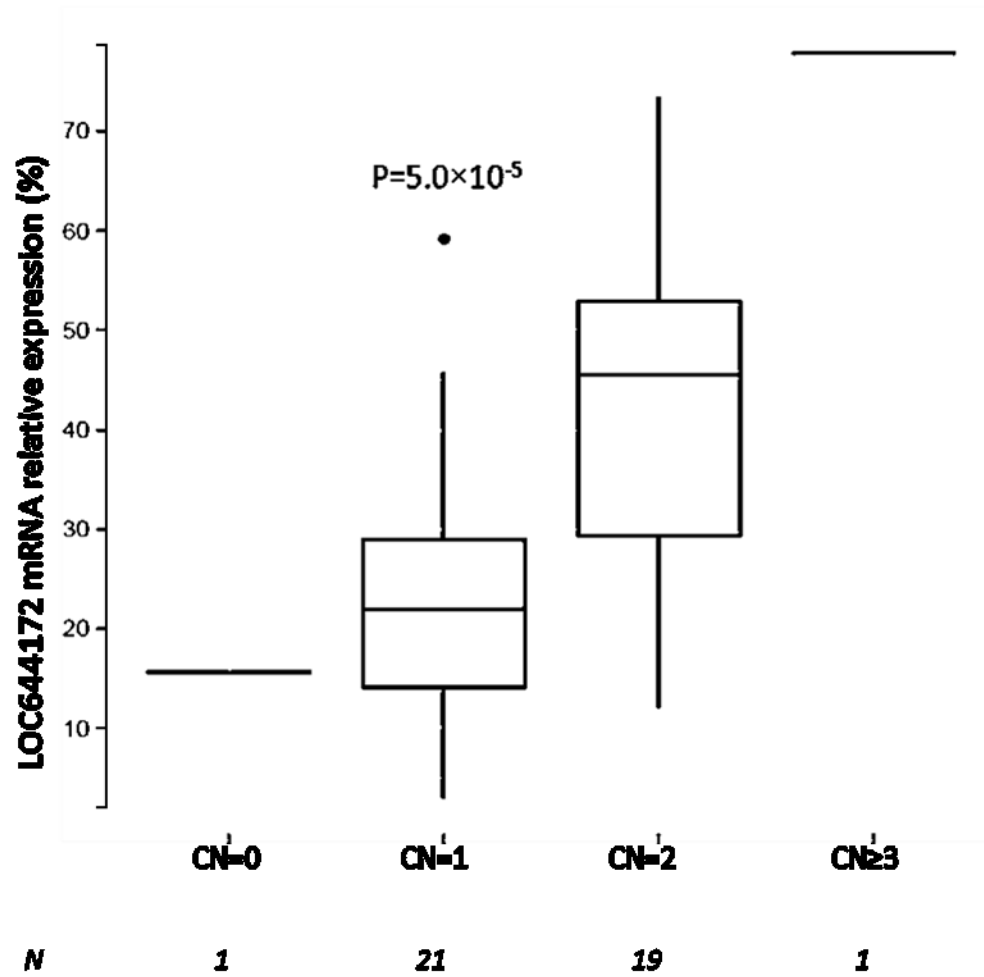


Figure S3. LOC644172 relative expression is modulated by copy-number change at the LOC644172 pseudogene locus. Our data showed a strong dose-dependent effect of the number of copies at the LOC644172 locus (CN) on the relative mRNA levels of LOC644172, in 42 CARTaGENE subjects. The CN effect on LOC644172 relative mRNA levels was estimated using a GLM.

CHAPITRE II : GWAS Identifies CNVRs Contributing to Hypertension in French Canadian Families

CNV risk score and hypertension

Mahine Ivanga, Alena Krajčoviechová, John Raelson, Francois Harvey, François-Christophe Marois-Blanchet, Gilles Godefroid, Renata Cifkova, John Chalmers, Stephen Harrap, Stephen McMahon, Michel Marre, Mark Woodward, Daniel Gaudet, Johanne Tremblay and Pavel Hamet

Key words: GWAS, CNVR, Hypertension, CNV risk score, Age at Hypertension onset, Family-based study

Manuscript submitted to Hypertension Journal

2.1- Abstract

Copy-number variants (CNVs) are known to modulate susceptibility to several complex diseases, however little is known about the effects of CNVs on hypertension. We performed a genome-wide association study (GWAS) of CNV regions (CNVRs) with hypertension in 181 French Canadians (FC) from 21 families from the Chicoutimi/Saguenay-Lac-St-Jean region of Quebec. A weighted CNV risk score based on the effects of selected CNVRs was estimated and corrected for over-fitting by bootstrap resampling methods. Familial transmission was also demonstrated. 6 CNVRs at 3q29, 8p23.1, 14q32.33, 17p11.3, 19q13.12 and 21q21.1 were identified with a mean inheritance of 82%. Area under the receiver characteristic curve (AUC) for each individual CNVR ranged from 0.54 to 0.64, while the AUC of the CNV score was 0.71. The CNV risk score was associated positively with prevalence of hypertension ($P=2\times 10^{-10}$) and negatively with age at hypertension onset ($\beta=-0.57$, $P=1.9\times 10^{-6}$). Decreasing copy-number at 19q13.12 (mapping to *FFAR3*) was associated with greater hypertension prevalence in FC ($P=0.009$) and in 187 individuals from the Czech Post-Monica study ($P=0.03$), odds ratio (OR) per copy-number loss 2.00. At the 3q29 locus (adjacent to *MUC20*), the duplication was pathogenic in FC (OR=3.55, $P=0.04$), while the deletion was protective in 3349 subject from the ADVANCE trial (OR=0.70, $P=0.006$). 6 highly heritable CNVRs contributing to risk/protection of hypertension and its early onset were identified in FC, 2 of which were validated in independent cohorts. The resulting CNV risk score appeared as powerful as Framingham hypertension risk score in predicting hypertension in individuals aged less than 25.

2.2- Introduction

Copy-number variants (CNVs) are sequences of at least 50bp^{1, 2} present in a variable number of copies within the genome and representing a source of genetic and phenotypic variability³. Several studies have implicated CNVs in the aetiology of complex human diseases such as schizophrenia and autism⁴⁻⁶; however, little is known about the effects of CNVs on blood pressure or hypertension (HT).

Population studies have shown variable heritability for hypertension^{7, 8}. Isolated populations with reduced genetic heterogeneity represent a valuable resource for genetic dissection of complex traits⁹⁻¹¹. We conducted a study using the founder French Canadian population of the Chicoutimi/Saguenay-Lac-St-Jean region of Quebec (CSLSJ), which represents a genetically isolated and homogeneous population characterized by large family size and rapid growth between the mid-19th and 20th centuries and having computerized genealogical records going back to the early 17th century¹².

The present study aims to identify copy-number variant regions (CNVRs) associated with increased or decreased risk for HT and to analyse their transmission, in a French Canadian family-based cohort. Replication was tested in two independent cohorts, namely Czech Post-Monica and ADVANCE.

2.3- Methods

2.3.1- Study cohorts

French Canadians (FC)

We selected 181 FC individuals from 21 multigenerational families¹³ from CSLSJ, ascertained by the presence of at least one sib pair affected by HT and dyslipidemia. The present study includes subjects aged from 19 to 92 years (mean age 50) with mean BMI of 27.2 Kg.m⁻². The sample consists of 52.5% hypertensive individuals and 46.4% males.

Czech Post-Monica (CPM)

We selected 187 individuals from the CPM cross-sectional survey¹⁴, ascertained by the presence/absence of albuminuria and/or metabolic syndrome (metS). The present sample consists of 47% hypertensives, 54% males, and includes subjects with mean age 51.8 (ranging from 26.2 to 65.7) and a mean BMI of 28.2 Kg.m⁻².

ADVANCE

We selected 3449 individuals from the Action in Diabetes and Vascular Disease: Peterax and Diamicron MR Controlled Evaluation (ADVANCE) study¹⁵. All subjects were diabetic aged from 47 to 87 (mean age 66.7) with a mean BMI of 30 Kg.m⁻². The sample consists of 60% hypertensive individuals and 64% males.

2.3.2- Phenotypes

French Canadians

Cohort selection and phenotyping have been previously described^{16, 17}. The population living in the CSLSJ is relatively isolated, displaying a founder effect for several conditions and with computerized genealogical records going back to 1608^{12, 13}. Affected sib pair inclusion criteria were HT (systolic blood pressure [SBP]>140 mmHg and/or diastolic

blood pressure [DBP]>90 mmHg measured on two occasions or the use of antihypertensive medication), dyslipidemia (plasma cholesterol ≥ 5.2 mmol/L and/or HDL cholesterol ≤ 0.9 mmol/L or the use of lipid-lowering medication), BMI < 35 kg/m² age 18–55 years, and Catholic French Canadian origin. Exclusion criteria included secondary hypertension, DBP > 110 mmHg in spite of therapy, diabetes mellitus, renal or liver dysfunction, malignancy, pregnancy, and substance abuse. Subsequent to affected sib pair selection, all first- and second-degree relatives aged > 18 years, independent of health status, were invited to participate in the study. The total recruited population included 897 subjects constituting 1,617 sib pairs within 120 families (mean size 7.3 persons; median size 5 persons). The study was approved by the Ethics Committees of Complexe hospitalier de la Sagamie (Chicoutimi, Quebec), Université du Québec à Chicoutimi, and the Centre hospitalier de l'Université de Montréal. All subjects gave informed consent. All phenotyping was performed by trained personnel following standard operating procedures as described previously^{16, 17}. Antihypertensive drugs were withdrawn for at least 1 week before the onset of the phenotyping protocol. A 1-year Framingham score¹⁸ was used to define risk of hypertension.

Czech Post-Monica

We examined a representative population sample (n=3612) aged 25-64 years for the presence of previously described¹⁹ major cardiovascular risk factors as part of the Czech Post-Monica cross-sectional survey (2007-2009). Urinary albumin excretion was determined using immunoturbidimetry in overnight spot urine samples. Albuminuria was defined as an albumin/creatinine ratio (ACR) ≥ 1.9 mg/mmol in males, and ≥ 2.8 mg/mmol in females¹⁹. The metabolic syndrome (metS) was defined according to a joint interim statement²⁰. We selected 94 cases ascertained by presence of albuminuria and matched

them by age and sex to 94 controls using following criteria: a) 63 cases presenting albuminuria and metS were matched to 63 controls without albuminuria or metS, b) 31 cases presenting ACR from the upper extreme of the distribution curve were matched to 31 controls without albuminuria, irrespective of the presence of metS. Individuals with blood pressure $>140/90$ mmHg, fasting plasma glucose >6 mmol/L, current use of antihypertensive and/or glucose lowering medication and manifesting cardiovascular diseases were not eligible as controls. For the present study, HT was defined as SBP >140 mmHg and/or DBP >90 mmHg measured on two occasions or the use of antihypertensive medication.

ADVANCE

The ADVANCE (Action in Diabetes and Vascular Disease: Peterax and Diamicron MR Controlled Evaluation) study is a factorial, multicentre, randomised controlled clinical trial of 11,140 participants recruited from 215 centers in 20 countries¹⁵. The ADVANCE genetic substudy was designed to identify the genomic signatures of complications in type 2 diabetes (T2D) and to determine the pre-symptomatic predictors of the risk of myocardial infarction, stroke and nephropathy in patients with T2D. Hypertension status was defined by history of hypertension and treatment with blood pressure-lowering drugs used at baseline, including: calcium antagonists or angiotensin II receptor antagonists or other ACE inhibitors or perindopril or beta-blockers or other antihypertensive agents or thiazide-like or other diuretics.

2.3.3- The sibling risk ratio (λ_s) in French Canadians

The sibling risk ratio (λ_s) of disease is a standard parameter used to estimate the statistical power for detection of a disease locus. Details about λ_s calculation and results are presented in methods of the supplemental appendix.

2.3.4- CNV detection and selection

182 FC and 188 CPM subjects were genotyped using the Affymetrix 6.0 Human SNP Array, while 3629 ADVANCE subjects were genotyped with Affymetrix 5.0 (~30%) and 6.0 arrays according to the manufacturer's protocol. CNV calls in FC and ADVANCE subjects were made using PennCNV²¹ and further analyzed with the ParseCNV algorithm²². We used the Birdsuite tool set²³ in combination with Plink²⁴ as an independent method of calling CNVs in FC and for calling CNVs in CPM subjects. Details about CNVs calling and CNVR selection are presented in methods of the supplemental appendix.

2.4- Statistical analyses

Analyses of the selected CNVRs were conducted using R software (version 2.15.3).

Association analyses used a Generalized Estimating Equation (GEE) model with a binomial distribution and an exchangeable working correlation matrix to account for familial connection for FC and a Generalized Linear Model (GLM) for other subjects. The models included BMI, sex and age (or the first two components of population stratification PCA, the genotyping chip and sex for ADVANCE subjects) as fixed effects, hypertension status and the copy-number (CN) change/presence of CNV.

Results for SBP and DBP were tested using the above models with a Gaussian distribution where explanatory variables were scaled to obtain the beta coefficients (β) so that one SD variation in CNV/CN change was associated with $\beta \times \text{SD}$ in quantitative trait unit.

The threshold for statistical significance was fixed at 0.05.

2.4.1- Construction of a weighted CNV score.

A *Weighted CNV Risk Score* was constructed in order to assess the combined impact of selected CNVs with independent effects on hypertension risk, in FC. We adapted the formula used by Lin *et al.*²⁵:

$$\text{Weighted CNV Score} = w_i \times \text{CN}_i$$

$$\text{Combined Weighted CNV Score} = w_1 \times \text{CN}_1 + \dots + w_k \times \text{CN}_k$$

For each FC individual, we constructed a weighted CNV score where CN_i was the number of risk copies (0, 1, 2, 3 or 4) or the presence of risk CNV (0 or 1) for the specific locus, and w was the appropriate determined weight. The weighted score was based on the assumption that the k selected CNVs have independent effects on hypertension and contribute to the log risk of hypertension in an additive fashion.

$$= \sum_{i=1}^k \log(\text{OR}_i) \times \text{CN}_i$$

The odds ratios (ORs) were based on CNV association analyses with hypertension performed in the present FC population sample. To avoid overestimations, we used the lower value of the confidence interval for increased effects and the upper value of the confidence interval for protection effects.

Internal validation of the CNV risk score was carried out using a bootstrap method involving 1000 replications to adjust model parameters for potential over-fitting²⁶.

Area under the receiver characteristic curve (AUC) assesses the extent to which predicted risks discriminate between individuals who will develop hypertension from those who will not²⁷, with AUC ranging from 0.5 (total lack of discrimination) to 1.0 (perfect discrimination). Curves were drawn using the ROCplot function in R.

2.4.2- Transmission analysis

We used PEDSTATS²⁸ to check for Mendelian inheritance of copy number. This tool provides summary of genetic data of large pedigrees and allows up to 4 genotypes as entries. Similar to Wang *et al.*,²⁹ we inferred chromosome-specific copy-numbers, namely 1/1, 1/2, 2/2 or 1/3, 2/3 and 3/3 for the given total copy-number 0, 1, 2, 3 and 4, respectively. We assigned chromosome-specific copy-numbers to parents, enabling the prediction of the more plausible chromosome-specific copy-numbers in offspring. Reported Mendelian inconsistencies in pedigrees were considered as *de novo* mutations.

Replication

We investigated the association with hypertension of the CNVRs initially identified in FC, in two additional cohorts namely CPM and ADVANCE.

2.5- Results

The family-based cohort showed increased λ s for hypertension, particularly in younger individuals (*Figure S1, supplemental appendix*). Study design is summarized in *Figure S2* (supplemental appendix) and results are displayed in *Table 1*. Two CNVR duplications were associated with increased risk of hypertension, while one duplication was associated with protection against hypertension. All CNVR deletions were associated with increased susceptibility to hypertension, two of which showed quantitative-effects. The six identified CNVRs, whose frequencies are presented in *Figure S5 (supplemental appendix)*, appear to map principally to human QTLs related to immune and inflammatory responses (*Figure S6, supplemental appendix*). We also analyzed 114 individuals from 18 families for the inheritance of these CNVRs. Families ranged in size from 3 to 24 (average 6), with 72

individuals having both parents genotyped. The identified CNVRs appeared mainly inherited, with a mean inheritance of 82% (*Table 2*).

CNV scores were constructed using CNVR independent effects, the combined CNV risk score ranging from -26% to 36%. The CNV risk score association with prevalence of hypertension reached $P=2\times 10^{-10}$ (genome-wide significant $P < 5\times 10^{-4}$)²² and showed an adjusted AUC of 0.71, making the additive model a more powerful discriminator (*Figure 1*). Deletions at 19q13.12 and 17p11.3 were the best individual discriminators with AUCs of 0.61 and 0.64, respectively.

Associations with quantitative traits are shown in *Table 3*. CN gain at 21q21.1 was associated with younger age of hypertension onset (-8.3 yrs, $P=0.005$). CN loss at 19q13.12 was associated with increased SBP (+5.75 mmHg per CN, $\beta=0.13$, $P=0.02$) as was gain at 3q29 (+8.88 mmHg, $P=0.0006$). Most importantly, the CNV risk score was positively associated with both SBP and DBP, adjusted β of 0.16 ($P=0.02$) and 0.13 ($P=0.04$) respectively, and negatively with age at hypertension onset (adjusted $\beta=-0.57$, $P=1.94\times 10^{-6}$).

In FC individuals younger than 25 years old, the CNV risk score showed an adjusted AUC of 0.94 and the HT Framingham risk score¹⁸ AUC was 0.97 (*Figure 2*).

Replication results are presented in *Figure S8 (supplemental appendix)*. The effect of the 19q13.12 deletion was confirmed in CPM, OR for hypertension per CN loss of 2.08 (95% CI 1.05-4.12, $P=0.03$). Particularly, SBP increased (+4.31 mmHg) per CN loss

($\beta=0.13$, $P=0.04$). In ADVANCE, hypertension prevalence also increased with CN at 3q29 (OR=1.18, 95% CI 1.00-1.40, $P<0.05$), 3q29-loss showing a protective effect (OR=0.70, 95% CI 0.55-0.90, $P=0.006$). However, in ADVANCE, gains at 14q32.33 appeared pathogenic (OR=1.74, 95% CI 1.16-2.61, $P=0.008$).

2.6- Discussion

We report six different loci contributing to hypertension risk, in French Canadian (FC) families from CSLSJ. Deletions 17p11.3 and 19q13.12 showed dose effects with mean OR per CN loss for hypertension of 1.80 and 2, respectively. Moreover, a CN loss on 21.q21.1 was found to be pathogenic with an OR of 2.52 for hypertension. Gains on chromosomes 3q29 and 8p23.1 presented ORs for hypertension of 3.55 and 2.44, respectively, while a duplication 14q32.33 was protective with an OR for hypertension of 0.16.

We evaluated the capacity of the six selected CNVRs to discriminate between individuals with and without hypertension. Deletions at 19q13.12 and 17p11.3 appeared as better discriminators. Although, the independent effect of each individual CNVR is small, combining them notably improved the predictive ability and the resulting CNV risk score appeared robustly associated with prevalence of hypertension and age at onset. Indeed with an additive impact to known risk factors such as smoking or obesity.

Obesity is known to be associated with increased risk for hypertension, a collection of metabolic disorders and cardiovascular diseases^{8, 30}. We identified two CNVRs localized at 14q32.33 and 19q13.12 that map to human body weight QTLs, while two CNVRs at 8p23.1 and 17p11.3 coincide with human blood-pressure QTLs, further suggesting their

implication in the aetiology of hypertension. Almost all reported CNVRs lie within QTLs related to immune and inflammatory responses. Activated immunity and inflammation represent an important trigger of hypertension pathogenesis and associated metabolic traits such as T2D³¹.

Z49979 mRNA (the closest coding sequence to the 21q21.1 deletion) is similar to mRNAs of ETS transcription factor family members, such as ESE-1 which plays an important counter-regulatory role in balancing Ang II-induced vascular inflammation and remodeling³². Zhan *et al.*³³ showed that ESE-1 gene deletion in mice is associated with vascular inflammation and remodeling alterations, along with exaggerated SBP in response to Ang II. We can hypothesize 21q21.1 CN losses may affect blood pressure through similar mechanisms involving modulation of Z49979 expression. We show that loss at 21q21.1 is associated with increased prevalence of hypertension and early-onset hypertension in FC.

CNVRs at 17q11.3 overlap several genes which could potentially have an impact on blood pressure including *LGALS9C*, which is similar to galectin-9 (GAL-9) which was shown to have anti-inflammatory properties in mice³⁴. Deletions of *LGALS9C* might lead to increased inflammatory responses involved in hypertension pathogenesis³¹. Our results reveal that decreased CN at 17q11.3 is associated with increased prevalence of hypertension in FC.

The CNVR at 14q32.33 maps to the Ig heavy chain variable region locus. Pandey³⁵ hypothesized that immunoglobulin GM genes - genetic markers of IgG heavy chains located on chromosome 14 - are functional risk/protective factors for Alzheimer's disease.

Similarly, our data show that duplication at 14q32.33 is protective against hypertension. Conversely, in ADVANCE diabetic subjects, the duplication at 14q32.33 appears to be pathogenic. According to Glessner *et al.*²², CNVs that map to the Ig heavy chain variable region, which are frequently found in multiple studies, are not factual. However, Schooling *et al.*³⁶ reported a case of paradoxical association where alanine transaminase is positively associated with ischemic heart disease (IHD) related to diabetes, but negatively associated with IHD unrelated to diabetes. Some aspects of IgG heavy chain function might underlie the opposite associations with diabetes and non-diabetes associated hypertension.

The 8p23.1 region is characterized by complex rearrangements and duplications³⁷. A pseudogene of family with sequence similarity 90, member A25 (*FAM90A25P*) maps to this region. Bosch *et al.*³⁷ have shown that *FAM90A* is exclusive to primates and represents an example of gene expansion by duplication across the hominoid lineage. In addition, *FAM90A* genes have been shown to have highly variable in CN between populations and between individuals³⁸. *FAM86B*, localized within this region, also belongs to a family of variants whose duplication is observed only in primates³⁹. A pseudogene of beta defensin 109 (*DEFB109P1*) is also found on chromosome 8p23.1. Defensins play important roles in adaptative immunity and beta-defensin is among the main host defenses produced by organs and tissues against microbial infection⁴⁰. In humans, the 8p22-23 defensin locus seems to have evolved by successive duplications and subsequent divergence to give a cluster of diverse paralogous genes⁴¹. Gains on chromosome 8p23.1 are associated with increased risk for hypertension in our FC population sample. Taken together, we can hypothesize the observed association may be the consequence of genetic adaptations specific to primates and particularly humans.

The free fatty acid receptor 3 gene, *FFAR3* (or *GPR41*), surrounds the 19q13.12 deletion. Propionate is a short-chain fatty acid (SCFA) produced by the gut microbiota and Pluznick *and al.*⁴² demonstrated that in mice, a portion of propionate hypotensive effect is mediated by *Ffar3*. Most importantly they identified a cross-talk between the gut microbiota and the renal-cardiovascular system. Our results show that prevalence of hypertension rises with CN loss at this 19q13.12 locus, particularly, SBP increases with CN loss. We validated the dose effect of 19q13.12 on hypertension in the CPM sample, ascertained by presence/absence of albuminuria and/or metS. It is likely that the 19q13.12 CNVR affects blood pressure in humans, in a quantitative way, via modulation of *FFAR3* expression in response to SCFAs stimulation.

Some segments of the CNVR on chromosome 3q29 overlap the *MUC20* gene. We show that the duplication at 3q29 is associated with increased risk for hypertension and higher SBP in FC. We validated that CN at 3q29 increased with hypertension prevalence, in ADVANCE diabetic subjects in which the 3q29-loss protective effect was emphasized. These effects may be mediated by *MUC20* which is associated with renal damaged and repair⁴³ and up-regulated in the renal tissues of patients with IgA nephropathy⁴⁴, whose early-onset symptoms include hypertension⁴⁵. Additional experiments are needed to clarify these different hypotheses.

We then investigated the inherited predisposition to hypertension, in our FC familial cohort. The three most frequent CNVRs, namely deletions at 17p11.3, 19q13.12 and 21.q21.1, appeared to be highly inherited. Nguyen *et al.*⁴⁶ argued that CNVs might be driven to higher frequency as a consequence of their adaptive benefit, in humans. This is well illustrated by the salivary amylase 1 genes (*AMY1*) with extensive CN variability,

reflecting positive selection imposed by diet on amylase copy-number during evolution⁴⁷. Similarly, *FFAR3* CN variation at 19q13.12 might be derived from dietary habits involving SCFA production by the gut microbiota. Gains on chromosome 3q29 were usually inherited, while losses appeared to be mainly *de novo*. Variations on chromosome 8p23.1 seemed principally inherited, with gains showing greater inheritance. Finally, most of the variations on chromosome 14q32.33 appeared inherited, although only one gain was available for the transmission analysis.

The Framingham score¹⁸ developed for predicting 1-year risk for new onset of hypertension includes BMI, sex, age, smoking, DBP, SBP and parental hypertension history. The CNV risk score was as powerful as the Framingham HT risk score, in predicting hypertension in individuals aged less than 25. The CNV risk score is likely to add precision over parental history on predicting early-onset hypertension and benefits over Framingham HT risk score in the identification of presymptomatic at risk individuals.

The results of this study require further validation. A limitation of the study is that the methods of CNV detection are limited to the detection of total copy-number, rather than copy-number in each of the two homologous chromosomes²⁹. Some identified associations may be restricted to FC from CSLSJ, since CNV distributions have been shown to be highly population-specific⁴⁸. However, Chen *et al.*⁴⁸ demonstrated that more frequent CNVRs are more likely to be common to many populations, and indeed two identified associations were also observed in two independent validation cohorts. Although precise boundary determinations are subject to some technical uncertainty, the CNVR-approach most likely reflects recurrent CN changes at the same locus. The major strength of the present analysis is the use of two different algorithms of detection and three different tests

for selection of true associations with hypertension. Moreover, we used a homogeneous family-based cohort with increased λ_s , enabling identification of Mendelian loci contributing to hypertension.

Perspectives

We identified six highly heritable loci, varying in copy-number and contributing to risk/protection of hypertension in our FC population sample, two of which were associated with higher SBP and further validated in two additional independent cohorts. The resulting combined CNV risk score is strongly associated with prevalence of hypertension and age at diagnosis of hypertension. This CNV risk score appeared as powerful as Framingham hypertension risk score in predicting hypertension in young people. This study provides support for an additive model in which multiple CNVRs appear involved in the risk for essential hypertension. Our findings suggest these CNVs influence genes mainly related to immune and inflammatory responses and renal damaged and repair, of which some are exclusive to primates and/or may have been selected through the hominoid lineage, based on their adaptive benefit.

Sources of Fundings

This work was supported by grants from Genome Quebec (“Genomics research in human health - translational stream”) and CIHR (ISO-106797 and MOP-93629), and by Prognomix Inc. and Les Laboratoires Servier. M.I. received doctoral research awards from La Société Québécoise d’Hypertension Artérielle and CIHR with Canadian Hypertension Society & Pfizer Canada Inc (area of hypertension prevention).

Disclosures

P.H. and J.T. are members of the Collège International de recherche Servier and were consultants for Prognomix Inc. J.C., M.W., S.H., M.M. received speaker fees from Les Laboratoires Servier.

2.7- References

1. Alkan C, Coe BP, Eichler EE. Genome structural variation discovery and genotyping. *Nat Rev Genet.* 2011;12:363-376
2. Valsesia A, Mace A, Jacquemont S, Beckmann JS, Kutalik Z. The growing importance of cnvs: New insights for detection and clinical interpretation. *Front Genet.* 2013;4:92
3. Fanciulli M, Petretto E, Aitman TJ. Gene copy number variation and common human disease. *Clin Genet.* 2010;77:201-213
4. Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature.* 2008;455:237-241
5. Pinto D, Pagnamenta AT, Klei L, et al. Functional impact of global rare copy number variation in autism spectrum disorders. *Nature.* 2010;466:368-372
6. Stefansson H, Meyer-Lindenberg A, Steinberg S, et al. Cnvs conferring risk of autism or schizophrenia affect cognition in controls. *Nature.* 2014;505:361-366
7. El Shamieh S, Visvikis-Siest S. Genetic biomarkers of hypertension and future challenges integrating epigenomics. *Clin Chim Acta.* 2012;414:259-265
8. Pausova Z, Gossard F, Gaudet D, Tremblay J, Kotchen TA, Cowley AW, Hamet P. Heritability estimates of obesity measures in siblings with and without hypertension. *Hypertension.* 2001;38:41-47
9. Peltonen L. Positional cloning of disease genes: Advantages of genetic isolates. *Hum Hered.* 2000;50:66-75
10. Shifman S, Darvasi A. The value of isolated populations. *Nat Genet.* 2001;28:309-310
11. Wright AF, Carothers AD, Pirastu M. Population choice in mapping genes for complex diseases. *Nat Genet.* 1999;23:397-404
12. Bouchard G, Roy R, Casgrain B, Hubert M. [population files and database management: The balsac database and the ingres/ingrid system]. *Hist Mes.* 1989;4:39-57
13. Hamet P, Merlo E, Seda O, et al. Quantitative founder-effect analysis of french canadian families identifies specific loci contributing to metabolic phenotypes of hypertension. *Am J Hum Genet.* 2005;76:815-832
14. Cifkova R, Skodova Z, Bruthans J, Holub J, Adamkova V, Jozifova M, Galovcova M, Wohlfahrt P, Krajcoviechova A, Petrzilkova Z, Lanska V. Longitudinal trends in cardiovascular mortality and blood pressure levels, prevalence, awareness, treatment, and control of hypertension in the czech population from 1985 to 2007/2008. *Journal of hypertension.* 2010;28:2196-2203
15. Patel A, MacMahon S, Chalmers J, et al. Effects of a fixed combination of perindopril and indapamide on macrovascular and microvascular outcomes in patients with type 2 diabetes mellitus (the advance trial): A randomised controlled trial. *Lancet.* 2007;370:829-840
16. Orlov SN, Pausova Z, Gossard F, Gaudet D, Tremblay J, Kotchen T, Cowley A, Larochelle P, Hamet P. Sibling resemblance of erythrocyte ion transporters in french-canadian sibling-pairs affected with essential hypertension. *J Hypertens.* 1999;17:1859-1865
17. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: Analysis and visualization of ld and haplotype maps. *Bioinformatics.* 2005;21:263-265

18. Parikh NI, Pencina MJ, Wang TJ, Benjamin EJ, Lanier KJ, Levy D, D'Agostino RB, Sr., Kannel WB, Vasan RS. A risk score for predicting near-term incidence of hypertension: The framingham heart study. *Ann Intern Med.* 2008;148:102-110
19. Warram JH, Gearin G, Laffel L, Krolewski AS. Effect of duration of type i diabetes on the prevalence of stages of diabetic nephropathy defined by urinary albumin/creatinine ratio. *Journal of the American Society of Nephrology : JASN.* 1996;7:930-937
20. Alberti KG, Eckel RH, Grundy SM, Zimmet PZ, Cleeman JI, Donato KA, Fruchart JC, James WP, Loria CM, Smith SC, Jr. Harmonizing the metabolic syndrome: A joint interim statement of the international diabetes federation task force on epidemiology and prevention; national heart, lung, and blood institute; american heart association; world heart federation; international atherosclerosis society; and international association for the study of obesity. *Circulation.* 2009;120:1640-1645
21. Wang K, Li M, Hadley D, Liu R, Glessner J, Grant SF, Hakonarson H, Bucan M. Penncnv: An integrated hidden markov model designed for high-resolution copy number variation detection in whole-genome snp genotyping data. *Genome Res.* 2007;17:1665-1674
22. Glessner JT, Li J, Hakonarson H. Parsecnv integrative copy number variation association software with quality tracking. *Nucleic Acids Res.* 2013;41:e64
23. Korn JM, Kuruvilla FG, McCarroll SA, et al. Integrated genotype calling and association analysis of snps, common copy number polymorphisms and rare cnvs. *Nat Genet.* 2008;40:1253-1260
24. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC. Plink: A tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 2007;81:559-575
25. Lin X, Song K, Lim N, et al. Risk prediction of prevalent diabetes in a swiss population using a weighted genetic score--the colaus study. *Diabetologia.* 2009;52:600-608
26. Steyerberg EW, Bleeker SE, Moll HA, Grobbee DE, Moons KG. Internal and external validation of predictive models: A simulation study of bias and precision in small samples. *J Clin Epidemiol.* 2003;56:441-447
27. Cook NR. Use and misuse of the receiver operating characteristic curve in risk prediction. *Circulation.* 2007;115:928-935
28. Wigginton JE, Abecasis GR. Pedstats: Descriptive statistics, graphics and quality assessment for gene mapping data. *Bioinformatics.* 2005;21:3445-3447
29. Wang K, Chen Z, Tadesse MG, Glessner J, Grant SF, Hakonarson H, Bucan M, Li M. Modeling genetic inheritance of copy number variations. *Nucleic Acids Res.* 2008;36:e138
30. Murphy NF, MacIntyre K, Stewart S, Hart CL, Hole D, McMurray JJ. Long-term cardiovascular consequences of obesity: 20-year follow-up of more than 15 000 middle-aged men and women (the renfrew-paisley study). *Eur Heart J.* 2006;27:96-106
31. Xiao L, Liu Y, Wang N. New paradigms in inflammatory signaling in vascular endothelial cells. *Am J Physiol Heart Circ Physiol.* 2014;306:H317-325
32. Oliver JR, Kushwah R, Hu J. Multiple roles of the epithelium-specific ets transcription factor, ese-1, in development and disease. *Laboratory investigation; a journal of technical methods and pathology.* 2012;92:320-330

33. Zhan Y, Yuan L, Kondo M, Oettgen P. The counter-regulatory effects of ese-1 during angiotensin ii-mediated vascular inflammation and remodeling. *Am J Hypertens.* 2010;23:1312-1317
34. Tsuboi Y, Abe H, Nakagawa R, Oomizu S, Watanabe K, Nishi N, Nakamura T, Yamauchi A, Hirashima M. Galectin-9 protects mice from the shwartzman reaction by attracting prostaglandin e2-producing polymorphonuclear leukocytes. *Clin Immunol.* 2007;124:221-233
35. Pandey JP. Immunoglobulin gm genes as functional risk and protective factors for the development of alzheimer's disease. *J Alzheimers Dis.* 2009;17:753-756
36. Schooling CM, Kelvin EA, Jones HE. Alanine transaminase has opposite associations with death from diabetes and ischemic heart disease in nhanes iii. *Ann Epidemiol.* 2012;22:789-798
37. Bosch N, Caceres M, Cardone MF, Carreras A, Ballana E, Rocchi M, Armengol L, Estivill X. Characterization and evolution of the novel gene family fam90a in primates originated by multiple duplication and rearrangement events. *Human molecular genetics.* 2007;16:2572-2582
38. Bosch N, Escaramis G, Mercader JM, Armengol L, Estivill X. Analysis of the multi-copy gene family fam90a as a copy number variant in different ethnic backgrounds. *Gene.* 2008;420:113-117
39. Cloutier P, Lavallee-Adam M, Faubert D, Blanchette M, Coulombe B. A newly uncovered group of distantly related lysine methyltransferases preferentially interact with molecular chaperones to regulate their activity. *PLoS genetics.* 2013;9:e1003210
40. Schutte BC, McCray PB, Jr. [beta]-defensins in lung host defense. *Annual review of physiology.* 2002;64:709-748
41. Semple CA, Rolfe M, Dorin JR. Duplication and selection in the evolution of primate beta-defensin genes. *Genome biology.* 2003;4:R31
42. Pluznick JL, Protzko RJ, Gevorgyan H, et al. Olfactory receptor responding to gut microbiota-derived signals plays a role in renin secretion and blood pressure regulation. *Proceedings of the National Academy of Sciences of the United States of America.* 2013;110:4410-4415
43. Li G, Zhang H, Lv J, Hou P, Wang H. Tandem repeats polymorphism of muc20 is an independent factor for the progression of immunoglobulin a nephropathy. *Am J Nephrol.* 2006;26:43-49
44. Waga I, Yamamoto J, Sasai H, Munger WE, Hogan SL, Preston GA, Sun HW, Jennette JC, Falk RJ, Alcorta DA. Altered mrna expression in renal biopsy tissue from patients with iga nephropathy. *Kidney Int.* 2003;64:1253-1264
45. Wang W, Chen N. Treatment of progressive iga nephropathy: An update. *Contrib Nephrol.* 2013;181:75-83
46. Nguyen DQ, Webber C, Hehir-Kwa J, Pfundt R, Veltman J, Ponting CP. Reduced purifying selection prevails over positive selection in human copy number variant evolution. *Genome Res.* 2008;18:1711-1723
47. Santos JL, Saus E, Smalley SV, Cataldo LR, Alberti G, Parada J, Gratacos M, Estivill X. Copy number polymorphism of the salivary amylase gene: Implications in human nutrition research. *J Nutrigenet Nutrigenomics.* 2012;5:117-131
48. Chen W, Hayward C, Wright AF, Hicks AA, Vitart V, Knott S, Wild SH, Pramstaller PP, Wilson JF, Rudan I, Porteous DJ. Copy number variation across european populations. *PLoS One.* 2011;6:e23087

Novelty and significance

What Is New?

- 6 CNVRs were identified in a GWAS of hypertension performed in a family-based cohort and a CNV risk score was calculated.
- The predictive power of hypertension-associated CNVRs was evaluated.

What Is Relevant?

- The effects of 2 CNVRs were confirmed in independent cohorts; however results of this study require further validation.
- The CNV risk score reached genome-wide significance and was a statistically significant predictor of hypertension in this FC population sample.
- The CNV risk score was as powerful discriminator as Framingham hypertension risk score in individuals less than 25 years of age.

Summary

The association of CNVs with hypertension risk is consistent with an effect on age at hypertension onset.

<i>CNVR</i>	<i>Cytoband location</i>	<i>Event</i>	<i>Freq. in cases</i>	<i>Freq. in controls</i>	<i>OR per CN gain/loss (95% CI)</i>	<i>P-value</i>	<i>OR per event (95% CI)</i>	<i>P-value</i>	<i>Closest Gene(s)</i>
chr3:195442787-195446922	3q29	Gain	16%	6%	1.54 (0.61,3.86)	0.360	3.55 (1.03, 12.17)	0.044	SDHAP2, MIR570, MUC20, MUC4
chr8:12241505-12247486	8p23.1	Gain	16%	5%	1.18 (0.77, 1.81)	0.453	2.44 (1.11, 5.38)	0.027	[FAM66A,..., LOC100506990]*
chr14:106893273-106918649	14q32.33	Gain	3%	12%	0.70 (0.32, 1.51)	0.362	0.16 (0.04, 0.55)	0.003	AbParts
chr17:18426461-18465221	17p11.3	Loss	69%	46%	1.80 (1.16-2.79)	0.008	1.48 (0.89,2.45)	0.125	[KRT16P1,..., CCDC144B]**
chr19:35855353-35861848	19q13.12	Loss	39%	19%	2.00 (1.19-3.36)	0.009	1.88 (1.04, 3.41)	0.036	FFAR3 (or GRP41)
chr21:23655764-23664658	21q21.1	Loss	18%	7%	NA	NA	2.52 (1.69, 3.75)	5.67×10⁻⁰⁶	Z49979, Z49981

Table1. GWAS identified six CNV regions nominally associated with hypertension in French Canadians. Details of Copy-Number (CN) Variant Regions (CNVR), cytoband location, event, frequency (Freq.) in cases and controls and P-values are indicated. Positions are given relative to UCSC Hg19. Results were adjusted for BMI, sex and age. Nominally significant results are shown in bold. *FAM66A, LOC649352, DEFB109P1, FAM90A25P, FAM86B2, LOC729732 LOC100506990. **KRT16P1, LGALS9C, USP32P2, FAM106A, CCDC144.

CNV locus	Cytoband location	Duplication inheritance	Deletion inheritance
chr3:195442787-195446922	3q29	100% (13/13)	25% (2/8)
chr8:12241505-12247486	8p23.1	80% (8/10)	72.7% (16/22)
chr14:10689327-106918649	14q32.33	100% (1/1)	80% (12/15)
chr17:18426461-18465221	17p11.3	-	100% (44/44)
chr19:35855353-35861848	19q13.12	-	82.3% (14/17)
chr21:23655764-23664658	21q21.1	-	100% (9/9)

Table 2. Percentage of Inheritance of the identified CNVRs. Positions are given relative to UCSC Hg19. Results are based on 72 individuals having both parents genotyped.

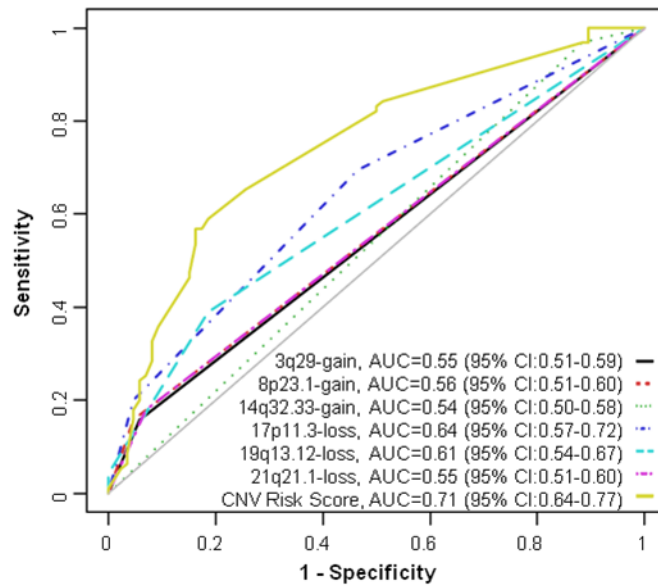


Figure 1. ROC plots for models containing each CNVR effect and the resulting CNV Risk Score. The CNV risk score was constructed using the six identified CNVR effects and adjusted for over-fitting.

Event	Age at Hypertension Onset		SBP		DBP	
	β	<i>P-value</i>	β	<i>P-value</i>	β	<i>P-value</i>
Gains						
3q29	-0.0318	0.7645	0.1233	0.0006	0.1214	0.0887
8p23.1	-0.0159	0.8617	0.0427	0.2377	0.0210	0.7558
14q32.33 [§]	NA	NA	-0.0948	0.1136	-0.0357	0.1523
Losses						
17p11.3*	0.0515	0.5663	0.0334	0.3477	0.0531	0.1921
19q13.12*	-0.0801	0.4626	0.1297	0.0232	0.1595	0.0517
21q21.1	-0.1726	0.0051	0.0697	0.3984	0.0651	0.3600
CNV risk score	-0.5687	1.94×10⁻⁶	0.1585	0.0252	0.1338	0.0427

Table 3. Associations with quantitative traits related to blood pressure. The beta coefficients, β , are given per *copy-number loss or presence of risk/protective CNV or score increase. The CNV risk score was constructed using the six identified CNVR effects and adjusted for over-fitting. Results of association of gains, losses or CNV risk score with drug corrected SBP and DBP were adjusted for BMI, sex and age. [§]Age at hypertension onset information was only available for 1 gain-carrier. Significant results are shown in bold.

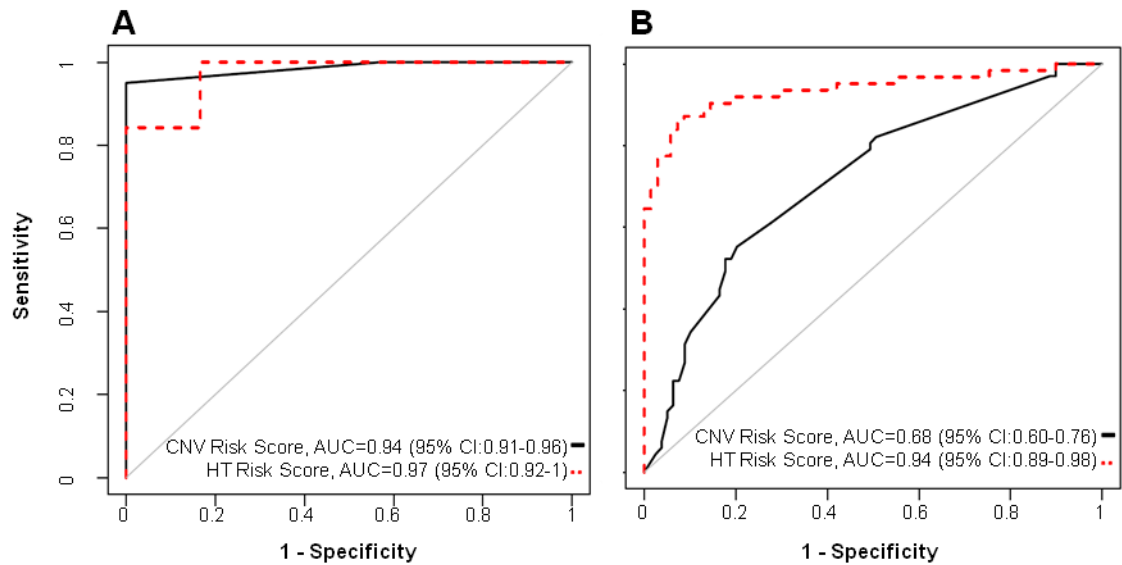


Figure 2. The CNV risk score was as powerful as Framingham hypertension risk score in predicting hypertension in individuals less than 25 years of age. The CNV Risk Score was constructed using the six identified CNVR effects and adjusted for over-fitting. The HT Risk Score was defined using a 1-year Framingham score¹⁸. A- Individuals aged (or diagnosed hypertensive at) less than 25. B- Individuals aged 25 or older, excluding those diagnosed hypertensive at less than 25 years old.

DISCUSSION

Cette étude présente, d'une part, une analyse pangénomique identifiant deux nouveaux loci de région de CNVs (CNVRs) à effets quantitatifs sur 17q21.31, en association avec l'hypertension et le diabète de type 2 chez les individus SLSJ et en association avec l'hypertension chez les diabétiques ADVANCE. L'approche statistique adoptée, nous a permis de souligner le rôle essentiel du locus CNVR1 sur l'insulino-résistance, la prévalence de l'hypertension et du diabète et le risque cardiovasculaire. Par la suite, des études fonctionnelles effectuées sur la cohorte CARTaGENE ont mené à l'identification du biomarqueur, *LOC644172*, régulé par CNVR1 et impliqué dans les complications du diabète de type 2.

D'autre part, nous avons identifié six régions de CNVs associées à la prévalence de l'hypertension chez des familles du SLSJ et évalué leur héritabilité. Plus particulièrement, un nouveau locus de CNVR à effet quantitatif identifié chez les individus SLSJ, au niveau du gène *FFAR3*, et négativement associé à la prévalence de l'hypertension a été validé dans la cohorte indépendante MONICA. De plus, au locus 3q29 (proche de *MUC20*), la duplication était pathogénique chez les sujets SLSJ alors que la délétion est apparue protectrice chez les sujets ADVANCE. Cette étude comprend également l'une des premières analyses utilisant des courbes de caractéristique de performance (courbes ROC) permettant d'évaluer la puissance prédictive des CNVs sur l'hypertension.

Les stratégies d'analyse

Les maladies complexes sont influencées par de nombreux facteurs de risque, à la fois génétiques et environnementaux. Cependant, les études d'association à l'échelle du génome entier évaluent communément l'effet d'un seul variant génétique à la fois, ignorant les effets des autres variants ainsi que les interactions entre variants. Cette approche aurait tendance à sous-estimer l'ampleur réelle de l'effet de certains variants génétiques sur les phénotypes qualitatifs [147]. En effet, un locus de CNVR, considéré isolément, ne peut suffire à expliquer la majorité des cas d'hypertension artérielle observés. Nos résultats montrent que combiner les effets de plusieurs loci de CNVRs, permet de distinguer avec une meilleure sensibilité/spécificité, les cas des contrôles (courbes ROC). De plus, le fait d'intégrer deux loci de CNVRs et leur interaction dans le même modèle d'association, nous a permis de révéler l'effet notable d'un locus de CNVR sur l'hypertension et le diabète de type 2, et de ce fait sur le risque cardiovasculaire.

Si les dernières générations de biopuces de SNPs comprennent des sondes spécifiques à la détection de CNVs [98], il faut noter que les études initiales utilisaient des biopuces de SNPs non-élaborées à la base pour la détection de la variabilité dans le nombre de copies [97]. Ainsi, l'utilisation de TagSNPs, pour capturer l'effet phénotypique des CNVs, pourrait également expliquer la difficulté des précédentes études d'association, réalisées à l'échelle génomique, à identifier de nouveaux CNVs influençant la susceptibilité à l'hypertension artérielle. En effet, de nombreux déséquilibres de liaison entre des SNPs présents sur les plateformes de génotypage commerciales et des CNVs communs ont été détectés. Ce qui implique qu'une proportion importante de l'effet phénotypique des CNVs

analysés par cette méthode, aurait déjà été indirectement capturée par la plupart des études d'association pangénomiques utilisant des SNPs [126,148,149]. De plus, cette approche TagSNPs cible généralement les délétions ou les duplications, négligeant l'impact de l'effet quantitatif exercé par de nombreux CNVs sur la variation phénotypique humaine.

La cohorte de découverte : une cohorte familiale

L'observation d'une agrégation familiale constitue une première étape vers l'estimation de la composante génétique d'une pathologie donnée. Plus particulièrement, les cohortes incluant des cas de développement précoce de maladies complexes seraient enrichies en marqueurs génétiques hérités [150,151]. De plus, ce type de cohortes faciliterait la détection de marqueurs génétiques à effets relativement importants [152,153]. Dans ce contexte, l'utilisation de la cohorte familiale du Saguenay-Lac Saint-Jean (SLSJ) comme cohorte de découverte a permis l'identification de loci de CNVRs hautement hérités et associés à une protection ou un risque, relativement accru(e), contre/pour l'hypertension artérielle. En effet, la présente étude montre que cette population, essentiellement homogène, se caractérise par un risque relatif familial (λ_s) lié à l'hypertension qui double, chez les sujets de 45 ans et moins ($\lambda_s \sim 4$) comparativement à ceux de plus de 45 ans ($\lambda_s \sim 2$). Plus particulièrement, la prévalence de l'hypertension est environ 27 fois plus grande chez les moins de 25 ans dans notre cohorte, comparativement à la même classe d'âge au SLSJ (voir figure S1- Supplemental data- Chapter II). Un risque relatif familial élevé reflète une susceptibilité génétique supérieure à la population générale. Ainsi, nous avons identifié six CNVRs hautement hérités et associés à l'hypertension chez les sujets SLSJ. Le score de risque des effets pondérés de ces CNVRs est apparu positivement et étroitement relié à la

prévalence de l'hypertension ($P=2\times 10^{-10}$) et négativement associé à l'âge de diagnostic de l'hypertension ($\beta=-0,57$, $P=1,94\times 10^{-06}$). Ce score de risque présente une statistique C de l'aire sous la courbe de 0,71 pour l'hypertension, dans notre population SLSJ. Le score de risque Framingham pour l'hypertension artérielle (HTA) [154] est basé sur l'histoire familiale et plusieurs critères cliniques incluant les pressions artérielles systolique et diastolique. Chez les sujets SLSJ, la performance du score de risque pondéré des CNVs est similaire à celle du score de risque HTA Framingham, chez les moins de 25 ans. Le score de risque pondéré des effets des CNVs, développé dans cette étude, est plus susceptible que le score de risque HTA Framingham de permettre le diagnostic présymptomatique des sujets à haut risque de développer l'hypertension artérielle.

La répliation des résultats dans des cohortes indépendantes

Des associations détectées dans la cohorte de découverte du SLSJ, à 3 loci de CNVR (17q21.31, 19q13.12 et 3q29), ont pu être effectivement répliquées dans la cohorte ADVANCE ou MONICA.

Le nombre de copies au locus CNVR1, sur le chromosome 17q21.31, est apparue positivement associée à la prévalence du phénotype diabète et hypertension chez les individus SLSJ et à la prévalence de l'hypertension chez les diabétiques ADVANCE. Plus particulièrement, la duplication au locus CNVR1 a également été reliée à une résistance à l'insuline élevée (SLSJ), un développement relativement plus précoce du diabète et de ce fait à une durée plus longue de la pathologie ainsi qu'à un risque cardiovasculaire accru (ADVANCE). En effet, l'hyperinsulinisme secondaire à l'insulino-résistance précède, de

plusieurs années, l'insulino-déficience à l'origine de l'hyperglycémie du diabète de type 2. D'autre part, une hyperglycémie chronique et prolongée affecte dramatiquement l'intégrité de nombreux tissus cibles incluant le tissu vasculaire. En conséquence, l'hypertension artérielle des porteurs de la duplication découlerait de l'insulino-résistance, ce qui expliquerait son association avec le diabète de type 2. De plus, l'utilisation de la réponse insulinique aigüe au glucose (AIRg) pour estimer l'insulino-résistance, suggère que CNVR1 affecterait la fonction cellulaire bêta pancréatique. CNVR1 est intergénique et hautement enrichie en signature H3K27Ac associée aux éléments régulateurs actifs. Des études fonctionnelles menées sur des échantillons de la cohorte CARTaGENE ont permis d'identifier un nouveau biomarqueur, *LOC644172*, régulé par CNVR1 et dont le dosage apparaît étroitement associé au risque cardiovasculaire et à l'âge vasculaire. En effet, *LOC644172* est également associé à la prévalence du diabète de type 2, de l'hypertension, à la pression systolique et au ratio cholestérol-HDL. Un pseudogène est une copie incomplète d'un gène, à laquelle il manque des éléments régulateurs (par ex. promoteurs ou introns). *LOC644172* est le pseudogène fonctionnel (transcrit mais non traduit) de *MAPK8IP1* qui encode la protéine IB1 [155]. *MAPK8IP1* et IB1 ont été reliés à la résistance à l'insuline [156], au diabète de type 2 (DT2) [157] et à la dégénérescence cellulaire observée dans Alzheimer et le DT2 [158]. IB1 est principalement exprimée dans les îlots de Langerhans et le cerveau où elle apparaît comme un régulateur clef du sentier de signalisation JNK dans les cellules neuronales et bêta pancréatiques [155,158]. IB1 est un transactivateur de l'expression de la protéine GLUT2 dont le gène, *SLC2A2*, a été associé au risque cardiovasculaire [159]. Des études ont montré qu'un pseudogène fonctionnel peut affecter la régulation de l'expression du gène fonctionnel « parent » en agissant par exemple comme

transcrit antisens, miRNA decoy ou via la production de petits ARNs interférences (siRNA) [160,161]. On peut donc envisager que les porteurs de duplication au locus CNVR1 développent précocement une anomalie de la fonction bêta pancréatique et de l'insulino-résistance, dues à un dosage élevé de *LOC644172* qui perturberait, en retour, la régulation de *MAPK8IP1*.

Un autre locus de CNVR, sur le chromosome 3q29, a été identifié comme ayant un nombre de copies positivement associé à la prévalence de l'hypertension, chez les individus SLSJ et ADVANCE. Chez les individus SLSJ, les porteurs de la duplication présentent un risque accru pour l'hypertension artérielle (Odds ratio, OR=3,55). Alors que chez les individus ADVANCE, les porteurs de la délétion ont tendance à être protégés contre l'hypertension (OR=0,70). Ce CNVR est localisé près du gène *MUC20* dont la répétition en tandem a précédemment été associée aux lésions et réparations rénales [162]. *MUC20* est également surexprimé dans le tissu rénal des patients atteints de néphropathie à immunoglobulines A (IgA) [163] ou maladie de Berger, qui développent aussi une hypertension artérielle [164]. L'hypertension paraît donc ici secondaire au développement de la néphropathie.

Le nombre de copies du locus de CNVR situé sur le chromosome 19q13.12 est apparu négativement associé à la prévalence de l'hypertension artérielle, chez les individus SLSJ et MONICA. Ainsi, les porteurs de la délétion à ce locus présentent un risque augmenté pour l'hypertension (OR=2,00 par perte de copie) et essentiellement une pression systolique plus élevée ($\beta=0,13$ par perte de copie), dans les deux cohortes. Ce CNVR englobe en partie le gène codant le récepteur 3 des acides gras libres, appartenant à la famille des récepteurs

couplés aux protéines G (*FFAR3/GRP41*). Le propionate est un acide gras à courte chaîne, produit par la flore microbienne de l'intestin. Il a été démontré que l'action hypotensive du propionate est en partie médiée par *Ffar3*, chez la souris. Notamment, une interférence entre la flore intestinale et les systèmes cardiovasculaire et rénal, a été identifiée [165]. Il apparaît qu'une variation dans le nombre de copie à ce locus de CNVR affecterait naturellement, et de façon quantitative, la pression sanguine humaine à travers la modulation de l'expression de *FFAR3*, en réponse à une stimulation du propionate. De plus, ce CNVR est localisé dans un QTL relié à l'indice de masse corporelle, ce qui suggère une influence sur le stockage des graisses. La variabilité en nombre de copies observée au niveau de *FFAR3* pourrait résulter d'une pression sélective imposée par la diète au cours de l'évolution. En effet, ces 5000 dernières années, la pression sélective aurait atteint un rythme 100 fois supérieur à tout autre période de l'évolution humaine, principalement du fait de changements démographiques importants [166]. Aussi, de nombreux ajustements génétiques ont accompagné les transformations alimentaires survenus lors de l'avènement de l'agriculture. Par exemple, le nombre de copies du gène de l'amylase salivaire (*AMY1*) est plus élevé chez les populations ayant été exposées à une alimentation riche en amidon durant l'évolution [116], ce qui suggère une influence de ce CNV sur la réponse glycémique suivant la consommation d'amidon [167]. D'autre part, des mutations dans le gène de la lactase ont été associées à la persistance de cette enzyme chez les Européens et certaines populations pastorales [168,169].

CNVs et QTLs

Les CNVRs identifiés dans cette étude, influencent des gènes ou sont situés dans des QTLs principalement reliés aux réponses inflammatoires et immunitaires, au système rénal ainsi qu'aux lésions/réparations rénales. Certains gènes sont également spécifiques aux primates et/ou semblent avoir évolué par duplications successives et spécifiques à l'évolution humaine. Schlattl *et al.* [93] ont observé que les CNVs sont plus fortement corrélés à l'expression des gènes que les SNPs, et suggèrent que les CNVs seraient souvent la cause de l'expression des QTLs. En ce qui a trait aux QTLs identifiés par Hamet *et al.* [151] dans la même population du SLSJ, ils apparaissent majoritairement reliés à des facteurs anthropométriques et hémodynamiques, ce qui expliquerait la non-correspondance avec les CNVRs identifiés ici. L'ensemble des résultats présentés dans cette étude suggèrent que les CNVs associés à l'hypertension dériveraient d'une pression sélective imposée par l'environnement (notamment la diète et les maladies infectieuses), au cours de l'évolution humaine.

Les limites de l'étude

La présente étude comporte également plusieurs limites, principalement d'ordre technique. En ce qui concerne la transmission des CNVs, les méthodes de détection des CNVs, employées ici, sont limitées à la détection du nombre de copies global et ne fournissent pas d'information sur le nombre de copies présent sur chaque chromosome homologue. En ce qui a trait à la région 17q21.31, l'utilisation de méthodes de validation ne tenant pas compte de la grande complexité de la région, ne permet pas d'évaluer avec précision le nombre de copies en présence [170]. En effet, la région d'intérêt sur 17q21.31,

analysée ici, comprend des familles de gènes hautement dupliqués et à fort pourcentage d'identité. Néanmoins, les sondes TaqMan ont été construites de façon à éviter toute hybridation compétitive. De plus, l'analyse d'association n'a pas pris en compte l'effet potentiel des haplotypes (H1 et H2) sur les loci CNVR1 et CNVR2. Toutefois, nos résultats ont montré une faible corrélation entre le nombre de copies au locus CNVR1 et l'haplotype H2. Finalement, dans cette étude, la petite taille des cohortes (particulièrement SLSJ) limite l'atteinte d'un seuil de significativité statistique à l'échelle génomique.

CONCLUSION ET PERSPECTIVES

En somme, les différents tests statistiques employés, l'utilisation de la cohorte familiale du Saguenay-Lac-St-Jean (SLSJ) comme cohorte de découverte et des cohortes ADVANCE et MONICA, comme cohortes de réplification, ont permis de révéler de nouvelles régions de CNVs communs associés à l'hypertension ou à l'hypertension associée au diabète de type 2.

Essentiellement, l'étude pangénomique menée sur l'hypertension associée au diabète (SLSJ), et sa réplification (ADVANCE) couplée à des études fonctionnelles (CARTaGENE) ont conduit à l'identification d'un biomarqueur, *LOC644172*, pour les complications du diabète de type 2. Il reste à valider l'effet du *LOC644172* dans une large population et éventuellement à estimer si ce biomarqueur peut effectivement permettre d'améliorer la stratification du risque et l'orientation du traitement des pathologies cardiovasculaires. De plus, des études additionnelles sont nécessaires afin de déterminer les mécanismes par lesquelles le pseudogène *LOC644172* régulerait l'expression de *MAPK8IP1*.

D'autre part, le score de risque pondéré des CNVRs élaboré dans la population SLSJ devrait être testé dans une population générale afin de déterminer s'il permet effectivement de détecter les sujets plus à risque de développer l'hypertension précocement. De plus, l'étude d'association pangénomique avec l'hypertension (SLSJ) a permis l'identification de six régions de CNVs incluant le locus 19q13.12 (comprenant *FFAR3*) dont l'effet a été répliqué chez les sujets MONICA. Une étude plus approfondie de *FFAR3* nécessite

l'analyse de phénotypes supplémentaires, l'utilisation d'une plus large cohorte de réplication et des études fonctionnelles.

Cette étude suggère que l'étiologie de l'hypertension ou de l'hypertension associée au diabète de type 2 est affectée par des effets additifs ou interactifs de CNVs. Ces CNVs auraient été fixés du fait d'une pression sélective imposée par l'environnement, au cours de l'évolution humaine. Leur relation avec la précocité du développement de l'hypertension artérielle ou du diabète de type 2 souligne l'effet génétique indépendamment de l'âge. Plus spécifiquement, ces résultats suggèrent l'origine causale de ces CNVs sur un vieillissement vasculaire précoce. Les futurs travaux d'analyse requièrent l'utilisation de méthodes à haute résolution pour la détection et la validation des segments de CNVs.

BIBLIOGRAPHIE

1. Messerli FH, Williams B, Ritz E (2007) Essential hypertension. *Lancet* 370: 591-603.
2. Coffman TM (2011) Under pressure: the search for the essential mechanisms of hypertension. *Nature medicine* 17: 1402-1409.
3. Folkow B (1987) Structure and function of the arteries in hypertension. *American heart journal* 114: 938-948.
4. Hadrava V, Tremblay J, Sekaly RP, Hamet P (1992) Accelerated entry of aortic smooth muscle cells from spontaneously hypertensive rats into the S phase of the cell cycle. *Biochemistry and cell biology = Biochimie et biologie cellulaire* 70: 599-604.
5. Hadrava V, Kruppa U, Russo RC, Lacourciere Y, Tremblay J, et al. (1991) Vascular smooth muscle cell proliferation and its therapeutic modulation in hypertension. *American heart journal* 122: 1198-1203.
6. Adji A, O'Rourke MF, Namasivayam M (2011) Arterial stiffness, its assessment, prognostic value, and implications for treatment. *American journal of hypertension* 24: 5-17.
7. Alreja G, Joseph J (2011) Renin and cardiovascular disease: Worn-out path, or new direction. *World journal of cardiology* 3: 72-83.
8. Parati G, Esler M (2012) The human sympathetic nervous system: its relevance in hypertension and heart failure. *European heart journal* 33: 1058-1066.
9. Sandberg K, Ji H (2012) Sex differences in primary hypertension. *Biology of sex differences* 3: 7.
10. Antchouey A-M, Goulet L, Hamet P (2011) L'hypertension artérielle en Afrique subsaharienne: déterminants intrinsèques, comportementaux et socio-économiques. *Cardiologie Tropicale* 33: 21-29
11. Geronimus AT, Bound J, Keene D, Hicken M (2007) Black-white differences in age trajectories of hypertension prevalence among adult women and men, 1999-2002. *Ethnicity & disease* 17: 40-48.
12. Wrobel MJ, Figge JJ, Izzo JL, Jr. (2011) Hypertension in diverse populations: a New York State Medicaid clinical guidance document. *Journal of the American Society of Hypertension : JASH* 5: 208-229.
13. Mayet J, Shahi M, Foale RA, Poulter NR, Sever PS, et al. (1994) Racial differences in cardiac structure and function in essential hypertension. *BMJ* 308: 1011-1014.
14. Gretler DD, Fumo MT, Nelson KS, Murphy MB (1994) Ethnic differences in circadian hemodynamic profile. *American journal of hypertension* 7: 7-14.
15. Rahman M, Douglas JG, Wright JT, Jr. (1997) Pathophysiology and treatment implications of hypertension in the African-American population. *Endocrinology and metabolism clinics of North America* 26: 125-144.
16. Manunta P, Messaggio E, Ballabeni C, Sciarrone MT, Lanzani C, et al. (2001) Plasma ouabain-like factor during acute and chronic changes in sodium balance in essential hypertension. *Hypertension* 38: 198-203.
17. DiBona GF (2000) Nervous kidney. Interaction between renal sympathetic nerves and the renin-angiotensin system in the control of renal function. *Hypertension* 36: 1083-1088.

18. Guyton AC (1990) Long-term arterial pressure control: an analysis from animal experiments and computer and graphic models. *The American journal of physiology* 259: R865-877.
19. Zhao Q, Gu D, Hixson JE, Liu DP, Rao DC, et al. (2011) Common variants in epithelial sodium channel genes contribute to salt sensitivity of blood pressure: The GenSalt study. *Circulation Cardiovascular genetics* 4: 375-380.
20. Weinberger MH, Miller JZ, Luft FC, Grim CE, Fineberg NS (1986) Definitions and characteristics of sodium sensitivity and blood pressure resistance. *Hypertension* 8: II127-134.
21. Fagard RH (2001) Exercise characteristics and the blood pressure response to dynamic physical training. *Medicine and science in sports and exercise* 33: S484-492; discussion S493-484.
22. Endre T, Mattiasson I, Hulthen UL, Lindgarde F, Berglund G (1994) Insulin resistance is coupled to low physical fitness in normotensive men with a family history of hypertension. *Journal of hypertension* 12: 81-88.
23. O'Keefe JH, Bhatti SK, Bajwa A, Dinicolantonio JJ, Lavie CJ (2014) Alcohol and Cardiovascular Health: The Dose Makes the Poison...or the Remedy. *Mayo Clin Proc* 89: 382-393.
24. Albert MA, Glynn RJ, Ridker PM (2003) Alcohol consumption and plasma concentration of C-reactive protein. *Circulation* 107: 443-447.
25. Beilin LJ (1995) Alcohol, hypertension and cardiovascular disease. *Journal of hypertension* 13: 939-942.
26. Grassi GM, Somers VK, Renk WS, Abboud FM, Mark AL (1989) Effects of alcohol intake on blood pressure and sympathetic nerve activity in normotensive humans: a preliminary report. *Journal of hypertension Supplement : official journal of the International Society of Hypertension* 7: S20-21.
27. Kojima S, Kawano Y, Abe H, Sanai T, Yoshida K, et al. (1993) Acute effects of alcohol ingestion on blood pressure and erythrocyte sodium concentration. *Journal of hypertension* 11: 185-190.
28. Giannattasio C, Mangoni AA, Stella ML, Carugo S, Grassi G, et al. (1994) Acute effects of smoking on radial artery compliance in humans. *Journal of hypertension* 12: 691-696.
29. Nikpay M, Seda O, Tremblay J, Petrovich M, Gaudet D, et al. (2012) Genetic mapping of habitual substance use, obesity-related traits, responses to mental and physical stress, and heart rate and blood pressure measurements reveals shared genes that are overrepresented in the neural synapse. *Hypertension research : official journal of the Japanese Society of Hypertension*.
30. Hamet P, Thorin-Trescases N, Moreau P, Dumas P, Tea BS, et al. (2001) Workshop: excess growth and apoptosis: is hypertension a case of accelerated aging of cardiovascular cells? *Hypertension* 37: 760-766.
31. Vlachopoulos C (2012) Progress towards identifying biomarkers of vascular aging for total cardiovascular risk prediction. *Journal of hypertension* 30 Suppl: S19-26.
32. Bielecka-Dabrowa A, Aronow WS, Rysz J, Banach M (2011) The Rise and Fall of Hypertension: Lessons Learned from Eastern Europe. *Current cardiovascular risk reports* 5: 174-179.

33. Kearney PM, Whelton M, Reynolds K, Muntner P, Whelton PK, et al. (2005) Global burden of hypertension: analysis of worldwide data. *Lancet* 365: 217-223.
34. Cooper R, Rotimi C, Ataman S, McGee D, Osotimehin B, et al. (1997) The prevalence of hypertension in seven populations of west African origin. *American journal of public health* 87: 160-168.
35. Sekikawa A, Hayakawa T (2004) Prevalence of hypertension, its awareness and control in adult population in Japan. *Journal of human hypertension* 18: 911-912.
36. Gu D, Reynolds K, Wu X, Chen J, Duan X, et al. (2002) Prevalence, awareness, treatment, and control of hypertension in china. *Hypertension* 40: 920-927.
37. Johnson RJ, Tittle S, Cade JR, Rideout BA, Oliver WJ (2005) Uric acid, evolution and primitive cultures. *Seminars in nephrology* 25: 3-8.
38. Helmchen LA, Henderson RM (2004) Changes in the distribution of body mass index of white US men, 1890-2000. *Annals of human biology* 31: 174-181.
39. Rodgers A, Ezzati M, Vander Hoorn S, Lopez AD, Lin RB, et al. (2004) Distribution of major health risks: findings from the Global Burden of Disease study. *PLoS medicine* 1: e27.
40. Kannel WB (1976) Some lessons in cardiovascular epidemiology from Framingham. *The American journal of cardiology* 37: 269-282.
41. Lawes CM, Vander Hoorn S, Rodgers A (2008) Global burden of blood-pressure-related disease, 2001. *Lancet* 371: 1513-1518.
42. Murray CJ (1994) Quantifying the burden of disease: the technical basis for disability-adjusted life years. *Bulletin of the World Health Organization* 72: 429-445.
43. Prentki M, Nolan CJ (2006) Islet beta cell failure in type 2 diabetes. *The Journal of clinical investigation* 116: 1802-1812.
44. Ferrannini E (2005) Insulin and blood pressure: connected on a circumference? *Hypertension* 45: 347-348.
45. Sarwar N, Danesh J, Eiriksdottir G, Sigurdsson G, Wareham N, et al. (2007) Triglycerides and the risk of coronary heart disease: 10,158 incident cases among 262,525 participants in 29 Western prospective studies. *Circulation* 115: 450-458.
46. Steinberg D, Gotto AM, Jr. (1999) Preventing coronary artery disease by lowering cholesterol levels: fifty years from bench to bedside. *JAMA : the journal of the American Medical Association* 282: 2043-2050.
47. Johnson ML, Pietz K, Battleman DS, Beyth RJ (2004) Prevalence of comorbid hypertension and dyslipidemia and associated cardiovascular disease. *The American journal of managed care* 10: 926-932.
48. Wilson PW, D'Agostino RB, Sullivan L, Parise H, Kannel WB (2002) Overweight and obesity as determinants of cardiovascular risk: the Framingham experience. *Archives of internal medicine* 162: 1867-1872.
49. Francischetti EA, Genelhu VA (2007) Obesity-hypertension: an ongoing pandemic. *International journal of clinical practice* 61: 269-280.
50. Goodfriend TL, Calhoun DA (2004) Resistant hypertension, obesity, sleep apnea, and aldosterone: theory and therapy. *Hypertension* 43: 518-524.
51. Pausova Z, Gaudet D, Gossard F, Bernard M, Kaldunski ML, et al. (2005) Genome-wide scan for linkage to obesity-associated hypertension in French Canadians. *Hypertension* 46: 1280-1285.

52. Fox CS, Coady S, Sorlie PD, D'Agostino RB, Sr., Pencina MJ, et al. (2007) Increasing cardiovascular disease burden due to diabetes mellitus: the Framingham Heart Study. *Circulation* 115: 1544-1550.
53. Bruno RM, Penno G, Daniele G, Pucci L, Lucchesi D, et al. (2012) Type 2 diabetes mellitus worsens arterial stiffness in hypertensive patients through endothelial dysfunction. *Diabetologia*.
54. Yamagishi S (2011) Cardiovascular disease in recent onset diabetes mellitus. *Journal of cardiology* 57: 257-262.
55. Dai S, Robitaille C, Bancej C, Loukine L, Waters C, et al. (2010) Executive summary--report from the Canadian Chronic Disease Surveillance System: hypertension in Canada, 2010. *Chronic diseases in Canada* 31: 46-47.
56. Alberti KG, Eckel RH, Grundy SM, Zimmet PZ, Cleeman JI, et al. (2009) Harmonizing the metabolic syndrome: a joint interim statement of the International Diabetes Federation Task Force on Epidemiology and Prevention; National Heart, Lung, and Blood Institute; American Heart Association; World Heart Federation; International Atherosclerosis Society; and International Association for the Study of Obesity. *Circulation* 120: 1640-1645.
57. Ford ES, Giles WH, Dietz WH (2002) Prevalence of the metabolic syndrome among US adults: findings from the third National Health and Nutrition Examination Survey. *JAMA : the journal of the American Medical Association* 287: 356-359.
58. Alberti KG, Zimmet PZ (1998) Definition, diagnosis and classification of diabetes mellitus and its complications. Part 1: diagnosis and classification of diabetes mellitus provisional report of a WHO consultation. *Diabetic medicine : a journal of the British Diabetic Association* 15: 539-553.
59. Alberti KG, Zimmet P, Shaw J (2005) The metabolic syndrome--a new worldwide definition. *Lancet* 366: 1059-1062.
60. Zhao Q, Kelly TN, Li C, He J (2013) Progress and Future Aspects in Genetics of Human Hypertension. *Current hypertension reports*.
61. Ayman D (1934) Heredity in arteriolar (essential) hypertension: a clinical study of blood pressure in 1,524 members of 277 families. *Arch Intern Med* 53:792-803
62. McIlhenny ML, Shaffer JW, Hines EA, Jr. (1975) The heritability of blood pressure: an investigation of 200 pairs of twins using the cold pressor test. *The Johns Hopkins medical journal* 136: 57-64.
63. Annett JL, Sing CF, Biron P, Mongeau JG (1979) Familial aggregation of blood pressure and weight in adoptive families. I. Comparisons of blood pressure and weight statistics among families with adopted, natural, or both natural and adopted children. *American journal of epidemiology* 110: 479-491.
64. Abifadel M, Varret M, Rabes JP, Allard D, Ouguerram K, et al. (2003) Mutations in PCSK9 cause autosomal dominant hypercholesterolemia. *Nature genetics* 34: 154-156.
65. Brown MS, Goldstein JL (1986) A receptor-mediated pathway for cholesterol homeostasis. *Science* 232: 34-47.
66. Garcia CK, Wilund K, Arca M, Zuliani G, Fellin R, et al. (2001) Autosomal recessive hypercholesterolemia caused by mutations in a putative LDL receptor adaptor protein. *Science* 292: 1394-1398.

67. Lehrman MA, Schneider WJ, Sudhof TC, Brown MS, Goldstein JL, et al. (1985) Mutation in LDL receptor: Alu-Alu recombination deletes exons encoding transmembrane and cytoplasmic domains. *Science* 227: 140-146.
68. Lifton RP (2004) Genetic dissection of human blood pressure variation: common pathways from rare phenotypes. *Harvey Lect* 100: 71-101.
69. Reich DE, Cargill M, Bolk S, Ireland J, Sabeti PC, et al. (2001) Linkage disequilibrium in the human genome. *Nature* 411: 199-204.
70. (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 447: 661-678.
71. Ehret GB, Munroe PB, Rice KM, Bochud M, Johnson AD, et al. (2011) Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature* 478: 103-109.
72. Ehret GB, Caulfield MJ (2013) Genes for blood pressure: an opportunity to understand hypertension. *Eur Heart J* 34: 951-961.
73. Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American journal of human genetics* 32: 314-331.
74. Weissenbach J, Gyapay G, Dib C, Vignal A, Morissette J, et al. (1992) A second-generation linkage map of the human genome. *Nature* 359: 794-801.
75. Sachidanandam R, Weissman D, Schmidt SC, Kakol JM, Stein LD, et al. (2001) A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* 409: 928-933.
76. Haga H, Yamada R, Ohnishi Y, Nakamura Y, Tanaka T (2002) Gene-based SNP discovery as part of the Japanese Millennium Genome Project: identification of 190,562 genetic variations in the human genome. *Single-nucleotide polymorphism. Journal of human genetics* 47: 605-610.
77. Lejeune J, Gautier M, Turpin R (1959) [Study of somatic chromosomes from 9 mongoloid children]. *Comptes rendus hebdomadaires des seances de l'Academie des sciences* 248: 1721-1722.
78. Ford CE, Jones KW, Polani PE, De Almeida JC, Briggs JH (1959) A sex-chromosome anomaly in a case of gonadal dysgenesis (Turner's syndrome). *Lancet* 1: 711-713.
79. Jacobs PA, Strong JA (1959) A case of human intersexuality having a possible XXY sex-determining mechanism. *Nature* 183: 302-303.
80. Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, et al. (2004) Detection of large-scale variation in the human genome. *Nature genetics* 36: 949-951.
81. Sebat J, Lakshmi B, Troge J, Alexander J, Young J, et al. (2004) Large-scale copy number polymorphism in the human genome. *Science* 305: 525-528.
82. Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, et al. (2006) Global variation in copy number in the human genome. *Nature* 444: 444-454.
83. Feuk L, Carson AR, Scherer SW (2006) Structural variation in the human genome. *Nature reviews Genetics* 7: 85-97.
84. Alkan C, Coe BP, Eichler EE (2011) Genome structural variation discovery and genotyping. *Nature reviews Genetics* 12: 363-376.
85. McCarroll SA (2008) Extending genome-wide association studies to copy-number variation. *Human molecular genetics* 17: R135-142.

86. Fanciulli M, Petretto E, Aitman TJ (2010) Gene copy number variation and common human disease. *Clinical genetics* 77: 201-213.
87. de Smith AJ, Walters RG, Coin LJ, Steinfeld I, Yakhini Z, et al. (2008) Small deletion variants have stable breakpoints commonly associated with alu elements. *PloS one* 3: e3104.
88. Gu W, Zhang F, Lupski JR (2008) Mechanisms for human genomic rearrangements. *PathoGenetics* 1: 4.
89. Hastings PJ, Ira G, Lupski JR (2009) A microhomology-mediated break-induced replication model for the origin of human copy number variation. *PLoS genetics* 5: e1000327.
90. Zhang F, Gu W, Hurles ME, Lupski JR (2009) Copy number variation in human health, disease, and evolution. *Annual review of genomics and human genetics* 10: 451-481.
91. Kidd JM, Cooper GM, Donahue WF, Hayden HS, Sampas N, et al. (2008) Mapping and sequencing of structural variation from eight human genomes. *Nature* 453: 56-64.
92. van Binsbergen E (2011) Origins and breakpoint analyses of copy number variations: up close and personal. *Cytogenetic and genome research* 135: 271-276.
93. Schlattl A, Anders S, Waszak SM, Huber W, Korbel JO (2011) Relating CNVs to transcriptome data at fine resolution: assessment of the effect of variant size, type, and overlap with functional regions. *Genome research* 21: 2004-2013.
94. Conrad DF, Pinto D, Redon R, Feuk L, Gokcumen O, et al. (2010) Origins and functional impact of copy number variation in the human genome. *Nature* 464: 704-712.
95. Pinkel D, Seagraves R, Sudar D, Clark S, Poole I, et al. (1998) High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. *Nature genetics* 20: 207-211.
96. Nemos C, Bursztejn AC, Jonveaux P (2008) [Management of the CNVs in constitutional human genetics using array CGH]. *Pathologie-biologie* 56: 354-361.
97. Komura D, Shen F, Ishikawa S, Fitch KR, Chen W, et al. (2006) Genome-wide detection of human copy number variations using high-density DNA oligonucleotide arrays. *Genome research* 16: 1575-1584.
98. Yau C, Holmes CC (2008) CNV discovery using SNP genotyping arrays. *Cytogenetic and genome research* 123: 307-312.
99. Korbel JO, Urban AE, Affourtit JP, Godwin B, Grubert F, et al. (2007) Paired-end mapping reveals extensive structural variation in the human genome. *Science* 318: 420-426.
100. Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America* 74: 5463-5467.
101. Medvedev P, Stanciu M, Brudno M (2009) Computational methods for discovering structural variation with next-generation sequencing. *Nature methods* 6: S13-20.
102. Ansorge WJ (2009) Next-generation DNA sequencing techniques. *New biotechnology* 25: 195-203.

103. Wu YL, Savelli SL, Yang Y, Zhou B, Rovin BH, et al. (2007) Sensitive and specific real-time polymerase chain reaction assays to accurately determine copy number variations (CNVs) of human complement C4A, C4B, C4-long, C4-short, and RCCX modules: elucidation of C4 CNVs in 50 consanguineous subjects with defined HLA genotypes. *Journal of immunology* 179: 3012-3025.
104. Pedersen K, Wiechec E, Madsen BE, Overgaard J, Hansen LL (2010) A simple way to evaluate self-designed probes for tumor specific Multiplex Ligation-dependent Probe Amplification (MLPA). *BMC research notes* 3: 179.
105. White SJ, Vink GR, Kriek M, Wuyts W, Schouten J, et al. (2004) Two-color multiplex ligation-dependent probe amplification: detecting genomic rearrangements in hereditary multiple exostoses. *Human mutation* 24: 86-92.
106. Charbonnier F, Raux G, Wang Q, Drouot N, Cordier F, et al. (2000) Detection of exon deletions and duplications of the mismatch repair genes in hereditary nonpolyposis colorectal cancer families using multiplex polymerase chain reaction of short fluorescent fragments. *Cancer research* 60: 2760-2763.
107. Stranger BE, Forrest MS, Dunning M, Ingle CE, Beazley C, et al. (2007) Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science* 315: 848-853.
108. Cahan P, Li Y, Izumi M, Graubert TA (2009) The impact of copy number variation on local gene expression in mouse hematopoietic stem and progenitor cells. *Nature genetics* 41: 430-437.
109. Schuster-Bockler B, Conrad D, Bateman A (2010) Dosage sensitivity shapes the evolution of copy-number varied regions. *PloS one* 5: e9474.
110. Henrichsen CN, Chaignat E, Reymond A (2009) Copy number variants, diseases and gene expression. *Human molecular genetics* 18: R1-8.
111. Poptsova M, Banerjee S, Gokcumen O, Rubin MA, Demichelis F (2013) Impact of constitutional copy number variants on biological pathway evolution. *BMC evolutionary biology* 13: 19.
112. Banerjee S, Oldridge D, Poptsova M, Hussain WM, Chakravarty D, et al. (2011) A computational framework discovers new copy number variants with functional importance. *PloS one* 6: e17539.
113. Clop A, Vidal O, Amills M (2012) Copy number variation in the genomes of domestic animals. *Animal genetics* 43: 503-517.
114. van Duyvenvoorde HA, Lui JC, Kant SG, Oostdijk W, Gijsbers AC, et al. (2013) Copy number variants in patients with short stature. *European journal of human genetics* : EJHG.
115. Korbelt JO, Kim PM, Chen X, Urban AE, Weissman S, et al. (2008) The current excitement about copy-number variation: how it relates to gene duplications and protein families. *Current opinion in structural biology* 18: 366-374.
116. Perry GH, Dominy NJ, Claw KG, Lee AS, Fiegler H, et al. (2007) Diet and the evolution of human amylase gene copy number variation. *Nature genetics* 39: 1256-1260.
117. Hardwick RJ, Machado LR, Zuccherato LW, Antolinos S, Xue Y, et al. (2011) A worldwide analysis of beta-defensin copy number variation suggests recent

- selection of a high-expressing DEFB103 gene copy in East Asia. *Human mutation* 32: 743-750.
118. Flint J, Hill AV, Bowden DK, Oppenheimer SJ, Sill PR, et al. (1986) High frequencies of alpha-thalassaemia are the result of natural selection by malaria. *Nature* 321: 744-750.
 119. May J, Evans JA, Timmann C, Ehmen C, Busch W, et al. (2007) Hemoglobin variants and disease manifestations in severe falciparum malaria. *JAMA : the journal of the American Medical Association* 297: 2220-2226.
 120. Dumas L, Kim YH, Karimpour-Fard A, Cox M, Hopkins J, et al. (2007) Gene copy number variation spanning 60 million years of human and primate evolution. *Genome research* 17: 1266-1277.
 121. Popesco MC, Maclaren EJ, Hopkins J, Dumas L, Cox M, et al. (2006) Human lineage-specific amplification, selection, and neuronal expression of DUF1220 domains. *Science* 313: 1304-1307.
 122. Dumas LJ, O'Bleness MS, Davis JM, Dickens CM, Anderson N, et al. (2012) DUF1220-domain copy number implicated in human brain-size pathology and evolution. *American journal of human genetics* 91: 444-454.
 123. Nguyen DQ, Webber C, Ponting CP (2006) Bias of selection on human copy-number variants. *PLoS genetics* 2: e20.
 124. Stankiewicz P, Lupski JR (2002) Molecular-evolutionary mechanisms for genomic disorders. *Current opinion in genetics & development* 12: 312-319.
 125. Bruder CE, Piotrowski A, Gijsbers AA, Andersson R, Erickson S, et al. (2008) Phenotypically concordant and discordant monozygotic twins display different DNA copy-number-variation profiles. *American journal of human genetics* 82: 763-771.
 126. McCarroll SA, Kuruvilla FG, Korn JM, Cawley S, Nemes J, et al. (2008) Integrated detection and population-genetic analysis of SNPs and copy number variation. *Nature genetics* 40: 1166-1174.
 127. Locke DP, Sharp AJ, McCarroll SA, McGrath SD, Newman TL, et al. (2006) Linkage disequilibrium and heritability of copy-number polymorphisms within duplicated regions of the human genome. *American journal of human genetics* 79: 275-290.
 128. Piotrowski A, Bruder CE, Andersson R, Diaz de Stahl T, Menzel U, et al. (2008) Somatic mosaicism for copy number variation in differentiated human tissues. *Human mutation* 29: 1118-1124.
 129. Weir BA, Woo MS, Getz G, Perner S, Ding L, et al. (2007) Characterizing the cancer genome in lung adenocarcinoma. *Nature* 450: 893-898.
 130. (2008) Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* 455: 1061-1068.
 131. Carvalho CM, Lupski JR (2008) Copy number variation at the breakpoint region of isochromosome 17q. *Genome research* 18: 1724-1732.
 132. Gu W, Lupski JR (2008) CNV and nervous system diseases--what's new? *Cytogenetic and genome research* 123: 54-64.
 133. Wafaei JR, Choy FY (2005) Glucocerebrosidase recombinant allele: molecular evolution of the glucocerebrosidase gene and pseudogene in primates. *Blood cells, molecules & diseases* 35: 277-285.

134. Steglich C, Bunge S, Hulsebos T, Beck M, Brandt NJ, et al. (1993) Molecular analysis in patients with mucopolysaccharidosis type II suggests that DXS466 maps within the Hunter gene. *Human genetics* 92: 179-182.
135. Figueiredo MS, Tavella MH, Simoes BP (1994) Large DNA inversions, deletions, and TaqI site mutations in severe haemophilia A. *Human genetics* 94: 473-478.
136. Tamer L, Calikoglu M, Ates NA, Yildirim H, Ercan B, et al. (2004) Glutathione-S-transferase gene polymorphisms (GSTT1, GSTM1, GSTP1) as increased risk factors for asthma. *Respirology* 9: 493-498.
137. Fanciulli M, Nordsworthy PJ, Petretto E, Dong R, Harper L, et al. (2007) FCGR3B copy number variation is associated with susceptibility to systemic, but not organ-specific, autoimmunity. *Nature genetics* 39: 721-723.
138. Willcocks LC, Lyons PA, Clatworthy MR, Robinson JI, Yang W, et al. (2008) Copy number of FCGR3B, which is associated with systemic lupus erythematosus, correlates with protein expression and immune complex uptake. *The Journal of experimental medicine* 205: 1573-1582.
139. Aitman TJ, Dong R, Vyse TJ, Nordsworthy PJ, Johnson MD, et al. (2006) Copy number polymorphism in *Fcgr3* predisposes to glomerulonephritis in rats and humans. *Nature* 439: 851-855.
140. Gonzalez E, Kulkarni H, Bolivar H, Mangano A, Sanchez R, et al. (2005) The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility. *Science* 307: 1434-1440.
141. Jeon JP, Shim SM, Nam HY, Ryu GM, Hong EJ, et al. (2010) Copy number variation at leptin receptor gene locus associated with metabolic traits and the risk of type 2 diabetes mellitus. *BMC genomics* 11: 426.
142. Jacquemont S, Reymond A, Zufferey F, Harewood L, Walters RG, et al. (2011) Mirror extreme BMI phenotypes associated with gene dosage at the chromosome 16p11.2 locus. *Nature* 478: 97-102.
143. Walters RG, Jacquemont S, Valsesia A, de Smith AJ, Martinet D, et al. (2010) A new highly penetrant form of obesity due to deletions on chromosome 16p11.2. *Nature* 463: 671-675.
144. Bochukova EG, Huang N, Keogh J, Henning E, Purmann C, et al. (2010) Large, rare chromosomal deletions associated with severe early-onset obesity. *Nature* 463: 666-670.
145. Willer CJ, Speliotes EK, Loos RJ, Li S, Lindgren CM, et al. (2009) Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. *Nature genetics* 41: 25-34.
146. Kosta K, Sabroe I, Goke J, Nibbs RJ, Tsanakas J, et al. (2007) A Bayesian approach to copy-number-polymorphism analysis in nuclear pedigrees. *American journal of human genetics* 81: 808-812.
147. Stringer S, Wray NR, Kahn RS, Derks EM (2011) Underestimated effect sizes in GWAS: fundamental limitations of single SNP analysis for dichotomous phenotypes. *PloS one* 6: e27964.
148. Craddock N, Hurles ME, Cardin N, Pearson RD, Plagnol V, et al. (2010) Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls. *Nature* 464: 713-720.

149. Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, et al. (2010) A map of human genome variation from population-scale sequencing. *Nature* 467: 1061-1073.
150. Thacker EL, Ascherio A (2008) Familial aggregation of Parkinson's disease: a meta-analysis. *Movement disorders : official journal of the Movement Disorder Society* 23: 1174-1183.
151. Hamet P, Merlo E, Seda O, Broeckel U, Tremblay J, et al. (2005) Quantitative founder-effect analysis of French Canadian families identifies specific loci contributing to metabolic phenotypes of hypertension. *American journal of human genetics* 76: 815-832.
152. Agopian AJ, Eastcott LM, Mitchell LE (2012) Age of onset and effect size in genome-wide association studies. *Birth defects research Part A, Clinical and molecular teratology* 94: 908-911.
153. Wilk JB, Djousse L, Arnett DK, Hunt SC, Province MA, et al. (2004) Genome-wide linkage analyses for age at diagnosis of hypertension and early-onset hypertension in the HyperGEN study. *American journal of hypertension* 17: 839-844.
154. Parikh NI, Pencina MJ, Wang TJ, Benjamin EJ, Lanier KJ, et al. (2008) A risk score for predicting near-term incidence of hypertension: the Framingham Heart Study. *Ann Intern Med* 148: 102-110.
155. Mooser V, Maillard A, Bonny C, Steinmann M, Shaw P, et al. (1999) Genomic organization, fine-mapping, and expression of the human islet-brain 1 (IB1)/c-Jun-amino-terminal kinase interacting protein-1 (JIP-1) gene. *Genomics* 55: 202-208.
156. Brajkovic S, Marenzoni R, Favre D, et al. Evidence for tuning adipocytes ICER levels for obesity care. *Adipocyte* 2012;1(3):157-60.
157. Waeber G, Delplanque J, Bonny C, Mooser V, Steinmann M, et al. (2000) The gene MAPK8IP1, encoding islet-brain-1, is a candidate for type 2 diabetes. *Nature genetics* 24: 291-295.
158. Beeler N, Riederer BM, Waeber G, Abderrahmani A (2009) Role of the JNK-interacting protein 1/islet brain 1 in cell degeneration in Alzheimer disease and diabetes. *Brain research bulletin* 80: 274-281.
159. Borglykke A, Grarup N, Sparso T, Linneberg A, Fenger M, et al. (2012) Genetic variant SLC2A2 [corrected] Is associated with risk of cardiovascular disease - assessing the individual and cumulative effect of 46 type 2 diabetes related genetic variants. *PLoS One* 7: e50418.
160. Pink RC, Wicks K, Caley DP, Punch EK, Jacobs L, et al. (2011) Pseudogenes: pseudo-functional or key regulators in health and disease? *RNA* 17: 792-798.
161. Li W, Yang W, Wang XJ (2013) Pseudogenes: pseudo or real functional elements? *J Genet Genomics* 40: 171-177.
162. Li G, Zhang H, Lv J, Hou P, Wang H (2006) Tandem repeats polymorphism of MUC20 is an independent factor for the progression of immunoglobulin A nephropathy. *American journal of nephrology* 26: 43-49.
163. Waga I, Yamamoto J, Sasai H, Munger WE, Hogan SL, et al. (2003) Altered mRNA expression in renal biopsy tissue from patients with IgA nephropathy. *Kidney international* 64: 1253-1264.

164. Wang W, Chen N (2013) Treatment of progressive IgA nephropathy: an update. *Contributions to nephrology* 181: 75-83.
165. Pluznick JL, Protzko RJ, Gevorgyan H, Peterlin Z, Sipos A, et al. (2013) Olfactory receptor responding to gut microbiota-derived signals plays a role in renin secretion and blood pressure regulation. *Proc Natl Acad Sci U S A* 110: 4410-4415.
166. Hawks J, Wang ET, Cochran GM, Harpending HC, Moyzis RK (2007) Recent acceleration of human adaptive evolution. *Proceedings of the National Academy of Sciences of the United States of America* 104: 20753-20758.
167. Santos JL, Saus E, Smalley SV, Cataldo LR, Alberti G, et al. (2012) Copy number polymorphism of the salivary amylase gene: implications in human nutrition research. *Journal of nutrigenetics and nutrigenomics* 5: 117-131.
168. Gerbault P, Moret C, Currat M, Sanchez-Mazas A (2009) Impact of selection and demography on the diffusion of lactase persistence. *PloS one* 4: e6369.
169. Gallego Romero I, Basu Mallick C, Liebert A, Crivellaro F, Chaubey G, et al. (2012) Herders of Indian and European cattle share their predominant allele for lactase persistence. *Molecular biology and evolution* 29: 249-260.
170. Sudmant PH, Kitzman JO, Antonacci F, Alkan C, Malig M, et al. (2010) Diversity of human copy number variation and multicopy genes. *Science* 330: 641-646.

SUPPLEMENTAL APPENDIX

GENOME-WIDE ASSOCIATION STUDY IDENTIFIES SIX COPY-NUMBER VARIANT REGIONS CONTRIBUTING TO HYPERTENSION IN FRENCH CANADIAN FAMILIES

Short title: CNV risk score and hypertension

Mahine Ivanga, Alena Krajčoviechová, John Raelson, Francois Harvey, François-Christophe Marois-Blanchet, Gilles Godefroid, Renata Cifkova, John Chalmers, Stephen Harrap, Stephen McMahon, Michel Marre, Mark Woodward, Daniel Gaudet, Johanne Tremblay and Pavel Hamet

Supplemental Appendix Chapter II- METHODS

The sibling risk ratio (λ_s) in French Canadians

λ_s by age groups was calculated, where λ_s indicates the increased risk of hypertension in siblings of affected cases (of our family-based cohort), compared with the risk of hypertension in the general Saguenay-Lac-St-Jean population as a component of the Canada heart health study^{1, 2}.

$\lambda_s = \text{Number of hypertensive siblings of affected cases} / \text{Expected number of affected cases of the same age category}$

CNV detection and selection

We performed a principal component analysis (PCA) using 147,088 SNPs to quantify the genetic divergence in ADVANCE subjects of European ancestry, which outputs each individual's coordinates along the principal axes of variation. The two first principal components (PC1 and PC2) corresponding to the Northwest and Southeast Europe axes of variation were obtained using smartpca (EIGENSOFT 3.0³). These two PCs were used as covariates in subsequent association analysis.

In FC and ADVANCE subjects, CNV calls were made using PennCNV⁴ and further analyzed with the ParseCNV algorithm⁵. PennCNV implements a Hidden Markov Model (HMM) that integrates multiple sources of information, including Log R ratio (LRR; a measure of total probe signal intensity) and B allele frequency (BAF; a measure of the relative intensity ratio of the two allelic probes), to infer copy-number from individual genotype samples. A quality control (QC) filtering step was applied at both the sample genotype and CNV level. Duplicate samples and samples with bad contrast readings were removed from cell files. The remaining cell files were analyzed using the Affymetrix power

tools and only files with a QC call-rate higher than 86% were retained. Individuals with sex-mismatch, ethnicity or relatedness issues were also removed, leaving a data set of 181 FC and 3449 ADVANCE genotyped individuals. Raw CNV calls were further filtered to retain only those CNVs with a CNV quality score greater than 10 and with detection by a minimum of 6 probes and a minimum CNV call length of 500bp.

ParseCNV⁵ creates probe-based statistics for the cleaned CNVs called by PennCNV, which are then used to determine Copy-number Variant Regions (CNVRs). The probe-based statistics are then tested for case control association by Plink⁶ using a Fisher exact test and probes without nominal significance ($P < 0.05$) are discarded from further association testing. We modified this procedure for ADVANCE subjects to test the probe based statistics in Plink using logistic regression with covariates namely sex, PCs and genotyping chip. The output from the probe-based association tests is then used by ParseCNV to merge significant probes into CNVRs based on probe proximity (<1 MB) and comparable significance ($\pm 1 \log P$ -value) of neighbouring probes. CNVRs are defined in a dynamic manner that allows for a complex CNV overlap while maintaining precise state-specific association region and allows flexible boundaries for complex CNVs. Finally, p-values for case/control association of the CNVRs are determined separately for duplications and deletions using two-tailed Fisher exact tests.

We used the Birdsuite tool set⁷ in combination with Plink⁶ as an independent method of calling CNVs in FC and for calling CNVs in CPM subjects. To infer copy-number from individual genotype samples; the Birdseye algorithm from Birdsuite also implements a HMM, including LRR and BAF, while a second Birdsuite algorithm, Canary, uses a one-dimensional Gaussian mixture model to detect common CNVs. The outputs from Birdseye and Canary were combined to produce a unified call of CNVs and these two algorithms are hereafter referred to as one algorithm. A QC filtering step was also applied and arrays with an overall call rate less than 86% were discarded from further analysis. One Czech Post-Monica individual was excluded based on the coarse grain population stratification using Hapmap phase 3 (only 78% Caucasian genome structure), leaving a data set of 187 Czech Post-Monica genotyped individuals.

The Plink package⁶ was used to perform association analyses of the CNVRs (determined by ParseCNV) with hypertension, in taking in consideration families, when appropriate. Analyses were realized separately for duplications and deletions. P-values were estimated using 1-sided and 2-sided permutation-based tests (100,000) for hypertension vs. normotension at each CNVR. CNVRs with p-values less than 5% were selected for further testing.

Supplement appendix Chapter II- RESULTS

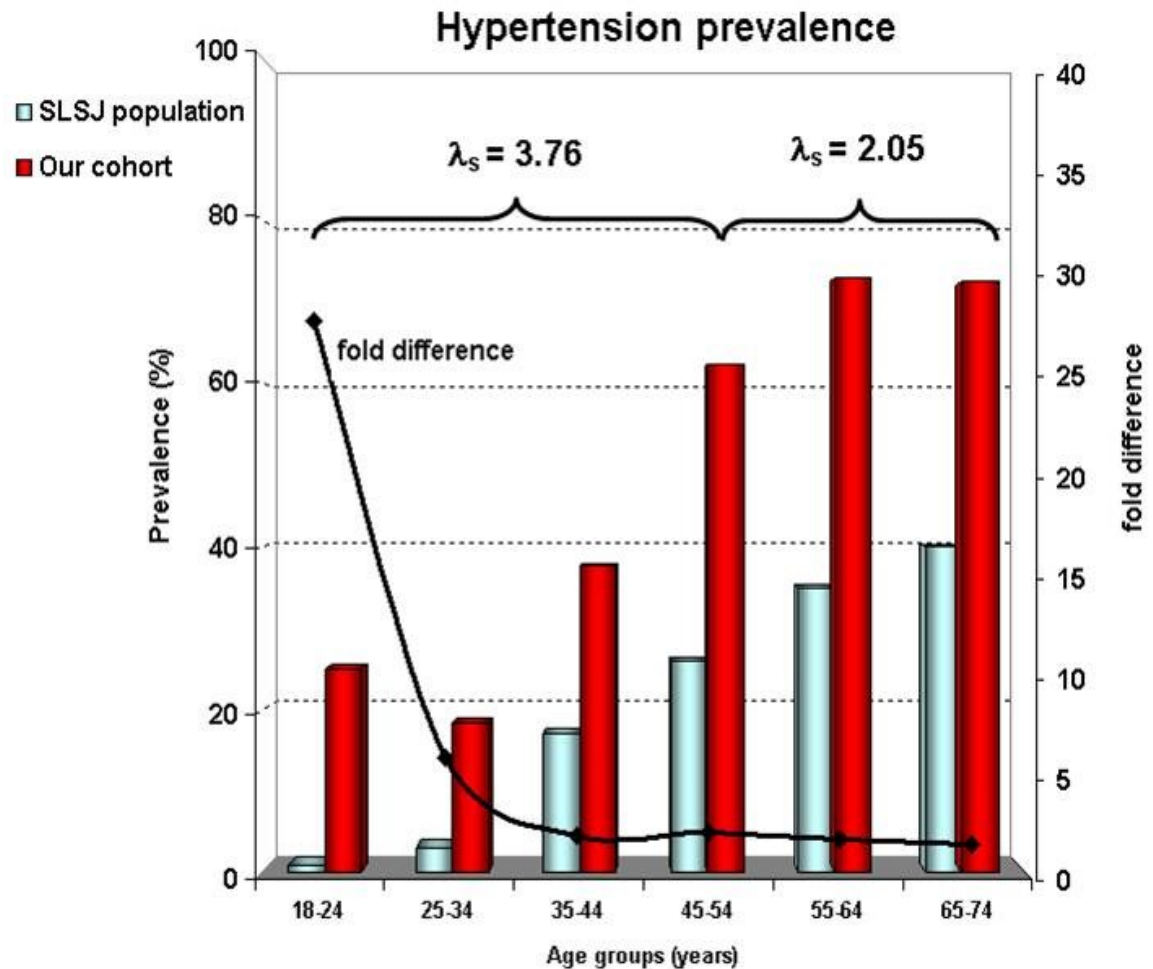


Figure S1. The sibling risk ratio (λ_s) of hypertension per age group in French Canadians. We observed a strong familial clustering in younger individuals, the highest λ_s being for cases occurring between 18 and 24 years, ~ 27 . Augmented λ_s for hypertension reflects increased genetic susceptibilities of our cohort, compared to the general population of Saguenay-Lac-St-Jean (SLSJ). However, the siblings may differ by environmental risk factors for hypertension, thus increasing the risk of hypertension independently of genetic effects.

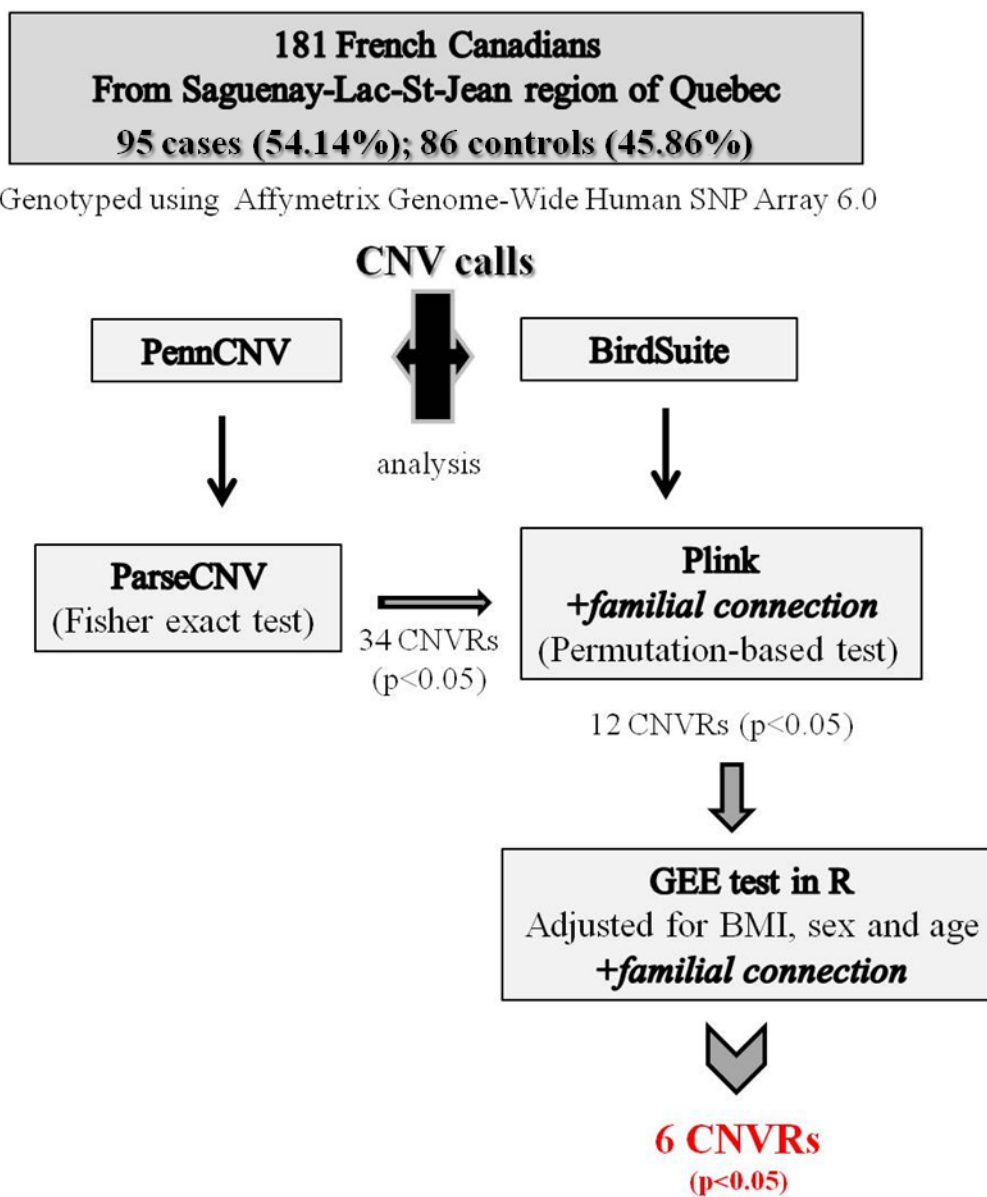


Figure S2. Summary of study design

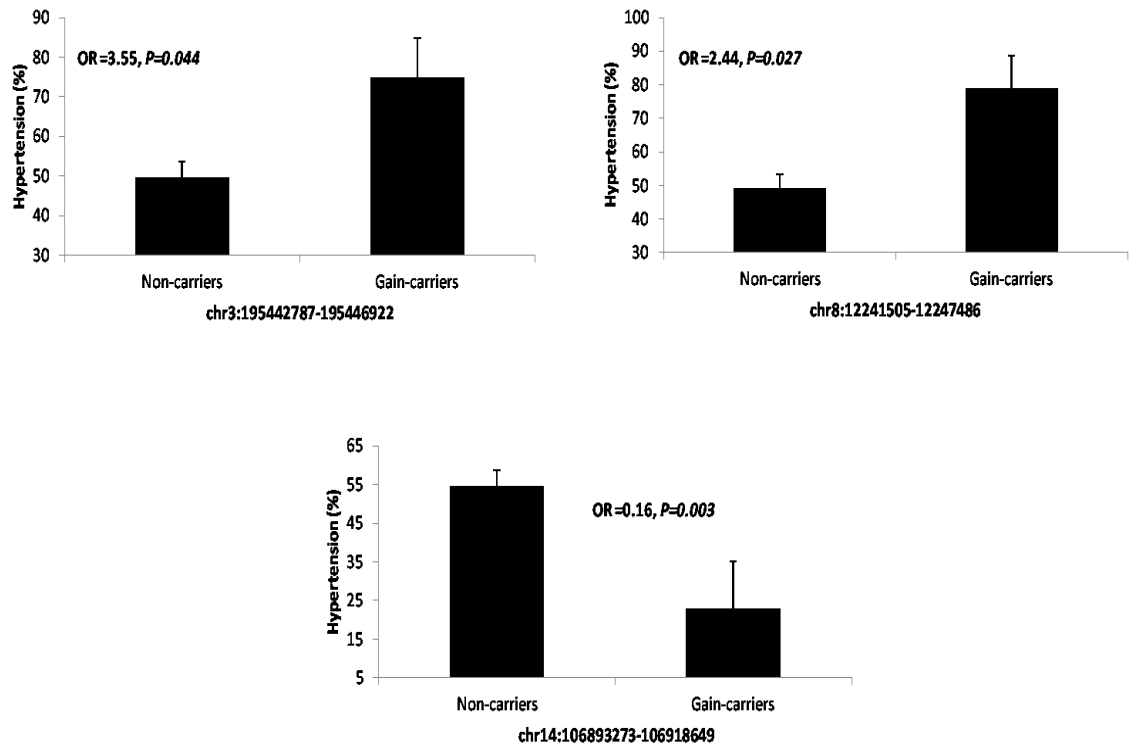


Figure S3. Prevalence of Hypertension and Duplication-CNVRs identified in FC.
Copy-Number:CN; Gain-carriers: carry CN>2.

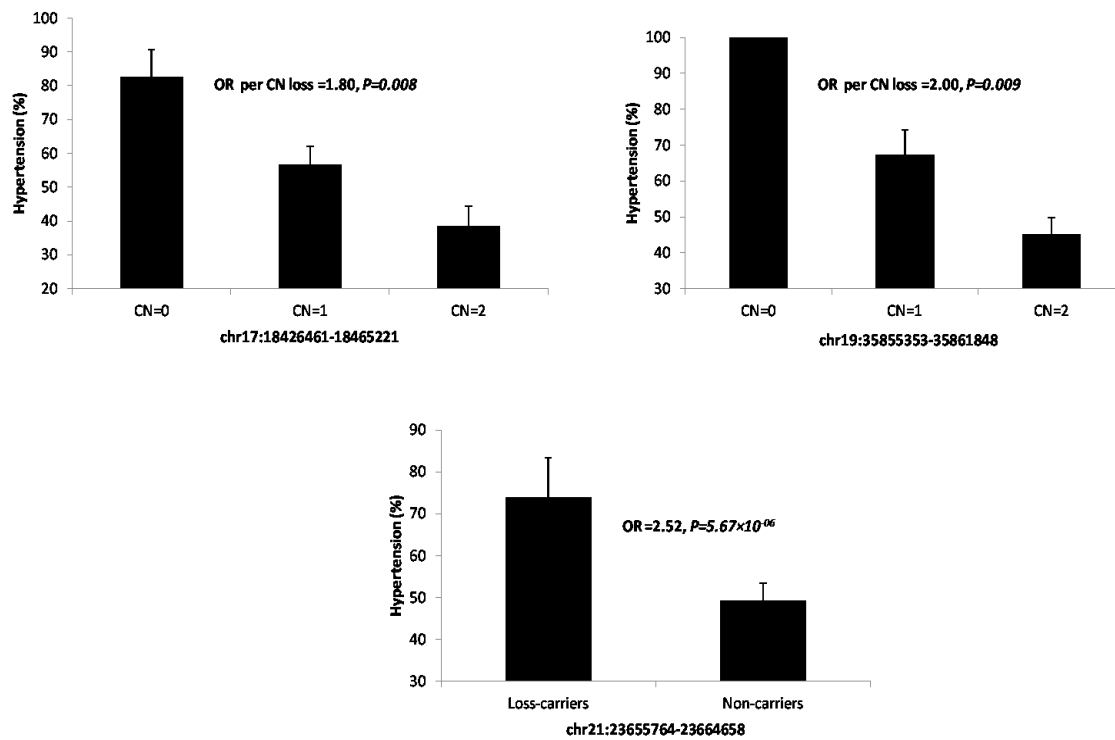


Figure S4. Prevalence of Hypertension and Deletion-CNVs identified in FC. Copy-Number:CN; Loss-carriers: carry CN<2.

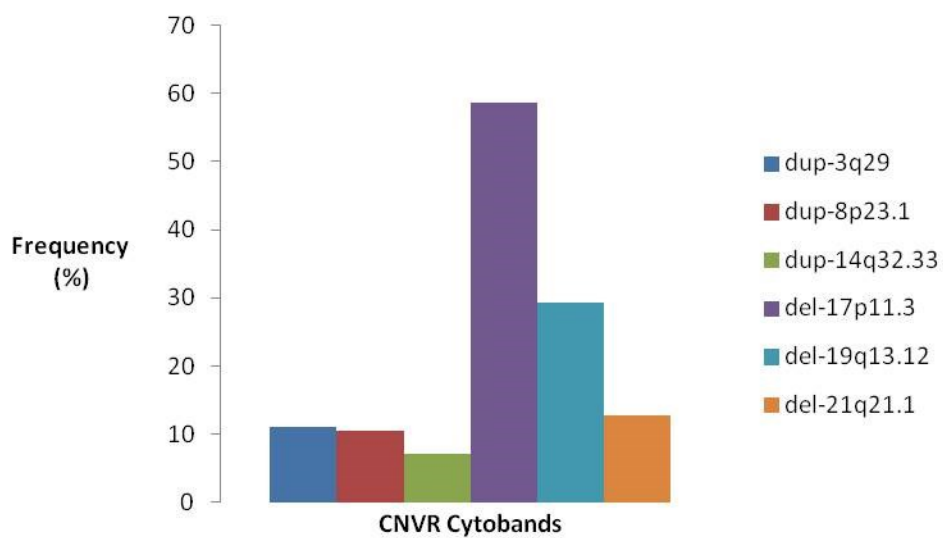


Figure S5. Frequency of selected CNVRs. Frequencies of all identified CNVRs are represented; dup=duplication; del=deletion.

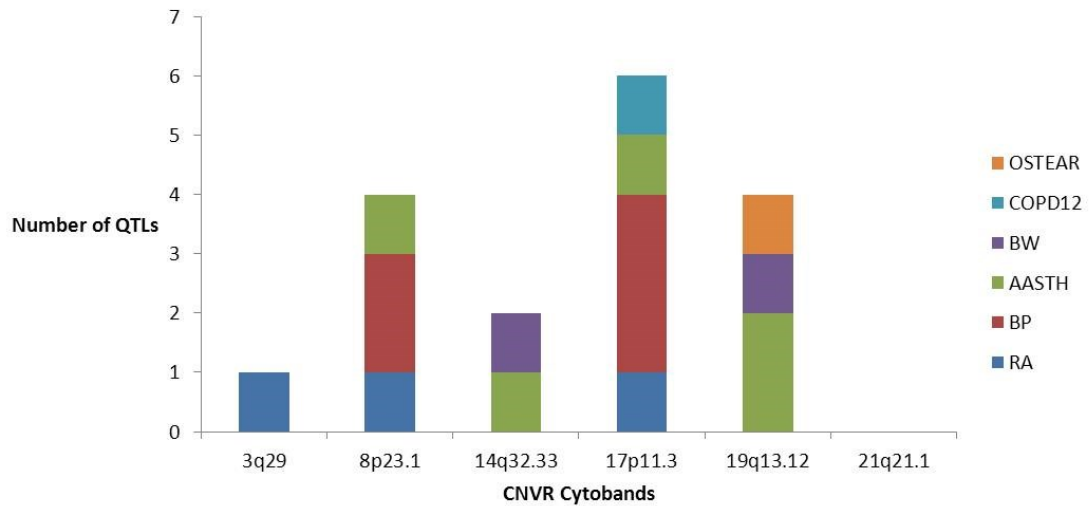


Figure S6. Most selected CNVRs map to human QTLs. Identified CNVRs according to the QTLs they overlap. Most reported CNVRs lie mainly within human QTLs related to immune and inflammatory responses, although the CNVR at 21q21.1 does not correspond to any QTL. AASTH=Allergic/Atopic Asthma, OSTEAR=Osteoarthritis, RA=Rheumatoid Arthritis and COPD12=Chronic Obstructive Pulmonary disease, BW=Body Weight, BP=Blood Pressure.

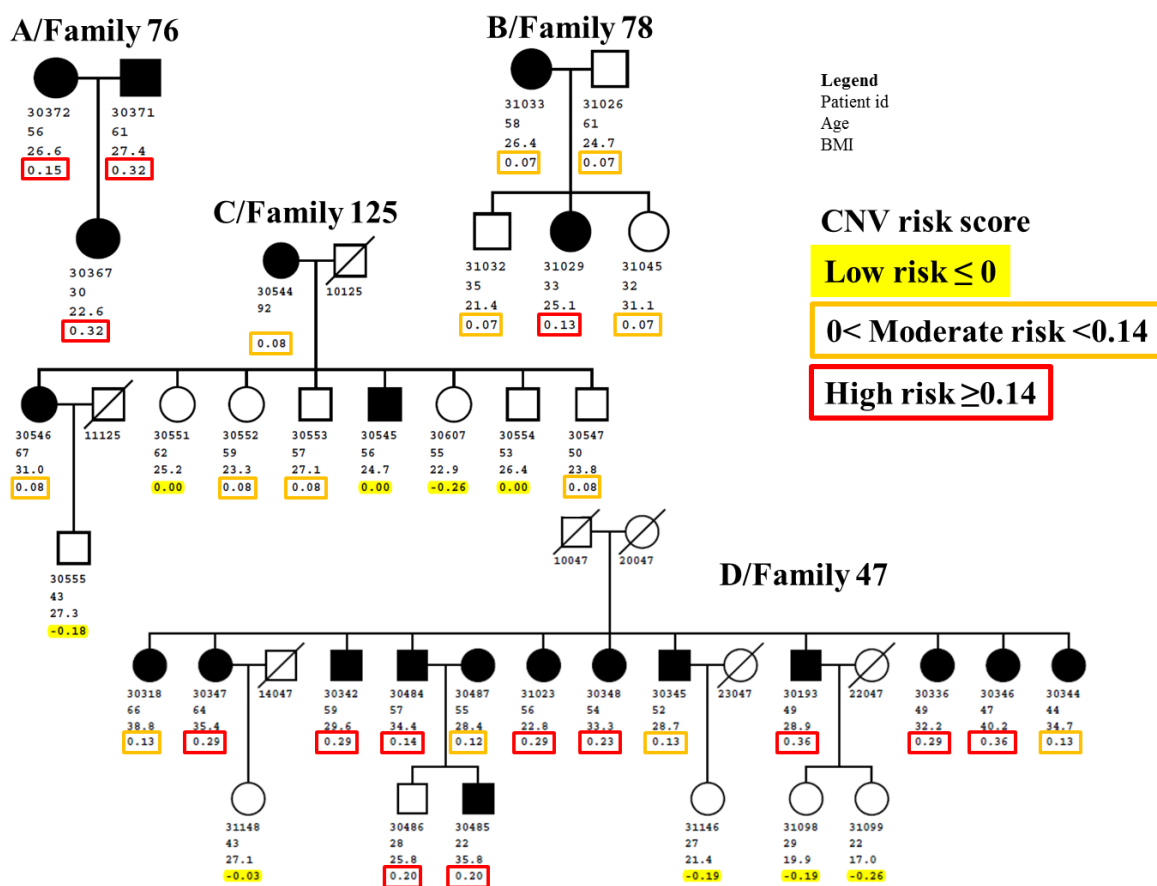


Figure S7. The CNV risk score in pedigrees. The CNV risk score ranges from -26% to 36%. A/Members of family 76 were at high risk for hypertension and appeared all hypertensives; B/Members of family 78 carried the deletion at 17p11.3 and were mostly at moderate risk for hypertension, with 31029 having the highest CNV risk score, as she carried the complete loss. She was at high risk for hypertension and affected by hypertension at 33 years old; B/Members of families 125 showed relatively low and moderate risk for hypertension. They appeared mainly normotensives; D/Members of families 47 showed rather high CNV scores and appeared mainly hypertensives. However some members were at moderate risk, while all low risk members were normotensives. *Squares indicate male, circles indicate female and affected family members are represented by filled black square or circle. A cross through a circle or a square specifies the individual is either deceased or ungenotyped.*

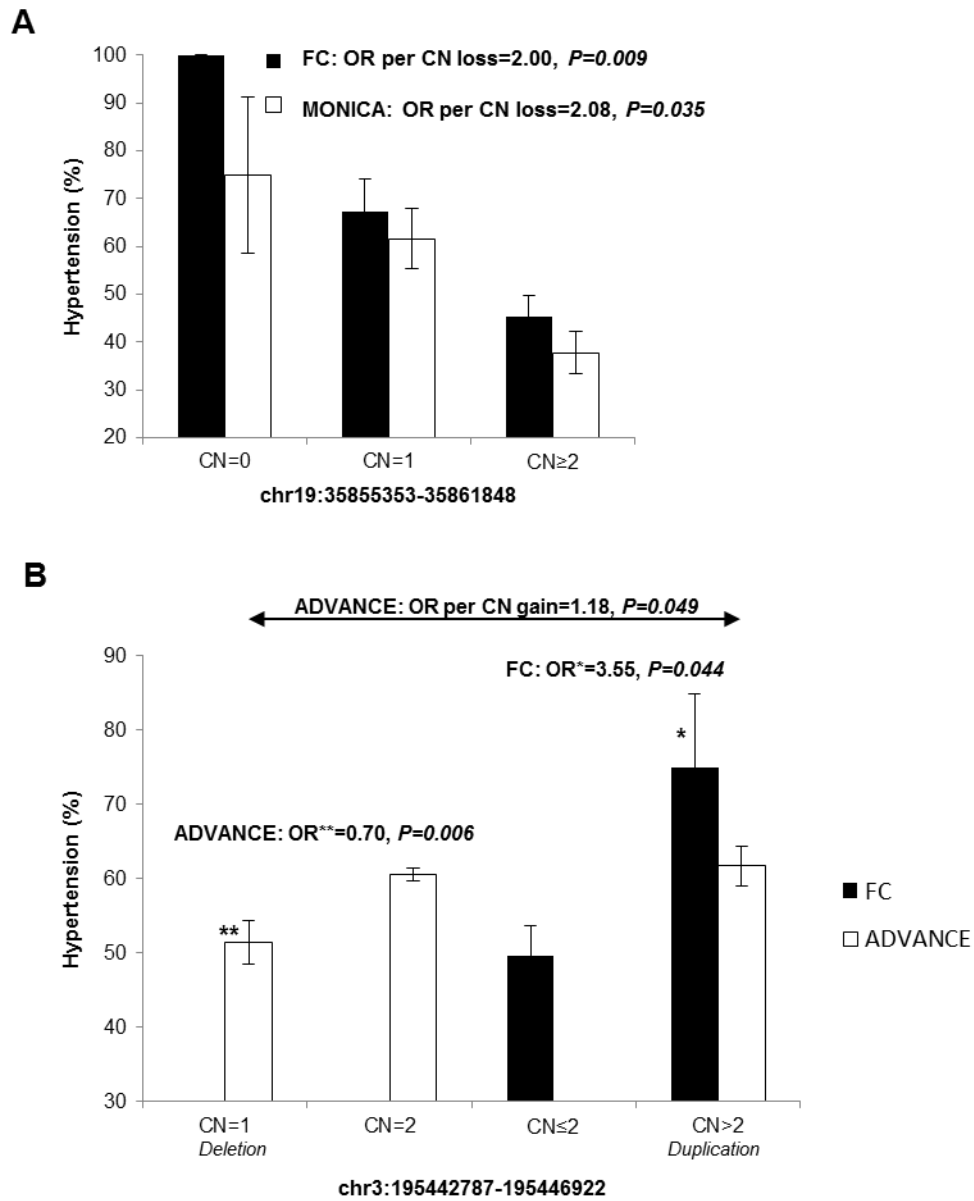


Figure S8. Replication in Czech Post-MONICA (CPM) and ADVANCE. A- The dose effect of 19q13.12-loss identified in FC, was replicated in CPM subjects. B- In ADVANCE, we confirmed that CN gain at 3q29 increased the risk for hypertension, 3q29-loss showing a strong protective effect. Copy Number=CN; Gain-carriers: carry CN>2; Loss-carriers: carry CN<2. Positions are given relative to UCSC Hg19.

Data Supplement Chapter- II REFERENCES

1. Santé Québec, Daveluy C, Chénard L, Levasseur M, Émond A. Et votre Coeur, Ça Va? Rapport de l'Enquête Québécoise sur la Santé Cardio-Vasculaire 1990. Ministère de la santé et des services sociaux, gouvernement du Québec. 1994; chapitre 3: 21-38
2. Joffres MR, Hamet P, MacLean DR, L'Italien G J, Fodor G. Distribution of blood pressure and hypertension in canada and the united states. *Am J Hypertens.* 2001;14:1099-1105
3. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet.* 2006;38:904-909
4. Wang K, Li M, Hadley D, Liu R, Glessner J, Grant SF, Hakonarson H, Bucan M. Penncnv: An integrated hidden markov model designed for high-resolution copy number variation detection in whole-genome snp genotyping data. *Genome Res.* 2007;17:1665-1674
5. Glessner JT, Li J, Hakonarson H. Parsecnv integrative copy number variation association software with quality tracking. *Nucleic Acids Res.* 2013;41:e64
6. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC. Plink: A tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 2007;81:559-575
7. Korn JM, Kuruvilla FG, McCarroll SA, et al. Integrated genotype calling and association analysis of snps, common copy number polymorphisms and rare cnvs. *Nat Genet.* 2008;40:1253-1260
8. Parikh NI, Pencina MJ, Wang TJ, Benjamin EJ, Lanier KJ, Levy D, D'Agostino RB, Sr., Kannel WB, Vasan RS. A risk score for predicting near-term incidence of hypertension: The framingham heart study. *Ann Intern Med.* 2008;148:102-110