

Université de Montréal

Variants génétiques associés à l'adiposité et à la pression artérielle : Une réplication

Par Danick Goulet

Département de médecine sociale et préventive
École de santé publique de l'Université de Montréal

Mémoire présenté
en vue de l'obtention du grade de Maîtrise
en Épidémiologie

Août 2020

© Danick Goulet 2020

Université de Montréal
Département de médecine sociale et préventive, École de santé publique de l'Université de
Montréal

Ce mémoire intitulé

Variants génétiques associés à l'adiposité et à la pression artérielle : Une réplication

Présenté par

Danick Goulet

A été évalué par un jury composé des personnes suivantes

Helen Trottier

Président-rapporteur

Marie-Pierre Sylvestre

Directeur de recherche

Bouchra Nasri

Membre du jury

RÉSUMÉ

Introduction: L'hypothèse de l'architecture génétique commune de phénotypes d'adiposité et de pression artérielle a été testée chez les adolescents par Melka et al. (*Journal of Clinical Endocrinology & Metabolism*, 2012) qui ont identifié cinq variants associés à l'indice de masse corporelle, rs16933812 (*PAX5*), rs7638110 (*MRPS22*), rs9930333 (*FTO*), rs7120548 (*MTCH2*) et rs17773430 (*MC4R*), dont trois (*PAX5*, *MRPS22* et *FTO*) l'était aussi avec la pression artérielle systolique.

Objectif : Investiguer si les associations entre les cinq variants et l'adiposité et la pression artérielle sont répliquées dans une population d'âge et de situation géographique similaire.

Méthodes: Les données proviennent de l'étude Nicotine Dependence In Teens, une étude longitudinale d'adolescents suivis à partir de 12 ans. L'ADN de 713 adolescents d'ascendance européenne a été génotypé et des mesures anthropométriques et de pression artérielle ont été effectuées lorsqu'ils avaient en moyenne 13, 15, 17 et 24 ans. Les associations entre les cinq variants avec l'indice de masse corporelle et la pression artérielle systolique ont été estimées à l'aide de modèles linéaires mixtes.

Résultats: Les associations des variants sur les gènes *FTO* et *MTCH2* avec l'indice de masse corporelle ont été répliquées, mais pas celles correspondantes aux variants sur *MRPS22*, *PAX5* et *MC4R*. Aucune des associations avec la pression artérielle systolique n'a été répliquée.

Conclusion: Les facteurs proposés pour expliquer la réplification partielle des résultats comprennent une différence de structure génétique entre les populations étudiées, une définition différente du phénotype de pression artérielle et la surestimation de certains effets estimés dans Melka et al.

Mots-clés : Adiposité, pression artérielle, association génétique, réplification, adolescence

ABSTRACT

Introduction : Melka et al. (*Journal of Clinical Endocrinology & Metabolism*, 2012) tested the hypothesis that adiposity and blood pressure share a common genetic architecture in adolescents. They identified five single nucleotide variants, rs16933812 (*PAX5*), rs7638110 (*MRPS22*), rs9930333 (*FTO*), rs7120548 (*MTCH2*) and rs17773430 (*MC4R*), associated with body mass index, three of which (*MRPS22*, *PAX* and *FTO*) were also associated with systolic blood pressure.

Objective: To investigate if the reported associations between the five variants and adiposity and blood pressure can be replicated in a population similar in age and geographical situation.

Methods: Data were drawn from the Nicotine Dependence In Teens study, a longitudinal study of adolescents followed-up from 12 years old. The DNA of 713 adolescents of European ancestry was genotyped and anthropometric and systolic blood pressure measurements were taken at age 13, 15, 17 and 24 on average. The associations between the five variants and both body mass index and systolic blood pressure were estimated using linear mixed models.

Results: We replicated the associations of variants at genes *FTO* and *MTCH2* with body mass index, but not those corresponding to *MRPS22*, *PAX* or *MC4R*. Further, none of the associations with systolic blood pressure were replicated.

Conclusion: Explanations for the partial replication of results include differences in genetic structures across study populations, different definitions of the blood pressure phenotype and the overestimation of specific effects initially estimated by Melka et al.

Keywords: Adiposity, blood pressure, genetic associations, replication, adolescence

TABLE DES MATIÈRES

RÉSUMÉ	1
ABSTRACT.....	2
LISTE DES TABLEAUX.....	6
LISTE DES FIGURES	7
ABBREVIATIONS	8
REMERCIEMENTS.....	9
CHAPITRE 1 - INTRODUCTION	10
CHAPITRE 2 – OBJECTIF DU MÉMOIRE	14
CHAPITRE 3 – RECENSION DES ÉCRITS	15
3.1 Pression artérielle.....	15
3.1.1 Définition conceptuelle.....	15
3.1.2 Méthodes de mesure de la pression artérielle	16
3.1.3 Effets néfastes d’une pression artérielle élevée	17
3.1.4 Génétique de la pression artérielle.....	20
3.2 Adiposité.....	23
3.2.1 Définition.....	23
3.2.2 Méthodes de mesure de l’adiposité.....	23
3.2.3 Effets néfastes de l’excès d’adiposité	25
3.2.4 Génétique de l’adiposité	27
3.3 Relation entre l’adiposité et la pression artérielle.....	29
3.3.1 Association entre l’adiposité et la pression artérielle	29
3.3.2 Relation génétique entre l’adiposité et la pression artérielle.....	30
3.3.3 Article par Melka et al.	34
3.4 Variants génétiques à répliquer.....	36
3.4.1 rs16933812 (PAX5).....	36
3.4.2 rs7638110 (MRPS22)	37
3.4.3 rs9930333 (FTO)	38
3.4.4 rs17773430 (MC4R)	39
3.4.5 rs7120548 (MTCH2)	40
3.5 Réplication en études d’association génétique	41

CHAPITRE 4 – MÉTHODES	45
4.1 Source des données	45
4.1.1 Étude NDIT	45
4.1.2 Échantillon analytique et pertes au suivi	46
4.2 Variables d'étude	47
4.2.1 Variables dépendantes	47
4.2.2 Variables indépendantes	48
4.2.3 Variables additionnelles.....	48
4.3 Analyses statistiques	50
4.3.1 Analyses descriptives.....	50
4.3.2 Analyses longitudinales	50
4.3.3 Taille d'échantillon et puissance statistique	52
4.3.4 Analyses de sensibilité.....	52
4.4 Comparaisons des études	53
CHAPITRE 5 – RÉSULTATS	55
5.1 Présentation du manuscrit.....	55
5.2 Manuscrit	56
5.3 Résultats supplémentaires.....	78
5.3.1 Comparaison de l'échantillon analytique NDIT avec celui de la SYS	78
5.3.2 Choix du modèle pour l'effet du temps sur l'IMC et la PAS	79
5.3.3 Analyses de sensibilité.....	80
CHAPITRE 6 – DISCUSSION	82
6.1 Interprétation des résultats	83
6.1.1 Variants dont l'association avec l'IMC est connue	83
6.1.2 Variants nouvellement identifiés comme associés à l'IMC.....	87
6.3 Explications possibles de non-réplication.....	88
6.3.1 Erreurs aléatoires	88
6.3.2 Effets non comparables entre les études	89
6.3.3 Risques de biais.....	94
6.3.4 Revue des explications.....	99
6.4 Forces et limites	100
6.5 Contributions du mémoire et futures directions.....	101

6.6 Conclusion	102
BIBLIOGRAPHIE	104
ANNEXE A. Définitions et concepts pour études d'association génétique	120
ANNEXE B. Description des gènes reliés aux variants à l'étude	128
ANNEXE C. Description détaillée des variables du questionnaire NDIT	135
ANNEXE D. Tableaux additionnels	138

LISTE DES TABLEAUX

Tableau 1. Variants associés à l'adiposité et leur association avec la PAS. Melka et al. (2012) [†]	36
Tableau 2. Recommandations adaptées de la littérature [†] concernant la réplication d'études d'association génétique.	44
Tableau 3. Caractéristiques au cycle 1 des participants inclus et exclus, NDIT 1999 [†]	46
Tableau 4. Description des pertes au suivi et du nombre d'observations disponibles pour l'IMC et la PAS, NDIT 1999-2012	47
Tableau 5. Comparaison des méthodes utilisées dans NDIT et par la SYS	54
Tableau 6. Description des observations provenant de NDIT et du SYS, combinaison de tous les cycles NDIT	79
Tableau 7. Choix du modèle pour l'effet du temps sur l'IMC et la PAS	80

LISTE DES FIGURES

Figure 1. Illustration d'un biais de sélection pour l'association entre X et Y.	95
Figure 2. Illustration d'un biais de sélection pour l'association entre X et Y en raison de la perte au suivi.	96

ABBREVIATIONS

PAS: Pression artérielle systolique

PAD : Pression artérielle diastolique

IMC : Indice de masse corporelle

GWAS : Étude d'association pangénomique, *Genome-wide association study*

NDIT: Nicotine Dependance In Teens study

IMCZ: Score Z de l'indice de masse corporelle

SYS : Saguenay Youth Study

MAF: Fréquence de l'allèle mineur, *Minor allele frequency*

LD: Déséquilibre de liaison, *Linkage disequilibrium*

REMERCIEMENTS

Je dois d'abord remercier Marie-Pierre Sylvestre pour son encadrement tout au long de la complétion de mon mémoire. J'ai été très chanceux de pouvoir profiter de son expertise et de ses conseils qui m'ont grandement aidé à compléter ma rédaction.

Je me dois également de remercier tous les étudiants et membres de notre équipe au centre de recherche pour me permettre de partager mon travail et m'aider à améliorer certaines parties de mon mémoire. Je souligne également l'excellente capacité d'adaptation de toute l'équipe pour s'être ajustée à un événement inattendu, une pandémie.

Aussi, merci à toute l'équipe de professeur à l'École de santé publique qui m'a offert une qualité d'apprentissage hors pair. Les connaissances et les compétences qui m'ont été transmises me seront utiles tout au long de la poursuite de mes études et bien au-delà.

Finalement, je ne peux oublier le soutien de ma famille et mes amis tout au long de cette aventure. Bien que l'arrivée inopportune d'une certaine pandémie ait réduit considérablement nos interactions en personne, les encouragements et distractions offertes par téléphone ou en ligne ne sont pas passés inaperçus. Un gros merci à mes parents, en particulier pour leur appui ces deux dernières années, mais aussi pour le tout support offert depuis si longtemps.

CHAPITRE 1 - INTRODUCTION

L'hypertension est le problème de santé responsable du plus grand nombre de mortalité et de handicap dans le monde (1). Selon les données disponibles pour l'année 2015, l'hypertension était liée à environ 10 millions de décès et à la perte de 211 millions d'années de vie en bonne santé dans le monde (2). Le Canada se compare favorablement aux autres pays développés, ayant une prévalence d'hypertension parmi les plus faibles et un niveau de traitement et de contrôle parmi les plus élevés (3). Toutefois, l'hypertension affecte tout de même près de 25% des adultes canadiens (4) et on estime que cette maladie compte pour 10% du budget canadien en soins de santé (5).

L'hypertension se produit lorsque le sang exerce une pression élevée de façon persistante sur les vaisseaux sanguins. Au Canada, une personne est diagnostiquée comme hypertendue à la suite de l'observation d'une pression artérielle élevée ($\geq 135/85$ mm Hg) en clinique de façon automatisée et confirmée de façon ambulatoire (6). Ce seuil est déterminé par le niveau de pression artérielle où les bénéfices liés au traitement surpassent les inconvénients (7). Le fardeau pour la santé associé à l'hypertension est relié au développement de maladies cardiovasculaires, incluant la cardiopathie ischémique et l'accident vasculaire cérébral (2). L'apparition de problèmes de santé peut toutefois survenir à des niveaux de pression artérielle inférieurs à ceux définissant l'hypertension. En effet, on estime qu'à partir de 115/75 mm Hg, chaque élévation de 20 mm Hg de pression artérielle systolique (PAS) et de 10 mm Hg de pression artérielle diastolique (PAD) double le risque de mortalité lié aux maladies du système cardiovasculaire (8).

L'hypertension se développe sur le long terme et pourrait prendre racine dès l'enfance et l'adolescence. De 2012 à 2015, environ 7% des adolescents canadiens âgés de 6 à 19 ans avaient une pression artérielle élevée (6-17 ans : $\geq 90^{\text{e}}$ percentile ou 120/80 mm Hg, 18-19 ans : $\geq 120/80$ mm Hg) (9, 10), comparativement aux 4% observés de 2009 à 2011 (10). Les enfants et les adolescents présentant une préhypertension (90^{e} au 95^{e} percentile) ou hypertension ($\geq 95^{\text{e}}$ percentile) ont un risque plus important d'être hypertensifs à l'âge adulte (11) et présentent un risque plus important de développer des troubles de santé comme l'hypertrophie du ventricule gauche (12), l'athérosclérose (13) et l'épaississement des artères carotides (14) qui sont des facteurs de risque de maladies cardiovasculaires à l'âge adulte (15). Il est donc essentiel de pouvoir

identifier les enfants et les adolescents qui sont plus à risque de développer une pression artérielle élevée durant l'enfance et l'adolescence afin de prévenir ses effets néfastes à court et à long terme.

La pression artérielle est un trait multifactoriel qui peut être influencé à la fois par des facteurs de risques non modifiables comme la génétique et par des facteurs de risque modifiables comme un niveau élevé d'adiposité, une haute consommation d'alcool, un faible niveau d'activité physique et une consommation excessive de sel (16). Plus spécifiquement, le surplus de poids est l'une des causes principales de l'hypertension et contribue à 60% à 70% des cas chez les adultes (17). Une méta-analyse de données longitudinales suggère que pour chaque augmentation de cinq unités d'indice de masse corporelle (IMC) le risque d'hypertension augmente de 49% (18). Des études effectuées sur des modèles animaux et des humains suggèrent que le système nerveux sympathique, le système rénine-angiotensine aldostérone, la compression rénale et la résistance au leptine jouent un rôle dans l'hypertension induite par le surplus de poids (19). Ainsi, l'occurrence grandissante du surpoids et de l'obésité au sein de la population canadienne observée au cours des dernières années (20) est probablement une des raisons principales en lien à l'augmentation de la prévalence de la pression artérielle (21).

Les données épidémiologiques s'accordent également pour soutenir que le lien entre l'adiposité et la pression artérielle est présent dès l'enfance et l'adolescence. Notamment, les enfants en surpoids et en situation d'obésité montrent une augmentation de PAS d'environ 5 mm Hg et 7 mm Hg, respectivement, comparativement à des enfants ayant un poids normal (22). Également, pour chaque augmentation d'un décile d'IMC chez les adolescents, on estime que la PAS augmente de 10 mm Hg (23). On observe que les jeunes ayant un niveau d'adiposité élevé durant l'enfance et l'adolescence, de même ceux présentant une importante augmentation de l'adiposité sont davantage à risque de développer une PAS élevée à l'âge adulte (24-27).

Le lien observé entre l'adiposité et la PAS pourrait être expliquée en partie par la présence de facteurs génétiques affectant les niveaux d'adiposité et de pression artérielle de façon conjointe, comme souligné dans l'hypothèse de variants communs/multiples maladies introduite par Becker (28). Cette hypothèse fait référence à ce que certains variants génétiques communs aient un rôle dans plus d'un trait relié. Cette hypothèse est supportée par les résultats d'études qui suggèrent que la corrélation élevée observée entre l'adiposité et la pression artérielle chez des jumeaux ou à l'intérieur de famille aurait une explication génétique (29). Une étude précédente effectuée par

Melka et al. (30) a identifié cinq variants génétiques associés avec l'IMC, dont deux nouveaux découverts avec une analyse d'association pangénomique (*genome-wide association study*, GWAS), au sein d'un échantillon de 598 adolescents âgés de 12 à 18 ans de la région du Saguenay-Lac-St-Jean au Québec, Canada. Trois de ces variants étaient également associés avec la PAS dans le même échantillon. Ces résultats n'ont toutefois pas été répliqués au sein d'autres échantillons indépendants.

Examiner l'architecture génétique sous-jacente à l'adiposité et la pression artérielle chez les jeunes est important puisqu'on ne peut pas assumer qu'elle peut être inférée directement aux enfants et adolescents (31). Par exemple, les résultats d'une étude longitudinale sur des jumeaux suggèrent que les facteurs génétiques associés à la pression artérielle sont exprimés différemment pendant l'enfance et le début de l'âge adulte (32). De plus, examiner la susceptibilité génétique chez les jeunes, qui sont rarement traités avec des médicaments antihypertenseurs et qui sont moins affectés que les adultes par l'effet cumulatif de l'exposition à l'environnement, pourrait aider à identifier les facteurs génétiques qui affectent l'adiposité et la pression artérielle (33, 34). Cependant, les études examinant les facteurs génétiques qui influencent l'adiposité et la pression artérielle chez les jeunes sont peu nombreuses et elles ont généralement de plus petites tailles d'échantillon comparativement aux études sur les adultes (35, 36), illustrant la nécessité à répliquer les résultats précédemment rapportés.

L'importance de la réplication des études d'association génétique est aussi supporté par l'impact du *Winner's Curse* (37), décrit brièvement ici et en plus de détails dans la revue de la littérature ([Chapitre 3.5 Réplication en études d'association génétique](#)). Ce phénomène fait référence à ce que l'effet de variants génétiques nouvellement découverts dans une GWAS est souvent surestimé dans l'étude ayant effectué la découverte (38). Ce phénomène s'applique particulièrement aux GWAS puisque les variants sont identifiés dans ces études à l'aide d'un seuil de significativité très strict. Ainsi, les variants identifiés pour réplifications ultérieures sont ceux dont les résultats sont extrêmes dans le jeu de données et ceux-ci sont en fait souvent surestimés (39). Souvent, ce problème survient lorsque le même échantillon est utilisé pour identifier des variants génétiques associés à un trait et pour estimer l'effet (40), et que les variants identifiés le sont en raison d'une significativité statistique (41). Globalement, ce phénomène complique la réplication de résultats initiaux, particulièrement si le calcul de la taille d'échantillon dans une

étude de réplication est basé sur les surestimations des tailles d'effet de l'étude initiale, ce qui peut résulter en une puissance non suffisante dans l'étude de réplication (42).

Ce mémoire vise à apporter davantage de connaissances concernant les facteurs génétiques affectant l'adiposité et la pression artérielle chez les jeunes. L'approche privilégiée est celle de la réplication d'analyses déjà publiées par Melka et al. (30) qui ont identifié cinq variants génétiques associés avec l'IMC sur les gènes *FTO* (rs9930333, $p = 1,9 \times 10^{-4}$), *MTCH2* (rs7120548, $p = 1,9 \times 10^{-4}$), *MC4R* (rs17773430, $5,8 \times 10^{-6}$), *PAX5* (rs16933812, $p = 5,2 \times 10^{-6}$) et *MRPS22* (rs7638110, $p = 4,6 \times 10^{-8}$), desquels rs9930333 ($p = 0,01$), rs16933812 ($p = 0,0007$) et rs7638110 ($p = 0,02$) étaient également associés avec la PAS dans le même échantillon de 598 adolescents âgés de 12 à 18 ans de la région du Saguenay-Lac-St-Jean. Le mémoire capitalise sur l'existence de données mesurées sur des adolescents d'âge semblable pour tenter de répliquer les résultats de Melka et al. (30). Les données en question proviennent de l'étude Nicotine Dependence In Teens (NDIT) (43) et sont tirées d'étudiants âgés de 12 ans suivis longitudinalement jusqu'à l'âge adulte dans la région de Montréal au Québec, Canada. Contraster les résultats de deux études dont les populations sont semblables réduit les facteurs pouvant causer l'inconsistance de résultats, et permet ainsi de mieux valider la plausibilité des résultats observés dans l'étude initiale. Le mémoire est organisé en sept chapitres. Le chapitre 2 présente l'objectif spécifique du mémoire. Le chapitre 3 contient la revue de la littérature. Le chapitre 4 décrit les méthodes utilisées. Le chapitre 5 présente les principaux résultats sous forme de manuscrit d'un court article scientifique (*brief research report*) en plus de présenter des résultats supplémentaires. Le chapitre 6 propose une discussion des résultats obtenus et la conclusion du mémoire.

CHAPITRE 2 – OBJECTIF DU MÉMOIRE

La réplication de résultats obtenus dans une étude antérieure représente une bonne avenue afin de renforcer les connaissances concernant l'identification de variants génétiques affectant l'adiposité et la pression artérielle. Melka et al. (30) ont détecté cinq variants associés à l'IMC sur les gènes *PAX5*, *MRPS22*, *FTO*, *MTCH2* et *MC4R*, dont deux nouveaux (*PAX5* et *MRPS22*) identifiés par GWAS, chez 598 adolescents âgés de 12 à 18 ans du Saguenay-Lac-Saint-Jean. Les variants sur les gènes *PAX5*, *MRPS22* et *FTO* étaient également associés à la PAS dans le même échantillon.

L'objectif spécifique du mémoire consiste à répliquer les associations observées par Melka et al. (30) au sein d'un échantillon d'adolescents semblables suivis longitudinalement à partir de 12 ans. On considère comme une réplication des effets statistiquement significatifs et dans la même direction que dans l'étude de Melka et al. (30). Plus spécifiquement, les associations à répliquer sont les suivantes :

- Association des variants rs16933812 (*PAX5*), rs7638110 (*MRPS22*), rs9930333 (*FTO*), rs7120548 (*MTCH2*) et rs17773430 (*MC4R*) avec l'IMC
- Association des variants rs16933812 (*PAX5*), rs7638110 (*MRPS22*) et rs9930333 (*FTO*) avec la PAS

Nous anticipons que certaines associations seront répliquées, mais pas toutes. Plus spécifiquement, puisque les gènes *FTO*, *MTCH2* et *MC4R* sont déjà connus pour leur lien avec l'adiposité, les variants investigués sur ces gènes devraient être répliqués pour l'association avec l'IMC.

CHAPITRE 3 – RECENSION DES ÉCRITS

La recension des écrits présente en premier lieu les deux traits d'intérêt du mémoire, soit la pression artérielle et l'adiposité. Pour chacun d'eux, on s'intéressera à leur définition, les méthodes standards pour les mesurer, leur impact sur la santé et leur composante génétique respective. Un aperçu des écrits concernant la relation épidémiologique et génétique unissant la pression artérielle et l'adiposité est ensuite présenté, incluant la place qu'occupe l'étude de Melka et al. (30) dans cette littérature. On propose ensuite une description des variants génétiques investigués et des gènes qui leur sont associés afin d'offrir une mise à jour des connaissances relatives au rôle possible qu'ils peuvent jouer dans la régulation de la pression artérielle et de l'adiposité. Finalement, l'importance de la réplcation dans les études d'association génétique est discutée. Les lecteurs moins familiers avec certains termes et méthodes utilisés en épidémiologie génétique peuvent se référer à [l'Annexe A](#) qui détaille certaines définitions et concepts de base des études d'association génétique.

3.1 Pression artérielle

3.1.1 Définition conceptuelle

La pression artérielle se définit par la pression exercée par le sang sur les vaisseaux sanguins. On distingue la PAS de la PAD. La PAS correspond à la pression exercée sur les vaisseaux lors de la systole, soit la phase de contraction du cœur. La PAD fait référence à la pression lors de la diastole qui correspond à la phase de remplissage du cœur (44). La pression artérielle est influencée par le débit cardiaque et la résistance des vaisseaux sanguins périphériques. Le débit cardiaque est fonction de la fréquence cardiaque et du volume d'éjection systolique, lui-même influencé par le volume sanguin. Plus le débit cardiaque est élevé, plus la pression exercée par le sang sur les vaisseaux sanguins est grande. La résistance vasculaire est principalement influencée par le diamètre des vaisseaux sanguins. Plus le diamètre des vaisseaux est petit, plus la résistance vasculaire est grande et plus la pression artérielle est élevée. La résistance vasculaire est régulée par le système nerveux autonome et des hormones vasomotrices (44).

La pression artérielle est un trait physiologique continu qui varie entre les individus. À court terme, la pression artérielle d'un individu peut varier selon le moment de la journée, la position du corps ou l'exercice physique (45). Bien que ces variations à court terme de pression artérielle comportent des conséquences cardiovasculaires non négligeables (45), le mémoire se

focalise sur l'élévation chronique de la pression artérielle. Celle-ci survient lors du dérèglement à long terme de la régulation du débit cardiaque ou de la résistance vasculaire périphérique (44). La régulation à long terme de la pression artérielle passe principalement par l'homéostasie de l'eau et du sel. Dans plusieurs cas, une augmentation de la pression artérielle chronique est établie en réponse à un dysfonctionnement rénal nécessitant une pression efférente plus élevée afin de conserver l'équilibre entre la rétention et l'excrétion d'eau et de sel (46). La sensibilité au sel, le système Rénine-Angiotensine Aldostérone, le remodelage structurel des vaisseaux sanguins, le stress oxydatif et le système nerveux autonome contribuent également à la régulation au long terme de la pression artérielle (46). Le surpoids et l'obésité, une consommation élevée en sodium et faible en potassium, un faible niveau d'activité physique et la consommation d'alcool représentent les comportements modifiables les plus solidement associés à l'augmentation de la pression artérielle (16).

2.1.2 Méthodes de mesure de la pression artérielle

Le sphygmomanomètre au mercure est considéré comme l'étalon d'or pour mesurer avec précision la pression artérielle. Or, l'utilisation de ce type d'outil se fait de plus en plus rare en raison de son interdiction d'utilisation en milieu clinique dû aux risques pour la santé et pour l'environnement causé par le mercure (47). Cet outil a largement été remplacé en clinique et en recherche par les tensiomètres anéroïdes et les oscillomètres électroniques. Le tensiomètre anéroïde requiert l'auscultation comme son homologue utilisant le mercure, mais est reconnu pour être moins précis et nécessitant des calibrations fréquentes (48). L'oscillomètre fonctionne en détectant les oscillations dans le volume sanguin pulsatile dans la gaine posée au bras. Cet outil est de plus en plus utilisé en raison de la facilité de son utilisation qui en fait un bon outil de mesure de pression artérielle hors clinique (48).

Les mesures de pression artérielle peuvent être effectuées soit en milieu clinique ou bien hors clinique. Selon les lignes directrices 2020 pour le diagnostic, l'évaluation des risques, la prévention et le traitement de l'hypertension d'Hypertension Canada (6), les mesures de pression artérielle en milieu clinique peuvent être effectuées de façon manuelle ou préférablement de façon automatique à l'aide d'un oscillomètre électronique. L'utilisation d'appareils automatiques permet d'obtenir une plus grande quantité de mesures et donc d'estimer la pression artérielle de manière

plus précise qu'avec une mesure ponctuelle. Aussi, l'utilisation d'appareils automatiques ne requiert pas la présence d'un professionnel de la santé lors de la mesure. Cela permet d'éviter l'inflation des valeurs de pression artérielle par l'effet de la blouse blanche. Cet effet correspond à l'élévation de la pression artérielle chez un patient dû à l'anxiété en réponse à la présence du professionnel de la santé (48). Les mesures automatiques devraient être prises à 1 à 2 minutes d'intervalle, la gaine placée sur le bras au niveau du cœur, et le patient placé en position assise, les jambes non croisées et sans parler. L'utilisation de mesures automatiques ne requiert pas une période de repos avant les mesures, comparativement aux mesures manuelles qui nécessitent une période de 5 minutes de repos préalables aux mesures. La première mesure manuelle doit être écartée et les deux mesures suivantes moyennées. Il est également important d'utiliser une gaine de taille adéquate, surtout chez les patients en surplus de poids. L'utilisation d'une gaine trop étroite chez ces patients peut mener à des mesures de pression artérielle surélevées (6).

Les méthodes de mesures sont en majorité les mêmes chez les enfants et les adolescents que chez les adultes. Les mesures de pression artérielle en clinique peuvent être effectuées avec un sphygmomanomètre anéroïde ou un oscillomètre électronique validé pour l'âge visé et devraient être effectuées dès l'âge de trois ans (49). Puisque les données relatives à la pression artérielle normalisée ont majoritairement été obtenues à l'aide d'un sphygmomanomètre et qu'encore peu d'appareils oscillométriques ont été validés chez les enfants, les mesures de pression artérielle anormalement élevée avec un oscillomètre devraient être validées par auscultation (50). Les mesures devraient également être effectuées après un repos de 5 minutes et en position assise. La mesure devrait être effectuée au bras droit puisque les données normatives pour ce groupe d'âge ont été récoltées au bras droit des patients (50). Une attention particulière devrait également être portée pour le choix de diamètre de la gaine du tensiomètre utilisé chez les enfants afin d'éviter des erreurs de mesure (51).

3.1.3 Effets néfastes d'une pression artérielle élevée

Comptant pour environ 10 millions de morts, l'élévation de la pression artérielle était le premier facteur responsable de morts prématurées en 2015 (2). La pression artérielle influence la mortalité et la morbidité en étant un facteur de risque important pour plusieurs événements liés au système cardiovasculaire. Par exemple, la majorité de la perte des années de vie corrigées par

l'incapacité associées avec l'élévation de la PAS est reliée à la maladie cardiovasculaire ischémique et l'accident vasculaire cérébral (2). De plus, l'élévation de la PAS et de la PAD est un facteur de risque pour d'autres troubles cardiovasculaires incluant l'insuffisance cardiaque, la maladie artérielle périphérique, et l'infarctus du myocarde (8, 52).

La conséquence directe de l'élévation chronique de la pression artérielle est l'hypertension. Celle-ci se produit lorsque le sang exerce une pression élevée de façon persistante sur les vaisseaux sanguins et fait référence à une élévation à un niveau où les bénéfices du traitement, par médicaments ou comportemental, dépassent les risques documentés dans les essais cliniques (7). Les lignes directrices cliniques en matière de pression artérielle au Canada définissent le seuil d'hypertension à 140 mm Hg pour la PAS et à 90 mm Hg pour la PAD pour une pression mesurée en clinique de façon manuelle (6). Les lignes directrices européennes (7) s'accordent en grande partie avec celles du Canada, alors que celles des États-Unis (16) ont récemment diminué les seuils des différents niveaux de pression artérielle en 2017. Puisque la pression artérielle des enfants et des adolescents (1 à 18 ans) est influencée par l'âge, le sexe et la taille, les lignes directrices proposent donc des seuils basés sur des tables de distribution de pression artérielle selon l'âge, le sexe et la taille (6). Ainsi, les lignes directrices canadiennes proposent un diagnostic d'hypertension advenant l'observation à trois occasions d'une pression artérielle supérieure ou égale au 95^e percentile ajusté pour l'âge, le sexe et la grandeur, ou supérieure à 120/80 chez les 6 à 11 ans et 130/85 chez les 12 à 17 ans (6).

La classification de la pression artérielle selon des seuils peut être utile afin d'identifier des individus particulièrement à risque de développer des troubles de santé, mais des niveaux de pression artérielle sous les seuils d'hypertension peuvent également présenter des risques importants pour la santé. Plusieurs études épidémiologiques ont montré une association linéaire entre l'élévation de la pression artérielle et le risque de développer des maladies cardiovasculaires. Par exemple, la Prospective Studies Collaboration (8) a combiné les données d'environ 1 million de participants provenant de 61 études observationnelles effectuées entre 1950 et 1990. La pression artérielle à des niveaux aussi bas que 115/75 mm Hg y était associée de façon log linéaire avec la mortalité liée aux accidents vasculaires cérébraux, la maladie cardiaque ischémique et d'autres causes vasculaires comme l'insuffisance cardiaque, l'athérosclérose ou l'anévrisme aortique (8). Une étude plus récente a évalué l'incidence de 12 troubles cardiovasculaires en relation avec la

pression artérielle en analysant le dossier électronique de 1,25 million de personnes de 1997 à 2010 avec le programme CARDiovascular research using LInked Bespoke studies and Electronic health Records. L'analyse montre une augmentation linéaire du risque pour les 12 troubles cardiovasculaires investigués incluant l'angine, l'infarctus du myocarde, l'insuffisance cardiaque et l'accident vasculaire cérébral (53). La relation linéaire entre la pression artérielle et les troubles cardiovasculaires est observée 30 à 89 ans (8, 53). Toutefois, le risque relatif de troubles cardiovasculaires associé avec l'élévation de la pression artérielle diminue graduellement avec l'âge (8, 53).

La PAS et la PAD sont des facteurs de risques indépendants des troubles cardiovasculaires. Examinées séparément, tant la PAS que la PAD sont de bons prédicteurs d'événements cardiovasculaires (8, 52, 53). Toutefois, la PAD est moins utilisée dans l'évaluation des risques liés à l'élévation de la pression artérielle chez les patients de plus de 50 ans puisque celle-ci tend à diminuer vers la fin de l'âge adulte (7). Une étude évaluant le risque d'accident vasculaire cérébral associé avec la PAS et a PAD chez 68 551 sujets âgés de 19 à 78 ans, montrait que le risque d'accident vasculaire cérébral associé à l'élévation de la PAD diminuait avec l'âge (54). Pour une PAD supérieure à 71 mm Hg, le risque d'accident vasculaire cérébral associé à une augmentation de PAD de 5 mm Hg était de 1,19 (IC95% 1,08 ; 1,31) comparativement à 1,08 (IC95% 1,01 ; 1,16) entre 40 et 49 ans et 1,04 (IC95% 0,99 ; 1,09) entre 50 et 59 ans. , De plus, à partir de 46 ans l'association de la PAD au risque d'accident vasculaire cérébral était moindre que la PAS (54). L'hypertension systolique isolée, caractérisée par une PAS supérieure à 140 mm de Hg et une PAD inférieure à 90 mm Hg, est la forme la plus commune d'hypertension chez tous les groupes d'âge (55). Plus spécifiquement, dans un échantillon de plus de 21 000 adolescents américains âgés de 10 à 19 ans l'hypertension systolique isolée représentait 92% des cas d'hypertension observés (56). Le mécanisme expliquant l'élévation de la PAS pourrait différer en fonction de l'âge. À l'âge adulte, la perte d'élasticité des vaisseaux sanguins est le facteur physiologique principal expliquant l'élévation de la PAS. Même si certains adolescents présentant une hypertension systolique isolée montrent aussi une perte d'élasticité vasculaire, le facteur principal les distinguant des individus ayant une pression artérielle normale est une augmentation du volume d'éjection systolique et du débit cardiaque. Ces deux aspects sont des marqueurs importants du début du processus de persistance de l'élévation de la pression artérielle et suggèrent donc une évolution vers une hypertension permanente à l'âge adulte (55).

Comparativement à l'âge adulte, la prévalence de l'hypertension est moins élevée chez les enfants et les adolescents. Au Canada, environ 7% des jeunes âgés de 6 à 19 ans présentent une hypertension ou une pression artérielle normale élevée (PAS ou PAD \geq 90^e percentile ou 120/80 mm Hg) en 2015 (9). Les effets négatifs de l'élévation de la pression artérielle durant l'enfance sont reliés à la progression vers l'hypertension à l'âge adulte et au début de dommages au système cardiovasculaire. Présenter une pression artérielle élevée ou une hypertension à l'enfance ou à l'adolescence est un facteur de risque pour l'hypertension à l'âge adulte (11, 57-59). L'élévation de la pression artérielle à l'enfance est également associée à plusieurs marqueurs intermédiaires des troubles cardiovasculaires. Notamment, l'hypertrophie du ventricule gauche (12), l'athérosclérose (13) et l'épaisseur des artères carotides (14) sont reliées à une élévation de la pression artérielle durant l'enfance. Ces dommages au niveau du cœur et des artères sont fortement liés aux troubles cardiovasculaires à l'âge adulte (60) et suggèrent donc un lien entre une haute pression artérielle à l'enfance et le développement de maladies cardiovasculaires au cours de la vie.

3.1.4 Génétique de la pression artérielle

La pression artérielle comporte une composante génétique complexe incorporant l'action de plusieurs gènes. L'héritabilité de la variation de pression artérielle est estimée entre 30% et 50% (61). La génétique peut affecter la pression artérielle de façon monogénique ou polygénique. Les formes monogéniques correspondent à ce qu'un gène spécifique provoque une élévation de la pression artérielle marquée généralement par des niveaux de rénine très bas, une histoire familiale d'hypertension, un début tôt des caractéristiques cliniques (61). Quelques formes monogéniques d'hypertension ont été identifiées, incluant l'aldostéronisme rémissible par glucocorticoïde, le syndrome de Liddle ou bien le syndrome de Gordon (61). Ces pathologies n'expliquent qu'une faible portion des cas d'hypertension. Les formes polygéniques font référence à ce que plusieurs variants génétiques à petits effets contribuent collectivement à l'élévation de la pression artérielle. Comme l'objet du mémoire porte sur des variants génétiques communs, les connaissances concernant ce type de variants relatif à la pression artérielle sont développées davantage dans les prochaines lignes.

Les GWAS ont été utilisées au cours des 10 dernières années afin d'identifier ces variants génétiques communs ayant des effets individuels faibles sur la pression artérielle. La première GWAS effectuée afin d'identifier des variants associés avec la pression artérielle a utilisé un devis cas-témoins. Le Wellcome Trust Case Control Consortium n'était composé que de 3000 contrôles communs pour l'étude de sept maladies différentes dont l'hypertension avec 2000 cas d'ascendance européenne. L'hypertension était le seul phénotype dans le Wellcome Trust Case Control Consortium à ne montrer aucune association significative (62). La première GWAS portant sur la pression artérielle comme variable continue a été réalisée par la Framingham Heart Study (63). L'étude incluant 1400 participants n'a pas non plus identifié de variants associés avec la PAS ou la PAD. Ainsi, ces études ont permis de souligner l'importance d'avoir un échantillon de grande taille afin d'identifier les effets modestes des variants sur la pression artérielle.

L'utilisation de consortiums réunissant plusieurs cohortes est devenue la norme afin d'obtenir l'échantillon de plus grande taille nécessaire aux GWAS. Ainsi, les premières études à identifier des variants associés avec la pression artérielle étaient composées de consortiums. Par exemple, le consortium Cohorts for Heart and Aging Research in Genomic Epidemiology (64) et le consortium Global Blood Pressure Genetics (65) ont chacun identifié huit loci associés avec la PAS ou la PAD, dont trois communs entre les études. Depuis, multiples méta-analyses et études réunissant de grandes tailles d'échantillon ont été effectuées (66-81). La GWAS avec la plus grande taille d'échantillon portant sur la pression artérielle a été publiée par Evangelou et al. en 2018 (67). L'étude incorporait des participants du UK Biobank (n=458 577) et du International Consortium for Blood Pressure (n=299 024) dans une méta-analyse initiale de 757 601 individus d'origine européenne. L'analyse initiale a permis d'identifier 325 loci qui ont ensuite été répliqués dans une méta-analyse combinant les données du UK Biobank, International Consortium for Blood Pressure, US Million Veteran Program et Estonian Genome Centre, totalisant 1 006 863 individus. De plus, 210 nouveaux loci ont été identifiés et répliqués à l'interne dans le UK Biobank et le International Consortium for Blood Pressure, pour un total de 535 nouveaux loci identifiés dans l'étude (67). On y confirme aussi l'association avec des variants sur 274 loci préalablement identifiés et répliqués dans des GWAS effectuées auparavant et on y confirme l'association avec 92 loci déjà rapportés, mais non validés. En combinant les 535 nouveaux loci aux 274 loci préalablement validés et aux 92 loci confirmés, ceux-ci expliquent 5,7% de la variabilité de la PAS (67). La différence observée avec l'héritabilité de la pression artérielle rapportée pourrait

s'expliquer par l'existence de variants rares non identifiés, par l'existence de variants communs à effet modeste nécessitant de plus grandes tailles d'échantillon pour identification ou bien par la présence de facteurs environnementaux menant à la surestimation de l'héritabilité rapportée dans les études sur les familles (82).

Bien que la littérature concernant la génétique de la pression artérielle ait considérablement évolué au cours des dernières années, celle-ci concerne principalement les adultes. La majorité des études d'association génétique pour la pression artérielle portant sur des enfants ou des adolescents ont examiné si certains variants identifiés chez les adultes étaient également associés à la pression artérielle chez les enfants et les adolescents. Certaines études (83-85) montrent que des scores de risque génétiques construits à partir de variants identifiés chez les adultes sont également associés avec la pression artérielle chez les enfants et les adolescents, mais que ce ne sont pas tous ces variants qui sont individuellement associés à la pression artérielle chez les enfants et les adolescents (86). Depuis la publication de ces études, multiples variants supplémentaires ont été identifiés pour leur association avec la pression artérielle chez les adultes et n'ont pas encore été validés pour leur association chez les enfants et les adolescents (86).

Seulement trois GWAS (33, 87, 88) ont été effectués afin d'identifier des variants associés à la pression artérielle chez les enfants et les adolescents, ne permettant l'identification que de deux variants associés avec la PAS dans une des trois études. L'étude de Parmar et al. (33) a été effectuée chez des enfants et des adolescents d'ascendance européenne âgés de 4 à 20 ans du consortium Early Genetics and Lifecourse Epidemiology. Les variants rs1563894 et rs872256 étaient associés avec la PAS lors de la prépuberté (4 à 7 ans, n=10 090) et de la puberté (8 à 12 ans, n=8423) respectivement (33). La différence de PAS par allèle était de $-0,093$ écarts types (IC95% $-0,126$; $-0,060$) pour rs1563894 et de $0,096$ écarts types (IC95% $0,063$; $0,129$) pour rs872256. Les deux autres GWAS ont été effectués à l'aide de 815 participants âgés de 10 et 11 ans d'ascendance africaine (88) et à l'aide d'une population pédiatrique de 1617 patients atteints d'anémie falciforme (87). Après correction pour tests multiples, ces deux études n'observaient aucun variant associé avec la pression artérielle. On note que ces études chez les jeunes ont des tailles d'échantillon bien en deçà de ce que l'on peut retrouver dans les études effectuées chez des adultes. Cela pourrait expliquer en partie que les variants identifiés chez les adultes ne soient pas systématiquement répliqués chez les jeunes.

3.2 Adiposité

3.2.1 Définition

L'adiposité correspond au degré d'accumulation de réserves de gras corporel qui survient lorsqu'il y a un déséquilibre de l'équilibre énergétique, c'est-à-dire lorsque les dépenses énergétiques sont inférieures à la consommation d'énergie par l'alimentation (89). Les dépenses énergétiques comportent le métabolisme basal, la thermogénèse alimentaire et les dépenses en énergie liée aux activités physiques du corps. Le métabolisme basal est responsable de la majorité des dépenses énergétiques et inclut notamment la régulation de la température, de la circulation sanguine, de la respiration et de la prolifération cellulaire. Le métabolisme basal peut différer entre les individus, entre autres, la quantité de masse maigre, l'âge et le sexe peuvent influencer sur la quantité d'énergie dépensée au repos. Lorsque l'énergie consommée sous forme de nourriture dépasse l'énergie nécessaire au fonctionnement du corps, le surplus est stocké à environ 60-80% sous forme de triglycérides dans les cellules adipeuses (90). Les causes directes d'un surplus d'adiposité correspondent à un faible métabolisme basal, à une consommation d'énergie élevée ou à un faible niveau de dépense énergétique (91).

3.2.2 Méthodes de mesure de l'adiposité

La mesure la plus fréquemment utilisée pour représenter l'adiposité est l'IMC, une mesure anthropométrique principalement utilisée en raison de sa facilité d'utilisation. Seules la masse et la taille des individus sont nécessaires afin de calculer l'index selon la formule suivante : $IMC = \text{Masse (kg)} / \text{Taille}^2 (\text{m}^2)$. Cependant, cette mesure possède certaines lacunes incluant le fait de ne pas distinguer la masse maigre (tissu musculaire, tissu osseux, eau) de la masse grasse (tissu adipeux), ainsi que de ne pas informer sur la localisation corporelle de la masse grasse. Ainsi, pour une même taille et un même IMC, deux individus peuvent différer en masse maigre et en masse grasse et ainsi présenter un degré d'adiposité différent. L'IMC étant un indicateur de l'adiposité totale, il ne permet pas de représenter la distribution de la matière grasse dans le corps. Or, le tissu adipeux viscéral qui s'accumule à l'abdomen présente davantage de risques pour la santé comparativement à l'accumulation de gras sous-cutané (92). Malgré ces inconvénients, l'IMC reste utile afin d'identifier les individus à risque de complications de santé en raison d'un surplus de poids puisqu'il est un indicateur de maladies reliées à l'excès de poids (93). Selon la

classification de l'Organisation mondiale de la Santé, un poids corporel normal se situe entre 18,5 et 24,9 kg/m², alors qu'un IMC entre 25,0 kg/m² et 29,9 kg/m² indique un surpoids et qu'un IMC supérieur à 30,0 kg/m² fait référence à l'obésité (94).

En complémentarité avec l'IMC, la mesure du tour de taille est aussi utilisée afin de mesurer l'adiposité. Le tour de taille permet de représenter l'adiposité abdominale plus précisément que l'IMC. Le tour de taille est mesuré à l'aide d'un ruban à mesurer positionné sur le plan horizontal à une distance équivalente entre la dernière côte et la crête iliaque des hanches (93). D'autres endroits sont proposés où appuyer le ruban à mesurer, mais le point milieu entre la dernière côte et la crête iliaque est la mesure du tour de taille la plus précise pour mesurer l'adiposité viscérale et le changement d'adiposité dans le temps (95). L'excès de gras au niveau abdominal est associé avec un risque indépendant de l'IMC pour les désordres métaboliques et cardiovasculaires (94). Le tour de taille permet ainsi l'identification d'individus à risque ayant un IMC considéré comme normal, mais un tour de taille suggérant un excès de matière grasse abdominale (96). Un tour de taille supérieur à 94 cm chez les hommes et à 80 cm chez les femmes est le seuil utilisé par l'Organisation mondiale de la Santé et la Fédération internationale du diabète pour identifier les individus ayant un risque plus élevé de complications métaboliques (97). Un tour de taille supérieur à 102 cm pour les hommes et à 88 cm pour les femmes est également utilisé pour identifier les individus ayant un risque encore plus important de complications de santé (97).

La mesure la plus utilisée afin de représenter l'adiposité chez les enfants et les adolescents est l'IMC. Cependant, puisque la composition corporelle varie tout au long du développement et selon le sexe, une valeur d'IMC standardisée est recommandée à cette période de la vie. La mesure d'adiposité la plus utilisée chez les individus de moins de 20 ans est le score Z de l'IMC (IMCZ) développé par l'Organisation mondiale de la Santé à l'aide de grilles de références de développement standardisées pour l'âge et le sexe (98). Cependant, tout comme chez les adultes, l'IMCZ ne mesure pas directement l'adiposité et ne permet donc pas la distinction entre la masse grasse et la masse maigre (99). L'IMCZ permet l'identification d'individus à risque de problèmes de santé en raison d'un excès d'adiposité. Selon la classification de l'Organisation mondiale de la Santé pour les individus de 5 à 19 ans, une valeur d'IMCZ entre -2 et 2 est considérée comme normal, alors qu'une valeur de 2 à 3 correspond à un surpoids et qu'une valeur supérieure à 3 indique l'obésité (98, 100). Notons que la classification du Centers for Disease Control and

Prevention qui qualifie le degré d'adiposité selon le percentile d'IMC relatif à l'âge et au sexe est également fréquemment utilisée. Les enfants présentant un IMC allant du 85^e au 94^e percentile selon les chartes de développement du Centre sont considérés en surpoids, alors que ceux ayant un IMC supérieur au 94^e percentile sont classifiés obèses (101).

3.2.3 Effets néfastes de l'excès d'adiposité

L'adiposité peut avoir des répercussions importantes au niveau de la santé, tant sur la mortalité que sur la morbidité. En 2015, l'adiposité était responsable d'environ 4 millions de morts et de la perte de 120 millions d'années de vie en bonne santé dans le monde (102). Le lien entre l'adiposité et la mortalité est bien documenté. Une méta-analyse combinant 44 études montre qu'une augmentation de 5 kg/m² est associée à un risque relatif de 1,18 (IC95% 1,15 ; 1,21) pour la mortalité de toutes causes chez les non-fumeurs (103). Une deuxième méta-analyse combinant 189 études montre que les non-fumeurs en surpoids (IMC de 25,0 à 29,9 kg/m²) et obèses (IMC de 30,0 à 34,9 kg/m²) sont plus à risque de mortalité de toutes causes comparativement aux individus au poids normal (IMC de 18,5 à 24,9 kg/m²) avec un risque relatif de 1,11 (IC95% 1,10 ; 1,11) et 1,44 (IC95% 1,41 ; 1,47) respectivement (104). . Aussi, une étude de cohorte populationnelle de presque 2 millions d'individus a été réalisée à l'aide de données cliniques et de registres nationaux de mortalité britanniques afin d'évaluer le risque de mortalité relié à des causes spécifiques et toutes causes confondues (105). On y observait que les individus en surpoids, obèses, et ayant un poids insuffisant avaient une espérance de vie réduite à partir de 40 ans. L'augmentation de cinq unités d'IMC y était associée avec une augmentation du risque de mortalité de 21% (IC95% 1,20 ; 1,22) chez les individus avec un IMC supérieur à 25 kg/m² (105). Ces données supportent l'observation que l'association entre le niveau d'adiposité et la mortalité se présente généralement par une forme en « J ». Le risque de mortalité augmente chez les individus en surpoids (IMC de 25,0 à 29,9 kg/m²) et obèses (IMC \geq 30,0 kg/m²), mais aussi chez les individus ayant un poids insuffisant (IMC \leq 18,5) (103-105).

La majorité de la mortalité en lien avec l'adiposité est reliée au fait que le niveau d'adiposité est un facteur de risque pour un large éventail de maladies. Par exemple, 69% et 14% de la mortalité liée à l'IMC est dus aux maladies cardiovasculaires et au diabète de type II respectivement (102). On retrouve l'association en forme de « J » entre l'adiposité mesurée par

l'IMC et la mortalité liée spécifiquement à certains types de cancers, ainsi qu'à plusieurs maladies cardiovasculaires, respiratoires, musculosquelettiques, digestives et métaboliques (105). Le niveau d'adiposité est notamment associé avec le diabète de type II (106-109), l'hypertension (109), la dyslipidémie (109), la cardiopathie ischémique (110, 111), plusieurs types de cancers (112), l'insuffisance cardiaque (109, 110, 113), la maladie coronarienne (106, 114, 115), les accidents vasculaires cérébraux (110) et la fibrillation auriculaire (116-118).

Bien que moins prévalente à l'enfance et à l'adolescence qu'elle ne l'est à l'âge adulte, l'excès d'adiposité y est aussi problématique. La prévalence de l'obésité a notamment plus que doublé à cette période de la vie au cours des 40 dernières années (119). Ainsi, environ 1 jeune Canadien sur 7 est maintenant obèse et près de 1 sur 3 est en situation de surpoids ou d'obésité (119). De nombreuses études montrent que l'excès d'adiposité à l'enfance et à l'adolescence est un facteur de risque pour l'excès d'adiposité à l'âge adulte (120-123). Notamment, une méta-analyse récente de 15 cohortes réunissant 200 000 participants montre que les enfants et adolescents obèses ont environ 5 fois le risque d'être obèses à l'âge adulte comparativement à ceux non obèses (risque relatif 5,21; IC95% 4,50 ; 6,02). On y observe que 70% des adolescents obèses le seront également à l'âge de 30 ans (121). Le fait d'avoir un excès de poids à l'enfance et à l'adolescence augmente également le risque de mortalité et de morbidité à l'âge adulte. Notamment, certaines études montrent qu'une augmentation de l'adiposité mesurée par l'IMC est associée avec la mortalité prématurée (124), la mortalité liée aux maladies cardiovasculaires (125), la maladie rénale en phase terminale (126), le diabète de type II (127) et plusieurs types de cancers à l'âge adulte (128).

Les effets d'un excès d'adiposité à l'enfance et à l'adolescence peuvent également s'observer directement à cette période. L'excès de poids y est lié à certains facteurs de risque de maladies cardiovasculaires et d'anomalies cardiométaboliques. Les enfants et adolescents en surplus de poids présentent des niveaux plus élevés de pression artérielle, de cholestérol, de triglycérides, de protéines C-réactives et d'insuline (22, 129) et ont une incidence plus élevée d'hypertension, de dyslipidémie, d'hyperinsulinisme et de résistance à l'insuline comparativement à ceux n'ayant pas d'excès de poids (128). Ainsi, les effets néfastes d'un surplus d'adiposité peuvent prendre racine dès l'enfance et l'adolescence en influençant l'adiposité et les comorbidités à l'âge adulte et en augmentant dès le jeune âge les niveaux de facteurs de risque cardiovasculaires.

3.2.4 Génétique de l'adiposité

La différence de poids corporel entre les individus comporte une composante génétique qui influence l'accumulation de matières grasses. Les récentes estimations de l'héritabilité de l'IMC sont d'environ 40%, dont plus de la moitié est reliée aux variants génétiques communs (130). L'apport de la génétique au développement de l'obésité peut se présenter sous une forme monogénique ou polygénique. Les formes monogéniques d'obésité correspondent à l'altération d'un seul gène menant à une perte de fonction biologique qui résulte en l'accumulation de gras. Ces mutations se retrouvent principalement dans les voies de régulation de l'appétit et du métabolisme énergétique. Par exemple le système de régulation leptine-mélanocortine (35). La forme monogénique affecte environ 5% de la population et est généralement caractérisée par un gain de poids extrême et un début tôt dans la vie (35).

Malgré l'apport non négligeable de l'aspect monogénique de l'obésité, la majorité des cas d'obésité sont en lien avec une combinaison de plusieurs facteurs génétiques et l'environnement de l'individu. Les formes polygéniques représentent l'accumulation de mutations communes affectant individuellement légèrement la prise de poids (131). L'introduction des technologies nécessaires à la conduite de GWAS a permis de solides avancées dans l'identification des variants communs ayant un effet modeste sur l'adiposité. L'IMC étant la mesure de l'obésité la plus largement utilisée due à la facilité de son utilisation, la majorité des GWAS ayant comme objectif d'identifier des variants associés avec l'adiposité l'utilise comme trait d'intérêt. Le premier variant à avoir été identifié pour son association avec l'IMC (rs9939609) était situé sur le gène *FTO* (132). Suivant la vague dans le domaine des GWAS d'augmenter les tailles d'échantillons des études afin d'avoir la puissance nécessaire d'identifier davantage de variants, certains consortiums ont également été mis à contribution dans les GWAS portant sur l'IMC. Notamment, deux études utilisant les données provenant du consortium Genetic Investigation of ANthropometric Traits et du UK Biobank ont assemblé des échantillons de plus de 300 000 individus permettant d'identifier 97 et 941 variants associés à l'IMC (133, 134). Cependant, les variants identifiés par l'étude par Yengo et al. réunissant la plus grande taille d'échantillon (n=681 275) n'expliquent que 6% de la variance de l'IMC (134). Bien que l'héritabilité de l'IMC soit estimée à environ 40% (130), la partie attribuable aux variants communs serait plutôt de l'ordre de 25% et la présence de facteurs

environnementaux dans les études d'héritabilité pourrait mener à surestimer la variation attribuable aux facteurs génétiques (82, 135).

Les facteurs génétiques peuvent influencer l'adiposité dès l'enfance. En effet, la contribution des facteurs génétique à la variation des niveaux d'adiposité est plus importante à l'enfance et à l'adolescence qu'à l'âge adulte (136). La littérature concernant les facteurs de risque génétiques reliés à l'adiposité est pourtant moins fréquente chez les enfants et les adolescents comparativement aux adultes. La plupart des variants associés à des mesures de l'adiposité dans des études portant sur des enfants ou des adolescents ont été initialement identifiés chez les adultes. Plusieurs études portant spécifiquement sur des variants déjà connus pour leur association chez les adultes ont validé leur association chez les enfants et les adolescents (35). Seulement quelques GWAS de traits reliés à l'adiposité ont été effectuées chez des enfants ou des adolescents (34, 137-143). Ces GWAS ont permis de répliquer plusieurs associations observées chez les adultes tout en identifiant quelques nouveaux variants.

Aussi, les résultats de certaines études longitudinales menées sur des enfants et des adolescents suggèrent que l'effet de certains variants sur l'IMC pourrait varier selon l'âge (142, 144-146). Par exemple, une étude montre une association positive entre un variant sur *FTO* et l'IMC après cinq ans et demi. Pour un allèle mineur le pourcentage de changement d'IMC était de 0,7% (IC95% 0,3 ; 1,1), 1,0% (IC95% 0,6 ; 1,3) et 1,3% (IC95% 0,6 ; 2,0) pour les 5.5 à 7 ans, 7 à 9 ans et 9 à 11 ans respectivement (146). Au contraire, l'étude montre une association inverse avant 2 ans et demi avec un pourcentage de changement d'IMC de -0,3% (IC95% -0,5 ; 0,0) entre 1,5 et 2,5 ans (146). Une autre étude (144) montre une interaction entre l'âge et le génotype du variant rs9939609 sur *FTO* dans l'association avec l'IMC, où l'effet du variant sur l'IMC augmente avec l'âge entre 8 et 17 ans. Les enfants homozygotes pour l'allèle d'effet avaient un IMC plus élevé de 0,7 kg/m² à huit ans comparativement aux hétérozygotes et homozygotes pour l'allèle protecteur, alors que cet effet était de 1,6 kg/m² à 17 ans (144). Dans la même optique, les résultats de Hardy et al. (145) suggèrent que l'effet de variants génétiques sur *FTO* et *MC4R* augmentent durant l'adolescence jusqu'à 20 ans, pour ensuite diminuer au cours de l'âge adulte. Le variant rs9939609 (*FTO*) était associé à une augmentation d'IMC standardisé de 0,007 par allèle par année (IC95% 0,003 ; 0,010) de 2 à 20 ans, puis à une diminution de 0,003 par allèle par année (IC95% 0,001 ; 0,005) à l'âge adulte. Le variant rs17782313 était associé à une

augmentation de poids corporel standardisé de 0,005 par allèle par année (IC95% 0,001; 0,008) de 2 à 20 ans, puis à une diminution de 0,002 (IC95% 0,000 ; 0,004) à l'âge adulte (145). En somme, ces observations suggèrent que les mêmes variants génétiques de l'adiposité pourraient être présent à l'âge adulte et chez les enfants et adolescents, que l'effet de ces variants peut changer au cours de la vie, mais aussi que l'architecture génétique de l'adiposité chez les enfants et les adolescents pourrait être davantage explorée.

3.3 Relation entre l'adiposité et la pression artérielle

Les sections qui suivent couvrent la littérature concernant l'association entre l'adiposité et la pression artérielle et le rôle de la génétique dans cette association. Plus spécifiquement, on examine d'abord l'impact d'une masse adipeuse élevée comme facteur de risque de l'augmentation de la pression artérielle chez les adultes et chez les jeunes. Ensuite, on examine les écrits concernant l'impact des facteurs génétiques sur l'évolution conjointe des niveaux d'adiposité et de pression artérielle, ainsi que les études ayant identifié des variants génétiques spécifiques associés aux deux traits. Puisque l'adiposité est un facteur de risque de la pression artérielle, on examine également la littérature concernant le potentiel de l'adiposité à modifier l'effet de variants génétiques sur la pression artérielle. Une description élargie de l'article de Melka et al. (30) est finalement présentée pour distinguer son positionnement dans la littérature.

3.3.1 Association entre l'adiposité et la pression artérielle

Les individus présentant un niveau élevé d'adiposité sont plus à risque d'avoir une pression artérielle élevée et donc une hypertension. Des données longitudinales provenant du Framingham Hearth Study montrent que comparativement aux individus au poids normal, les individus en surpoids ont un risque relatif pour le développement de l'hypertension de 1,48 (IC95% 1,24 ; 1,75) pour les hommes et de 1,70 (IC95% 1,48 ; 1,94) pour les femmes et les individus obèses ont un risque relatif de 2,23 (IC95% 1,75 ; 2,79) pour les hommes et de 2,63 (IC95% 2,20 ; 3,15) pour les femmes (147, 148). Également, une méta-analyse combinant les résultats de 57 études prospectives indépendantes montre que le risque d'hypertension augmente de façon continue en fonction de l'élévation de multiples mesures anthropométriques, incluant l'IMC, la circonférence

de taille, le ratio taille-hanche et le poids corporel (18). Pour chaque augmentation de 5 unités d'IMC, de 10 cm de circonférence de taille et de 0.1 unité de ratio taille-hanche, le risque d'hypertension augmente respectivement de 49% (IC95% 1,41 ; 1,58), 27% (IC95% 1,15 ; 1,39) et 37% (IC95% 1,24 ; 1,51) (18).

Les données s'accordent également sur le fait que l'obésité à l'enfance et à l'adolescence est reliée au développement de l'hypertension à cette même période, mais aussi à l'âge adulte (51). Notamment, une méta-analyse de 63 études portant sur des enfants âgés de 5 à 15 ans montre une augmentation de PAS de 4,54 mm Hg (IC99% 2,44 ; 6,64) chez les enfants en surpoids et de 7,49 mm Hg (IC99% 3,36 ; 11,62) chez les enfants obèses, comparativement aux enfants en poids normal (22). Une étude transversale portant sur plus de 700 000 adolescents âgés de 16 à 20 ans montre que chaque augmentation d'une unité d'IMC est associée à un risque accru d'avoir une PAS supérieure à 130 mm Hg chez les hommes (OR = 1,108 ; IC95% 1,107 ; 1,110) et les femmes (OR = 1,114 ; IC95% 1,139 ; 1,146) (23). Un IMC plus élevé lors de l'enfance et l'adolescence est également relié à une pression artérielle plus élevée à l'âge adulte. Une méta-analyse combinant 14 études montre une association positive entre l'IMC à l'enfance et la PAS et la PAD à l'âge adulte avec un coefficient de corrélation de transformé de Fisher de 0,11 (IC95% 0,07 ; 0,14) pour la PAS et la PAD (26). Le profil de l'IMC au cours de l'enfance et de l'adolescence peut également informer sur le risque subséquent d'avoir une pression artérielle élevée. En effet, des trajectoires d'IMC correspondant à une augmentation de l'adiposité durant l'enfance sont associées à une pression artérielle plus élevée à l'âge adulte (24, 25, 27).

3.3.2 Relation génétique entre l'adiposité et la pression artérielle

L'hypothèse de variant commun/multiples maladies référant aux maladies complexes communes dans la population pourrait jouer un rôle dans la relation entre l'adiposité et la pression artérielle. Cette hypothèse, initialement développée par Becker (28), indique que certains variants génétiques communs puissent jouer un rôle dans l'étiologie de traits reliés dans des contextes génétiques et environnementaux différents. Ainsi la présence de certains variants génétiques communs pourrait expliquer en partie la corrélation importante observée entre les niveaux d'adiposité et de pression artérielle.

Les premières études examinant la relation génétique entre l'adiposité et la pression artérielle étaient effectuées au sein de familles où l'on examinait la part des facteurs génétiques dans la corrélation entre les deux traits. Plusieurs études ont examiné le rôle de la génétique dans la corrélation et la covariance de diverses mesures d'adiposité et de pression artérielle chez les jumeaux (149), au sein de familles (150-155) et au sein de fratries (156). La majorité de ces études (149, 151, 152, 154-156) parmi ces dernières ont obtenu des résultats suggérant que l'impact conjoint de facteurs génétiques sur l'adiposité et la pression artérielle représente une partie significative de l'évolution conjointe des deux traits. Carmelli et al. (149), Livshits et al. (152), Cui Jisheng et al. (151) et Schork et al. (154) ont observé à différents degrés une composante génétique significative affectant la covariance entre la pression artérielle et plusieurs mesures d'adiposité incluant l'IMC. Pausova et al. (156) ont observé que les fratries présentant une hypertension avaient des niveaux plus élevés d'adiposité et une héritabilité des niveaux d'adiposité plus élevée que les fratries sans hypertension. À l'aide d'analyses génétiques bivariées, l'étude récente de Man et al. (155) a observé une corrélation génétique de 0,42 entre l'IMC et la PAS supérieure à la corrélation environnementale (0.31). Les résultats de deux études effectuées dans la Québec Family Study (153) et la Heritage Family Study (150) utilisant des modèles de corrélation familiale multivariés suggéraient que des facteurs environnementaux plutôt que génétiques affectaient le lien entre l'adiposité et la pression artérielle. Finalement, Cadby et al. (157) a examiné la corrélation génétique entre des mesures anthropométriques et des facteurs de risque cardiométaboliques incluant la pression artérielle au sein d'un échantillon populationnel combinant des individus provenant de mêmes familles et d'individus non reliés. La PAS et la PAD montraient une corrélation génétique significative avec plusieurs mesures anthropométriques incluant l'IMC (157). Bien que la majorité des études mentionnées plus haut suggèrent que des facteurs génétiques communs ont un impact sur l'évolution conjointe de l'adiposité et de la pression artérielle, ce type d'étude ne permet pas d'identifier les facteurs génétiques spécifiques à l'origine de cet impact.

Certains variant génétiques ont été identifiés pour leur association avec l'adiposité et la pression artérielle dans des jeux de données différents. Par exemple, parmi 535 nouveaux variants identifiés dans une récente GWAS de pression artérielle, une dizaine étaient en déséquilibre de liaison (LD) ($r^2 > 0.80$) avec des variants déjà identifiés pour une association avec des traits relatifs à l'adiposité (67). Bien que moins de données soient disponibles chez les jeunes, une étude a

observé que des variants associés à l'IMC pouvaient également influencer le niveau de pression artérielle chez les enfants. Howe et al. (83) ont observé qu'un score de risque génétique composé de 32 variants identifiés dans un GWAS de l'IMC chez des adultes était associé avec la PAS à l'âge de six ans (0,107 mm Hg par allèle de risque ; $p = 0,037$), et ce de façon comparable à un score de risque composé de 29 variants de la PAS (0,097 mm Hg par allèle de risque ; $p = 0,013$).

D'autres études ont examiné s'il était possible d'identifier des variants génétiques spécifiques affectant conjointement la pression artérielle et l'adiposité dans un même jeu de données. Premièrement, l'étude de Melka et al. (30) fait partie de ces études ayant identifié par GWAS cinq variants associés avec l'IMC et la masse adipeuse totale qui étaient ensuite testés pour leur association avec la PAS. Trois variants sur les gènes *FTO*, *MRPS22* et *PAX5* étaient également associés avec la PAS. Cette étude sera détaillée dans la prochaine section puisqu'elle est l'objet de la réplique présentée dans le mémoire.

D'autres études ont permis d'identifier des variants sur certains gènes incluant *FTO*, *SLC39A8* et *ACE* affectant à la fois l'adiposité et la pression artérielle dans un même jeu de données. Ces associations sont toutefois peu nombreuses et inconsistantes pour un même variant étudié dans des études différentes. Olza et al. (158) ont examiné 52 variants sur *FTO* et ont identifié un variant (rs8061518) significativement associé à l'IMC (-1,27 kg/m² par allèle mineur ; IC95% -2,05 ; -0,50) et à la PAD (-2,23 mm Hg ; IC95% -3,80 ; -0,65) chez 534 enfants (6-15 ans) d'ascendance européenne. À l'aide d'une méta-analyse totalisant plus de 37 000 individus d'ascendance européenne, Vliet-Ostapchouk et al. (159) ont examiné 32 loci préalablement associés avec l'IMC et ont identifié un variant (rs13107325) sur le gène *SLC39A8* associé avec l'IMC (0,057 écarts types ; $p = 2,3 \times 10^{-4}$) et la PAD (-0,058 écarts types ; $p = 2,5 \times 10^{-4}$), mais dans une direction opposée.

Eisenmann et al. (160) et Lemes et al. (161) ont étudié spécifiquement un polymorphisme sur le gène *ACE* chez des enfants. Eisenmann et al. ont observé que le polymorphisme était associé à l'IMC, mais pas à la pression artérielle (160), alors que Lemes et al. ont observé une association avec l'IMC et l'hypertension, mais seulement chez les garçons (161).

Pausova et al. (162), Luczynski et al. (163) et Garcia-Solis et al. (164) ont identifié une association entre le variant rs9939609 (*FTO*) et l'adiposité conjointement à la PAS et la PAD chez des jeunes. Les associations observées dans ces dernières études étaient toutefois inconsistantes

concernant le rôle de l'adiposité comme variable intermédiaire dans l'association entre le variant et la pression artérielle. Après ajustement pour l'IMC, le variant n'était plus associé avec la PAS (p non-ajusté = $2,0 \times 10^{-5}$; p ajusté = 0,14) et la PAD (p non-ajusté = 0,006 ; p ajusté = 0,61) dans Luczynski et al. (163), comparativement à Pausova et al. (162) où le variant demeurait associé à la PAS après ajustement pour l'IMC (p non-ajusté = 0,002 ; p ajusté = 0,01), et à Garcia-Solis et al. (164) où après ajustement pour l'IMC le variant demeurait associé à la PAS (p non-ajusté < 0,001 ; p ajusté = 0,0015) et la PAD (p non-ajusté < 0,01 ; p ajusté = 0,034) (162, 164). L'utilisation de population d'étude provenant de régions différentes (Canada, Pologne et Mexique), ainsi que l'utilisation de différentes variables pour représenter l'adiposité (IMC et IMCZ) pourrait expliquer ces résultats divergents.

Considérant qu'un niveau élevé d'adiposité est un facteur de risque pour l'élévation de la pression artérielle, certaines études ont traité de la relation génétique entre la pression artérielle et l'adiposité en examinant si le niveau d'adiposité pouvait modifier l'association entre des variants et la pression artérielle. En général, ces études sont peu nombreuses, alors que deux études (165, 166) montrent certains variants qui interagissent avec le niveau d'adiposité pour affecter la PAS, une autre (167) n'a pas été en mesure d'obtenir de tels résultats. Plus spécifiquement, Xi et al. (166) ont directement examiné l'effet de l'obésité sur l'association entre 6 variants et la pression artérielle chez des enfants obèses et au poids normal. L'association entre 3 variants et la PAS y était observable dans le groupe obèse ($rs17249754$: $p = 0,02$; $rs1378942$: $p = 0,003$; $rs1004467$: $p = 0,04$), mais pas chez les participants en surpoids ou au poids normal ($p > 0,05$ pour les six variants) suggérant que le niveau d'adiposité pouvait influencer la relation entre certains variants et la pression artérielle. L'étude par Kim et al. (165) a identifié par GWAS le variant $rs13390641$ près du gène *TMEM182* significativement associé avec la PAS en considérant une interaction entre le variant et l'IMC ($p_{\text{variant}} = 3,83 \times 10^{-8}$; $p_{\text{interaction}} = 5,28 \times 10^{-8}$). Lim et al. (167) ont construit un score de risque génétique incluant 42 variants validés pour leur association avec la pression artérielle et ont examiné l'interaction entre le score de risque génétique et des facteurs environnementaux incluant l'IMC pour l'association avec la pression artérielle. L'IMC ne modifiait pas l'association entre le score de risque ou les variants considérés individuellement et la pression artérielle. En bref, ces études suggèrent qu'il est possible que l'effet de certains variants sur la pression artérielle soit modifié par le niveau d'adiposité, mais que de tels phénomènes ne sont pas facilement identifiables.

En résumé, la relation génétique entre l'adiposité et la pression artérielle représente un bon exemple de l'hypothèse de variant commun/multiples maladies qui stipule que certains allèles de susceptibilité ne sont pas spécifiques à une maladie ou un trait, mais peuvent affecter conjointement plus d'un trait relié. Des études au sein de familles et de jumeaux ont montré que des facteurs génétiques sous-jacents sont du moins en partie responsables de la corrélation et de la covariation des niveaux d'adiposité et de pression artérielle. Ce type d'étude ne permet toutefois pas l'identification spécifique de variants génétiques affectant ces deux traits. Certains variants ont été identifiés dans ces études pour leur association simultanée avec la pression artérielle et l'adiposité après avoir examiné spécifiquement des variants déjà connus pour leur association avec l'un ou l'autre des deux traits. Toutefois, ceux-ci sont peu nombreux et les études ayant identifié ces variants obtiennent des résultats inconsistants qui n'ont pas été répliqués. Également, quelques études suggèrent que l'adiposité puisse modifier l'association entre certains variants génétiques et la pression artérielle. Ainsi, la pression artérielle et l'adiposité ont une architecture génétique complexe qui comporte à la fois des variants génétiques affectant individuellement chacun des traits, ainsi que des variants génétiques communs aux deux traits, soit de façon indépendante ou par l'intermédiaire de l'effet de l'adiposité sur la pression artérielle.

3.3.3 Article par Melka et al.

L'étude de Melka et al. (30) a été effectuée sous l'optique que l'obésité est un facteur de risque important pour le développement de l'hypertension et que ces deux conditions sont présentement en émergence à l'adolescence. Considérant que l'obésité et l'hypertension sont deux phénotypes affectés par une composante génétique importante et que des études familiales suggèrent que des facteurs génétiques affectent ces deux phénotypes conjointement (155), ceux-ci ont émis l'hypothèse que certains variants associés avec des mesures de l'adiposité pouvaient l'être également avec la pression artérielle. Afin de tester cette hypothèse, ceux-ci ont effectué une GWAS au sein d'un échantillon de 598 adolescents âgés de 12 à 18 ans du Saguenay Youth Study (SYS). Afin de simplifier la lecture, l'étude de Melka et al. (30) sera dorénavant désignée par la SYS et la présente étude par NDIT. Les variants sélectionnés pour leur association avec l'IMC et la masse adipeuse totale ont ensuite été testés pour leur association avec la PAS.

La GWAS pour l'IMC et la masse grasse totale suivait un modèle de régression linéaire multivarié sous un codage additif des variants en incluant l'âge, le sexe et la taille (masse adipeuse totale uniquement) comme covariables. Les variants rs16933812 (*PAX5*) et rs7638110 (*MRPS22*) ont passé le seuil de signification pangénomique de 5×10^{-8} pour leur association avec l'IMC ou la masse adipeuse totale. De plus, 33 loci identifiés préalablement par GWAS pour l'IMC ont été testés pour leur association avec l'IMC et la masse adipeuse totale. Huit variants étaient associés avec l'IMC et la masse adipeuse totale ($p < 0,05$) et avaient la même direction d'effet que dans l'étude préalable. Seuls les variants rs17773430 (*MC4R*), rs7120548 (*MTCH2*) et rs9930333 (*FTO*) ayant une valeur p inférieure à 5×10^{-4} ont été ajoutés aux analyses sur la PAS. Ces trois variants étaient en haut LD ($D' > 0,9$) dans la SYS avec les variants identifiés dans des GWAS préalables (rs9939609 (*FTO*) (168), rs3817334 (*MTCH2*) (169) et rs17782313 (*MC4R*) (170)).

L'association des cinq variants avec la PAS était testé avec un modèle linéaire mixte sous un codage génotypique des variants et en ajustant pour des covariables (l'âge, le sexe et la taille) et liens familiaux entre les participants. Les mesures de PAS étaient obtenues lors d'un protocole de 52 minutes comportant des mesures à différentes postures (assis, debout et couché) et à différents niveaux de stress (effort mental, récupération d'effort). Les mesures de PAS obtenues de ces cinq sections étaient incluses dans les modèles linéaires mixtes en les considérant comme des mesures répétées pour chaque participant. Seuls les variants sur *PAX5*, *MRPS22* et *FTO* étaient significativement associés avec la PAS. Ainsi, la SYS a identifié trois variants associés à la fois avec des mesures de l'adiposité et avec la PAS au sein du même échantillon. Ceux-ci concluent qu'il est possible que seulement certains types génétiquement définis d'obésité puissent être responsables de certains cas d'hypertension reliés à la masse adipeuse. Les effets des variants sur l'IMC et la pression artérielle étaient rapportés selon les différences d'IMC et de PAS entre les individus homozygotes pour l'allèle mineur, hétérozygotes ou homozygotes pour l'allèle majeur. Ainsi, le Tableau 1 présente une adaptation des résultats de la SYS pour les variants investigués et leur association avec l'IMC et la PAS. Notons que les intervalles de confiance n'étaient pas rapportés pour l'association avec l'IMC des variants sur les gènes *FTO*, *MTCH2* et *MC4R*.

Tableau 1. Variants associés à l'adiposité et leur association avec la PAS. Melka et al. (2012)[†]

Variant	Gène	MAF	Allèle mineur	IMC		PAS	
				Différence (Kg/m ²) ^a	Valeur pb	Différence (mm Hg) ^a	Valeur p
rs16933812	<i>PAX5</i>	0,22	G	3,3 [1,9 ; 4,6]	5,2×10⁻⁶	6,68 [2,80 ; 10,57]	0,0007
rs7638110	<i>MRPS22</i>	0,06	T	2,8 [1,9 ; 3,6]	4,6×10⁻⁸	3,30 [0,49 ; 6,12]	0,02
rs17773430	<i>MC4R</i>	0,27	C	2,4 [NA]	5,8×10⁻⁶	-0,05 [-3,44 ; 3,34]	0,98
rs7120548	<i>MTCH2</i>	0,29	C	-1,6 [NA]	1,9×10⁻⁴	-0,51 [-3,86 ; 2,84]	0,77
rs9930333	<i>FTO</i>	0,38	G	1,8 [NA]	1,9×10⁻⁴	3,63 [0,73 ; 6,53]	0,01

Présentation adaptée des résultats de l'étude de Melka et al. (30).

[†] Différence d'IMC et de PAS entre homozygotes pour l'allèle mineur comparativement aux homozygotes pour l'allèle majeur. NA = intervalle de confiance non-rapporté. Les valeurs p en **gras** signifient une association considérée statistiquement significative. La valeur p rapportée pour l'IMC correspond au GWAS sous un codage additif des allèles.

3.4 Variants génétiques à répliquer

La SYS a identifié cinq variants associés avec l'IMC et la masse grasse totale. Trois de ces variants étaient déjà connus pour leur association avec l'adiposité, alors que les deux autres ont été associés pour une première fois avec ces mesures de l'adiposité. Trois de ces variants étaient également associés avec la PAS dans le même échantillon. Les lignes qui suivent présentent un aperçu du rôle biologique des gènes associés à ces variants et de leur relation possible avec l'adiposité et la pression artérielle. Une description plus détaillée de ces aspects est également disponible à [l'Annexe B](#).

3.4.1 rs16933812 (PAX5)

Le variant rs16933812 se trouve dans un intron du gène *PAX5* situé sur le chromosome 9p13 (171). Ce gène code la protéine *paired box protein Pax-5*, un facteur de transcription de la famille Paired Box qui joue un rôle dans le processus de différenciation des cellules lymphatiques B (172). Les *common lymphoid progenitors* sont les précurseurs des cellules lymphatiques B, T et *natural killer*. *PAX5* fait partie des facteurs nécessaires à la différenciation de ces précurseurs en cellules lymphatiques B matures tout en inhibant leur passage vers d'autres types de cellules (173, 174). Étant exprimé au niveau du système nerveux central et des testicules, *PAX5* pourrait également avoir un rôle dans le développement neuronal et la spermatogenèse (172).

PAX5 est principalement connu pour son rôle dans le développement de divers cancers hématologiques. Des niveaux d'expression anormaux de *PAX5* ont été observés dans plusieurs cas de ce type de cancer, incluant leucémies et lymphomes (174). En ce qui concerne l'adiposité et la pression artérielle, une seule GWAS récente a identifié un variant (rs3013733) sur *PAX5* associé à l'IMC dans un échantillon de plus de 450 000 adultes du UK Biobank (175). On note toutefois que cette étude ne comportait pas d'analyse de réplique interne et que ce variant présente un faible LD ($r^2 = 0,306$) avec le variant rs16933812 identifié par la SYS dans la population comparable du 1000 Genomes Project Phase 3 (176) (résidents de l'Utah provenant de l'Europe de l'Ouest et du Nord (CEU)). Aucune GWAS n'a pour l'instant identifié une association entre un variant sur *PAX5* et des phénotypes liés à la pression artérielle.

Son rôle principal étant au niveau de la différenciation des cellules lymphatiques B il est difficile d'établir un lien précis entre *PAX5* et l'adiposité ou la pression artérielle. L'inflammation en lien avec le gain en masse adipeuse est reconnue pour avoir un rôle à jouer dans le développement de l'hypertension (177). Toutefois, ce rôle est principalement médié par les lymphocytes de type T qui ont leur effet pro-inflammatoire potentialisé par un environnement riche en noradrénaline en lien avec une élévation de l'afflux sympathique (177). On note que *PAX5* est exprimé dans une région du système nerveux sympathique ayant un rôle dans l'influx sympathique vers le tissu adipeux (178). Le gène pourrait donc influencer la lipolyse dans les cellules adipeuses et contribuer à l'accumulation de matière grasse.

3.4.2 rs7638110 (MRPS22)

Le variant rs7638110 se trouve sur un intron du gène *MRPS22* qui se trouve à la position 3q23 (171). Ce gène code pour la *protéine 28S ribosomal protein S22 (mitochondrial)* (172). Cette protéine fait partie de la sous-unité 28S des ribosomes et aide à la synthèse des protéines mitochondriales qui participent aux processus de phosphorylation oxydative dans la mitochondrie nécessaire à la production d'ATP (179).

Les déficiences sur *MRPS22* sont principalement associées avec des troubles mitochondriaux (179). Notamment, des mutations sur *MRPS22* ont déjà été associées avec une hypotonie musculaire à la naissance (180), une acidose lactique à la naissance (180, 181), une hyperammoniémie à la naissance (180), des phénotypes semblables au syndrome Cornelia de

Lange (182), des anomalies au cerveau et une cardiopathie hypertrophique (181, 182). Depuis la publication de la SYS aucune autre étude, que ce soit chez les adultes ou les jeunes, n'a identifié d'association entre le variant rs7638110 et l'adiposité ou la pression artérielle (183). Cette affirmation reste vraie en considérant l'ensemble du gène *MRPS22*.

La protéine mitochondriale ribosomique S22 participant activement à la phosphorylation oxydative (182), il est envisageable que *MRPS22* puisse être lié au métabolisme énergétique. Une altération de *MRPS22* pourrait donc altérer le métabolisme énergétique et ainsi affecter l'accumulation de graisse. Puisqu'aucun lien avec la pression artérielle ne soit apparent, une relation potentielle avec la pression artérielle pourrait passer par l'intermédiaire d'un effet sur le niveau d'adiposité.

3.4.3 rs9930333 (FTO)

Le variant rs9930333 est situé sur un intron du gène *FTO* positionné sur le chromosome 16q12.2 (171). Ce gène code pour la protéine *Alpha-Ketoglutarate Dependent Dioxygenase* qui est une enzyme du noyau cellulaire faisant partie de la famille d'enzymes *Alpha-Ketoglutarate Dependent Dioxygenase AlkB-like* (172). Ce type d'enzyme sert à réparer l'ARN et l'ADN alkylé par déméthylation oxydative. *FTO* est exprimé dans plus de 25 tissus chez l'humain dont majoritairement dans le cerveau (172). Ce gène est principalement connu pour arborer la région génétique la plus fortement associée à l'IMC (184).

FTO a été le premier gène identifié associé à l'adiposité par GWAS. Frayling et al. (132) ont effectué une GWAS pour le diabète de type II dans une population de 1924 cas et 2938 témoins d'origine européenne. Ils ont observé que l'association du variant rs9939609 avec le diabète passait par son association avec l'IMC. Ce variant, ainsi que plusieurs autres sur *FTO*, ont par la suite été répliqués à multiples reprises chez les adultes (35) et les enfants et adolescents (137). Plus spécifiquement, le variant investigué rs9930333 a déjà été associé à l'IMC par GWAS (185) et était avait un LD élevé ($D' > 0,90$) dans la SYS avec le variant rs9939609 qui fait partie d'un groupement de variants sur l'intron 1 de *FTO* fréquemment associé au poids corporel (186).

Une incertitude réside toujours en ce qui concerne le mécanisme biologique par lequel *FTO* puisse influencer le niveau d'adiposité. Une étude (187) suggère que l'influence de variants

génétiqes sur *FTO* sur l'adiposité passe par un autre gène. Ceux-ci ont montré que les séquences non codantes associées avec l'adiposité sur *FTO* sont connectées et interagissent avec les promoteurs de gène *IRX3*. En effet, les variants sur *FTO* sont associés avec l'expression d'*IRX3* et non pas *FTO* dans le cerveau humain. De plus, l'étude montrait que des souris privées du gène *IRX3* présentaient un poids corporel réduit de 25 à 30% (187). Cela suggère que l'effet des variants sur *FTO* sur l'accumulation de gras passe donc vraisemblablement par son effet sur l'expression du gène *IRX3*.

Alors que le lien entre *FTO* et l'adiposité est donc bien établi, un tel lien est toutefois moins bien défini avec la pression artérielle. Seulement quelques GWAS ont identifié des variants affectant des traits liés à la pression artérielle. Des relations ont été observés avec la PAS et PAD (77), et avec la pression artérielle en prenant en compte une interaction avec le statut de fumeur (75) et la consommation d'alcool (68). D'autres études ayant examiné des variants génétiques spécifiques sur *FTO* ont montré des résultats inconsistants (188). Puisque les rôles de *FTO* en dehors du métabolisme énergétique sont toujours méconnus, il est difficile d'élaborer sur un lien précis avec la régulation de la pression artérielle. Cependant, une étude de Pausova et al. (162) suggère que *FTO* pourrait avoir un effet sur la pression artérielle et que cet effet pourrait passer par l'influx sympathique vasomoteur.

3.4.4 rs17773430 (MC4R)

Le variant rs17773430 est positionné sur le gène *MC4R* situé sur le chromosome 18q21.32 (171). Ce gène code pour la protéine *melanocortin receptor 4*, un récepteur de membrane cellulaire de la famille des récepteurs de mélanocortine (172). *MC4R* est exprimé principalement dans le système nerveux central, plus particulièrement dans le noyau paraventriculaire de l'hypothalamus (189). Le récepteur fait partie du système de mélanocortine qui comprend plusieurs agonistes, antagonistes et récepteurs. Les agonistes de *MC4R* font principalement partie de la famille des mélanocortines et de l'hormone adénocorticotropique qui sont dérivés du pro-opiomélanocortine. Les agonistes et antagonistes de *MC4R* sont sécrétés en fonction des signaux provenant du système digestif ou du tissu adipeux (190, 191). Le système mélanocortine agit sur le métabolisme énergétique en rapport avec le contrôle de la satiété, la dépense d'énergie, la pression artérielle et la fréquence cardiaque. En général, l'activation de *MC4R* agit comme activateur du métabolisme

en ayant comme effet d'augmenter la satiété, la dépense énergétique et la fréquence cardiaque (191).

Des mutations sur *MC4R* sont responsables pour la plus grande partie des formes monogéniques d'obésité. *MC4R* fait partie d'un groupe de huit gènes ayant un rôle dans le système mélanocortine qui sont responsables pour environ 10% des cas d'obésité extrêmes tôt dans la vie. Plus de 200 mutations sur *MC4R* ont été identifiées pour des formes monogéniques d'obésité (191). Ces mutations ont comme effet d'augmenter la masse grasse et la masse maigre et favoriser l'hyperphagie. Aussi *MC4R* est l'un des premiers gènes à avoir été associé avec des mesures de l'adiposité par GWAS. La première étude du genre à avoir identifié un variant sur *MC4R* était le consortium international Genetic Investigation of Anthropometric Traits (170). Plusieurs autres variants sur *MC4R* ont ensuite été identifiés pour une association avec une variété de mesures anthropométriques chez les adultes (35) et ont ensuite été validés chez les enfants et les adolescents (137).

Bien que la relation entre *MC4R* et l'adiposité soit bien établie, il est moins clair si des variants génétiques sur ce gène influencent également la pression artérielle. Un seul GWAS (67) et quelques études examinant spécifiquement *MC4R* ont identifié des variants sur ce gène associés à la pression artérielle systolique ou l'hypertension (192-194). Globalement ces résultats n'ont pas été répliqués de façon consistante à-travers la littérature.

3.4.5 rs7120548 (*MTCH2*)

Le variant rs7120548 se trouve sur un intron du gène *MTCH2* situé sur le chromosome 11p11.2 (171). Ce gène code pour la protéine *mitochondrial carrier homolog 2*, un membre de la famille de transporteurs nucléiques SLC25 sont localisés dans la membrane externe des mitochondries (172). Cette protéine occupe notamment des fonctions dans la régulation de l'apoptose des mitochondries (195), le métabolisme énergétique incluant la phosphorylation oxydative (196), ainsi que le métabolisme lipidique (197). *MTCH2* est exprimé dans plus de 25 tissus, dont le foie, les reins, le cœur, le tissu musculaire et le tissu adipeux (197).

Comme *FTO* et *MC4R*, *MTCH2* est parmi les premiers gènes à avoir été identifié par GWAS pour un lien avec l'adiposité (35). Des variants sur ce gène ont ensuite été associés à

plusieurs reprises avec des mesures de l'adiposité incluant notamment l'IMC (133, 168, 169, 198), l'IMC prenant en considération le statut de fumeur (199) ou le niveau d'activité physique (185) et le pourcentage de gras (200). Les études examinant cette association chez les enfants et les adolescents sont toutefois moins nombreuses et plusieurs n'observent pas d'association avec l'adiposité (201-206). Une étude examinant la relation entre l'expression de certains gènes et l'obésité a observé que le niveau d'expression de *MTCH2* dans le tissu adipeux était effectivement associé avec l'obésité (207). Quelques études effectuées sur des modèles animaux montrent également un lien entre l'expression de *MTCH2* et l'accumulation de gras dans le tissu adipeux (197, 208).

Au niveau de la pression artérielle, la grande majorité des GWAS pour la pression artérielle n'ont pas identifié de variants sur *MTCH2*. Seulement un variant (rs7107356) sur ce gène a été associé avec la PAS, la PAD et la pression pulsée dans deux GWAS. Hoffman et al. (71) ont identifié rs7107356 pour une association avec les trois mesures de la pression artérielle. L'étude portait sur une méta-analyse d'individus de diverses origines ethniques. Le même variant a été identifié en 2018 par Kichaev et al. (209) en effectuant GWAS sur 422 000 individus d'ascendance européenne. Le variant rs7107356 y était associé avec la PAS. Un lien possible entre *MTCH2* et la pression artérielle est toutefois moins clair qu'avec l'obésité. Sachant que l'obésité affecte la pression artérielle, il est possible d'envisager que l'obésité agisse comme intermédiaire entre *MTCH2* et la pression artérielle, une possibilité qui n'a pas été vérifiée dans les deux études citées précédemment.

3.5 Réplication en études d'association génétique

La réplication des résultats de recherche est une partie essentielle de la démarche scientifique. L'objectif de la réplication en épidémiologie est d'évaluer les résultats obtenus dans une étude préalable à l'aide de données supplémentaires afin d'augmenter ou de diminuer la crédibilité de l'existence d'une association particulière (210). L'observation répétée d'une association dans diverses études de qualité effectuées par plusieurs groupes de chercheurs et à travers diverses populations et différentes méthodes constitue un indicateur important que l'association ne soit pas due à la chance ou à un biais (210). Le manque de réplication entre les études peut être attribuable à des résultats faux positifs lors des études initiales, à des résultats faux

négatifs ou un manque de puissance dans les études subséquentes ou bien à de véritables variations des associations entre les populations (211).

Les études d'association génétiques sont particulièrement susceptibles à l'observation de résultats inconsistants entre une étude initiale et subséquente. Dès les premières années des études d'association génétique, l'absence de réplification entre les études initiales et subséquentes était plus souvent la norme qu'une aberration. À titre d'exemple, Hirschhorn et al. (212) ont examiné 166 associations génétiques, montrant que seulement six d'entre elles étaient répliquées de façon consistante. Plus récemment, Border et al. (213) ont évalué 18 gènes identifiés par études par gènes candidats pour leur relation avec la dépression dans des échantillons allant de 62 138 à 443 264 participants. Aucun des 18 variants étudiés ne montrait d'effet significatif sur la dépression. Les 18 gènes dans lesquels se trouvaient ces variants sont parmi les plus vastement présents dans la littérature des associations par gènes candidats avec la dépression. Ces exemples illustrent bien l'importance d'effectuer des études de réplification dans le domaine de l'épidémiologie génétique afin d'évaluer la validité des résultats observés au préalable.

Plusieurs facteurs peuvent contribuer à l'inconsistance de résultats provenant d'études d'association génétique, incluant notamment l'utilisation de tailles d'échantillon trop petites, la dépendance en des seuils de significativité trop faibles, la stratification de population, les interactions gène-environnement ou les différences dans les populations d'étude, la définition des traits étudiés et les méthodes utilisées (210, 214).

Un autre défi inhérent aux études d'association génétique est la gestion du phénomène du *Winner's Curse* d'abord introduit dans le domaine de l'économie et des enchères (215, 216), et ensuite illustré à plusieurs reprises en épidémiologie génétique (37, 39, 217, 218). Ce phénomène correspond à ce que l'effet estimé d'un variant génétique est souvent plus grand dans l'étude qui rapporte l'effet pour la première fois que dans les études subséquentes qui tentent de répliquer le résultat initial (218). À titre d'exemple, dans une revue de 301 études d'association génétique, 24 des 25 loci étudiés montraient une estimation de l'effet génétique plus grand dans l'étude initiale que dans les études subséquentes (39).

L'exemple d'une enchère est une manière efficace et communément utilisée afin d'illustrer le *Winner's Curse*. Considérons une situation où tous les participants à une enchère placent des mises qui sont une estimation imprécise de la valeur d'un objet. La mise gagnante reflète une

surestimation de la valeur de l'objet conditionnellement à ce que cette mise est l'estimation la plus élevée de l'ensemble des mises non biaisées (39). Le mécanisme expliquant le *Winner's Curse* dans les études d'association génétique est lié à la sélection des variants à répliquer. Ceux-ci le sont uniquement en raison d'une valeur élevée de significativité statistique, sans tenir compte d'une hypothèse scientifique préalable (219). Également, l'estimation de l'effet génétique sur le trait étudié est directement corrélée avec la significativité statistique (217). Combiné à la méthode de sélection des variants à rapporter, la corrélation entre l'effet génétique et la significativité statistique suggère que la faible valeur de test observée le soit en raison d'un effet mesuré surestimé (217). Cela peut mener à la non-réplication de résultats initiaux en raison de la sous-estimation de la taille d'échantillon nécessaire pour détecter l'association dans les études subséquentes (37).

En réponse aux problématiques de réplication des études d'association génétique, le domaine a vu apparaître des rapports, recommandations ou critères concernant la façon de rapporter une association et le processus pour établir des réplifications solides. Le lecteur peut se référer au Tableau 2 pour une adaptation des recommandations soulevées dans la littérature par Bush et al. (220), Kraft et al. (210) et par un groupe de travail du National Cancer Institute et du National Human Genome Research Institute (221) concernant les bonnes pratiques concernant la réplication des résultats d'études d'association génétique. En bref, on souligne que la réplication d'association génétique fait référence à la réplication de l'effet d'un variant précis sur un trait particulier sous un modèle analytique spécifique (220). Ainsi, les études de réplication devraient d'abord être effectuées au sein d'une population d'étude semblable et indépendante de la population de l'étude originale afin d'éviter d'observer des différences dans l'association étudiée qui seraient dues à la différence entre les populations d'étude. Le ou les traits étudiés devraient également suivre une classification et des méthodes de mesures se rapprochant le plus possible de celles utilisées dans l'étude initiale. Idéalement, les variants étudiés initialement devraient être les mêmes dans les tentatives de réplifications. Or, dans l'éventualité où ceux-ci ne sont pas disponibles, il est possible d'utiliser des substituts en corrélation avec les variants initiaux. Une taille d'échantillon appropriée afin de détecter l'association d'intérêt devrait minimalement être plus grande afin de prévoir une possible surestimation de l'effet génétique dans l'étude initiale.

Tableau 2. Recommandations adaptées de la littérature[†] concernant la réplication d'études d'association génétique.

Aspect	Recommandation
Population d'étude	Les répliques devraient être effectuées dans des populations semblables, mais indépendantes afin de prévenir des différences au niveau de l'ascendance. Une fois l'association répliquée dans une population semblable, celle-ci peut être étudiée dans un échantillon provenant d'une différente population. Puisqu'une association génétique peut varier d'une population à une autre, une telle réplique peut ajouter davantage de crédibilité. Inversement, un échec de réplique dans une population différente ne peut pas à lui seul invalider l'association initiale qui pourrait être spécifique à une certaine sous-population.
Traits	Le trait examiné dans les études de réplique devrait être le plus identique possible à celui examiné lors de l'étude initiale. Utiliser une classification différente ou des différences dans les méthodes de mesure peut rendre l'interprétation des résultats plus complexe.
Variants	Les variants examinés devraient idéalement être les mêmes dans les études de réplique que dans l'étude initiale. Advenant que le variant étudié initialement ne puisse être utilisé, seul les variant en haut degré de LD devraient être utilisés comme substituts.
Modèle analytique	Idéalement, le même modèle analytique choisi dans l'étude initiale devrait être utilisé dans les études de réplique afin d'éviter toute variation due à une modélisation différente. Par exemple, advenant qu'un codage additif fût considéré pour le génotype, cela devrait également être le cas dans les études subséquentes.
Taille d'échantillon	Les études de réplique devraient avoir une taille d'échantillon suffisante afin d'identifier l'association à répliquer. On devrait également prévoir qu'il soit possible que l'effet estimé dans les études initiales soit en inflation dû au phénomène du <i>Winner's Curse</i> . Il est donc raisonnable d'ajuster la taille d'échantillon en anticipant un tel phénomène et ainsi conserver une taille d'échantillon suffisante.

[†]Tableau adapté à partir d'informations assemblées dans la littérature (210, 220, 221).

CHAPITRE 4 – MÉTHODES

4.1 Source des données

4.1.1 Étude NDIT

Les données utilisées proviennent de l'étude NDIT (43), une étude longitudinale ayant comme objectif d'examiner le cours naturel de la dépendance à la nicotine et les déterminants liés à l'utilisation de la cigarette chez les adolescents. Une technique d'échantillonnage par école a été utilisée afin d'identifier les participants débutant en 1999 (43). Un total de 13 écoles secondaires de la région de Montréal sélectionnées afin d'inclure des écoles francophones et anglophones, des écoles de milieux ruraux, urbains et suburbains et des écoles de milieux socioéconomiques diversifiés (43). Deux écoles ont été exclues en raison du faible niveau de retour de consentement parental et une école a été retirée en raison d'une non-garantie de participation après la première année (43). Parmi les 10 écoles, 1294 sur 2325 élèves éligibles en première secondaire (56%) ont accepté de participer à l'étude (43).

Un formulaire de consentement a été signé par les parents au recrutement des élèves prenant part à NDIT, puis par les participants à l'atteinte de la majorité et ultérieurement pour le prélèvement des échantillons de sang et de salive. L'étude a été approuvée par le Comité d'éthique de la recherche du Centre hospitalier de l'Université de Montréal (2017-6855), par le comité d'éthique du département de santé publique de Montréal et par le comité d'examen institutionnel de la faculté de médecine de l'Université McGill.

Les données collectées proviennent principalement de questionnaires autoadministrés s'étalant sur 23 cycles, 20 effectués de la 1^{ère} à la 5^e secondaire (4 questionnaires à intervalle de 3 mois au cours des 10 mois de l'année scolaire) et les 3 derniers effectués lorsque les participants avaient en moyenne 20, 24 et 30 ans (collecte de données en cours) (43). Outre les objectifs principaux de l'étude NDIT, ces données permettent aussi l'étude de l'activité physique, les comportements sédentaires, l'alimentation, l'utilisation d'alcool et d'autres substances illicites, la santé mentale et le sommeil (43). Des questionnaires supplémentaires pour les parents et les administrateurs des écoles ont également été effectués au cours de l'étude (43). En ce qui concerne l'objet du mémoire, des échantillons de sang et de salive ont été collectés en 2002 ou de 2007 à 2012 en vue d'analyses génétiques. Les données relatives aux mesures anthropométriques et à la pression artérielle ont été collectées lorsque les participants avaient en moyenne 13 (cycle 1), 15

(cycle 12), 17 (cycle 19), 24 ans (cycle 22) et 30 ans (cycle 23) (43). Afin de limiter les analyses aux adolescents et jeunes adultes, les données du cycle 30 ont été exclues des analyses du mémoire.

4.1.2 Échantillon analytique et pertes au suivi

Des 1294 participants de NDIT, 943 (73%) ont accepté de fournir un échantillon de sang ou de salive en vue d'analyses génétiques. Parmi les 943 échantillons disponibles, 40 ont été retirés en raison d'une faible qualité ou de discordance relative au sexe et 35 ont été exclus en raison d'un pourcentage de génotypage inférieur à 98%. Utilisant une analyse par composante principale, 726 participants ont été identifiés comme ayant une ascendance européenne. Neuf participants ont été exclus en raison de liens familiaux (frères et sœurs, jumeaux, triplés et mère). Finalement, quatre participants ont été exclus parce qu'ils n'avaient aucune mesure d'IMC et de PAS disponible. Ainsi, l'échantillon analytique regroupe 713 participants, soit 55% de l'échantillon total de NDIT et 76% de ceux ayant accepté de fournir un échantillon de sang ou de salive.

Le Tableau 3 présente les caractéristiques au cycle 1 des participants de NDIT inclus et exclus des analyses. Ceux-ci ont une PAS et un IMC semblable. Les participants exclus sont plus âgés en moyenne et proviennent davantage de familles monoparentales comparativement aux participants inclus.

Tableau 3. Caractéristiques au cycle 1 des participants inclus et exclus, NDIT 1999[†]

Variable	Inclus (n=713)	Exclus (n=581)	Valeur p
Âge, moyenne (ET)	12,7 (0,5)	12,8 (0,6)	<0,0001
Femme, %	53,7	49,5	0,1444
Mère a une éducation universitaire, %	46,4	40,8	0,1092
Famille monoparentale, %	6,3	13,1	0,0001
Déjà consommé de l'alcool, %	45,4	40,0	0,0647
Déjà fumé, %	29,0	32,2	0,2650
Taille, moyenne (ET)	156,3 (7,7)	156,2 (7,9)	0,9380
IMC, moyenne (ET)	20,2 (3,8)	20,0 (3,9)	0,2720
PAS, moyenne (ET)	105,4 (10,1)	105,1 (10,3)	0,7190

[†] Valeur p calculée par Chi-carré pour les variables catégorielles et par ANOVA pour les variables continues normalement distribuées. Les valeurs p en **gras** sont significatives à $p < 0,05$.

Dès 713 participants de l'échantillon analytique, 601 (84%) étaient toujours présents aux suivis au cycle 22. Le nombre d'observations d'IMC et de PAS disponibles par cycle est présenté au Tableau 4. Au total, 2350 et 2349 observations sont disponibles pour analyses pour l'IMC et la PAS respectivement.

Tableau 4. Description des pertes au suivi et du nombre d'observations disponibles pour l'IMC et la PAS, NDIT 1999-2012

Cycle	Participants suivis	%	IMC manquant	PAS manquante	Observations IMC	Observations PAS
1	713	100,0	48	49	665	664
12	712	99,9	106	104	606	608
19	688	96,5	166	167	522	521
22	601	84,3	44	45	557	556

4.2 Variables d'étude

4.2.1 Variables dépendantes

La PAS a été mesurée à l'aide d'un appareil oscillométrique (Dinamap XL, modèle CR9340, Critikon Co, Tampa, Fla) calibré à un sphygmomanomètre au mercure au début de chaque suivi de collecte de données relatives à la PAS et aux mesures anthropométriques. Toutes les mesures ont été effectuées par des techniciens ayant reçu une formation appropriée au préalable. À tous les suivis, trois mesures de pression artérielle étaient prises à intervalle d'une minute, au bras droit, en position assise et après un repos de cinq minutes. La première mesure était systématiquement exclue afin d'éviter le risque potentiel d'un effet de blouse blanche. La valeur rapportée pour le suivi correspondait à la moyenne entre la deuxième et la troisième mesure. Si la deuxième et troisième mesure différaient de plus de 20 mm de Hg pour la PAS ou de 10 mm Hg pour la PAD, une quatrième et une cinquième mesure étaient effectuées et la valeur de pression artérielle rapportée était la moyenne des deux mesures les plus semblables.

La taille et le poids des participants ont été mesurés avec un stadiomètre (modèle 214 Road Rod; Seca Corp., Hanover, MD, USA) et à l'aide d'une balance (modèle 761, Seca Corp., Hanover, MD, USA). Deux mesures étaient effectuées pour ces deux variables à chaque suivi. Une troisième mesure était prise dans la situation où la différence entre les mesures était supérieure à 0,2 kg pour le poids et à 0,5 cm pour la taille. La valeur finale conservée pour le suivi correspondait à la moyenne des deux mesures les plus proches. Les coefficients de fidélité interévaluateur étaient de

0,99 pour la taille et le poids. L'IMC était calculé comme la masse divisée par le carré de la taille (m²).

4.2.2 Variables indépendantes

Les variables indépendantes à l'étude sont les variants rs16933812 (*PAX5*), rs7638110 (*MRPS22*), rs9930333 (*FTO*), rs17773430 (*MC4R*) et rs7120548 (*MTCH2*) identifiés par la SYS pour leur association statistiquement significative avec l'IMC, et testés pour leur association avec la PAS. Les variants ont été génotypés par Génome Québec à l'aide de la puce Illumina Global Screening Array-24 v1.0. L'imputation des variants a été effectuée avec le logiciel Minimac3 (222) utilisant les haplotypes du panel de référence 1000 Genomes Phase 3 (176). Le variant rs17773430 sur *MC4R* n'étant pas disponible dans le panel de variants de NDIT, il a été remplacé par un proxy en haut LD ($r^2=0,976$), rs111638368 (*MC4R*), dans une population comparable du 1000 Genomes Project Phase 3 (176) (résidents de l'Utah avec ascendance de l'Europe du Nord et de l'Ouest (CEU)). Afin de comparer les effets estimés dans NDIT à ceux de la SYS, le génotype des individus en rapport aux variants à l'étude est codé selon un codage génotypique, c'est-à-dire de façon catégorique (homozygote allèle mineur, hétérozygote et homozygote allèle majeur). Une analyse secondaire sera effectuée selon un codage additif (0, 1 ou 2 allèles) afin de modéliser un effet linéaire des allèles sur l'IMC et la PAS, codage généralement plus puissant qu'un codage génotypique.

4.2.3 Variables additionnelles

Certaines variables additionnelles sont utiles afin de décrire la population d'étude. Les variables sociodémographiques utilisées incluent l'âge des participants (variable continue), le statut familial (monoparental ou non, mesuré par le nombre d'adultes à la maison), le sexe des participants (femme ou homme), le pays de naissance (Canada ou autre) et l'éducation des mères des participants (éducation universitaire ou non). Le nombre de séances d'activité physique modérée ou vigoureuse effectuées par semaine (variable continue), la consommation d'alcool (consommation ou non au cours des trois derniers mois), le statut de fumeur (jamais ou déjà fumé), ainsi que le nombre de cigarettes fumées au cours des derniers trois mois chez ceux ayant déjà fumé (variable continue) ont également été utilisés afin de décrire la population d'étude. Une description plus élargie de ces variables incluant les cycles de données disponibles, l'énoncé du

questionnaire, les choix de réponses possibles et le formatage des réponses pour analyse si nécessaire est disponible à [l'Annexe C](#). Les variables qui variaient dans le temps ont été décrites aux cycles 1, 12, 19 et 22, à l'exception du statut familial qui n'avait pas de données disponibles pour le cycle 22. Les variables qui ne variaient pas dans le temps ont été considérées au questionnaire du cycle 1 pour le sexe et le pays de naissance et au questionnaire destiné aux mères lorsque les participants avaient 15-16 ans pour l'éducation des mères.

Il est également important de déterminer si certains facteurs de confusion pourraient biaiser l'association entre les variants étudiés et l'IMC et la PAS. Une variable peut être considérée comme une variable de confusion potentielle si elle affecte à la fois la variable indépendante et la variable dépendante, sans être dans le chemin causal entre ces deux dernières. Dans NDIT, les variables indépendantes étudiées sont des variants génétiques déterminés à la naissance qui surviennent temporellement avant l'influence d'autres variables. Ces variants ne sont donc pas influencés par des facteurs de risque de l'IMC ou de la PAS et ne seront donc pas considérés comme des variables de confusion. L'ascendance des participants est une exception et pourrait influencer à la fois les fréquences alléliques des variants génétiques investigués et l'IMC et la PAS. Bien que l'échantillon analytique ait été réduit aux participants d'ascendance européenne, une sous-structure de population peut tout de même persister et introduire de la confusion résiduelle. Les variables représentant des facteurs de risque pour les variables dépendantes peuvent tout de même être introduites dans les analyses statistiques afin d'augmenter la variation expliquée des variables dépendantes et ainsi améliorer la puissance statistique, dans la mesure où ces variables ne se retrouvent pas sur le chemin causal entre les variables indépendantes et dépendantes. Ainsi, l'âge au début de l'étude, le sexe et la taille (pour la PAS seulement) ont été considérés comme covariables dans les analyses statistiques. En ce qui concerne l'analyse pour la PAS, il est possible que l'IMC soit une variable intermédiaire et ne sera donc pas ajoutée initialement comme covariable. Advenant l'observation d'une association statistiquement significative entre un variant investigué et la PAS, une analyse de médiation par l'IMC pourra être effectuée.

4.3 Analyses statistiques

4.3.1 Analyses descriptives

Toutes les analyses statistiques, incluant les analyses descriptives et longitudinales ont été effectuées à l'aide du logiciel statistique R version 3.5.3 (223). Une description des participants a été effectuée pour les quatre suivis ayant des données disponibles pour l'IMC et la PAS (cycle 1, 12, 19 et 22). La description des participants comprend la moyenne et l'écart-type des variables continues normalement distribuées, la médiane et l'écart interquartile des variables continues qui ne sont pas normalement distribuées et la fréquence des variables catégorielles. L'échantillon analytique pour chaque cycle de mesure correspond au nombre d'individus étant actifs aux suivis au cycle correspondant. Les données manquantes sont également décrites pour chaque variable aux quatre cycles étudiés. Une donnée pouvait être manquante en raison d'un questionnaire non rempli, de questions spécifiques non répondues dans un questionnaire, ou bien d'une mesure anthropométrique ou de pression artérielle manquée spécifiquement.

Les caractéristiques de l'échantillon analytique au cycle 1 ont également été comparées entre les participants perdus au suivi au cours de l'étude (n=112) et les participants toujours disponibles au cycle 22 (n=601). Cette analyse a été effectuée avec ANOVA pour les variables normalement distribuées, Wilcoxon pour les variables non normalement distribuées et chi-carré pour les variables catégoriques.

4.3.2 Analyses longitudinales

La PAS et l'IMC sont considérés comme des variables continues et ne sont donc pas catégorisées selon les critères de classification de l'hypertension et de l'obésité. L'analyse de l'association entre les cinq variants à l'étude avec l'IMC et la PAS a été effectuée à l'aide de modèles linéaires mixtes avec effets aléatoires pour tenir compte de la corrélation entre les mesures sur les mêmes individus (224). On décrit en premier lieu la stratégie de modélisation pour le modèle ayant l'IMC comme variable dépendante. Une structure de corrélation par symétrie composée a été utilisée afin de tenir compte de l'intervalle non équivalent entre les mesures anthropométriques et de pression artérielle (0, 36, 57 et 144 mois). Un modèle sans effet génétique, mais incluant l'âge au début de l'étude, le sexe et le temps écoulé depuis le cycle 1 (en mois) comme covariables a d'abord été construit afin de bien modéliser l'effet du temps sur l'IMC.

D'abord, une pente aléatoire a été considérée pour l'effet du temps et ce modèle est représenté à la formule (A). L'hypothèse de linéarité de l'effet du temps a ensuite été évaluée graphiquement et à l'aide d'une paramétrisation polynomiale, de la catégorisation pour le temps et par lissage. Le choix du modèle général représentant le mieux l'effet du temps sur l'IMC a été effectué en comparant l'Akaike Information Criterion (AIC) de chaque modèle. Une analyse des résidus a ensuite été effectuée afin d'évaluer les présuppositions (homogénéité de la variance et distribution normale des résidus) du modèle linéaire mixte retenu. Une fois le modèle sans effet génétique déterminé, les variants ont été ajoutés aux modèles. On s'assure ainsi que les modèles restent constants entre l'analyse de chaque variant et que les différences observées entre les modèles ne représentent que l'ajout des variants. Afin de comparer les résultats obtenus avec ceux de la SYS les variants ont été incorporés aux modèles selon un codage génotypique qui est présenté à la formule (B).

$$IMC_{ij} = \beta_0 + \beta_1 T_{ij} + \beta_2 Sex_i + \beta_3 AgeB_i + S_{0i} + S_{1i} T_{ij} + \varepsilon_{ij} \quad (A)$$

$$\text{Avec : } \varepsilon_{ij} \sim N(0, \sigma^2), S_{0i} \sim N(0, \sigma_{S_0}^2) \text{ et } S_{1i} \sim N(0, \sigma_{S_1}^2)$$

La covariance entre S_{0i} et S_{1i} est $\sigma_{S_0 S_1}$

$$IMC_{ij} = \beta_0 + \beta_1 T_{ij} + \beta_2 Sex_i + \beta_3 AgeB_i + \beta_4 Aa_i + \beta_5 aa_i + S_{0i} + S_{1i} T_{ij} + \varepsilon_{ij} \quad (B)$$

Où pour (A) et (B) :

- $i = 1, \dots, n$ individus et $j=1, \dots, 4$ occasions de mesure
- T : Temps écoulé depuis le cycle 1
- $AgeB$: Âge au début de l'étude
- Aa : groupe des hétérozygote (individus possédant un allèle de risque)
- aa : groupe des homozygote pour l'allèle mineur (individus possédant deux allèles de risque)

La même stratégie a été appliquée aux modèles ayant la PAS pour variable dépendante. Cependant, la taille a aussi été considérée comme une covariable en plus de l'âge en début d'étude, le temps écoulé depuis le cycle 1 et du sexe.

4.3.3 Taille d'échantillon et puissance statistique

La présente étude est une étude secondaire de NDIT donc la puissance statistique pour détecter des tailles d'effets plausibles pour les variants en tenant compte de la taille d'échantillon disponible ($n=713$) a été calculée avec le logiciel GLIMPSE (225). Les moyennes de PAS de 110 mm Hg (écart-type = 11 mm Hg) et d'IMC de 22 kg/m² (écart-type = 4 kg/m²) pour le cycle 1 de NDIT ont été utilisées pour le calcul de puissance. L'augmentation moyenne entre les garçons et les filles a été utilisée à partir des données du Centers for Disease Control and Prevention (226) pour déterminer l'augmentation de la PAS et de l'IMC avec l'âge. Des fréquences alléliques entre 15% et 45% pour l'allèle mineur ont été considérées afin de couvrir l'étendue des fréquences alléliques des variants investigués. Une fréquence allélique entre 5 et 10% peut être envisagée pour rs7638110 (*MRPS22*), mais une telle fréquence ne peut être représentée en raison de contraintes du logiciel utilisé. La puissance calculée varie de 75% à 80% pour détecter un effet de 0,5 mm Hg et de 89% à 93% pour détecter un effet de 0,25 kg/m² par allèle de risque pour des fréquences d'allèle mineur supérieures à 20%. En raison du faible niveau d'exposition pour une fréquence de l'allèle mineur fixée à 15%, la puissance y est réduite à 68% et 83% pour la PAS et l'IMC respectivement. Les estimations obtenues pour rs7638110 (*MRPS22*) devront donc être interprétées avec prudence. Les valeurs détaillées pour le calcul de puissance sont disponibles dans [l'Annexe D](#) (Tableaux A1 et A2).

4.3.4 Analyses de sensibilité

Des analyses de sensibilité supplémentaires ont été effectuées. D'abord, un modèle examinant l'effet des variants sur l'IMCZ comme variable dépendante a été considéré afin de tenir compte du fait que l'IMCZ est habituellement la mesure d'adiposité utilisée chez la population pédiatrique. Puisque les participants sont âgés d'en moyenne 24 ans au cycle 22, cette analyse a été restreinte aux trois premières collectes de données anthropométriques. Les analyses longitudinales de l'effet des variants sur la PAS et l'IMC ont également été effectuées en omettant

le cycle 22 où les participants étaient de jeunes adultes (24 ans en moyenne) du modèle afin de se rapprocher du groupe d'âge utilisé par la SYS. Comme mentionné plus haut, l'ascendance des participants pourrait représenter une variable de confusion dans l'association entre les variants génétiques et les traits examinés. Ainsi, une analyse par composantes principales incluant les cinq premières composantes a donc été ajoutées aux modèles statistiques. Aussi, un terme d'interaction entre les variants et le temps mesurant si l'effet des variants sur l'IMC et la PAS est modifié par le temps a été considéré afin de s'assurer de la constance de l'effet du variant sur l'IMC et la PAS et ainsi obtenir une meilleure comparabilité avec les résultats obtenus par la SYS. Finalement, un terme d'interaction entre les variants et l'IMC a été considéré afin d'évaluer l'influence de l'IMC sur la relation entre les variants et la PAS. Ces effets conditionnels des variants sur l'IMC et la PAS ont été représentés graphiquement à l'aide du module R *interplot* (227).

4.4 Comparaisons des études

Pour faciliter la lecture de l'étude de réplication, le Tableau 5 offre une comparaison sommaire de l'échantillon analytique, des variants génétiques et des traits investigués, des méthodes de mesure et des analyses statistiques entre NDIT et la SYS.

Tableau 5. Comparaison des méthodes utilisées dans NDIT et par la SYS

	NDIT (2020)	SYS (2012)
Échantillon analytique	713	598
Provenance de l'échantillon analytique	Participants de l'étude longitudinale NDIT recrutés dans 10 écoles secondaires de Montréal	Participants de l'étude SYS recrutés dans des écoles secondaires du Saguenay-Lac-St-Jean
Âge au recrutement	12 à 13 ans en 1999-2000	12 à 18 ans entre 2003 et 2009
Ascendance	Européenne (validé par analyse de composantes principales)	Européenne (grands-parents maternels et paternels d'origine canadienne-française)
Devis	NDIT est une étude longitudinale.	Transversal
Traits investigués	Adiposité (IMC) Pression artérielle (PAS)	Adiposité (Masse grasse totale, IMC) Pression artérielle (PAS)
Méthode de mesure IMC	-Taille et poids mesurés avec un stadiomètre et une balance. -Deux mesures par collecte de données. -Troisième mesure lorsque la différence entre les deux premières mesures supérieures à 0.2 kg et 0.5 cm. -La moyenne des deux mesures les plus proches était rapportée.	-Taille et poids mesurés avec une précision de 0.1 kg et de 1 mm
Méthode de mesure PAS	-PAS mesurée avec un appareil oscillométrique. -Trois mesures à chaque collecte de données. -Première mesure systématiquement retirée. -Deux mesures supplémentaires effectuées advenant un différence entre les mesures supérieure à 20 mm de Hg pour la PAS ou 10 mm Hg pour la PAD. -La valeur rapportée était la moyenne des deux mesures les plus semblables.	-PAS mesurée avec un moniteur hémodynamique continu pour les doigts. -Protocole de 52 minutes en cinq contextes : debout, assis, couché, stress mental et repos de stress. -La valeur rapportée était la moyenne de PAS pour chacune des parties.
Variants génétiques investigués	IMC, PAS : rs16933812, rs7638110, rs9930333, rs17773430 et rs7120548	IMC, masse adipeuse totale : GWAS PAS : variants associés aux mesures d'adiposité (rs16933812, rs7638110, rs9930333, rs111638368 et rs7120548)
Analyses statistiques	-IMC : modèle linéaire mixte avec effet aléatoire pour mesures rassemblées pour chaque participant. -PAS : modèle linéaire mixte avec effet aléatoire pour mesures rassemblées pour chaque participant.	-IMC : modèle linéaire simple avec composante de variance pour prendre en compte la corrélation entre individus provenant d'une même famille. -PAS : modèle linéaire mixte avec effet aléatoire pour les cinq mesures provenant du même individu à différents contextes et pour la corrélation entre mesures sur individus d'une même famille.
Covariables	Âge au début de l'étude, sexe, temps écoulé depuis le cycle 1 et taille (PAS uniquement)	Âge, sexe et taille (PAS uniquement)

CHAPITRE 5 – RÉSULTATS

5.1 Présentation du manuscrit

Ce chapitre comporte le manuscrit intitulé : *Common genetic variants associated with body-mass index and systolic blood pressure: replication and the Winner's Curse*. Celui-ci est sous la forme d'un *brief research report* et le journal cible est *Frontiers in Genetics* qui encourage la soumission de rapports traitant de la non-reproductibilité de résultats publiés antérieurement. Le manuscrit contient le corps de l'ouvrage, ainsi qu'une section sur les matériels supplémentaires qui seront inclus lors de la soumission au journal.

En tant que premier auteur et sous la supervision de Marie-Pierre Sylvestre, j'ai joué un rôle central à toutes les étapes de la préparation du manuscrit, incluant la conceptualisation, le développement des objectifs de recherche et de la méthodologie utilisée, l'analyse formelle des résultats et la rédaction du manuscrit. En plus d'agir comme ma superviseuse de recherche, Marie-Pierre Sylvestre a également révisé le manuscrit. Jennifer O'Loughlin a révisé le manuscrit et a été essentielle à l'acquisition des données utilisées dans l'étude en tant qu'investigateur principal de l'étude NDIT.

5.2 Manuscrit

Title: Common genetic variants associated with body-mass index and systolic blood pressure: replication and the Winner's Curse

Authors: Goulet D¹, O'Loughlin J^{1,2}, Sylvestre MP^{1,2}.

Affiliations: ¹Department of Social and Preventive Medicine, Université de Montréal, Montréal, Quebec, Canada; ²Centre de recherche du centre hospitalier de l'Université de Montréal (CRCHUM), Montréal, Quebec, Canada.

Correspondance : Marie-Pierre Sylvestre, Centre de recherche du centre hospitalier de l'Université de Montréal, 850 Saint-Denis (S03-458), Montreal, Quebec, Canada, H2X 0A9. Tel: 514-890-8000, extension 31474; Fax: 514-412-7953; Email: marie-pierre.sylvestre@umontreal.ca

Word count (main text): 2954 (Maximum 4000)

Word count (abstract): 198

Number of tables: 3

Number of figures: 0

Number of Supplementary materials: 8

Abbreviations: BMI – body mass index; SBP – systolic blood pressure; DBP – diastolic blood pressure; SNP – single nucleotide polymorphism; NDIT study – Nicotine Dependence In Teens study; LD – linkage disequilibrium; MAF – minor allele frequency.

ABSTRACT

The strong correlation between adiposity and blood pressure might in part be explained by shared genetic risk factors. A recent study identified five single nucleotide variants, rs16933812 (*PAX5*), rs7638110 (*MRPS22*), rs9930333 (*FTO*), rs7120548 (*MTCH2*) and rs17773430 (*MC4R*), associated with body mass index, three of which (*MRPS22*, *PAX5* and *FTO*) were also associated with systolic blood pressure in adolescents aged 12 to 18 years old. We attempted to replicate these findings in a similar sample of adolescents. A total of 713 adolescents were genotyped and had anthropometric and systolic blood pressure measured at age 13, 15, 17 and 24 on average. Using linear mixed models, we assessed associations of the five variants with both body mass index and systolic blood pressure. We replicated the effect of variants at genes *FTO* and *MTCH2* on body mass index, but not those on *MRPS22*, *PAX5* or *MC4R*. Further, none of the associations between the variants and systolic blood pressure was replicated. Explanations for the partial failure to replicate include differences in genetic structures across populations and in definitions of phenotype. However, our results also suggest that selection of variants in the earlier study might have been prone to the Winner's Curse phenomenon.

Key words: Adiposity, blood pressure, genetic association study, adolescence.

INTRODUCTION

In 2015, high blood pressure and high body mass index (BMI) were the first and fourth risk factors, respectively for disability-adjusted life-years (1). Blood pressure and adiposity are strongly correlated (2), and higher adiposity is an important risk factor for elevated blood pressure, both in adults (3) and adolescents (4). The common variants/multiple disease hypothesis first introduced by Becker (5) posits that the correlation between related traits, such as adiposity and blood pressure, may be explained by shared genetic risk factors. In a previous study, Melka et al. (6) identified three single nucleotide polymorphisms (SNPs), associated with both BMI and systolic blood pressure (SBP) in adolescents aged 12 to 18 years in Saguenay-Lac-St-Jean in Quebec, Canada. However, these results have not yet been replicated in an independent sample.

Investigating the genetic underpinning of adiposity and blood pressure in youth is important because it cannot be assumed the genetic architecture present in adults can be inferred to children and adolescents (7). For example, results from a longitudinal twin study suggested that genes associated with blood pressure are expressed differently in childhood and adulthood (8). Further, investigating genetic susceptibility in youth, who are rarely treated with antihypertensive medication and who are less affected than adults by the cumulative impact of exposure to the environment, might help identify genetic factors that influence both adiposity and blood pressure (9, 10). However, genetic association studies for adiposity and blood pressure on youth are scarce and generally have smaller sample size than those in adults (11, 12), highlighting the need for proper replication of previously reported findings, whether previously identified in adults or newly discovered in adolescents.

The importance of replication of genetic studies is further supported by the phenomenon of the Winner's Curse (13), in which the estimates of the associations between newly discovered genetic variants and phenotypes in genome-wide associations studies (GWAS) tend to be larger in the discovery sample than in subsequent replication studies. This complicates the replication of findings, especially if the sample size calculations for the replication study is based on overestimated effect sizes, resulting in an underpowered study (14). First described in the field of economics and auctions (15, 16), this phenomenon applies to GWAS because discoveries are identified based on a stringent significance level, often around 5×10^{-8} . A newly discovered variant is likely to have an overestimated effect because its discovery is conditional on it being the highest of all previous estimations (17). The problem often arises when the same sample is used for both discovery and effect estimation (18) and the SNPs of interests correspond to those that had the lowest p-values (19).

Among the three single SNPs reported as associated with both BMI and systolic blood pressure in adolescents by Melka et al, two corresponded to newly discovered variants of BMI that were next found to be associated with SBP in the same dataset. With considerations of both the Winner's Curse and the importance of replicating genetics studies in adolescents in mind, we sought to replicate Melka et al.'s findings by capitalizing on a similar sample in the Nicotine Dependence In Teens (NDIT) study (20), a longitudinal investigation of adolescents in Montreal in Quebec, Canada followed from age 12 (1999-2012) to 24.

METHODS

Study design and analytical sample

Study participants provided data in the Nicotine Dependence in Teens (NDIT) study, a longitudinal investigation of the natural course of nicotine dependence across adolescence (20). A total of 1294 grade 7 participants age 12 to 13 years were recruited in 1999-2000 in 10 Montreal-area high schools selected to include French and English schools, schools located in rural, urban and suburban areas, and schools serving socioeconomically diverse neighborhoods. Data was collected in self-administered questionnaires across 23 study cycles. Data from cycles 1 to 20 were collected in grade 7 to 11 (i.e., the last year of high school in Quebec) every three months during the 10-month school year over five years. Post high school data were collected in cycles 21 to 23 when participants were age 20, 24 and 31 years on average.

DNA samples were collected from 943 of 1294 NDIT participants (73%). A total of 713 participants were unrelated, of European descent, had passed genetic quality control and had at least one BMI or SBP measure available, and thus comprised the analytical sample. Anthropometric and blood pressure measurements used in NDIT were obtained at cycles 1, 12, 19, 22 and 23 when participants were age 13, 15, 17, 24 and 31 years on average, respectively. Analyses excluded survey cycle 23 to avoid a too severe drop in sample size due to attrition.

Consent form was signed by parents at inception, and then by students when they have attained legal age. Additional consent was obtained for the collection of blood and saliva samples. This study was approved by the ethics review committees of the Montréal Department of Public Health, the McGill University Faculty of Medicine, and the Centre de Recherche du Centre Hospitalier de l'Université de Montréal.

Outcome variables

Trained technicians measured height and weight (Seca Portable Stadiometer – Model 214 and Seca Scale – Model 761, Seca Corporation, Columbia, MD, USA) according to a standardized protocol (21). BMI was calculated as weight (kg) divided by height squared (m^2). Blood pressure was assessed three times at 1-min interval with an oscillometer (Dinamap XL, modèle CR9340, Critikon Co, Tampa, Fla) calibrated to a mercury sphygmomanometer. SBP was computed as the average of the second and third SBP measurements. Additional information on BMI and SBP measurement in NDIT has been reported previously (20) and can also be found in Supplementary text 1.

SNP genotyping

SNPs investigated were rs16933812 (*PAX5*), rs7638110 (*MRPS22*), rs9930333 (*FTO*), rs7120548 (*MTCH2*) and rs17773430 (*MC4R*). Genotyping was performed by Genome Québec using the Illumina Global Screening Array-24 v1.0 (GSA) and imputation was done with Minimac3 software (22) according to haplotypes from the 1000 Genomes Phase 3 reference panel (23). The SNP rs17773430 (*MC4R*) was not available within the genotyping array used and was therefore replaced with a proxy, rs111638368 (*MC4R*), in strong linkage disequilibrium (LD) ($r^2=0.976$) in a comparable population in the 1000 Genomes Project Phase 3 (24) (Utah residents with Northern and Western European ancestry (CEU)). Additional details about the genotyping process and genetic quality control is available in Supplementary text 2.

Statistical analysis

Characteristics of the 713 participants in the analytical sample were described at each cycle. Normally distributed variables were expressed as mean (SD), non-normally distributed variables as medians (IQR), and categorical variables as frequencies.

Melka et al, SNPs used two approaches to select SNPs associated with BMI that would then be tested for their association with SBP. First, they perform a genome-wide scan of their data and identified SNPs that passed the conventional genome-wide threshold ($p < 5 \times 10^{-8}$) for their association with either BMI or total fat mass. SNPs rs16933812 (*PAX5*) and rs7638110 (*MRPS22*) were selected this way. Second, they considered 33 SNPs detected in previously published GWAS of BMI and estimated the associations those SNPs had with BMI in their data. From this list, they selected SNPs that (i) were in high LD ($D' > 0.90$) with at least one of the 33 SNPs; (ii) had the same direction of association as in the previously published GWAS; and (iii) were associated with BMI in the SYS study with $p < 5 \times 10^{-4}$. Three

SNPs satisfied these three criteria, namely rs9930333 (*FTO*), rs7120548 (*MTCH2*) and rs17773430 (*MC4R*). These three SNPs were in high LD with SNPs previously reported in GWAS for BMI (rs9939609 (*FTO*) (25), rs3817334 (*MTCH2*) (26) and rs17782313 (*MC4R*) (27)). In order to detect differences in LD for those SNPs between Melka et al. and our study, we calculated the D' measure for the three pairs of SNPs in the NDIT sample using Haploview v4.2 (28).

Regression coefficients for the difference in BMI and SBP between minor and major allele homozygotes for each SNPs were estimated in linear mixed models to account for repeated measures. Models were adjusted for age at baseline, sex, time elapsed since baseline (0, 36, 57, 144, months) and height (SBP only). A compound symmetry correlation structure was assumed to consider the unequal time elapsed between BMI and SBP measurements across study cycles. Genotype was coded according to a genotypic model (minor allele homozygous, heterozygous or major allele homozygous) in order to compare results with Melka et al. All statistical analyses were conducted with R statistical software version 3.6.3 (29).

RESULTS

Table 1 shows participant characteristics at cycles 1, 12, 19 and 22. There were more females than males in the analytical sample across all study cycles (53.7%-55.9%). Mean values of BMI and SBP increased over time, although SBP plateaued from 17 to 24 years old. Due to attrition, sample size decreased from 713 at cycle 1 to 601 participants (84.3% of initial sample) at cycle 22. Compared to participants who were lost to follow-up before cycle 22, those remaining were younger, more likely to be female (56% vs 42%) and had a lower BMI (1.2 kg/m²) on average at baseline (Supplementary Table 1).

Table 1: Characteristics of analytical sample at cycles 1, 12, 19 and 22, NDIT 1999-2012†

Variable	Cycle 1 (n=713)	Missing values	Cycle 12 (n=712)	Missing values	Cycle 19 (n=688)	Missing values	Cycle 22 (n=601)	Missing values
Age, mean (SD)	12.7 (0.5)	14	15.1 (0.4)	89	17.0 (0.4)	139	24.0 (0.7)	44
Female, %	53.7	0	53.7	0	54.4	0	55.9	0
Mother university- educated, %	46.4	73	46.4	72	46.8	60	47.2	46
Canada born, %	98.5	0	98.5	0	98.7	0	98.8	0
Single-parent family, %	6.3	32	11.6	93	14.8	142		
MVPA/week, med (IQR)	11.0 (14.0)	33	11.0 (15.0)	140	9.0 (13.0)	233		
Alcohol consumption, %	45.4	37	54.9	102	77.2	145		
Ever smoked, %	29.0	38	47.4	92	50.9	140	76.5	1
No cigarettes/month, med (IQR)	0.2 (2.1)	46	0.2 (35.8)	101	0.3 (47.3)	149	0.5 (130.9)	8
Height, \mean (SD)	156.3 (7.7)	48	166.9 (7.9)	106	169.6 (8.6)	166	170.4 (9.0)	44
BMI, mean (SD)	20.2 (3.8)	48	21.9 (3.7)	106	22.7 (3.7)	166	24.6 (4.6)	44
SBP, mean (SD)	105.4 (10.1)	49	106.8 (10.4)	104	110.8 (11.3)	167	109.2 (11.5)	45

Note: n = sample size, SD = standard deviation, IQR = interquartile range, med = median, MVPA = moderate or vigorous physical activity.

† For each cycle n = number of participants remaining in the study. Data were missing because participants not completing a questionnaire or missing data on specific variables. At cycle 22, data for family composition, physical activity and alcohol consumption was not collected, while BMIz was not considered since participants were adults. Mother university-educated: measured in parental questionnaires when participants were 15-16 years old. Alcohol consumption: had any alcohol consumption in the past three months. MVPA/week: number of vigorous or moderate physical activities done for at least 10 minutes for the last week. No cigarettes/month: number of cigarettes smoked per month for the last three months among ever smokers.

Table 2 shows the LD between pairs of originally reported SNPs in GWAS studies (rs9939609 (*FTO*), rs3817334 (*MTCH2*) and rs17782313 (*MC4R*)) and SNPs investigated our study (rs9930333 (*FTO*), rs7120548 (*MTCH2*) and rs111638368 (*MC4R*)). Those pairs were reported to be in high LD (>0.90) in Melka et al. In our study, high LD was observed for the *FTO* ($D' = 0.96$) and *MTCH2* SNP pairs ($D' = 1.00$). In contrast, rs111638368 (*MC4R*) was in low LD with rs17782313 (*MC4R*) ($D' = 0.33$). In accordance with our objective to specifically replicate findings from Melka et al., we elected to pursue with rs111638368 despite its low LD with SNPs reported in GWAS.

Table 2. Measure of linkage disequilibrium (D') between SNPs previously reported in GWAS and Melka et al. and NDIT studies[†]

GWAS			Melka et al.		NDIT	
Gene	SNP	Reference	SNP	D'	SNP	D'
FTO	rs9939609	Willer et al. ²⁵	rs9930333	>0.90	rs9930333	0.96
MTCH2	rs3817334	Speliotes et al. ²⁶	rs7120548	>0.90	rs7120548	1.00
MC4R	rs17782313	Loos et al. ²⁷	rs17773430	>0.90	rs111638368	0.33

[†] SNPs are those investigated in the original GWAS study, in SYS and in NDIT.

Table 3 compares the results reported in Melka et al. and those obtained in NDIT for both phenotypes. MAFs were similar across studies except for rs16933812 (*PAX5*) which was higher in NDIT than in Melka et al. (37% vs. 22%). Table 3 further shows estimated adjusted regression coefficients with 95% confidence intervals (95% IC) for the difference in BMI and SBP between minor and major allele homozygotes for each of the five SNPs considered. SNPs rs9930333 (*FTO*) and rs7120548 (*MTCH2*) were the only two associated with BMI and the directions of these association were consistent with Melka et al. (Table 3). Mean BMI difference between minor allele homozygotes and major allele homozygotes for rs9930333 (*FTO*) was 1.18 (95% CI 0.42 , 1.94) kg/m². This difference was -0.89 (95% CI -1.76 , -0.02) kg/m² for rs7120548 (*MTCH2*). Neither SNP was associated with SBP.

The remaining three SNPs (rs111638368 (*MC4R*), rs16933812 (*PAX5*) and rs7638110 (*MRPS22*)), were not associated with either BMI or SBP (Table 3). The estimated difference in BMI for the *PAX5* SNP was near null with a narrow CI, and that difference was small and in the opposite direction compared to Melka et al. for the *MRPS22* SNP. Both SNPs were associated with SBP in Melka et al, but

the CI for their estimated effect on SBP included 0 and was in the opposite direction in NDIT. Although not statistically significant, the difference in BMI observed for the SNP at the *MC4R* SNP was greater than the previous two SNPs.

Table 3: Estimated difference in BMI and SBP between minor allele homozygous and major allele homozygous, NDI 1999-2012 and Melka et al. 2003-2009[†]

SNP information			Melka et al.			NDIT				
Gene	SNP	Minor allele	MAF	\hat{B} [CI]	P value	Proxy	Minor allele	MAF	\hat{B} [CI]	P value
BMI (n= 713/2350)										
FTO	rs9930333	G	0.38	1.8 [NA]	1.9×10^{-4}		G	0.43	1.18 [0.42, 1.94]	0.0025
PAX5	rs16933812	G	0.22	3.3 [1.9, 4.6]	5.2×10^{-6}		G	0.37	-0.03 [-0.88, 0.81]	0.9353
MTCH2	rs7120548	C	0.29	-1.6 [NA]	1.9×10^{-4}		C	0.35	-0.89 [-1.76, -0.02]	0.0445
MC4R	rs17773430	C	0.27	2.4 [NA]	5.8×10^{-6}	rs111638368	T	0.29	0.71 [-0.27, 1.68]	0.1565
MRPS22	rs7638110	T	0.06	2.8 [1.9, 3.6]	4.6×10^{-8}		T	0.07	-0.22 [-0.98, 0.53]	0.5623
SBP (n= 713/2347)										
FTO	rs9930333	G	0.38	3.63 [0.73, 6.53]	0.01		G	0.43	0.64 [-1.08, 2.35]	0.4679
PAX5	rs16933812	G	0.22	6.68 [2.80, 10.57]	0.0007		G	0.37	-0.55 [-2.44, 1.34]	0.5716
MTCH2	rs7120548	C	0.29	-0.51 [-3.86, 2.84]	0.77		C	0.35	0.31 [-1.65, 2.26]	0.7563
MC4R	rs17773430	C	0.27	-0.05 [-3.44, 3.34]	0.98	rs111638368	T	0.29	0.77 [-1.41, 2.95]	0.4880
MRPS22	rs7638110	T	0.06	3.30 [0.49, 6.12]	0.02		T	0.07	-1.52 [-3.20, 0.17]	0.0777

[†] Estimated differences in SBP and BMI levels of minor allele vs. major allele homozygous (*PAX5*, *MTCH2*, *MC4R* and *FTO*) and minor allele homozygous + heterozygous vs. major allele homozygous (*MRPS22*) with 95% CI. BMI 95% CI not reported for SNPs at *FTO*, *MTCH2* and *MC4R* genes in Melka et al. Sample size corresponds to n = number of participants / number of observations with available trait (BMI or SBP) data. P values were obtained from Student's T test. **Bold** p value signifies a statistically significant ($P < 0.05$) difference in BMI or SBP. BMI regressions are adjusted for age at baseline, sex and data collection cycle. SBP regressions are adjusted for age at baseline, sex, height and data collection cycle.

Results for sensitivity analyses are available in the Supplementary materials. Overall, these analyses revealed that our results are robust to the genotype coding model, the use of BMI compared to BMI Z-score as the adiposity measure, the use of young adults in our analytical sample, and to potential confounding by population stratification (Supplementary Tables 2-5 and Supplementary Figure 1).

DISCUSSION

The common variants/multiple disease hypothesis posits that shared genetic risk factors play a role in the strong correlation between related traits, such as BMI and blood pressure (5). Melka et al. (6) reported associations of novel SNPs rs16933812 (*PAX5*) and rs7638110 (*MRPS22*), and well-established adiposity SNPs rs9930333 (*FTO*), rs7120548 (*MTCH2*) and rs17773430 (*MC4R*), with BMI in

adolescents. Three of these (*FTO*, *MRPS22*, *PAX5*) were also associated with SBP in the same sample. We attempted to replicate these findings in NDIT, which includes an adolescent sample similar to the sample studied in Melka et al. We were able to replicate associations of SNPs on genes *FTO* and *MTCH2* with BMI, but not those of SNPs on *PAX5*, *MRPS22* and *MC4R*. Additionally, none of the three associations with SBP were replicated. Overall, these results do not support the hypothesis that these variants contribute to the correlation between BMI and SBP levels.

The successful replication of the association between rs9930333 (*FTO*) and BMI in NDIT adolescents is consistent with multiple studies reporting SNPs on *FTO* as susceptibility SNPs for increased BMI (30, 31), obesity (10, 32, 33), and higher BMIz (34) in children and adolescents. In contrast, although SNPs on *MTCH2* were among the first to be associated by GWAS with BMI among adults (11), many studies of children and adolescents did not show associations between these SNPs and BMI (30, 32, 35, 36). Replication of the association between rs7120548 (*MTCH2*) and BMI in NDIT adolescents goes against those studies, suggesting that the effect of SNPs on *MTCH2* on BMI may begin to manifest early in life.

We failed to replicate associations previously detected in Melka et al. between rs16933812 (*PAX5*), rs7638110 (*MRPS22*) and rs17773430 (*MC4R*) and BMI, and between rs16933812 (*PAX5*), rs7638110 (*MRPS22*) and rs9930333 (*FTO*) and SBP. Several possible explanations could underpin this lack of replication. First, differences in LD pattern between populations can result in inconsistencies in genetic effects, since most SNPs detected in genetic association studies are not causal for the trait investigated but are in high LD with a causal SNP. Although both samples were of European ancestry, participants in the Melka et al. study comprised adolescents from a founder population in Saguenay-Lac-St-Jean (37), in contrast to the admixed NDIT sample in Montréal. Founder populations usually show a higher degree of LD than admixed populations (38). In Melka et al., rs17773430 (*MC4R*) was in high LD ($D' > 0.9$) with rs17782313 (*MC4R*), the original SNP previously reported in a GWAS for BMI (27). However, LD measures between these two SNPs was much lower in the 1000 Genomes Project Phase 3 ($D' = 0.31$ using the Utah residents with Northern and Western European ancestry (CEU)) (24), which closely resembled the LD pattern between rs17782313 and its proxy in NDIT. This suggests the difference in LD observed between Melka et al. and NDIT likely reflects a population-specific LD pattern in the Saguenay-Lac-St-Jean founder population, rather than the use of a proxy in our study. We acknowledge that finding a proxy SNP in LD with the original GWAS SNP in NDIT sample may have led to replication of the effect with BMI.

Second, differences in blood pressure measurement techniques may contribute to failed replication of the associations between rs16933812 (*PAX5*), rs7638110 (*MRPS22*) and rs9930333 (*FTO*) and SBP. Average blood pressure in Melka et al. was measured in a 52-minute protocol with participants supine, standing, sitting, under mental stress and during mental stress recovery (37). In NDIT blood pressure was measured in sitting position only according to clinical guidelines (39). Short-term variation in blood pressure in response to position change or mental effort is regulated by multiple factors including sympathetic outflow, cardiopulmonary reflexes, blood vessel elasticity, blood viscosity and hormones, and the genetic architecture of this variation still remains mostly unknown (40). Inconsistencies in associations with SBP across studies may reflect heterogeneity of blood pressure sub-phenotypes related to differences in measurement.

However, because they were initially identified by GWAS in only 598 individuals in Melka et al. and never replicated elsewhere, non-replication of the *PAX5* and *MRPS22*-BMI associations most likely stems from the overestimation of genetic effects in Melka et al. (Winner's Curse phenomenon). Coefficients estimated in Melka et al. for the effect of novel SNPs rs16933812 (*PAX5*) and rs7638110 (*MRPS22*) on BMI were higher than those for rs9930333 (*FTO*), rs7120548 (*MTCH2*) and rs17773430 (*MC4R*). However, the newly identified SNPs on *PAX5* and *MRPS22* should have shown weaker estimated effects than those on *FTO*, *MTCH2* and *MC4R* which are among the first genes associated by GWAS with BMI that have been consistently replicated since (11). This illustrates the risk of publishing GWAS results without an internal replication to verify the accuracy effect estimates and sparse through false positive findings (41). Indeed, when accounting for the Winner's Curse, most results from GWAS internally replicate in the same continental ancestry as anticipated per their expected power to replicate (41). Acknowledging the scarcity of available genetic data from youth which hinders the possibility to systematically assess the replicability of genetic associations internally, external replication by independent investigators are essential in order to confirm the validity of previously reported genetic findings in children and adolescents.

Limitations to this study pertains to a lack of power to detect effect for rs7638110 (*MRPS22*) on SBP. Because of a low MAF (0.07), the power to detect an effect of 0.5 mm Hg per minor allele was below 70%. Also, as NDIT is a longitudinal investigation, a portion of participants are lost to follow-up. Although retention in the study was good (84% of baseline participants were still available at study cycle 22), differences in participants remaining and lost to follow-up may still induce a selection bias.

CONCLUSION

We replicated associations observed in an earlier study of Quebec adolescents of European ancestry between SNPs in the well-established adiposity genes *FTO* and *MTCH2* and BMI. However, we could not replicate SNPs at genes *PAX5*, *MRPS22* or *MC4R*. None of the SNPs showed an association with SBP. Our findings illustrate the difficulty in replicating genetic associations and provide examples of overestimation of genetic effects initially identified related to small sample size and differences in LD structure between founder and admixed populations in Quebec, Canada.

MANUSCRIPT REFERENCES

1. Forouzanfar MH, Afshin A, Alexander LT, Anderson HR, Bhutta ZA, Biryukov S, et al. Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015. *The Lancet*. 2016;388(10053):1659-724.
2. Man T, Nolte IM, Jaju D, Al-Anqoudi ZAM, Muñoz ML, Hassan MO, et al. Heritability and genetic correlations of obesity indices with ambulatory and office beat-to-beat blood pressure in the Oman Family Study. *Journal of Hypertension*. 2020;38(8):1474-80.
3. Jayedi A, Rashidy-Pour A, Khorshidi M, Shab-Bidar S. Body mass index, abdominal adiposity, weight gain and risk of developing hypertension: a systematic review and dose–response meta-analysis of more than 2.3 million participants. *Obesity Reviews*. 2018;19(5):654-67.
4. Chorin E, Hassidim A, Hartal M, Havakuk O, Flint N, Ziv-Baran T, et al. Trends in Adolescents Obesity and the Association between BMI and Blood Pressure: A Cross-Sectional Study in 714,922 Healthy Teenagers. *American Journal of Hypertension*. 2015;28(9):1157-63.
5. Becker KG. The common variants/multiple disease hypothesis of common complex genetic disorders. *Medical Hypotheses*. 2004;62(2):309-17.
6. Melka MG, Bernard M, Mahboubi A, Abrahamowicz M, Paterson AD, Syme C, et al. Genome-Wide Scan for Loci of Adolescent Obesity and Their Relationship with Blood Pressure. *The Journal of Clinical Endocrinology & Metabolism*. 2012;97(1):E145-E50.
7. Wang X, Xu X, Su S, Snieder H. Familial aggregation and childhood blood pressure. *Current hypertension reports*. 2015;17(1):509-.
8. Kupper N, Ge D, Treiber Frank A, Snieder H. Emergence of Novel Genetic Effects on Blood Pressure and Hemodynamics in Adolescence. *Hypertension*. 2006;47(5):948-54.
9. Parmar Priyakumari G, Taal HR, Timpson Nicholas J, Thiering E, Lehtimäki T, Marinelli M, et al. International Genome-Wide Association Study Consortium Identifies Novel Loci Associated With Blood Pressure in Children and Adolescents. *Circulation: Cardiovascular Genetics*. 2016;9(3):266-78.
10. Bradfield JP, Taal HR, Timpson NJ, Scherag A, Lecoeur C, Warrington NM, et al. A genome-wide association meta-analysis identifies new childhood obesity loci. *Nature Genetics*. 2012;44(5):526-31.
11. Albuquerque D, Stice E, Rodríguez-López R, Manco L, Nóbrega C. Current review of genetics of human obesity: from molecular mechanisms to an evolutionary perspective. *Molecular Genetics and Genomics*. 2015;290(4):1191-221.
12. Ahn S-Y, Gupta C. Genetic Programming of Hypertension. *Frontiers in pediatrics*. 2018;5:285-.
13. Kraft P. Commentary: Curses - Winner's and Otherwise - in Genetic Epidemiology. *Epidemiology*. 2008;19(5):649-51.
14. Poirier JG, Faye LL, Dimitromanolakis A, Paterson AD, Sun L, Bull SB. Resampling to Address the Winner's Curse in Genetic Association Analysis of Time to Event. *Genetic Epidemiology*. 2015;39(7):518-28.
15. Thaler RH. Anomalies: The Winner's Curse. *Journal of Economic Perspectives*. 1988;2(1):191-202.
16. Bazerman MH, Samuelson WF. I Won the Auction But Don't Want the Prize. *Journal of Conflict Resolution*. 1983;27(4):618-34.
17. Lohmueller KE, Pearce CL, Pike M, Lander ES, Hirschhorn JN. Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. *Nature Genetics*. 2003;33(2):177-82.
18. Göring HH, Terwilliger JD, Blangero J. Large upward bias in estimation of locus-specific effects from genomewide scans. *American journal of human genetics*. 2001;69(6):1357-69.

19. Faye LL, Sun L, Dimitromanolakis A, Bull SB. A flexible genome-wide bootstrap method that accounts for ranking and threshold-selection bias in GWAS interpretation and replication study design. *Statistics in Medicine*. 2011;30(15):1898-912.
20. O'Loughlin J, Dugas EN, Brunet J, DiFranza J, Engert JC, Gervais A, et al. Cohort Profile: The Nicotine Dependence in Teens (NDIT) Study. *International Journal of Epidemiology*. 2014;44(5):1537-46.
21. Evers SE, Hooper MD. Dietary intake and anthropometric status of 7 to 9 year old children in economically disadvantaged communities in Ontario. *J Am Coll Nutr*. 1995;14(6):595-603.
22. Das S, Forer L, Schönherr S, Sidore C, Locke AE, Kwong A, et al. Next-generation genotype imputation service and methods. *Nature Genetics*. 2016;48(10):1284-7.
23. Auton A, Abecasis GR, Altshuler DM, Durbin RM, Abecasis GR, Bentley DR, et al. A global reference for human genetic variation. *Nature*. 2015;526(7571):68-74.
24. Machiela MJ, Chanock SJ. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics (Oxford, England)*. 2015;31(21):3555-7.
25. Willer CJ, Speliotes EK, Loos RJJ, Li S, Lindgren CM, Heid IM, et al. Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. *Nature Genetics*. 2009;41(1):25-34.
26. Speliotes EK, Willer CJ, Berndt SI, Monda KL, Thorleifsson G, Jackson AU, et al. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nature Genetics*. 2010;42(11):937-48.
27. Loos RJJ, Lindgren CM, Li S, Wheeler E, Zhao JH, Prokopenko I, et al. Common variants near MC4R are associated with fat mass, weight and risk of obesity. *Nature Genetics*. 2008;40(6):768-75.
28. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics (Oxford, England)*. 2004;21(2):263-5.
29. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2019.
30. den Hoed M, Ekelund U, Brage S, Grøntved A, Zhao JH, Sharp SJ, et al. Genetic susceptibility to obesity and related traits in childhood and adolescence: influence of loci identified by genome-wide association studies. *Diabetes*. 2010;59(11):2980-8.
31. Felix JF, Bradfield JP, Monnereau C, van der Valk RJP, Stergiakouli E, Chesni A, et al. Genome-wide association analysis identifies three new susceptibility loci for childhood body mass index. *Human molecular genetics*. 2016;25(2):389-403.
32. Zhao J, Bradfield JP, Zhang H, Sleiman PM, Kim CE, Glessner JT, et al. Role of BMI-Associated Loci Identified in GWAS Meta-Analyses in the Context of Common Childhood Obesity in European Americans. *Obesity*. 2011;19(12):2436-9.
33. da Silva TER, Andrade NL, Cunha DdO, Leão-Cordeiro JAB, Vilanova-Costa CAST, Silva AMTC. The FTO rs9939609 polymorphism and obesity risk in teens: Evidence-based meta-analysis. *Obesity Research & Clinical Practice*. 2018;12(5):432-7.
34. Namjou B, Keddache M, Marsolo K, Wagner M, Lingren T, Cobb B, et al. EMR-linked GWAS study: investigation of variation landscape of loci for body mass index in children. *Frontiers in genetics*. 2013;4:268-.
35. Mei H, Chen W, Jiang F, He J, Srinivasan S, Smith EN, et al. Longitudinal replication studies of GWAS risk SNPs influencing body mass index over the course of childhood and adulthood. *PLoS One*. 2012;7(2):e31470-e.
36. Wang H-J, Hinney A, Song J-Y, Scherag A, Meng X-R, Grallert H, et al. Association of common variants identified by recent genome-wide association studies with obesity in Chinese children: a case-control study. *BMC medical genetics*. 2016;17(1):7.

37. Pausova Z, Paus T, Abrahamowicz M, Bernard M, Gaudet D, Leonard G, et al. Cohort Profile: The Saguenay Youth Study (SYS). *International Journal of Epidemiology*. 2017;46(2):e19-e.
38. Service S, DeYoung J, Karayiorgou M, Roos JL, Pretorius H, Bedoya G, et al. Magnitude and distribution of linkage disequilibrium in population isolates and implications for genome-wide association studies. *Nature Genetics*. 2006;38(5):556-60.
39. O'Loughlin J, Dugas EN, Brunet J, DiFranza J, Engert JC, Gervais A, et al. Cohort Profile: The Nicotine Dependence in Teens (NDIT) Study. *International Journal of Epidemiology*. 2015;44(5):1537-46.
40. Parati G, Ochoa JE, Lombardi C, Bilo G. Assessment and management of blood-pressure variability. *Nature Reviews Cardiology*. 2013;10(3):143-55.
41. Palmer C, Pe'er I. Statistical correction of the Winner's Curse explains replication variability in quantitative trait genome-wide association studies. *PLOS Genetics*. 2017;13(7):e1006916-e.

SUPPLEMENTARY MATERIAL

Supplementary text 1: Blood pressure and body-mass index measurement

Anthropometric and blood pressure measurements were obtained at study cycles 1, 12, 19 and 22 where participants were respectively 13, 15, 17 and 24 years old on average. Trained technicians measured height and weight (Seca Portable Stadiometer – Model 214 and Seca Scale – Model 761, Seca Corporation, Columbia, MD, USA) according to a standardized protocol (1). Two measures of height and weight were taken at each collection point. BMI was then calculated as weight (kg) divided by height squared (m^2). A third measure was obtained if discrepancies of 0.2 kg and 0.5 cm was observed and BMI was then computed as the mean of the two closest measures.

Blood pressure was assessed three times with an oscillometer (Dinamap XL, modèle CR9340, Critikon Co, Tampa, Fla) calibrated to a mercury sphygmomanometer at 1-min interval, at the right arm, sitting and after a five-minute rest period. SBP was computed as the average of the second and third SBP measurements. When discrepancies of 20 mm Hg SBP or 10 mm Hg DBP were observed, fourth and fifth measures were taken and SBP was computed as the average of the two closest measures.

Supplementary text 2: Genotyping and genetic quality control procedures

Genotypes for SNPs of interest were obtained by a genome-wide screening performed on samples from 943 NDIT participants by Genome Québec in 2017 using an Illumina Global Screening Array-24 v1.0 (GSA). Genetic imputations were performed using Minimac3 software (2) using the 1000 Genome Phase 3 reference panel (3). A total of 40 participants had to be removed because of DNA samples of poor quality or sex inconsistencies. From the remaining 903 participants, 868 passed quality control based on a sample call rate threshold of 98%. 726 of those were identified to be of European ancestry based on principal components analysis. Finally, 9 first degree relatives were excluded, bringing the available sample for our study to 717 participants with genotype data.

Table S1: Comparison of baseline characteristics of participants remaining and lost to follow-up, NDIT 1999[†]

Variable	Remaining (n=601)	Attrited (n=112)	p value
Age, mean (SD)	12.7 (0.4)	12.8 (0.6)	0.001
Female, %	55.9	42.0	0.009
Mother university-educated ¹ , %	47.2	41.2	0.357
Single family, %	6.2	6.8	1.000
Alcohol consumption ² , %	44.1	52.9	0.121
Depression symptoms, mean (SD)	2.1 (0.6)	2.1 (0.7)	0.946
Ever smoked, %	27.6	37.3	0.062
No cigarettes/month, med (IQR)	0.2 (2.5)	0.2 (0.4)	0.446
Height, mean (SD)	156.1 (7.5)	156.8 (8.8)	0.407
BMI, mean (SD)	20.0 (3.8)	21.2 (4.0)	0.003
SBP, mean (SD)	105.2 (9.8)	106.2 (11.6)	0.373

[†] Mother university-educated: measured in parental questionnaires when participants were 15-16 years old. Alcohol consumption: any alcohol consumption in the past three months. For categorical variables differences between included and excluded participants were assessed using chi-square. Differences in means of normally distributed continuous variables were assessed with ANOVA. Differences in means of non-normally distributed continuous variables (no cigarettes/month) were assessed with Wilcoxon-test. **Bold** p value signifies a statistically significant ($P < 0.05$) difference.

Table S2: Estimated effect of the five SNPs on BMI and SBP under an additive model, NDIT 1999-2012[†]

Description of SNP				BMI (n=713/2350)	SBP (n=713/2347)
Gene	rsID	Minor allele	MAF	\hat{B} [CI]	\hat{B} [CI]
FTO	rs9930333	G	0.43	0.61 [0.24, 0.99]	0.41 [-0.43, 1.26]
PAX5	rs16933812	G	0.37	-0.01 [-0.40, 0.38]	-0.29 [-1.17, 0.58]
MTCH2	rs7120548	C	0.35	-0.44 [-0.84, -0.05]	0.12 [-0.77, 1.00]
MC4R	rs111638368	T	0.29	0.36 [-0.05, 0.77]	0.12 [-0.80, 1.04]
MRPS22	rs7638110	T	0.07	-0.28 [-0.99, 0.43]	-1.52 [-3.10, 0.07]

[†] Estimated effects correspond to the additive effect of one minor allele increase on BMI and SBP levels with 95% CI. **Bold CI** signifies a statistically significant ($P < 0.05$) effect on BMI or SBP.

Table S3: Estimated effect of rs111638368 (*MC4R*) on SBP by time of measurement, NDIT 1999-2012[†]

	Cycle 1	Cycle 12	Cycle 19	Cycle 22
\hat{B} [IC]	0.81 [-0.29 , 1.90]	-0.47 [-1.67 , 0.73]	-1.00 [-2.35 , 0.36]	0.03 [-1.24 , 1.29]

[†] Estimated effects correspond to the additive effect of one minor allele increase on SBP levels with 95% CI. **Bold CI** signifies a statistically significant ($P < 0.05$) effect on BMIz, BMI or SBP.

Table S4: Estimated effect of the five SNPs investigated on BMIz, BMI and SBP excluding cycle 22 data, NDIT 1999-2005[†]

SNPs information			BMIz (n=704/1793)	BMI (n=704/1793)	SBP (n=704/1791)
Gene	rsID	Minor allele	\hat{B} [IC] ^a	\hat{B} [IC] ^a	\hat{B} [IC] ^a
PAX5	rs16933812	G	0.00 [-0.10,0.10]	-0.03 [-0.43,0.36]	-0.14 [-1.07,0.78]
MRPS22	rs7638110	T	-0.09 [-0.26,0.08]	-0.29 [-1.00,0.42]	-1.58 [-3.25,0.09]
MC4R	rs111638368	T	0.03 [-0.07,0.13]	0.37 [-0.04,0.78]	0.08 [-0.89,1.06]
MTCH2	rs7120548	C	-0.10 [-0.19,0.00]	-0.44 [-0.84,-0.04]	0.10 [-0.84,1.04]
FTO	rs9930333	G	0.13 [0.04,0.22]	0.63 [0.25,1.00]	0.50 [-0.39,1.39]

[†] Estimated effects correspond to the additive effect of one minor allele increase on BMIz, BMI and SBP levels with 95% confidence interval. **Bold CI** signifies a statistically significant ($P < 0.05$) effect on BMI or SBP.

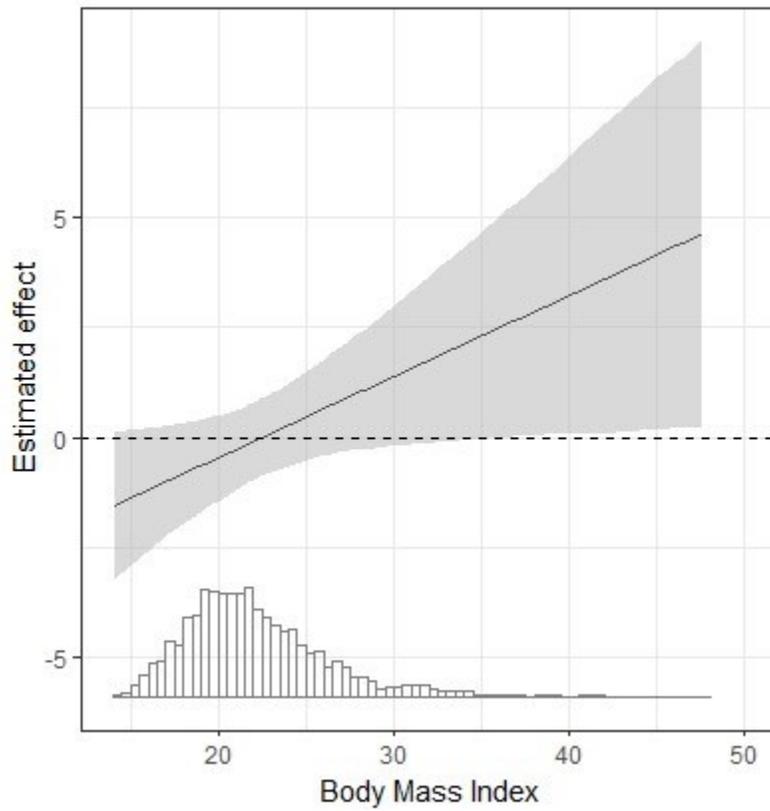
Table S5: Association analysis of the five SNPs investigated with BMI and SBP, adjusted for first five principal components, NDIT 1999-2012

SNPs information			BMI (n=713/2350)		SBP (n=713/2347)	
Gene	rsID	Minor allele	PC-adj \hat{B} [IC] ^a	Non-adj \hat{B} [IC] ^b	PC-adj \hat{B} [IC] ^a	Non-adj \hat{B} [IC] ^b
PAX5	rs16933812	G	0.01 [-0.39 , 0.40]	-0.01 [-0.40 , 0.38]	-0.31 [-1.20 , 0.57]	-0.29 [-1.17 , 0.58]
MRPS22	rs7638110	T	-0.31 [-1.02 , 0.41]	-0.28 [-0.99 , 0.43]	-1.67 [-3.26 , -0.09]	-1.52 [-3.10 , 0.07]
MC4R	rs111638368	T	0.35 [-0.06 , 0.76]	0.36 [-0.05 , 0.77]	0.12 [-0.80 , 1.04]	0.12 [-0.80 , 1.04]
MTCH2	rs7120548	C	-0.46 [-0.86 , -0.07]	-0.44 [-0.84 , -0.05]	0.19 [-0.70 , 1.08]	0.12 [-0.77 , 1.00]
FTO	rs9930333	G	0.62 [0.25 , 0.99]	0.61 [0.24 , 0.99]	0.37 [-0.47 , 1.21]	0.41 [-0.43 , 1.26]

a Models Adjusted for first five principal components. Estimated effects correspond to the additive effect of one minor allele increase on BMI and SBP levels with 95% confidence interval. **Bold CI** signifies a statistically significant ($P < 0.05$) effect on BMI or SBP.

b Models not adjusted for first five principal components. Estimated effects correspond to the additive effect of one minor allele increase on BMI and SBP levels with 95% confidence interval. **Bold CI** signifies a statistically significant ($P < 0.05$) effect on BMI or SBP.

Figure S1: Additive estimated effect of rs111638368 (*MC4R*) on SBP by BMI level, NDIT 1999-2012



SUPPLEMENTAL REFERENCES

1. Evers SE, Hooper MD. Dietary intake and anthropometric status of 7 to 9 year old children in economically disadvantaged communities in Ontario. *J Am Coll Nutr.* 1995;14(6):595-603.
2. Das S, Forer L, Schönherr S, Sidore C, Locke AE, Kwong A, et al. Next-generation genotype imputation service and methods. *Nature Genetics.* 2016;48(10):1284-7.
3. Auton A, Abecasis GR, Altshuler DM, Durbin RM, Abecasis GR, Bentley DR, et al. A global reference for human genetic variation. *Nature.* 2015;526(7571):68-74.

5.3 Résultats supplémentaires

Cette section comprend des résultats additionnels qui ne sont pas incorporés au manuscrit, mais restent pertinents au mémoire. On présente la comparaison des échantillons analytiques de l'étude NDIT et de celui de la SYS, ainsi que le processus de sélection du modèle sans variants génétiques pour modéliser l'effet du temps sur l'IMC et la PAS. Les analyses de sensibilité étant présentées dans le matériel supplémentaire du manuscrit, celles-ci sont décrites en plus de détails ci-bas.

5.3.1 Comparaison de l'échantillon analytique NDIT avec celui de la SYS

La comparaison directe des participants est compliquée par leurs devis différents. En effet, NDIT est une étude longitudinale et suit des adolescents qui avaient 12-13 ans au début de l'étude jusqu'à l'âge de 24 ans en moyenne au cycle 22, tandis que la SYS comprend des données provenant de l'étude transversale SYS et inclut des participants âgés de 12 à 18 ans. Il est toutefois possible de tirer des conclusions générales en colligeant les données de toutes les observations de NDIT et en tenant compte de la présence de mesures répétées sur les mêmes participants.

Le Tableau 6 présente une description de la moyenne et de l'erreur-type des caractéristiques des participants de l'échantillon analytique de NDIT et de celle de la SYS. En somme, les participants de NDIT et de la SYS sont semblables dans la proportion de femmes et pour l'IMC moyen. On remarque que les moyennes de PAS et de PAD sont plus élevées chez les participants provenant du SYS comparativement à ceux de NDIT. Cette différence d'environ 15 mm Hg pour la PAS peut s'expliquer par les méthodes de mesure de pression artérielle utilisées dans les deux études. En rappel, la PAS des participants du SYS était mesurée à l'aide d'un protocole de 52 minutes comportant cinq sections distinctes où les participants étaient successivement couchés, debout, assis, en effort mental et en récupération d'effort mental. La pression artérielle des participants du SYS rapportée au Tableau 6 correspond à la moyenne des mesures effectuées en position assise, mais faisant également partie du protocole complet dans des contextes différents, alors que les participants de NDIT avaient leur pression artérielle mesurée uniquement au repos. Ainsi, il est possible que les niveaux de pression artérielle observés dans l'échantillon utilisé par la SYS soient plus élevés en raison du contexte de la prise de mesure et non en raison à une différence relative aux populations étudiées.

Tableau 6. Description des observations provenant de NDIIT et du SYS, combinaison de tous les cycles NDIIT

	NDIT	n	SYS	n
Âge, moyenne (Erreur-type)	16,9 (0,2)	2428	15,1 (0,1)	590
Femme, %	53,7	2852	52,3	598
Taille, moyenne (Erreur-type)	165,2 (0,6)	2350	163,3 (0,4)	589
IMC, moyenne (Erreur-type)	22,2 (0,3)	2350	21,3 (0,2)	586
PAS, moyenne (Erreur-type)	107,9 (0,7)	2349	122,6 (0,6)	497
PAD, moyenne (Erreur-type)	58,6 (0,4)	2349	77,8 (0,4)	484

Note : n représente le nombre d'observations disponibles peu importe le cycle de mesure.

5.3.2 Choix du modèle pour l'effet du temps sur l'IMC et la PAS

Le modèle de base utilisé pour représenter l'effet du temps sur l'IMC et la PAS est un modèle linéaire mixte avec l'âge au début de l'étude, le sexe et la taille (PAS uniquement) comme covariables et suivant une structure de corrélation par symétrie composée. Le Tableau 7 montre l'AIC associé aux différentes spécifications de modèles de base pour modéliser les évolutions de l'IMC et de la PAS dans le temps. L'AIC permet de sélectionner les modèles optimaux pour l'IMC et la PAS, dans lesquels seront ensuite inclus les variants d'intérêt. Le modèle minimisant l'AIC est un modèle linéaire mixte avec pente aléatoire pour le cycle de mesure permettant de tenir compte de la corrélation entre les mesures sur les mêmes individus et d'un effet différent du cycle de mesure par individus, suivant une structure de corrélation par symétrie composée et modélisant le temps écoulé depuis le cycle 1 comme une variable catégorique. Pour un AIC identique à une modélisation par lissage, la modélisation catégorique du temps écoulé depuis le cycle 1 a été choisie pour faciliter l'interprétation des résultats.

Tableau 7. Choix du modèle pour l'effet du temps sur l'IMC et la PAS

	AIC (IMC)	AIC (PAS)
Modèle de base ^a	11 276	16 954
Pente aléatoire Temps ^b	11 070	16 950
Temps polynomial	10 850	16 937
Temps catégorique	10 842	16 875
Temps lissé	10 842	16 875
Hétérogénéité de la variance	10 848	16 881

a Modèle de base : modèle linéaire mixte incluant le nombre de mois écoulés depuis le cycle 1, l'âge au cycle 1 et le sexe comme covariables et suivant une structure de corrélation par symétrie composée.

b Temps : nombre de mois écoulés depuis le cycle 1

5.3.3 Analyses de sensibilité

Codification additive du génotype : Comme montré au [Chapitre 5](#) (Table S2), les résultats des modèles avec codification additive s'alignent à ceux obtenus sous un codage génotypique. Les variants sur les gènes *FTO* et *MTCH2* restent les deux seuls variants associés avec l'IMC. L'addition d'un allèle mineur de rs9930333 (*FTO*) est associée à une augmentation d'IMC de 0,61 (95% IC 0,24 ; 0,99) kg/m². L'addition d'un allèle mineur de rs7120548 (*MTCH2*) est associée à une diminution d'IMC de 0,44 (95% IC 0,05 ; 0,84) kg/m². Puisque le modèle avec codage additif montrait un AIC inférieur au modèle avec codage génotypique, ce modèle était préféré pour le reste des analyses de sensibilité.

Interaction variant X Temps : Une interaction statistiquement significative entre le variant rs111638368 (*MC4R*) et le temps a été identifiée pour la PAS. Cependant, l'intervalle de confiance de l'effet du variant sur la PAS inclus 0 à chaque temps, suggérant que le variant n'affecte pas la PAS de l'adolescence au début de l'âge adulte ([Chapitre 5](#), Table S3).

Interaction variant X IMC pour les modèles de PAS : Puisque des études précédentes (165, 166) suggéraient que l'adiposité puisse modifier l'effet de variants génétiques sur la pression artérielle, un terme d'interaction entre les variants et l'IMC a été considéré pour les associations avec la PAS. L'estimation du coefficient pour le terme d'interaction entre rs111638368 (*MC4R*) et l'IMC était statistiquement significatif. La Figure S1 au [Chapitre 5](#) montre une augmentation de l'effet estimé du variant sur la PAS conditionnellement à l'IMC. Cependant, l'intervalle de confiance de l'effet du variant (en gris) inclue 0 pour toutes les valeurs d'IMC sous 35 kg/m² qui correspond à seulement 25 individus (3,5%) de l'échantillon analytique. Ainsi, le variant sur

MC4R montre seulement un effet significatif sur la PAS pour un intervalle étroit d'IMC peu supporté par les données disponibles. On note également que ces analyses d'interaction devraient être considérées de nature exploratoire considérant que les analyses d'interactions nécessitent une taille d'échantillon supérieure à celle disponible. La représentation graphique de cette interaction a été effectuée à l'aide du module *interplot* (227) du logiciel R.

IMCZ comme mesure de l'adiposité : L'utilisation de l'IMCZ comparativement à l'IMC comme variable représentant l'adiposité n'a pas modifié les résultats, suggérant que nos analyses sont robustes à l'utilisation de l'IMC comme mesure de l'adiposité chez les adolescents ([Chapitre 5](#), Table S4).

Restriction des analyses aux adolescents : Afin d'examiner si l'inclusion de jeunes adultes (cycle 22) dans nos analyses affecterait la réplication des résultats de la SYS qui ne comportait que des adolescents, les analyses ont été réduites aux cycles 1, 12 et 19. Les associations étaient similaires pour tous les variants et les traits examinés, suggérant que les analyses sont robustes à l'inclusion de jeunes adultes dans nos données ([Chapitre 5](#), Table S4).

Analyse par composantes principales : Afin d'évaluer le potentiel de biais de confusion en raison de la présence de sous-populations dans l'échantillon analytique d'ascendance européenne de NDIT, les modèles d'IMC et de PAS ont été ajustés pour les cinq premières composantes principales. Les effets estimés ajustés et non ajustés pour composantes principales étaient similaires ([Chapitre 5](#), Table S5). On note que l'effet du variant rs7638110 (*MRPS22*) sur la PAS était statistiquement significatif dans le modèle ajusté. Cependant, la direction de l'effet reste opposée à celle observée dans la SYS.

CHAPITRE 6 – DISCUSSION

Ce mémoire avait comme objectif de répliquer les résultats d'une étude préalable effectuée par la SYS qui avait identifié cinq variants génétiques associés à l'adiposité chez des adolescents d'ascendance européenne âgés de 12 à 18 ans de la région du Saguenay-Lac-Saint-Jean (2003-2009), dont trois étaient conjointement associés à l'adiposité et à la pression artérielle. La réplification est une étape essentielle au processus scientifique, particulièrement dans l'identification d'associations génétiques où de nombreux résultats de nature exploratoire ne sont pas confirmés par les études ultérieures. Ainsi, le mémoire prend avantage de données longitudinales disponibles pour un échantillon d'adolescents semblable en temps (1999-2012), en âge (suivi de 12 à 24 ans) et en géographie (Montréal) à celui utilisé par la SYS afin de vérifier l'identification de variants génétiques spécifiques associés à l'adiposité et à la pression artérielle. L'utilisation de données provenant d'adolescents permet de mieux isoler l'effet de variants génétiques sur l'adiposité et la PAS de l'influence de facteurs comportementaux et environnementaux qui apparaissent sur le long terme. Également, la similarité des deux échantillons permet de réduire le risque que des différences entre les études induisent l'inconsistance de résultats.

À l'aide d'une GWAS, la SYS a découvert deux nouveaux variants génétiques associés à l'IMC (rs16933812, *PAX5*; rs7638110, *MRPS22*) ainsi que trois variants (rs9930333, *FTO*; rs7120548, *MTCH2*; rs17773430, *MC4R*) préalablement rapportés dans la littérature. Trois de ces variants (rs9930333, *FTO*; rs16933812, *PAX5*; rs7638110, *MRPS22*) étaient également associés à la PAS dans le même échantillon. En utilisant les données provenant de l'étude NDIT, les associations des variants sur les gènes *FTO* et *MTCH2* avec l'IMC sont répliquées, alors que celles pour les variants sur les gènes *MC4R*, *PAX5* et *MRPS22* ne le sont pas. Aucune des trois associations génétiques avec la PAS de la SYS n'a été répliquée dans NDIT. Ainsi, les présents résultats ne s'alignent pas avec les résultats de la SYS qui avaient observés des effets conjoints sur l'IMC et la PAS. La discussion présente d'abord une revue des principaux résultats observés et leur interprétation spécifique dans le contexte des connaissances relatives aux variants génétiques examinés. On discute ensuite d'hypothèses permettant d'expliquer les différences de résultats observées entre la SYS et NDIT, pour terminer avec les limites et les contributions de ce travail.

6.1 Interprétation des résultats

La revue des résultats est effectuée en examinant dans un premier temps les trois variants sur les gènes *FTO*, *MTCH2* et *MC4R* déjà connus de la littérature pour leur association avec l'IMC, puis les deux nouveaux variants sur *PAX5* et *MRPS22* qui ont été associés pour la première fois à l'IMC dans la SYS.

6.1.1 Variants dont l'association avec l'IMC est connue

FTO : L'association entre le variant rs9930333 (*FTO*) et l'IMC observée dans l'étude SYS a été répliquée dans l'étude NDIT. Spécifiquement, dans NDIT, la différence moyenne d'IMC entre les individus ayant deux allèles mineurs comparativement à deux allèles majeurs du variant est de 1,18 (IC95% 0,42 ; 1,94) kg/m². Cet effet est dans la même direction que celui estimé dans la SYS ($\beta_{FTO-SYS} = 1,8 \text{ kg/m}^2$; $p = 1,9 \times 10^{-4}$), quoiqu'atténué. Les intervalles de confiance pour l'estimation des coefficients des variants sur *FTO*, *MTCH2* et *MC4R* pour l'IMC n'étant pas rapportés dans l'étude SYS, il est impossible d'évaluer le chevauchement des intervalles de confiance entre les deux études pour évaluer s'ils diffèrent. Le variant rs9930333 fait partie d'une région sur les deux premiers introns du gène *FTO* associé à l'adiposité de façon robuste dans la littérature, tant chez les adultes (35) que chez les jeunes (137). Bien que le fonctionnement biologique derrière l'influence de *FTO* sur l'adiposité ne soit pas totalement élucidé, des résultats récents (184, 187) suggèrent une implication dans la dérégulation de la thermogénèse des cellules adipeuses.

L'étude SYS a également observé une association entre le variant rs9930333 et la PAS ($\beta_{FTO-SYS} = 3,63 \text{ mm Hg}$, IC95% 0,73 ; 6,53) que l'étude NDIT n'a pas répliquée ($\beta_{FTO-NDIT} = 0,64 \text{ mm Hg}$, IC95% -1,08 ; 2,35). Nos résultats s'alignent avec la littérature disponible concernant les études GWAS ayant examiné la PAS. Un examen de la base de données d'analyses pangénomiques du NHGRI-EBI GWAS Catalog (183) indique que la grande majorité des GWAS pour la pression artérielle n'ont pas détecté d'association avec *FTO*. Par exemple, l'analyse d'Evangelou et al. de 2018 (67) combinant plus d'un million de participants correspondant au plus grand GWAS de pression artérielle à ce jour, n'a détecté aucun variant sur *FTO* associé avec la PAS.

Des résultats discordants ont cependant été publiés. En effet, trois GWAS de plus petite taille d'échantillon ont identifié des variants sur *FTO* associés à la PAS. Une première étude suggère que le variant rs11642015 (*FTO*) est associé à la PAS dans un GWAS d'individus d'ascendance est-asiatique (β (SE) = 0,58 (0,08) mm Hg, $p = 1,9 \times 10^{-12}$) (77). Les deux autres études sont des études pangénomiques d'interactions entre gène et environnement. Une de ces études suggère que le même variant est associé à la PAS considérant une interaction avec le statut de fumeur dans un GWAS d'individus de multiples ascendances (β (SE) = 0,35 (0,04) mm Hg, $p_{\text{Joint}} = 9,91 \times 10^{-21}$) (75), alors que l'autre observe que le variant rs55872725 (*FTO*) est associé à la PAS en interaction avec la consommation d'alcool dans une GWAS d'individus d'ascendance européenne ($\beta = 0,69$ mm Hg, $p = 3,39 \times 10^{-9}$) (68). D'autres études devront confirmer si ces résultats sont spécifiques, que ce soit pour l'ascendance ou l'interaction avec le tabagisme ou l'alcool sont robustes.

De plus, une méta-analyse publiée en 2014 et comptant environ 50 000 cas et témoins (188) a détecté une association entre trois variants génétiques sur *FTO* et le risque d'hypertension (OR = 1,16, IC95% 1,07 ; 1,25, $p < 0.001$). Cependant, ces associations n'étant pas statistiquement significatives après ajustement pour l'IMC (OR = 1,04, IC95% 0,98 ; 1,10, $p = 0,162$), un facteur de risque important pour l'hypertension, suggérant que l'IMC soit une variable intermédiaire dans l'association entre les variants sur *FTO* et le risque d'hypertension. Ainsi, bien que cette hypothèse n'ait pas été investiguée dans la méta-analyse, il ne semblerait pas y avoir d'effet direct de *FTO* sur la pression artérielle hors de son effet sur l'adiposité. Alternativement, les résultats de l'étude de Pausova et al. (162) suggèrent que *FTO* puisse influencer la PAS de façon indépendante de l'adiposité. Le variant rs9939609 était associé à la PAS ($\beta = 4,4$ mm Hg pour AA vs AT/TT, $p = 0,002$), et ce après ajustement pour l'IMC ($\beta = 3,7$ mm Hg pour AA vs AT/TT, $p = 0,01$). Une analyse spectrale de la variabilité de fréquence cardiaque a également montré une association entre rs9939609 et l'influx sympathique vasomoteur ($\beta = 632$ mm Hg pour AA vs AT/TT, $p < 0,0001$) qui influence la PAS. Cette observation concorde avec le fait que *FTO* agisse au niveau de l'hypothalamus, une région également connue pour la régulation de l'influx nerveux sympathique (187). En résumé, les résultats de l'étude NDIIT concernant *FTO* s'alignent à ceux de la majorité des GWAS examinant la PAS en n'observant pas d'association entre *FTO* et la PAS. L'association observée dans l'étude SYS mérite tout de même d'être examinée davantage considérant l'existence de résultats suggérant des explications plausibles à l'association du gène avec la pression artérielle.

MTCH2 : Conformément à ce qui est observé dans l'étude SYS ($\beta_{\text{MTCH2-SYS}} = -1,6 \text{ kg/m}^2$, $p = 1,9 \times 10^{-4}$), l'étude NDIT a également observé que les porteurs de deux allèles mineurs sur rs7120548 (*MTCH2*) avaient en moyenne un IMC inférieur à ceux ayant deux allèles majeurs du variant ($\beta_{\text{MTCH2-NDIT}} = -0,89$, IC95% -1,76 ; -0,02 kg/m^2). Comme dans le cas de *FTO*, l'association est atténuée dans NDIT en comparaison avec ce qui est rapporté dans l'étude SYS. Des variants sur *MTCH2* ont été parmi les premiers à avoir été identifiés par GWAS pour leur association avec des mesures de l'adiposité et ont depuis été répliqués chez les adultes au sein de diverses ascendances (208). Toutefois, l'association entre *MTCH2* et l'adiposité n'a pas été observée de façon consistante chez les jeunes (201-204), suggérant l'hypothèse que l'effet du gène sur l'adiposité ne débute qu'à la transition vers l'âge adulte. Cette hypothèse est soutenue par les résultats de Graff et al. (138) qui observent que le variant rs3817334 sur *MTCH2* était associé au changement d'IMC ($\beta = 0,14 \text{ kg/m}^2$, $p = 9,23 \times 10^{-5}$) de la fin de l'adolescence jusqu'au début de l'âge adulte (16 à 25 ans). La réplification dans l'étude NDIT de l'association observée dans l'étude SYS pour *MTCH2* et l'IMC permet au contraire de suggérer que l'effet de *MTCH2* sur l'adiposité puisse être présent dès l'adolescence.

Le variant rs7120548 (*MTCH2*) faisait partie des variants associés à l'IMC dans l'étude SYS, et a donc également été testé pour une association avec la PAS. En observant une association entre le variant et l'IMC, mais pas avec la PAS ($\beta_{\text{MTCH2-NDIT}} = 0,31 \text{ mm Hg}$, IC95% -1,65 ; 2,26), l'étude NDIT concorde avec l'étude SYS ($\beta_{\text{MTCH2-SYS}} = -0,51 \text{ mm Hg}$, IC95% -3,86 ; 2,84). Des GWAS de grandes tailles (centaines de milliers de participants) ont récemment identifié une association entre un variant génétique sur *MTCH2* (rs7107356) et la PAS ($\beta = -0,309 \text{ mm Hg}$, $p = 4,0 \times 10^{-16}$) (71), indépendamment de l'IMC. Cette association a aussi été observée par Kichaev et al. (209) dans une étude du UK Biobank. Ces résultats récents suggèrent que *MTCH2* pourrait influencer la pression artérielle chez les adultes, mais avec une taille d'effet qui ne soit identifiable qu'avec des tailles d'échantillon beaucoup plus grandes que celles de l'étude SYS ou NDIT. Le rôle principalement connu du gène dans le processus de phosphorylation oxydative (196) et dans le métabolisme lipidique (197) suggère que l'adiposité ait un rôle d'intermédiaire dans son effet potentiel sur la PAS. Cependant, l'effet observé dans Hoffman et al. (71) était ajusté pour le niveau d'IMC des participants, suggérant que l'effet potentiel de *MTCH2* sur la PAS soit indépendant de l'IMC et fasse intervenir des fonctions biologiques actuellement méconnues.

MC4R : L'étude SYS avait observé une association entre rs17773430 (*MC4R*) et l'IMC ($\beta_{MC4R-SYS} = 2,4 \text{ kg/m}^2$, $p = 5,8 \times 10^{-6}$). Le gène *MC4R* est l'un des premiers à avoir été associée à l'adiposité (35) et cette association a été observée régulièrement chez les enfants et les adolescents (137). Le gène code pour un récepteur de la famille des récepteurs de mélanocortine qui fait partie du système mélanocortine qui agit sur le métabolisme énergétique. L'impact de *MC4R* sur l'adiposité repose sur son effet sur la satiété et la dépense énergétique (191). Le variant rs17773430 (*MC4R*) n'était pas disponible dans NDIT et le variant rs111638368 ($r^2 = 0,986$) a été utilisé comme proxy. Celui-ci ne montrait pas d'effet significatif sur l'IMC ($\beta_{MC4R-NDIT} = 0,71 \text{ kg/m}^2$, IC95% -0,27 ; 1,68) dans l'étude NDIT. Puisque l'association entre *MC4R* et l'adiposité est bien établie dans la littérature, il est probable que la non-association avec l'IMC dans l'étude NDIT représente un résultat faux négatif.

Faisant partie des variants associés à l'IMC par l'étude SYS, rs17773430 (*MC4R*) a également été testé pour une association avec la PAS. Le variant ne montrait pas d'effet significatif sur la PAS, résultat qui concorde avec l'étude NDIT pour le proxy rs111638368 qui n'était pas non plus associé à la PAS. Bien que quelques études antérieures ayant spécifiquement examiné des variants sur *MC4R* aient identifié une association avec la pression artérielle ou le risque d'hypertension (192-194), d'autres études (164, 228, 229) n'ont pas été en mesure de détecter une telle association. Ainsi, les résultats provenant de l'étude SYS et de l'étude NDIT ne supportent pas la présence d'une association entre rs17773430 (*MC4R*) et la PAS.

L'étude NDIT a toutefois identifié une interaction entre le variant rs111638368 (*MC4R*) et l'IMC affectant la PAS, ce qui suggère que l'IMC pourrait modifier l'effet du variant sur la PAS. Toutefois, l'identification d'une interaction statistiquement significative n'indique pas nécessairement que le variant ait un effet marginal significatif sur toute l'étendue des valeurs d'IMC (230). Selon les données de NDIT, l'effet du variant sur la PAS augmentait avec l'IMC, mais était seulement statistiquement significatif chez les individus ayant un IMC supérieur à 35, soit un niveau d'obésité de classe 2 représentant moins de 3% des participants. Ainsi, cette interaction est supportée de façon limitée par les données de NDIT. Finalement, il faut aussi souligner que l'étude NDIT n'avait pas comme objectif principal d'identifier ce phénomène et ne possède donc pas la puissance nécessaire afin de détecter une interaction de façon constante. Des

études subséquentes avec des tailles d'échantillon plus importantes seront nécessaires afin d'explorer la relation possible entre *MC4R*, l'adiposité et la pression artérielle.

6.2.2 Variants nouvellement identifiés comme associés à l'IMC

PAX5 et MRPS22 : L'étude SYS a identifié une association entre les variants génétiques rs16933812 (*PAX5*) et rs7638110 (*MRPS22*) et l'adiposité ($\beta_{PAX5-SYS} = 3,30 \text{ kg/m}^2$, IC95% 1,9 ; 4,6, $\beta_{MRPS22-SYS} = 2,80 \text{ kg/m}^2$, IC95% 1,9 ; 3,6), ainsi qu'avec la PAS ($\beta_{PAX5-SYS} = 6,68 \text{ mm Hg}$, IC95% 2,80 ; 10,57, $\beta_{MRPS22-SYS} = 3,30 \text{ mm Hg}$, IC95% 0,49 ; 6,12). L'étude NDIT n'a répliqué aucune de ces associations (adiposité : $\beta_{PAX5-NDIT} = -0,03 \text{ kg/m}^2$, IC95% -0,88 ; 0,81, $\beta_{MRPS22-NDIT} = -0,22 \text{ kg/m}^2$, IC95% -0,98 ; 0,53 et pression artérielle : $\beta_{PAX5-NDIT} = -0,55 \text{ mm Hg}$, IC95% -2,44 ; 1,34, $\beta_{MRPS22-NDIT} = -1,52 \text{ mm Hg}$, IC95% -3,20 ; 0,17). L'étude SYS a été la première à identifier de telles associations et aucune étude tentant de répliquer ces résultats n'a été publiée à ce jour. L'étude SYS a proposé une explication biologique basée sur les fonctions respectives des deux gènes. *PAX5* est exprimé dans une région du système nerveux sympathique responsable de l'influx sympathique vers le tissu adipeux (30) et pourrait donc influencer la régulation de la lipolyse. *MRPS22* est impliqué dans le processus de phosphorylation oxydative dans les mitochondries (174, 179) qui est responsable du transfert d'énergie des nutriments ingérés en molécules d'ATP. Ces deux gènes pourraient donc avoir un rôle plausible dans l'accumulation de matière grasse dans les tissus adipeux.

Cependant, la plausibilité biologique ne garantit pas la présence d'une véritable association génétique, comme Biedrzycki et al. (231) l'ont montré dans une étude sur l'interprétation de résultats empiriques d'études d'association génétique. Dans cette étude, une liste de variants et de traits reliés ou non reliés entre eux a été proposée à des annotateurs afin qu'ils évaluent la connexion biologique potentielle. Dans 55% des cas il était possible d'identifier une connexion biologique convaincante entre un variant et un trait, même lorsque ceux-ci étaient en réalité indépendants entre eux, suggérant qu'il est relativement facile de trouver un lien biologique plausible entre un variant génétique un trait même en l'absence d'une réelle association. Ainsi, bien que l'étude SYS identifie un lien biologique possible entre les gènes *MRPS22* et *PAX5* et l'adiposité, il est possible d'établir ce lien même lorsque le variant et le trait ne sont pas réellement associés. En effet, bien que les deux gènes en question puissent avoir un rôle dans la régulation de

l'adiposité, cela ne garantit pas que des variants sur ces gènes aient un effet causal sur le niveau d'adiposité. Fait à noter, l'effet des deux variants sur l'IMC dans l'étude NDIT se rapprochant de 0 suggère que cette association est nulle et non simplement non détectée en raison d'un manque de puissance.

6.3 Explications possibles de non-réplication

Comme suggéré auparavant, plusieurs associations présentées dans l'étude SYS n'ont pas été répliquées dans l'étude NDIT. L'inconsistance de résultats peut refléter qu'en raison d'erreurs aléatoires (erreurs de type 1 et 2) ou de biais, les résultats initiaux soient des faux positifs ou que les résultats de l'étude de réplication soient des faux négatifs. Il est aussi possible que les différences entre les protocoles des études soient telles que les quantités estimées ne soient pas comparables entre elles (211). Le terme hétérogénéité est communément utilisé dans la littérature relative aux méta-analyses afin de désigner les variations d'effets entre études qui excèdent les différences attendues par la chance (c.-à-d. dues aux risques de biais et différences entre les études) (214). On aborde d'abord les explications relatives aux erreurs aléatoires. Les sources d'hétérogénéité provoquant des effets non comparables entre l'étude de SYS et l'étude NDIT sont ensuite explorées pour conclure sur les sources d'hétérogénéité évoquant la différence de risque de biais dans chacune des études. La plausibilité que ces explications aient causé l'absence de réplication de certains résultats est évaluée à la lumière des résultats obtenus dans le mémoire.

6.3.1 Erreurs aléatoires

L'inconsistance de résultats entre deux études peut être le produit de l'observation de résultats faux-positifs lorsqu'une étude identifie un effet en l'absence d'une association réelle, ou faux-négatifs lorsqu'une association réelle n'est pas observée dans une étude (232). L'observation d'inférences erronées en raison de la chance (erreurs aléatoires) correspond à des erreurs de type 1 et de type 2 (233). Bien que les erreurs aléatoires ne puissent pas être complètement évitées, certains facteurs peuvent influencer la probabilité d'observer les erreurs de type 2.

Les facteurs influençant la puissance incluant la taille d'échantillon, la prévalence de l'exposition (fréquences alléliques dans les associations génétiques) ou la variabilité du trait étudié peuvent affecter la capacité à détecter l'effet d'intérêt et donc la probabilité d'obtenir une erreur

de type 2. Les écarts-types pour les valeurs d'IMC rapportés dans l'étude NDIT et l'étude SYS étaient semblables, alors que les écarts-types pour la pression artérielle étaient plus variables dans l'étude SYS. Ainsi, puisque la précision des méthodes de mesure utilisées dans la SYS semble équivalente ou moins précise que dans NDIT, il est peu probable que la précision des méthodes de mesure utilisées puisse expliquer la non-réplication des trois associations avec la PAS et de trois des cinq associations avec l'IMC. La puissance de l'étude NDIT était suffisante (0,83 - 0,92) afin de détecter adéquatement l'effet génétique des variants sur l'IMC. La puissance pour détecter l'effet des variants sur la PAS était adéquate pour les variants ayant une fréquence de l'allèle mineur (*minor allele frequency*, MAF) supérieure à 20% (0,75-0,80), mais diminuait pour les variants ayant une MAF sous 15% (0,68). La MAF de 0,07 du variant rs7638110 (*MRPS22*) observé dans l'échantillon étudié réduit ainsi la capacité à détecter un effet sur la PAS pour ce variant.

Les erreurs de type 1 surviennent en raison de la variation échantillonnale qui correspond aux différences aléatoires observées entre des échantillons indépendants provenant de la même population. La variation échantillonnale explique que des échantillons produisent des estimés différents, même si les variables sont mesurées de façon identique et les données analysées de la même façon. Bien qu'il soit impossible de vérifier si la variance échantillonnale a joué un rôle dans la non-réplication d'associations, compte tenu de la littérature concernant l'association de chacun des gènes investigués avec la pression artérielle et l'adiposité, il est plus probable que les associations rapportées par l'étude SYS avec les variant sur *PAX5* et *MRPS22* soient des erreurs de type 1 que de type 2 puisque ces variants n'ont jamais été détectés comme étant associés avec l'IMC et la PAS dans la littérature, même au sein d'études de très grandes tailles. Au contraire, il est plus probable qu'une erreur de type 2 soit en cause concernant la non-réplication de l'association entre le variant sur *MC4R* et l'IMC dans NDIT car cette association a été fréquemment rapportée dans la littérature.

6.3.2 Effets non comparables entre les études

Les sources d'hétérogénéité causant l'observation d'effets non comparables entre des études peuvent être multiples et référer aux populations étudiées, aux interactions gène-environnement, aux définitions des traits investigués ou encore aux différents devis d'étude (214,

234, 235). L'ascendance des populations d'étude, les interactions gènes-environnement, ainsi que l'opérationnalisation de l'IMC et de la PAS sont les sources d'hétérogénéité ayant la plus grande possibilité d'avoir entraîné de réelles différences d'effets dans la tentative de réplification présentée, et sont donc élaborées dans cette section.

Ascendance des populations d'étude: L'ascendance des populations composant les échantillons analysés est une considération importante dans les études génétiques. L'étude de Melka et al. (30) utilisait des données provenant de la SYS, une cohorte comportant des adolescents du Saguenay-Lac-Saint-Jean, alors que les participants de l'étude NDIT étaient recrutés de la région de Montréal. Les deux échantillons sont d'ascendance européenne, mais celui de la SYS provient d'une population fondatrice isolée (236) provenant d'environ 2000 individus du Saguenay-Lac-Saint-Jean au 17^e siècle (237). Deux phénomènes peuvent expliquer les différences de résultats obtenus lorsqu'on compare une population fondatrice à une population plus diversifiée, soit l'hétérogénéité génétique et les motifs de LD.

L'observation de résultats inconsistants dans différentes études peut être le produit de l'hétérogénéité allélique (plusieurs variants différents au sein d'un même gène peuvent causer un même phénotype) ou d'hétérogénéité de locus (un même phénotype peut être causé par différents variants à des loci différents) (214). Les traits complexes comme la pression artérielle et l'adiposité sont caractérisés par une étiologie génétique comprenant multiples variants différents. Or, l'identité des variants affectant un trait et leur fréquence allélique peut différer entre des populations d'ascendance différentes et ainsi mener à l'observation de tailles d'effet différentes entre des études examinant un variant et un trait spécifique (238). Le variant rs16933812 (*PAX5*) avait une fréquence allélique dans l'étude NDIT ($MAF_{NDIT} = 0,37$) comparativement à l'étude SYS ($MAF_{SYS} = 0,22$). La non-réplification des associations avec l'IMC et la PAS pour ce variant pourrait ainsi refléter une différence d'effets estimés en raison de la variation de fréquence allélique observée. Or, l'effet du variant sur l'IMC était nul ($\beta_{PAX5-NDIT} = -0,03 \text{ kg/m}^2$, IC95% -0,88 ; 0,81) et dans la direction opposée pour la PAS ($\beta_{PAX5-NDIT} = -0,55 \text{ mm Hg}$, IC95% -2,44 ; 1,34) dans l'étude NDIT, des observations qui se prêtent moins bien à cette explication.

Bien que la population du Saguenay-Lac-Saint-Jean et celle de Montréal sont toutes deux d'ascendance européenne, les populations isolées ont généralement une hétérogénéité génétique moins grande caractérisée par un éventail plus restreint de variants affectant un trait spécifique

comparativement aux populations génétiquement diversifiées (239). Puisque l'association entre un variant spécifique et un trait dépend de la fréquence allélique des autres variants associés à ce trait (214), il est plus facile de détecter l'effet des variants moins nombreux des populations fondatrices (239). Cela pourrait expliquer que l'étude SYS ait découvert deux nouveaux variants sur *PAX5* et *MRPS22* associés à l'IMC et à la PAS qui n'avaient pu être identifiés à l'aide d'études avec de grandes tailles d'échantillon de populations plus hétérogènes génétiquement. Cependant, sans réplication, on ne peut distinguer une véritable découverte d'une découverte due à la chance.

L'inconsistance entre les résultats obtenus de l'étude SYS et de l'étude NDIT peut aussi être expliquée par le degré de LD supérieur observé dans les populations fondatrices. La majorité des variants identifiés dans les études d'association génétique ne représentent pas des variants causaux pour le trait investigué. Ceux-ci représentent plutôt des variants en LD, ou corrélés, avec un variant causal. Un des facteurs affectant la taille d'effet observé pour un variant non causal identifié est le degré de LD qu'il présente avec le variant causal. Pour des fréquences alléliques semblables, l'effet du variant détecté évolue proportionnellement au produit de l'effet du variant causal et du LD (238, 240).

Le petit nombre de fondateurs, l'isolation et le temps écoulé depuis la fondation expliquent que les populations fondatrices présentent moins d'haplotypes (Texte supplémentaire 1) et de possibilité d'évènement de recombinaison génétique (Texte supplémentaire 1). Ces facteurs font en sorte que les populations fondatrices comme celle du Saguenay-Lac-Saint-Jean dont est composée la SYS sont connues pour avoir un niveau de LD systématiquement plus élevé à l'échelle du génome comparativement aux populations génétiquement plus diverses (241). Ainsi, on pourrait s'attendre à ce que les tailles d'effet estimées dans la population SYS fondatrice soient systématiquement plus grandes que celles observées pour la population NDIT (242). Ce phénomène pourrait s'appliquer aux diminutions des tailles d'effet observées dans l'étude NDIT comparativement à celles observées dans l'étude SYS. Par exemple, les associations entre les variants sur *FTO* et *MTCH2* avec l'IMC ont été répliquées, mais avaient des tailles d'effet inférieures dans l'étude NDIT comparativement à l'étude SYS. Les variants rs9930333 (*FTO*), rs7120548 (*MTCH2*) et rs17773430 (*MC4R*) associés à l'IMC dans l'étude SYS étaient considérés comme préalablement rapportés dans la littérature pour leur association avec l'IMC en raison d'une valeur de LD élevée ($D' > 0,9$) avec les variants rs9939609 (*FTO*), rs3817334 (*MTCH2*) et

rs17782313 (*MC4R*) qui avaient été associés à l'IMC dans des GWAS antérieures (168-170). Or, le LD entre ces paires de variants était inférieur dans l'étude NDIT ($r^2 = 0,82$ (*FTO*); 0,31 (*MTCH2*); 0,08 (*MC4R*)). Les baisses de LD entre les paires de variants sur les gènes *FTO* et *MTCH2* pourrait expliquer la diminution de la taille d'effet observée dans l'étude NDIT comparativement à l'étude SYS. Le LD très faible entre les variants du gène *MC4R* dans l'étude NDIT pourrait être la raison principale pour la non-réplication de l'effet de ce variant sur l'IMC. Bien que rs17773430 (*MC4R*) soit un bon proxy pour l'effet déjà connu de rs17782313 (*MC4R*) sur l'IMC dans la population fondatrice utilisée dans l'étude SYS, il semble que ce variant ne présente pas le même degré de LD dans les populations européennes plus diversifiées.

Interaction gène-environnement : L'inconsistance de résultats entre deux études peut aussi être expliquée par des interactions gène-environnement, qui réfèrent à ce que des caractéristiques environnementales peuvent modifier l'association entre des variants génétiques et des phénotypes. Par exemple, l'effet du variant rs9939609 (*FTO*) sur le risque d'obésité est atténué de 27% chez les individus physiquement actifs (OR = 1,22 par allèle de risque, IC95% 1,19 ; 1,25) comparativement aux individus inactifs (OR = 1,30 par allèle de risque, IC95% 1,24 ; 1,36) (243). Les tailles d'effet observées pour une association génétique au sein de deux populations pourraient varier en raison d'une interaction gène-environnement lorsque ces populations ont des proportions distinctes concernant des facteurs environnementaux et que l'interaction n'est pas prise en compte (232). Advenant que l'effet d'un variant génétique sur un trait diffère entre les deux strates d'une variable dichotomique, l'effet génétique observé pour l'échantillon complet, en ignorant l'interaction gène-environnement, correspondra à une moyenne pondérée des effets génétiques observés dans chaque strate. Ainsi, cet effet moyen différera selon la proportion de participants provenant de chacune des strates.

Certaines études suggèrent que des facteurs comportementaux pourraient interagir avec des variants génétiques sur *FTO* pour une association avec l'IMC. L'effet de variants sur l'IMC était modifié par le niveau d'activité physique (185, 244), la consommation d'alcool (244), la qualité de l'alimentation (244) et la durée du sommeil (244). De plus, la PAS des consommateurs d'alcool est davantage affectée par le variant rs55872725 (*FTO*) comparativement aux non-consommateurs (68). Ainsi, une proportion différente de participants inactifs, consommateurs d'alcool ou avec une courte durée de sommeil entre l'échantillon NDIT et du SYS peut être à l'origine de la diminution

de l'effet du variant sur *FTO* observée pour l'IMC ($\beta_{FTO-NDIT} = 1,18 \text{ kg/m}^2$, $\beta_{FTO-SYS} = 1,8 \text{ kg/m}^2$) et de la non-réplication de l'effet sur la PAS ($\beta_{FTO-NDIT} = 0,64 \text{ mm Hg}$, $\beta_{FTO-SYS} = 3,63 \text{ mm Hg}$). Toutefois, les tailles d'échantillon de l'étude NDIT et de l'étude SYS ne sont pas suffisantes afin d'examiner l'effet potentiel d'interactions avec ces facteurs comportementaux.

Définition des traits examinés: La non-réplication d'une étude subséquente peut être causée par l'utilisation de différentes définitions des traits examinés. Une étape importante à considérer dans la définition des traits est le processus d'opérationnalisation qui consiste à spécifier concrètement les observations capables de capturer empiriquement les concepts à l'étude (245). Une définition opérationnelle comprend un indicateur empirique, la méthode pour mesurer l'indicateur et la façon dont les valeurs de l'indicateur sont interprétées. Par exemple, l'étude NDIT et celle de l'étude SYS ont utilisé l'IMC comme indicateur de l'adiposité considéré continu dans les analyses et mesuré selon des méthodes semblables. Ainsi, l'opérationnalisation de l'adiposité ne devrait pas être une source d'hétérogénéité entre les deux études, comparativement à la pression artérielle qui pourrait représenter une source d'hétérogénéité. La pression artérielle des participants du SYS était mesurée selon un protocole de 52 minutes évaluant la pression artérielle en continu à l'aide d'un *Finometer* (FNS Finapres, Amsterdam, The Netherlands) validé pour la mesure de pression artérielle chez les adultes et les enfants de plus de 6 ans (246, 247). Cette méthode comportait cinq sections différentes (variant en position et en niveau d'effort mental) où les participants étaient couchés, assis, debout, à l'effort mental ou en récupération d'effort mental (236). Les moyennes de PAS à chacune des cinq sections étaient ensuite utilisées comme mesures répétées dans un modèle linéaire mixte. À titre comparatif, les participants de l'étude NDIT avaient quant à eux leur pression artérielle mesurée de trois à cinq reprises à une minute d'intervalle en position assise uniquement à l'aide d'un oscillomètre (Dinamap XL, modèle CR9340, Critikon Co, Tampa, Fla) validé pour mesurer la pression artérielle chez les enfants et les adolescents (248). Ces mesures ont été obtenues pour quatre temps différents et incorporées comme mesures répétées dans le temps dans un modèle linéaire mixte.

Ainsi, bien que l'étude NDIT et l'étude SYS examinent la PAS, opérationnellement les deux études ne mesurent pas exactement le même trait. La combinaison des cinq sections utilisées dans la SYS peut adéquatement représenter la variation de pression artérielle entre les individus, mais pourrait faire intervenir des fonctions biologiques diverses qui seraient relatives à un sous-

trait de la pression artérielle. En effet, la variation de pression artérielle à très court terme en réponse à un changement de position ou à un effort mental est régulée par multiples facteurs incluant l'influx sympathique central, les réflexes cardiopulmonaires, l'élasticité des vaisseaux sanguins, la viscosité du sang et certaines hormones et son architecture génétique demeure largement méconnue (45). Dans une investigation utilisant un échantillon réduit de la SYS (n=485), puisque moins récente, portant sur l'association entre rs9939609 (*FTO*) et la PAS, Pausova et al. (162) ont également examinés si l'effet du variant sur la PAS était modifiée par les sections de protocole de mesure de PAS. Ceux-ci ont observé une interaction significative ($p = 0,01$) entre le variant et les sections du protocole de PAS de la SYS (162), suggérant que l'apport de variants génétiques dans la variation de pression artérielle puisse différer selon diverses conditions posturales ou mentales. Ainsi, bien que les instruments de mesure utilisés dans les deux études soient individuellement validés pour mesurer la pression artérielle chez les adolescents, il est possible qu'elles représentent des sous-traités de la pression artérielle. Plus spécifiquement, la découverte des associations entre les variants sur les gènes *PAX5* et *MRPS22* et la PAS par l'étude SYS pourrait avoir été permise en raison de la présence des sections de PAS sous stress qui ne sont pas présentes dans la plupart des études d'association génétique avec la pression artérielle.

6.3.3 Risques de biais

Biais de sélection : Le biais de sélection survient lorsque la sélection des participants prenant part à une étude ou leur chance de rester dans l'étude provoque des associations systématiquement différentes de ce qui aurait été obtenu si toute la population cible était incluse dans l'étude (249). Le processus de sélection peut introduire un biais de collision qui peut mener à des estimations biaisées d'associations génétiques (250). Le biais de collision, aussi référé par le biais de sélection structurelle, survient lorsqu'on conditionne la participation à l'étude sur un effet commun (collisionneur) de (i) l'exposition et l'issue ou (ii) de deux autres variables – l'une associée à l'exposition et l'autre associée à l'issue (aussi référé par le biais en M) (249).

Cette instance peut être présente dans l'étude de Melka et al. (30) où l'échantillon analytique provenait du SYS où les participants étaient sélectionnés afin d'obtenir une proportion de 50% de participants exposés à la fumée de cigarette durant la grossesse (236). La Figure 1 illustre un biais de sélection où la sélection dans l'étude est un collisionneur dans l'association

entre les variants génétiques des participants et l'IMC et la PAS. L'exposition à la fumée secondaire durant la grossesse est associée à l'adiposité (251) et à la pression artérielle (252). Des variants génétiques de la mère affectant la propension à fumer seront également plus présents chez le participant. De cette façon, il est possible que la sélection des participants sur la base de leur exposition à la fumée secondaire durant la grossesse ait induit une association entre des variants génétiques affectant la propension à fumer et l'IMC et la PAS. Cependant, on remarque que les variants identifiés dans l'étude SYS ne sont pas connus pour influencer la propension à fumer, réduisant la probabilité qu'ils aient été identifiés en raison d'un biais de sélection.

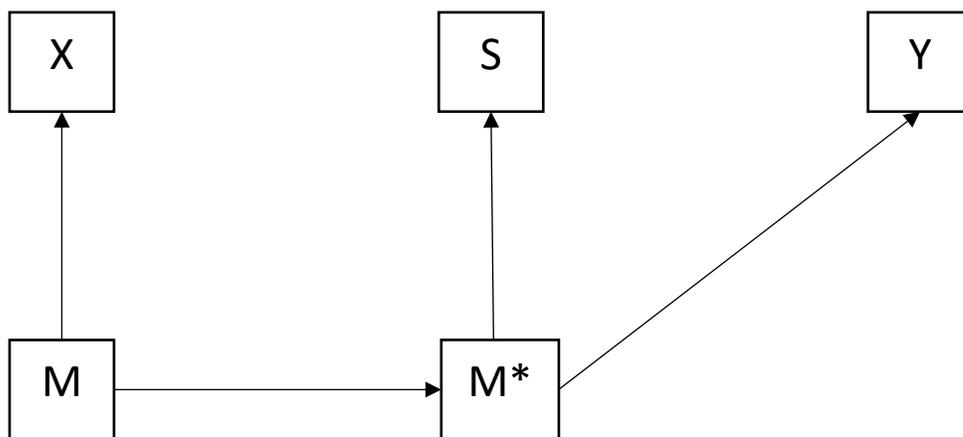


Figure 1. Illustration d'un biais de sélection pour l'association entre X et Y. X représente l'exposition (variants génétiques), Y représente l'issue (IMC/PAS), M et M* représentent des variables non mesurées (variants génétiques de la mère et l'exposition à la fumée secondaire durant la grossesse, respectivement), et S représente la sélection dans l'étude. Dans cet exemple, X ne cause pas Y, mais ils deviendront associés en conditionnant l'association sur S qui est un effet commun 1) de M* qui est associé à Y et 2) de M qui est associé à X.

Du côté de l'étude NDIT, le principal risque biais de sélection provient de la nature longitudinale des données collectées. Le biais de collision peut jouer un rôle dans les études longitudinales d'associations génétiques dans la mesure où conditionner sur la rétention dans l'étude peut introduire une corrélation artificielle entre un variant génétique et une issue (250). L'instance la plus simple à représenter consiste à ce que les variants investigués (X) soient associés à un phénotype connexe (Z) qui, comme l'issue d'intérêt (Y), influence la rétention dans l'étude

(S) (Figure 2). D'autres instances plus complexes de biais de collision incluant le biais en M peuvent survenir lorsque des données longitudinales sont utilisées pour investiguer des associations génétiques (250). L'association entre les variants génétiques étudiés et un phénotype connexe affectant la rétention dans l'étude est nécessaire à toutes ces instances. Des variants génétiques sur les gènes *PAX5* et *MTCH2* ont été respectivement associés à la dépression (253) et à la névrose (254), deux phénotypes possiblement en mesure d'affecter la rétention dans les études longitudinales (255). Toutefois, les symptômes de dépressions ne semblent pas affecter la rétention dans l'étude NDIIT ([Chapitre 5](#), Table S1).

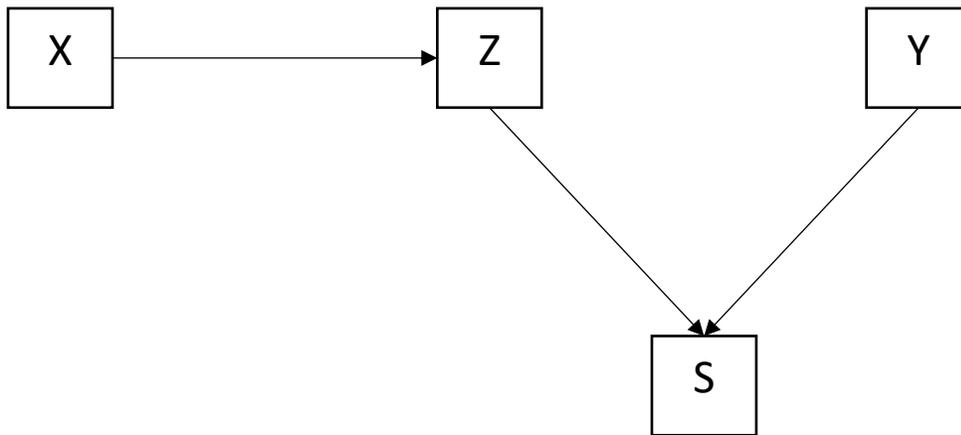


Figure 2. Illustration d'un biais de sélection pour l'association entre X et Y en raison de la perte au suivi. X représente l'exposition (variants génétiques), Y représente l'issue (IMC/PAS), Z représente un phénotype affecté par l'exposition, et S représente la sélection à l'étude (rétention dans l'étude). Dans cet exemple, X ne cause pas Y, mais ils deviendront associés en conditionnant l'association sur S qui est un effet commun de X et d'Y.

Biais de confusion : Les études d'association génétique sont moins vulnérables au biais de confusion comparativement aux études observationnelles n'ayant pas des variants génétiques comme exposition. Le biais de confusion survient lorsqu'une variable influence à la fois l'exposition et l'issue à l'étude, sans être dans le chemin causal et que cette variable est ignorée dans l'analyse. Or, il est peu probable qu'une variable affectant l'issue investiguée puisse aussi influencer la distribution des variants génétiques qui sont déterminés au hasard à la naissance.

Comme mentionné plus en détail à [l'Annexe A](#), la stratification de population peut introduire un biais de confusion dans les études d'association génétique. Ce concept fait référence aux différences de fréquences alléliques qui peuvent être observées dans une population d'étude en raison de la présence d'individus provenant de diverses ancestralités. Il est peu probable que ce facteur ait influencé les résultats de la SYS ou de l'étude NDI. Les participants recrutés dans la SYS devaient obligatoirement avoir des parents et grands-parents provenant d'origine canadienne-française nés dans la région du Saguenay-Lac-St-Jean, réduisant le risque de différences d'ancestralité dans l'échantillon (236). De plus, trois participants ont été exclus de l'étude en raison d'une analyse des composantes principales et les graphiques quantile-quantile effectués dans la SYS ne suggèrent pas la présence d'une stratification de population résiduelle (30). Dans l'étude NDI, l'ajout des cinq premières composantes principales aux modèles statistiques en [analyse de sensibilité](#) n'affectait pas les résultats de l'étude, indiquant un faible risque de biais de confusion en raison de la stratification de population dans NDI.

Biais d'information : Le biais d'information fait référence à une famille de biais se rapportant à des écarts comparativement à la vérité quant à la collection, le rappel ou le processus de gestion des données d'une étude (256). Un biais de rappel est peu probable dans la SYS ou l'étude NDI puisque l'exposition (génotype) et l'issue (PAS et IMC) ne sont pas des variables rapportées directement des participants, mais sont plutôt mesurées. Dans les études d'association génétique, un biais d'information peut provenir de la mauvaise classification du génotype des participants en raison des processus de génotypage utilisés qui ne sont pas précis à 100% des cas. Tous les participants inclus dans la SYS et l'étude NDI avaient un taux de génotypage des marqueurs supérieur à 97% et 98% respectivement, excluant les individus aux échantillons d'ADN plus difficiles à génotyper. Advenant tout de même la présence d'erreur de génotypage, ces erreurs de classifications seraient probablement non-différentielles puisqu'il est peu probable que cette erreur diffère selon les niveaux d'IMC ou de PAS des participants. Lorsque présent, ce genre d'erreur produit une diminution des effets estimés. Un biais d'information survenant en raison d'une erreur de classification des issues provenant de la connaissance préalable du statut d'exposition est également peu probable puisque les techniciens chargés des prises de mesure d'IMC et de PAS dans les deux études ne pouvaient être au courant du génotype des participants. En somme, il existe un faible risque de biais d'information dans les deux études en raison d'erreurs de génotypage qui ferait en sorte de sous-estimer la force des associations investiguées.

Surestimation d'effet génétique (*Winner's Curse*): Le processus de sélection de deux des cinq variants proposé par l'étude SYS pourrait avoir causé une surestimation de l'association génétique et subséquemment empêché sa réplication (phénomène du *Winners's Curse* décrit dans la [section 3.5](#)). En effet, rs16933812 (*PAX5*) et rs7638110 (*MRPS22*) n'ont pas été sélectionnés à partir de la littérature, mais plutôt parce qu'ils constituaient deux seuls variants parmi ~530 000 ayant atteint le seuil de signification pangénomique (5×10^{-8}) pour leur association avec l'IMC ou la masse adipeuse totale à partir des données du SYS. Compte tenu de l'échantillon modeste du SYS (n=598), il est peu probable que cette étude avait la puissance statistique nécessaire à la découverte de ces deux associations. Puisque l'effet estimé des variants sur les traits d'intérêt est directement corrélé à la significativité statistique, les associations découvertes sont enrichies de variants présentant un effet surestimé dans l'échantillon examiné dû à la chance (38). Les associations souffrant du *Winners's Curse* sont souvent non répliquées, ou si elles le sont, la taille de l'association répliquée est plus modeste que celle rapportée dans l'étude initiale.

Deux observations supportent l'hypothèse que les résultats associés avec rs16933812 (*PAX5*) et rs7638110 (*MRPS22*) et l'IMC soient des faux positifs dû au *Winner's Curse*. Premièrement, aucun variant sur les gènes *PAX5* ou *MRPS22* n'a été identifié dans les GWAS d'IMC publiées, incluant une méta-analyse ayant identifié 751 nouveaux variants chez un peu moins de 700 000 individus (134). Or, on s'attend à ce que les méta-analyses ayant une grande taille d'échantillon auraient été en mesure d'identifier les associations détectées au sein de l'échantillon inférieur de l'étude SYS (n=598). Deuxièmement, en comparant les résultats des modèles pour l'IMC dans l'étude SYS, on observe que les coefficients estimés pour les variants rs16933812 (*PAX5*) et rs7638110 (*MRPS22*) sont supérieurs à ceux des trois autres variants rs9930333 (*FTO*), rs7120548 (*MTCH2*) et rs17773430 (*MC4R*) ayant été sélectionnés à partir d'études antérieures. Or, advenant que l'estimation de l'effet pour ces deux variants ait été près de l'effet réel, ceux-ci auraient dû montrer des effets estimés plus faibles comparativement aux variants sur *FTO*, *MTCH2* et *MC4R*. Ces derniers ont été identifiés parmi les premiers associés à l'IMC par GWAS et ont été subséquemment rapportés à multiples reprises dans la littérature suggérant une taille d'effet plus grande et donc plus facile à identifier.

En ce qui concerne les variants sur *PAX5* et *MRPS22*, en plus d'avoir un IMC plus élevé (surestimation), il est possible que les individus porteurs des allèles de risque pour les deux variants

aient également des niveaux de PAS plus élevés comparativement aux non porteurs. L'IMC aurait un rôle de variable intermédiaire entre les variants et la PAS en raison du lien bien documenté entre l'adiposité et la PAS. Ainsi, l'association des variants sur PAX5 et MRPS22 avec et la PAS observée par l'étude SYS pourrait être reliée indirectement à la surestimation de leur effet sur l'IMC.

6.3.4 Revue des explications

D'abord, on doit reconnaître que plusieurs explications peuvent être conjointement responsables de la non-réplication des associations investiguées. Considérant l'observation d'un faible LD dans l'étude NDIT comparativement à l'étude SYS entre le proxy du variant rs17773430 (*MC4R*) et le variant original connu et validé (rs17782313) pour son association avec l'IMC chez les individus d'ascendance européenne, la non-réplication de l'association de ce variant et l'IMC représenterait davantage la présence d'un haut LD spécifique à la population fondatrice du Saguenay-Lac-Saint-Jean du variant investigué avec le variant validé.

Bien qu'il faille reconnaître les différences génétiques entre la population du Saguenay-Lac-Saint-Jean et celle de Montréal, la non-réplication relative aux variants sur *PAX5* et *MRPS22* reflète davantage des résultats faux-positifs dans l'étude SYS en raison de surestimation de leurs effets sur l'IMC et indirectement sur la PAS. Les deux conditions de base pour être susceptible d'une surestimation d'effets sont remplies en ayant basé le processus de sélection des variants sur le passage du seuil pangénomique au sein d'un échantillon insuffisant. L'indice principal pointant vers une surestimation d'effet reste l'observation d'effets supérieurs des variants sur *PAX5* et *MRPS22* comparativement aux variants sur les gènes bien établis dans la littérature pour leur association avec l'IMC. Cet exemple illustre bien la découverte de nouvelles associations génétiques due au *Winner's Curse*.

L'instance de non-réplication la plus ardue à expliquer reste l'association entre rs9930333 (*FTO*) et la PAS qui avait été identifiée par la SYS et non répliquée dans l'étude NDIT. La consommation d'alcool a déjà été rapportée comme modificateurs de l'effet d'un variant sur le gène *FTO* (68). Cette modification d'effet peut avoir généré des différences d'effet advenant une différence dans la proportion de consommateurs d'alcool entre les deux études. N'ayant pas la taille d'échantillon nécessaire, ces interactions n'ont pu être examinées en détail dans l'étude

NDIT. Également, les mesures de PAS à différentes postures corporelles et différents niveaux de stress sont une source d'inconsistance entre la SYS et l'étude NDIT, ainsi que la majorité des études mesurant la PAS au repos uniquement. Ainsi, il est difficile de conclure qu'une des deux études ont observé un faux résultat en raison de cette différence qui fait en sorte que les effets provenant des deux études pourraient être réellement différents. Des études supplémentaires utilisant une méthode de mesure de la PAS semblable à celle utilisée par la SYS sont nécessaires afin de souligner la complexité de l'influence génétique sur la PAS.

6.4 Forces et limites

La force principale concerne la similitude entre les échantillons provenant des études NDIT et SYS. Les participants de NDIT suivis de 12 à 24 ans entre 1999 et 2012 dans la région de Montréal sont semblables aux participants du SYS âgés de 12 à 18 ans recrutés au Saguenay-Lac-Saint-Jean de 2003 à 2012. Le suivi longitudinal des participants de l'étude NDIT a également permis l'obtention de plusieurs mesures anthropométriques et de pression artérielle pour chaque participant, permettant ainsi d'augmenter la puissance de l'étude.

Les limites de l'étude NDIT incluent la capacité à détecter l'effet génétique de variants avec une faible MAF. Bien que la puissance soit suffisante pour détecter l'effet des variants avec une MAF supérieure à 20% pour la PAS (0,75-0,80), on note une puissance inférieure pour les variants ayant une MAF sous 15% (0,68). La MAF de 0,07 du variant rs7638110 (*MRPS22*) observé dans l'échantillon étudié réduit ainsi la capacité à détecter un effet sur la PAS pour ce variant. Un biais de sélection correspond à la distorsion d'un effet estimé en raison de la manière de sélectionner les individus participant à une étude. Conceptuellement, l'effet estimé d'une exposition sur une issue peut être biaisé lorsqu'on conditionne la sélection dans une étude sur un effet commun de l'exposition et de l'issue. Comme expliqué davantage plus haut, un biais de sélection peut avoir modifié les résultats advenant la présence de phénotypes connexes affectés par les variants génétiques investigués, qui influencent également la perte au suivi (attrition). Les participants de NDIT perdus au suivi sont relativement semblables à ceux toujours présents au cycle 22 ([Chapitre 5](#), Table S1), à l'exception de l'âge et du sexe des participants qui ne sont pas susceptibles d'être influencés par des facteurs génétiques. Toutefois, on ne peut exclure que des phénotypes connexes non mesurés pourraient correspondre à cette définition. L'étude NDIT

présente une proportion élevée de suivi alors que près de 97% des participants inclus dans l'échantillon analytique étaient toujours présent au cycle 19 alors que 84% l'étaient toujours au cycle 22.

Il est également intéressant de se questionner sur certains aspects qui auraient permis d'améliorer la qualité générale de l'étude de réplication. D'abord, les études de réplication cherchent à se rapprocher le plus possible des caractéristiques des études initiales afin de réduire l'observation de résultats divergents en raison de différences méthodologiques. Ainsi, il aurait été optimal de pouvoir compter sur un échantillon d'étude provenant du Saguenay-Lac-Saint-Jean afin d'éviter de s'interroger sur les possibles différences entre cette population fondatrice et la population plus génétiquement diversifiée de Montréal. Dans la même optique, il aurait été intéressant d'avoir des données de PAS collectées de manière semblable à la SYS, afin de s'assurer de mesurer le même trait. Une taille d'échantillon plus grande aurait également permis d'examiner plus en détail le rôle possible de l'IMC comme modificateur de l'effet des variants génétiques sur la PAS.

6.5 Contributions du mémoire et futures directions

La contribution principale du mémoire est d'illustrer l'importance de la réplication d'études d'association génétique tout en soulignant les défis associés avec celle-ci. La réplication permet l'orientation des recherches futures en distinguant les résultats les plus crédibles de ceux qui nécessitent d'abord davantage de validation. Nous avons fait ressortir qu'une des difficultés rencontrées lors de la réplication d'associations génétique est d'identifier une population d'étude ayant des caractéristiques et des mesures semblables à la population d'étude initiale afin de déterminer si une absence apparente de réplication n'est en fait que le fruit de deux études trop différentes pour être comparées. Malgré la similitude entre les échantillons utilisés dans la SYS et de notre étude (populations québécoises d'âge semblable dont les données ont été colligées dans la même décennie), quelques différences (différences génétiques entre une population fondatrice et mixte, et différence dans la manière de mesurer la pression artérielle) pourraient tempérer la validité de la comparaison des résultats entre les deux études.

Une autre contribution du mémoire correspond à l'inventaire des phénomènes pouvant expliquer l'inconsistance de résultats entre des études examinant une même association génétique. Plusieurs facteurs peuvent agir de concert pour affecter la réplification d'associations génétiques et ces facteurs peuvent différer entre variants génétiques investigués. Entre autres, il a été possible de mettre de l'avant le phénomène du *Winner's Curse* en illustrant un bon exemple de surestimation d'un effet génétique en raison de la sélection de variants par le passage d'un seuil de significativité élevé.

Comme l'a souligné un groupe de travail sur la réplification d'association génétique du National Cancer Institute et du National Human Genome Research Institute (221), les chercheurs devraient s'assurer d'utiliser un échantillon d'étude et des méthodes de mesures et analytiques semblables à l'étude initiale, et d'étudier les mêmes traits et variants génétiques afin d'optimiser la capacité à répliquer une association génétique. Or, la moindre disponibilité de données génétiques pour les jeunes rend difficile l'identification d'études comparables examinant les facteurs génétiques affectant l'adiposité et la pression artérielle chez les jeunes. Les résultats d'associations génétiques chez les jeunes obtenus dans des contextes différents pourraient être mis à profit dans la mesure où ils peuvent refléter des phénomènes permettant d'améliorer les connaissances concernant certains processus biologiques. Par exemple, considérant qu'une différence d'architecture génétique soit suggérée par nos résultats en examinant seulement cinq variants, il est concevable que d'autres études pourraient identifier de telles différences en investiguant davantage de variants génétiques dans la population fondatrice du Saguenay-Lac-Saint-Jean et le reste de la population d'ascendance européenne présente au Québec.

6.6 Conclusion

Le mémoire contribue à souligner l'importance de la réplification en exposant certaines difficultés rencontrées lors de la réplification d'études d'association génétique chez les jeunes. Plus spécifiquement, une étude préalable avait identifié cinq variants génétiques associés à l'IMC chez des adolescents, dont trois étaient également associés à la PAS. L'étude de réplification présentée dans le mémoire a été en mesure de répliquer l'association de seulement deux variants avec l'IMC, alors qu'aucun des variants n'était associé à la PAS chez un échantillon indépendant d'adolescents. Un exposé complet des facteurs pouvant avoir affecté l'inconsistance des résultats entre les deux

études suggère que des différences concernant les populations étudiées et la définition de la PAS entre les deux études aient joué un rôle dans la non-réplication de certaines associations.

De plus, le mémoire illustre un bon exemple de surestimation de l'effet génétique de deux variants identifiés par GWAS en raison de leur sélection par le passage d'un haut seuil de significativité. Cela illustre le risque de rapporter des résultats provenant de GWAS sans réplication interne pour vérifier la précision de l'estimation des effets génétiques dans un échantillon indépendant (38). Considérant la pénurie de données génétiques disponibles chez les jeunes qui gêne la possibilité de systématiquement examiner la répliquabilité des associations génétiques à l'interne, la réplication externe par des investigateurs indépendant s'avère essentielle afin de confirmer la validité des résultats génétiques chez les jeunes.

BIBLIOGRAPHIE

1. Forouzanfar MH, Afshin A, Alexander LT, Anderson HR, Bhutta ZA, Biryukov S, et al. Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015. *The Lancet*. 2016;388(10053):1659-724.
2. Forouzanfar MH, Liu P, Roth GA, Ng M, Biryukov S, Marczak L, et al. Global Burden of Hypertension and Systolic Blood Pressure of at Least 110 to 115 mm Hg, 1990-2015. *JAMA*. 2017;317(2):165-82.
3. Collaboration NCDRF. Long-term and recent trends in hypertension awareness, treatment, and control in 12 high-income countries: an analysis of 123 nationally representative surveys. *Lancet* (London, England). 2019;394(10199):639-51.
4. DeGuire J, Clarke J, Rouleau K, Roy J, T B. Blood pressure and hypertension. *Health Reports*. 2019;30(02):14-21.
5. Padwal RS, Bienek A, McAlister FA, Campbell NRC. Epidemiology of Hypertension in Canada: An Update. *Canadian Journal of Cardiology*. 2016;32(5):687-94.
6. Rabi DM, McBrien KA, Sapir-Pichhadze R, Nakhla M, Ahmed SB, Dumanski SM, et al. Hypertension Canada's 2020 Comprehensive Guidelines for the Prevention, Diagnosis, Risk Assessment, and Treatment of Hypertension in Adults and Children. *Canadian Journal of Cardiology*. 2020;36(5):596-624.
7. Williams B, Mancia G, Spiering W, Agabiti Rosei E, Azizi M, Burnier M, et al. 2018 ESC/ESH Guidelines for the management of arterial hypertension: The Task Force for the management of arterial hypertension of the European Society of Cardiology (ESC) and the European Society of Hypertension (ESH). *European Heart Journal*. 2018;39(33):3021-104.
8. Prospective Studies Collaboration. Age-specific relevance of usual blood pressure to vascular mortality: a meta-analysis of individual data for one million adults in 61 prospective studies. *The Lancet*. 2002;360(9349):1903-13.
9. Canada S. Blood pressure of children and youth, 2012 to 2015 2016 [
10. Canada S. Blood pressure of Canadian children and youth, 2009 to 2011 2013 [
11. Falkner B, Gidding SS, Portman R, Rosner B. Blood Pressure Variability and Classification of Prehypertension and Hypertension in Adolescence. *Pediatrics*. 2008;122(2):238.
12. Brady TM, Fivush B, Flynn JT, Parekh R. Ability of Blood Pressure to Predict Left Ventricular Hypertrophy in Children with Primary Hypertension. *The journal of Pediatrics*. 2008;152(1):73-8.e1.
13. Juhola J, Magnussen Costan G, Berenson Gerald S, Venn A, Burns Trudy L, Sabin Matthew A, et al. Combined Effects of Child and Adult Elevated Blood Pressure on Subclinical Atherosclerosis. *Circulation*. 2013;128(3):217-24.
14. Lamotte C, Iliescu C, Libersa C, Gottrand F. Increased intima-media thickness of the carotid artery in childhood: a systematic review of observational studies. *European Journal of Pediatrics*. 2011;170(6):719-29.
15. Magnussen CG, Smith KJ. Pediatric Blood Pressure and Adult Preclinical Markers of Cardiovascular Disease. *Clinical medicine insights Blood disorders*. 2016;9:1-8.
16. Whelton PK, Carey RM, Aronow WS, Casey DE, Collins KJ, Dennison Himmelfarb C, et al. 2017 ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ASPC/NMA/PCNA Guideline for the Prevention, Detection, Evaluation, and Management of High Blood Pressure in Adults: A Report

- of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. *Journal of the American College of Cardiology*. 2018;71(19):e127-e248.
17. Henry SL, Barzel B, Wood-Bradley RJ, Burke SL, Head GA, Armitage JA. Developmental origins of obesity-related hypertension. *Clinical and Experimental Pharmacology and Physiology*. 2012;39(9):799-806.
18. Jayedi A, Rashidy-Pour A, Khorshidi M, Shab-Bidar S. Body mass index, abdominal adiposity, weight gain and risk of developing hypertension: a systematic review and dose–response meta-analysis of more than 2.3 million participants. *Obesity Reviews*. 2018;19(5):654-67.
19. Chrysant SG. Pathophysiology and treatment of obesity-related hypertension. *The Journal of Clinical Hypertension*. 2019;21(5):555-9.
20. Canada S. Overweight and obese adults, 2018 2019 [
21. Twells LK, Gregory DM, Reddigan J, Midodzi WK. Current and predicted prevalence of obesity in Canada: a trend analysis. *CMAJ open*. 2014;2(1):E18-E26.
22. Friedemann C, Heneghan C, Mahtani K, Thompson M, Perera R, Ward AM. Cardiovascular disease risk in healthy children and its association with body mass index: systematic review and meta-analysis. *BMJ : British Medical Journal*. 2012;345:e4759.
23. Chorin E, Hassidim A, Hartal M, Havakuk O, Flint N, Ziv-Baran T, et al. Trends in Adolescents Obesity and the Association between BMI and Blood Pressure: A Cross-Sectional Study in 714,922 Healthy Teenagers. *American Journal of Hypertension*. 2015;28(9):1157-63.
24. Buscot M-J, Thomson RJ, Juonala M, Sabin MA, Burgner DP, Lehtimäki T, et al. Distinct child-to-adult body mass index trajectories are associated with different levels of adult cardiometabolic risk. *European Heart Journal*. 2018;39(24):2263-70.
25. Munthali RJ, Kagura J, Lombard Z, Norris SA. Childhood adiposity trajectories are associated with late adolescent blood pressure: birth to twenty cohort. *BMC Public Health*. 2016;16:665-.
26. Umer A, Kelley GA, Cottrell LE, Giacobbi P, Jr., Innes KE, Lilly CL. Childhood obesity and adult cardiovascular disease risk factors: a systematic review with meta-analysis. *BMC Public Health*. 2017;17(1):683-.
27. Ziyab AH, Karmaus W, Kurukulaaratchy RJ, Zhang H, Arshad SH. Developmental trajectories of Body Mass Index from infancy to 18 years of age: prenatal determinants and health consequences. *Journal of Epidemiology and Community Health*. 2014;68(10):934-41.
28. Becker KG. The common variants/multiple disease hypothesis of common complex genetic disorders. *Medical Hypotheses*. 2004;62(2):309-17.
29. Rankinen T, Sarzynski MA, Ghosh S, Bouchard C. Are there genetic paths common to obesity, cardiovascular disease outcomes, and cardiovascular risk factors? *Circulation research*. 2015;116(5):909-22.
30. Melka MG, Bernard M, Mahboubi A, Abrahamowicz M, Paterson AD, Syme C, et al. Genome-Wide Scan for Loci of Adolescent Obesity and Their Relationship with Blood Pressure. *The Journal of Clinical Endocrinology & Metabolism*. 2012;97(1):E145-E50.
31. Wang X, Xu X, Su S, Snieder H. Familial aggregation and childhood blood pressure. *Current hypertension reports*. 2015;17(1):509-.
32. Kupper N, Ge D, Treiber Frank A, Snieder H. Emergence of Novel Genetic Effects on Blood Pressure and Hemodynamics in Adolescence. *Hypertension*. 2006;47(5):948-54.
33. Parmar Priyakumari G, Taal HR, Timpson Nicholas J, Thiering E, Lehtimäki T, Marinelli M, et al. International Genome-Wide Association Study Consortium Identifies Novel Loci Associated With Blood Pressure in Children and Adolescents. *Circulation: Cardiovascular Genetics*. 2016;9(3):266-78.

34. Bradfield JP, Taal HR, Timpson NJ, Scherag A, Lecoeur C, Warrington NM, et al. A genome-wide association meta-analysis identifies new childhood obesity loci. *Nature Genetics*. 2012;44(5):526-31.
35. Albuquerque D, Stice E, Rodríguez-López R, Manco L, Nóbrega C. Current review of genetics of human obesity: from molecular mechanisms to an evolutionary perspective. *Molecular Genetics and Genomics*. 2015;290(4):1191-221.
36. Ahn S-Y, Gupta C. Genetic Programming of Hypertension. *Frontiers in pediatrics*. 2018;5:285-.
37. Kraft P. Commentary: Curses - Winner's and Otherwise - in Genetic Epidemiology. *Epidemiology*. 2008;19(5):649-51.
38. Palmer C, Pe'er I. Statistical correction of the Winner's Curse explains replication variability in quantitative trait genome-wide association studies. *PLOS Genetics*. 2017;13(7):e1006916-e.
39. Lohmueller KE, Pearce CL, Pike M, Lander ES, Hirschhorn JN. Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. *Nature Genetics*. 2003;33(2):177-82.
40. Faye LL, Sun L, Dimitromanolakis A, Bull SB. A flexible genome-wide bootstrap method that accounts for ranking and threshold-selection bias in GWAS interpretation and replication study design. *Statistics in Medicine*. 2011;30(15):1898-912.
41. Göring HH, Terwilliger JD, Blangero J. Large upward bias in estimation of locus-specific effects from genomewide scans. *American journal of human genetics*. 2001;69(6):1357-69.
42. Poirier JG, Faye LL, Dimitromanolakis A, Paterson AD, Sun L, Bull SB. Resampling to Address the Winner's Curse in Genetic Association Analysis of Time to Event. *Genetic Epidemiology*. 2015;39(7):518-28.
43. O'Loughlin J, Dugas EN, Brunet J, DiFranza J, Engert JC, Gervais A, et al. Cohort Profile: The Nicotine Dependence in Teens (NDIT) Study. *International Journal of Epidemiology*. 2014;44(5):1537-46.
44. Seidel E, Scholl UI. Genetic mechanisms of human hypertension and their implications for blood pressure physiology. *Physiological Genomics*. 2017;49(11):630-52.
45. Parati G, Ochoa JE, Lombardi C, Bilo G. Assessment and management of blood-pressure variability. *Nature Reviews Cardiology*. 2013;10(3):143-55.
46. Hall ME, Hall JE. 5 - Pathogenesis of Hypertension. In: Bakris GL, Sorrentino MJ, editors. *Hypertension: A Companion to Braunwald's Heart Disease (Third Edition)*: Elsevier; 2018. p. 33-51.
47. Vischer AS, Burkard T. Principles of Blood Pressure Measurement – Current Techniques, Office vs Ambulatory Blood Pressure Measurement. In: Islam MS, editor. *Hypertension: from basic research to clinical practice*. Cham: Springer International Publishing; 2017. p. 85-96.
48. Ogedegbe G, Pickering T. Principles and techniques of blood pressure measurement. *Cardiology clinics*. 2010;28(4):571-86.
49. Nerenberg KA, Zarnke KB, Leung AA, Dasgupta K, Butalia S, McBrien K, et al. Hypertension Canada's 2018 Guidelines for Diagnosis, Risk Assessment, Prevention, and Treatment of Hypertension in Adults and Children. *Canadian Journal of Cardiology*. 2018;34(5):506-25.
50. Muntner P, Shimbo D, Carey Robert M, Charleston Jeanne B, Gaillard T, Misra S, et al. Measurement of Blood Pressure in Humans: A Scientific Statement From the American Heart Association. *Hypertension*. 2019;73(5):e35-e66.

51. Flynn JT, Kaelber DC, Baker-Smith CM, Blowey D, Carroll AE, Daniels SR, et al. Clinical Practice Guideline for Screening and Management of High Blood Pressure in Children and Adolescents. *Pediatrics*. 2017;140(3):e20171904.
52. Stamler J, Stamler R, Neaton JD. Blood Pressure, Systolic and Diastolic, and Cardiovascular Risks: US Population Data. *Archives of Internal Medicine*. 1993;153(5):598-615.
53. Rapsomaniki E, Timmis A, George J, Pujades-Rodriguez M, Shah AD, Denaxas S, et al. Blood pressure and incidence of twelve cardiovascular diseases: lifetime risks, healthy life-years lost, and age-specific associations in 1.25 million people. *Lancet (London, England)*. 2014;383(9932):1899-911.
54. Vishram Julie KK, Borglykke A, Andreassen Anne H, Jeppesen J, Ibsen H, Jørgensen T, et al. Impact of Age on the Importance of Systolic and Diastolic Blood Pressures for Stroke Risk. *Hypertension*. 2012;60(5):1117-23.
55. Palatini P, Rosei EA, Avolio A, Bilo G, Casiglia E, Ghiadoni L, et al. Isolated systolic hypertension in the young: a position paper endorsed by the European Society of Hypertension. *Journal of Hypertension*. 2018;36(6).
56. Cheung EL, Bell CS, Samuel JP, Poffenbarger T, Redwine KM, Samuels JA. Race and Obesity in Adolescent Hypertension. *Pediatrics*. 2017;139(5):e20161433.
57. Chen X, Wang Y. Tracking of Blood Pressure From Childhood to Adulthood. *Circulation*. 2008;117(25):3171-80.
58. Juhola J, Magnussen CG, Viikari JSA, Kähönen M, Hutri-Kähönen N, Jula A, et al. Tracking of Serum Lipid Levels, Blood Pressure, and Body Mass Index from Childhood to Adulthood: The Cardiovascular Risk in Young Finns Study. *The journal of Pediatrics*. 2011;159(4):584-90.
59. Sun SS, Grave GD, Siervogel RM, Pickoff AA, Arslanian SS, Daniels SR. Systolic Blood Pressure in Childhood Predicts Hypertension and Metabolic Syndrome Later in Life. *Pediatrics*. 2007;119(2):237.
60. Rao DP, Kropac E, Do MT, Roberts KC, Jayaraman GC. Childhood overweight and obesity trends in Canada. *Health promotion and chronic disease prevention in Canada : research, policy and practice*. 2016;36(9):194-8.
61. Ehret GB. 6 - Genetics of Hypertension. In: Bakris GL, Sorrentino MJ, editors. *Hypertension: A Companion to Braunwald's Heart Disease (Third Edition)*: Elsevier; 2018. p. 52-9.
62. Burton PR, Clayton DG, Cardon LR, Craddock N, Deloukas P, Duncanson A, et al. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*. 2007;447(7145):661-78.
63. Levy D, Larson MG, Benjamin EJ, Newton-Cheh C, Wang TJ, Hwang S-J, et al. Framingham Heart Study 100K Project: genome-wide associations for blood pressure and arterial stiffness. *BMC medical genetics*. 2007;8(1):S3.
64. Levy D, Ehret GB, Rice K, Verwoert GC, Launer LJ, Dehghan A, et al. Genome-wide association study of blood pressure and hypertension. *Nature Genetics*. 2009;41(6):677-87.
65. Newton-Cheh C, Johnson T, Gateva V, Tobin MD, Bochud M, Coin L, et al. Genome-wide association study identifies eight loci associated with blood pressure. *Nature Genetics*. 2009;41(6):666-76.
66. Ehret GB, Ferreira T, Chasman DI, Jackson AU, Schmidt EM, Johnson T, et al. The genetics of blood pressure regulation and its target organs from association studies in 342,415 individuals. *Nature Genetics*. 2016;48(10):1171-84.

67. Evangelou E, Warren HR, Mosen-Ansorena D, Mifsud B, Pazoki R, Gao H, et al. Genetic analysis of over 1 million people identifies 535 new loci associated with blood pressure traits. *Nature Genetics*. 2018;50(10):1412-25.
68. Feitosa MF, Kraja AT, Chasman DI, Sung YJ, Winkler TW, Ntalla I, et al. Novel genetic associations for blood pressure identified via gene-alcohol interaction in up to 570K individuals across multiple ancestries. *PLoS One*. 2018;13(6):e0198166.
69. Franceschini N, Fox E, Zhang Z, Edwards Todd L, Nalls Michael A, Sung Yun J, et al. Genome-wide Association Analysis of Blood-Pressure Traits in African-Ancestry Individuals Reveals Common Associated Genes in African and Non-African Populations. *The American Journal of Human Genetics*. 2013;93(3):545-54.
70. Giri A, Hellwege JN, Keaton JM, Park J, Qiu C, Warren HR, et al. Trans-ethnic association study of blood pressure determinants in over 750,000 individuals. *Nature Genetics*. 2019;51(1):51-62.
71. Hoffmann TJ, Ehret GB, Nandakumar P, Ranatunga D, Schaefer C, Kwok P-Y, et al. Genome-wide association analyses using electronic health records identify new loci influencing blood pressure variation. *Nature Genetics*. 2017;49(1):54-64.
72. Kato N, Loh M, Takeuchi F, Verweij N, Wang X, Zhang W, et al. Trans-ancestry genome-wide association study identifies 12 genetic loci influencing blood pressure and implicates a role for DNA methylation. *Nature Genetics*. 2015;47(11):1282-93.
73. Kraja AT, Cook JP, Warren HR, Surendran P, Liu C, Evangelou E, et al. New Blood Pressure-Associated Loci Identified in Meta-Analyses of 475 000 Individuals. *Circulation Cardiovascular genetics*. 2017;10(5):e001778.
74. Liu C, Kraja AT, Smith JA, Brody JA, Franceschini N, Bis JC, et al. Meta-analysis identifies common and rare variants influencing blood pressure and overlapping with metabolic trait loci. *Nature Genetics*. 2016;48(10):1162-70.
75. Sung YJ, Winkler TW, de las Fuentes L, Bentley AR, Brown MR, Kraja AT, et al. A Large-Scale Multi-ancestry Genome-wide Study Accounting for Smoking Behavior Identifies Multiple Significant Loci for Blood Pressure. *The American Journal of Human Genetics*. 2018;102(3):375-400.
76. Surendran P, Drenos F, Young R, Warren H, Cook JP, Manning AK, et al. Trans-ancestry meta-analyses identify rare and common variants associated with blood pressure and hypertension. *Nature Genetics*. 2016;48(10):1151-61.
77. Takeuchi F, Akiyama M, Matoba N, Katsuya T, Nakatochi M, Tabara Y, et al. Interethnic analyses of blood pressure loci in populations of East Asian and European descent. *Nature Communications*. 2018;9(1):5052.
78. Ehret GB, Munroe PB, Rice KM, Bochud M, Johnson AD, Chasman DI, et al. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature*. 2011;478(7367):103-9.
79. Wain Louise V, Vaez A, Jansen R, Joehanes R, van der Most Peter J, Erzurumluoglu AM, et al. Novel Blood Pressure Locus and Gene Discovery Using Genome-Wide Association Study and Expression Data Sets From Blood and the Kidney. *Hypertension*. 2017;70(3):e4-e19.
80. Wain LV, Verwoert GC, O'Reilly PF, Shi G, Johnson T, Johnson AD, et al. Genome-wide association study identifies six new loci influencing pulse pressure and mean arterial pressure. *Nature Genetics*. 2011;43(10):1005-11.

81. Warren HR, Evangelou E, Cabrera CP, Gao H, Ren M, Mifsud B, et al. Genome-wide association analysis identifies novel blood pressure loci and offers biological insights into cardiovascular risk. *Nature Genetics*. 2017;49(3):403-15.
82. Yang J, Zeng J, Goddard ME, Wray NR, Visscher PM. Concepts, estimation and interpretation of SNP-based heritability. *Nature Genetics*. 2017;49(9):1304-10.
83. Howe LD, Parmar PG, Paternoster L, Warrington NM, Kemp JP, Briollais L, et al. Genetic influences on trajectories of systolic blood pressure across childhood and adolescence. *Circulation Cardiovascular genetics*. 2013;6(6):608-14.
84. Oikonen M, Tikkanen E, Juhola J, Tuovinen T, Seppälä I, Juonala M, et al. Genetic variants and blood pressure in a population-based cohort: the Cardiovascular Risk in Young Finns study. *Hypertension (Dallas, Tex : 1979)*. 2011;58(6):1079-85.
85. Taal HR, Verwoert GC, Demirkan A, Janssens ACJW, Rice K, Ehret G, et al. Genome-wide profiling of blood pressure in adults and children. *Hypertension (Dallas, Tex : 1979)*. 2012;59(2):241-7.
86. Wang X, Snieder H. Assessing genetic risk of hypertension at an early age: future research directions. *Expert review of cardiovascular therapy*. 2017;15(11):809-12.
87. Bhatnagar P, Barron-Casella E, Bean CJ, Milton JN, Baldwin CT, Steinberg MH, et al. Genome-Wide Meta-Analysis of Systolic Blood Pressure in Children with Sickle Cell Disease. *PLoS One*. 2013;8(9):e74193.
88. Lule SA, Mentzer AJ, Namara B, Muwenzu AG, Nassanga B, Kizito D, et al. A genome-wide association and replication study of blood pressure in Ugandan early adolescents. *Molecular genetics & genomic medicine*. 2019;7(10):e00950-e.
89. Bays HE, Toth PP, Kris-Etherton PM, Abate N, Aronne LJ, Brown WV, et al. Obesity, adiposity, and dyslipidemia: A consensus statement from the National Lipid Association. *Journal of Clinical Lipidology*. 2013;7(4):304-83.
90. Oussaada SM, van Galen KA, Cooman MI, Kleinendorst L, Hazebroek EJ, van Haelst MM, et al. The pathogenesis of obesity. *Metabolism*. 2019;92:26-36.
91. Sharma AM, Padwal R. Obesity is a sign – over-eating is a symptom: an aetiological framework for the assessment and management of obesity. *Obesity Reviews*. 2010;11(5):362-70.
92. Koliaki C, Liatis S, Kokkinos A. Obesity and cardiovascular disease: revisiting an old relationship. *Metabolism*. 2019;92:98-107.
93. Douketis JD, Paradis G, Keller H, Martineau C. Canadian guidelines for body weight classification in adults: application in clinical practice to screen for overweight and obesity and to assess disease risk. *Canadian Medical Association Journal*. 2005;172(8):995.
94. Yumuk V, Tsigos C, Fried M, Schindler K, Busetto L, Micic D, et al. European Guidelines for Obesity Management in Adults. *Obesity Facts*. 2015;8(6):402-24.
95. Shi W, Neubeck L, Gallagher R. Measurement matters: A systematic review of waist measurement sites for determining central adiposity. *Collegian*. 2017;24(5):513-23.
96. Plourde G, Prud'homme D. Managing obesity in adults in primary care. *Canadian Medical Association Journal*. 2012;184(9):1039.
97. Alberti KGMM, Eckel Robert H, Grundy Scott M, Zimmet Paul Z, Cleeman James I, Donato Karen A, et al. Harmonizing the Metabolic Syndrome. *Circulation*. 2009;120(16):1640-5.
98. Hruby A, Hu FB. The Epidemiology of Obesity: A Big Picture. *PharmacoEconomics*. 2015;33(7):673-89.
99. Sweeting HN. Measurement and Definitions of Obesity In Childhood and Adolescence: A field guide for the uninitiated. *Nutrition Journal*. 2007;6(1):32.

100. de Onis M, Onyango AW, Borghi E, Siyam A, Nishida C, Siekmann J. Development of a WHO growth reference for school-aged children and adolescents. *Bulletin of the World Health Organization*. 2007;85(9):660-7.
101. Barlow SE. Expert Committee Recommendations Regarding the Prevention, Assessment, and Treatment of Child and Adolescent Overweight and Obesity: Summary Report. *Pediatrics*. 2007;120(Supplement 4):S164.
102. Collaborators GBDO, Afshin A, Forouzanfar MH, Reitsma MB, Sur P, Estep K, et al. Health Effects of Overweight and Obesity in 195 Countries over 25 Years. *The New England journal of medicine*. 2017;377(1):13-27.
103. Aune D, Sen A, Prasad M, Norat T, Janszky I, Tonstad S, et al. BMI and all cause mortality: systematic review and non-linear dose-response meta-analysis of 230 cohort studies with 3.74 million deaths among 30.3 million participants. *BMJ*. 2016;353:i2156.
104. Di Angelantonio E, Bhupathiraju SN, Wormser D, Gao P, Kaptoge S, de Gonzalez AB, et al. Body-mass index and all-cause mortality: individual-participant-data meta-analysis of 239 prospective studies in four continents. *The Lancet*. 2016;388(10046):776-86.
105. Bhaskaran K, dos-Santos-Silva I, Leon DA, Douglas IJ, Smeeth L. Association of BMI with overall and cause-specific mortality: a population-based cohort study of 3.6 million adults in the UK. *The Lancet Diabetes & Endocrinology*. 2018;6(12):944-53.
106. Riaz H, Khan MS, Siddiqi TJ, Usman MS, Shah N, Goyal A, et al. Association Between Obesity and Cardiovascular Outcomes: A Systematic Review and Meta-analysis of Mendelian Randomization Studies. *JAMA Network Open*. 2018;1(7):e183788-e.
107. Afzal S, Brøndum-Jacobsen P, Bojesen SE, Nordestgaard BG. Vitamin D concentration, obesity, and risk of diabetes: a mendelian randomisation study. *The Lancet Diabetes & Endocrinology*. 2014;2(4):298-306.
108. Holmes Michael V, Lange Leslie A, Palmer T, Lanktree Matthew B, North Kari E, Almoguera B, et al. Causal Effects of Body Mass Index on Cardiometabolic Traits and Events: A Mendelian Randomization Analysis. *The American Journal of Human Genetics*. 2014;94(2):198-208.
109. Fall T, Hägg S, Mägi R, Ploner A, Fischer K, Horikoshi M, et al. The Role of Adiposity in Cardiometabolic Traits: A Mendelian Randomization Analysis. *PLoS medicine*. 2013;10(6):e1001474.
110. Hägg S, Fall T, Ploner A, Mägi R, Fischer K, Draisma HHM, et al. Adiposity as a cause of cardiovascular disease: a Mendelian randomization study. *International Journal of Epidemiology*. 2015;44(2):578-86.
111. Nordestgaard BG, Palmer TM, Benn M, Zacho J, Tybjaerg-Hansen A, Davey Smith G, et al. The Effect of Elevated Body Mass Index on Ischemic Heart Disease Risk: Causal Estimates from a Mendelian Randomisation Approach. *PLoS medicine*. 2012;9(5):e1001212.
112. Goodarzi MO. Genetics of obesity: what genetic association studies have taught us about the biology of obesity and its complications. *The Lancet Diabetes & Endocrinology*. 2018;6(3):223-36.
113. Kenchaiah S, Evans JC, Levy D, Wilson PWF, Benjamin EJ, Larson MG, et al. Obesity and the Risk of Heart Failure. *New England Journal of Medicine*. 2002;347(5):305-13.
114. Cole CB, Nikpay M, Stewart AFR, McPherson R. Increased genetic risk for obesity in premature coronary artery disease. *European journal of human genetics : EJHG*. 2016;24(4):587-91.

115. Manson JE, Colditz GA, Stampfer MJ, Willett WC, Rosner B, Monson RR, et al. A Prospective Study of Obesity and Risk of Coronary Heart Disease in Women. *New England Journal of Medicine*. 1990;322(13):882-9.
116. Chatterjee NA, Giulianini F, Geelhoed B, Lunetta KL, Misialek JR, Niemeijer MN, et al. Genetic Obesity and the Risk of Atrial Fibrillation: Causal Estimates from Mendelian Randomization. *Circulation*. 2017;135(8):741-54.
117. Tedrow UB, Conen D, Ridker PM, Cook NR, Koplan BA, Manson JE, et al. The Long- and Short-Term Impact of Elevated Body Mass Index on the Risk of New Atrial Fibrillation: The WHS (Women's Health Study). *Journal of the American College of Cardiology*. 2010;55(21):2319-27.
118. Wang TJ, Parise H, Levy D, D'Agostino RB, Wolf PA, Vasan RS, et al. Obesity and the Risk of New-Onset Atrial Fibrillation. *JAMA*. 2004;292(20):2471-7.
119. Bancej C, Jayabalasingham B, Wall RW, Rao DP, Do MT, de Groh M, et al. Evidence Brief--Trends and projections of obesity among Canadians. *Health promotion and chronic disease prevention in Canada : research, policy and practice*. 2015;35(7):109-12.
120. Deshmukh-Taskar P, Nicklas TA, Morales M, Yang SJ, Zakeri I, Berenson GS. Tracking of overweight status from childhood to young adulthood: the Bogalusa Heart Study. *European Journal of Clinical Nutrition*. 2006;60(1):48-57.
121. Simmonds M, Llewellyn A, Owen CG, Woolacott N. Predicting adult obesity from childhood obesity: a systematic review and meta-analysis. *Obesity Reviews*. 2016;17(2):95-107.
122. Singh AS, Mulder C, Twisk JWR, Van Mechelen W, Chinapaw MJM. Tracking of childhood overweight into adulthood: a systematic review of the literature. *Obesity Reviews*. 2008;9(5):474-88.
123. The NS, Suchindran C, North KE, Popkin BM, Gordon-Larsen P. Association of Adolescent Obesity With Risk of Severe Obesity in Adulthood. *JAMA*. 2010;304(18):2042-7.
124. Reilly JJ, Kelly J. Long-term impact of overweight and obesity in childhood and adolescence on morbidity and premature mortality in adulthood: systematic review. *International Journal of Obesity*. 2011;35(7):891-8.
125. Twig G, Yaniv G, Levine H, Leiba A, Goldberger N, Derazne E, et al. Body-Mass Index in 2.3 Million Adolescents and Cardiovascular Death in Adulthood. *New England Journal of Medicine*. 2016;374(25):2430-40.
126. Vivante A, Golan E, Tzur D, Leiba A, Tirosh A, Skorecki K, et al. Body mass index in 1.2 million adolescents and risk for end-stage renal disease. *Archives of Internal Medicine*. 2012;172(21):1644-50.
127. Park MH, Falconer C, Viner RM, Kinra S. The impact of childhood obesity on morbidity and mortality in adulthood: a systematic review. *Obesity Reviews*. 2012;13(11):985-1000.
128. Weihrauch-Blüher S, Schwarz P, Klusmann J-H. Childhood obesity: increased risk for cardiometabolic disease and cancer in adulthood. *Metabolism*. 2019;92:147-52.
129. Maximova K, Kuhle S, Davidson Z, Fung C, Veugelers PJ. Cardiovascular Risk-Factor Profiles of Normal and Overweight Children and Adolescents: Insights From the Canadian Health Measures Survey. *Canadian Journal of Cardiology*. 2013;29(8):976-82.
130. Hoffmann TJ, Choquet H, Yin J, Banda Y, Kvale MN, Glymour M, et al. A Large Multiethnic Genome-Wide Association Study of Adult Body Mass Index Identifies Novel Loci. *Genetics*. 2018;210(2):499.
131. Albuquerque D, Nóbrega C, Manco L, Padez C. The contribution of genetics and environment to obesity. *British Medical Bulletin*. 2017;123(1):159-73.

132. Frayling TM, Timpson NJ, Weedon MN, Zeggini E, Freathy RM, Lindgren CM, et al. A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science (New York, NY)*. 2007;316(5826):889-94.
133. Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, Day FR, et al. Genetic studies of body mass index yield new insights for obesity biology. *Nature*. 2015;518(7538):197-206.
134. Yengo L, Sidorenko J, Kemper KE, Zheng Z, Wood AR, Weedon MN, et al. Meta-analysis of genome-wide association studies for height and body mass index in ~700000 individuals of European ancestry. *Human molecular genetics*. 2018;27(20):3641-9.
135. Yang J, Bakshi A, Zhu Z, Hemani G, Vinkhuyzen AAE, Lee SH, et al. Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. *Nature Genetics*. 2015;47(10):1114-20.
136. Elks CE, den Hoed M, Zhao JH, Sharp SJ, Wareham NJ, Loos RJJ, et al. Variability in the heritability of body mass index: a systematic review and meta-regression. *Frontiers in endocrinology*. 2012;3:29-.
137. Felix JF, Bradfield JP, Monnereau C, van der Valk RJP, Stergiakouli E, Chesi A, et al. Genome-wide association analysis identifies three new susceptibility loci for childhood body mass index. *Human molecular genetics*. 2016;25(2):389-403.
138. Graff M, Ngwa JS, Workalemahu T, Homuth G, Schipf S, Teumer A, et al. Genome-wide analysis of BMI in adolescents and young adults reveals additional insight into the effects of genetic loci over the life course. *Human molecular genetics*. 2013;22(17):3597-607.
139. Melén E, Granell R, Kogevinas M, Strachan D, Gonzalez JR, Wjst M, et al. Genome-wide association study of body mass index in 23 000 individuals with and without asthma. *Clinical and experimental allergy : journal of the British Society for Allergy and Clinical Immunology*. 2013;43(4):463-74.
140. Namjou B, Keddache M, Marsolo K, Wagner M, Lingren T, Cobb B, et al. EMR-linked GWAS study: investigation of variation landscape of loci for body mass index in children. *Frontiers in genetics*. 2013;4:268-.
141. Stergiakouli E, Gaillard R, Tavaré JM, Balthasar N, Loos RJ, Taal HR, et al. Genome-wide association study of height-adjusted BMI in childhood identifies functional variant in ADCY3. *Obesity*. 2014;22(10):2252-9.
142. Warrington NM, Howe LD, Paternoster L, Kaakinen M, Herrala S, Huikari V, et al. A genome-wide association study of body mass index across early life and childhood. *International Journal of Epidemiology*. 2015;44(2):700-12.
143. Bradfield JP, Voegele S, Felix JF, Chesi A, Helgeland Ø, Horikoshi M, et al. A trans-ancestral meta-analysis of genome-wide association studies reveals loci associated with childhood obesity. *Human molecular genetics*. 2019;28(19):3327-38.
144. Hallman DM, Friedel VC, Eissa MAH, Boerwinkle E, Huber JC, Jr., Harrist RB, et al. The association of variants in the FTO gene with longitudinal body mass index profiles in non-Hispanic white children and adolescents. *International journal of obesity (2005)*. 2012;36(1):61-8.
145. Hardy R, Wills AK, Wong A, Elks CE, Wareham NJ, Loos RJJ, et al. Life course variations in the associations between FTO and MC4R gene variants and body size. *Human molecular genetics*. 2010;19(3):545-52.
146. Sovio U, Mook-Kanamori DO, Warrington NM, Lawrence R, Briollais L, Palmer CNA, et al. Association between Common Variation at the FTO Locus and Changes in Body Mass Index from Infancy to Late Childhood: The Complex Nature of Genetic Association through Growth and Development. *PLOS Genetics*. 2011;7(2):e1001307.

147. Landsberg L, Aronne LJ, Beilin LJ, Burke V, Igel LI, Lloyd-Jones D, et al. Obesity-Related Hypertension: Pathogenesis, Cardiovascular Risk, and Treatment. *The Journal of Clinical Hypertension*. 2013;15(1):14-33.
148. Wilson PWF, D'Agostino RB, Sullivan L, Parise H, Kannel WB. Overweight and Obesity as Determinants of Cardiovascular Risk: The Framingham Experience. *Archives of Internal Medicine*. 2002;162(16):1867-72.
149. Carmelli D, Cardon LR, Fabsitz R. Clustering of hypertension, diabetes, and obesity in adult male twins: same genes or same environments? *American journal of human genetics*. 1994;55(3):566-73.
150. An P, Rice T, Gagnon J, Leon AS, Skinner JS, Wilmore JH, et al. Cross-trait familial resemblance for resting blood pressure and body composition and fat distribution: The HERITAGE family study. *American Journal of Human Biology*. 2000;12(1):32-41.
151. Cui J, Hopper John L, Harrap Stephen B. Genes and Family Environment Explain Correlations Between Blood Pressure and Body Mass Index. *Hypertension*. 2002;40(1):7-12.
152. Livshits G, Gerber Linda M. Familial Factors of Blood Pressure and Adiposity Covariation. *Hypertension*. 2001;37(3):928-35.
153. Rice T, Province M, Pérusse L, Bouchard C, Rao DC. Cross-trait familial resemblance for body fat and blood pressure: familial correlations in the Québec Family Study. *American journal of human genetics*. 1994;55(5):1019-29.
154. Schork NJ, Weder AB, Trevisan M, Laurenzi M. The contribution of pleiotropy to blood pressure and body-mass index variation: the Gubbio Study. *American journal of human genetics*. 1994;54(2):361-73.
155. Man T, Nolte IM, Jaju D, Al-Anqoudi ZAM, Muñoz ML, Hassan MO, et al. Heritability and genetic correlations of obesity indices with ambulatory and office beat-to-beat blood pressure in the Oman Family Study. *Journal of Hypertension*. 2020;38(8):1474-80.
156. Pausova Z, Gossard F, Gaudet D, Tremblay J, Kotchen Theodore A, Cowley Allen W, et al. Heritability Estimates of Obesity Measures in Siblings With and Without Hypertension. *Hypertension*. 2001;38(1):41-7.
157. Cadby G, Melton PE, McCarthy NS, Almeida M, Williams-Blangero S, Curran JE, et al. Pleiotropy of cardiometabolic syndrome with obesity-related anthropometric traits determined using empirically derived kinships from the Busselton Health Study. *Human Genetics*. 2018;137(1):45-53.
158. Olza J, Ruperez AI, Gil-Campos M, Leis R, Fernandez-Orth D, Tojo R, et al. Influence of FTO variants on obesity, inflammation and cardiovascular disease risk biomarkers in Spanish children: a case-control multicentre study. *BMC medical genetics*. 2013;14:123-.
159. van Vliet-Ostaptchouk JV, den Hoed M, Luan J, Zhao JH, Ong KK, van der Most PJ, et al. Pleiotropic effects of obesity-susceptibility loci on metabolic traits: a meta-analysis of up to 37,874 individuals. *Diabetologia*. 2013;56(10):2134-46.
160. Eisenmann JC, Sarzynski MA, Glenn K, Rothschild M, Heelan KA. ACE I/D genotype, adiposity, and blood pressure in children. *Cardiovascular Diabetology*. 2009;8(1):14.
161. Lemes VAF, Neves AL, Guazzelli IC, Frazzatto E, Nicolau C, Corrêa-Giannella ML, et al. Angiotensin converting enzyme insertion/deletion polymorphism is associated with increased adiposity and blood pressure in obese children and adolescents. *Gene*. 2013;532(2):197-202.
162. Pausova Z, Syme C, Abrahamowicz M, Xiao Y, Leonard Gabriel T, Perron M, et al. A Common Variant of the FTO Gene Is Associated With Not Only Increased Adiposity but Also

- Elevated Blood Pressure in French Canadians. *Circulation: Cardiovascular Genetics*. 2009;2(3):260-9.
163. Luczynski W, Zalewski G, Bossowski A. The association of the FTO rs9939609 polymorphism with obesity and metabolic risk factors for cardiovascular diseases in Polish children. *Journal of physiology and pharmacology : an official journal of the Polish Physiological Society*. 2012;63(3):241-8.
164. García-Solís P, Reyes-Bastidas M, Flores K, García OP, Rosado JL, Méndez-Villa L, et al. Fat mass obesity-associated (FTO) (rs9939609) and melanocortin 4 receptor (MC4R) (rs17782313) SNP are positively associated with obesity and blood pressure in Mexican school-aged children. *British Journal of Nutrition*. 2016;116(10):1834-40.
165. Kim YK, Kim Y, Hwang MY, Shimokawa K, Won S, Kato N, et al. Identification of a genetic variant at 2q12.1 associated with blood pressure in East Asians by genome-wide scan including gene-environment interactions. *BMC medical genetics*. 2014;15:65-.
166. Xi B, Zhao X, Chandak GR, Shen Y, Cheng H, Hou D, et al. Influence of Obesity on Association Between Genetic Variants Identified by Genome-Wide Association Studies and Hypertension Risk in Chinese Children. *American Journal of Hypertension*. 2013;26(8):990-6.
167. Lim JE, Kim HO, Rhee SY, Kim MK, Kim Y-J, Oh B. Gene-environment interactions related to blood pressure traits in two community-based Korean cohorts. *Genetic Epidemiology*. 2019;43(4):402-13.
168. Willer CJ, Speliotes EK, Loos RJF, Li S, Lindgren CM, Heid IM, et al. Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. *Nature Genetics*. 2009;41(1):25-34.
169. Speliotes EK, Willer CJ, Berndt SI, Monda KL, Thorleifsson G, Jackson AU, et al. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nature Genetics*. 2010;42(11):937-48.
170. Loos RJF, Lindgren CM, Li S, Wheeler E, Zhao JH, Prokopenko I, et al. Common variants near MC4R are associated with fat mass, weight and risk of obesity. *Nature Genetics*. 2008;40(6):768-75.
171. Yates AD, Achuthan P, Akanni W, Allen J, Allen J, Alvarez-Jarreta J, et al. Ensembl 2020. *Nucleic Acids Research*. 2019;48(D1):D682-D8.
172. Gene [Internet]. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information Bethesda (MD): National Library of Medicine (US); 2004 - [cited 2020/08/31] [August 1st 2020]. Available from: <https://www.ncbi.nlm.nih.gov/gene/>.
173. Cobaleda C, Schebesta A, Delogu A, Busslinger M. Pax5: the guardian of B cell identity and function. *Nature Immunology*. 2007;8(5):463-70.
174. Shahjahani M, Norozi F, Ahmadzadeh A, Shahrabi S, Tavakoli F, Asnafi AA, et al. The role of Pax5 in leukemia: diagnosis and prognosis significance. *Medical Oncology*. 2014;32(1):360.
175. Zhu Z, Guo Y, Shi H, Liu C-L, Panganiban RA, Chung W, et al. Shared genetic and experimental links between obesity-related traits and asthma subtypes in UK Biobank. *Journal of Allergy and Clinical Immunology*. 2020;145(2):537-49.
176. Auton A, Abecasis GR, Altshuler DM, Durbin RM, Abecasis GR, Bentley DR, et al. A global reference for human genetic variation. *Nature*. 2015;526(7571):68-74.
177. Case AJ, Zimmerman MC. Sympathetic-mediated activation versus suppression of the immune system: consequences for hypertension. *The Journal of physiology*. 2016;594(3):527-36.

178. Torlakovic E, Slipicevic A, Robinson C, DeCoteau JF, Alfsen GC, Vyberg M, et al. Pax-5 Expression in Nonhematopoietic Tissues. *American Journal of Clinical Pathology*. 2006;126(5):798-804.
179. Kılıç M, Oğuz K-K, Kılıç E, Yüksel D, Demirci H, Sağıroğlu MŞ, et al. A patient with mitochondrial disorder due to a novel mutation in MRPS22. *Metabolic Brain Disease*. 2017;32(5):1389-93.
180. Saada A, Shaag A, Arnon S, Dolfin T, Miller C, Fuchs-Telem D, et al. Antenatal mitochondrial disease caused by mitochondrial ribosomal protein (MRPS22) mutation. *Journal of medical genetics*. 2007;44(12):784-6.
181. Baertling F, Haack TB, Rodenburg RJ, Schaper J, Seibt A, Strom TM, et al. MRPS22 mutation causes fatal neonatal lactic acidosis with brain and heart abnormalities. *neurogenetics*. 2015;16(3):237-40.
182. Smits P, Saada A, Wortmann SB, Heister AJ, Brink M, Pfundt R, et al. Mutation in mitochondrial ribosomal protein MRPS22 leads to Cornelia de Lange-like phenotype, brain abnormalities and hypertrophic cardiomyopathy. *European Journal of Human Genetics*. 2011;19(4):394-9.
183. MacArthur J, Bowler E, Cerezo M, Gil L, Hall P, Hastings E, et al. The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Research*. 2017;45(D1):D896-D901.
184. Claussnitzer M, Dankel SN, Kim K-H, Quon G, Meuleman W, Haugen C, et al. FTO Obesity Variant Circuitry and Adipocyte Browning in Humans. *New England Journal of Medicine*. 2015;373(10):895-907.
185. Graff M, Scott RA, Justice AE, Young KL, Feitosa MF, Barata L, et al. Genome-wide physical activity interactions in adiposity — A meta-analysis of 200,452 adults. *PLOS Genetics*. 2017;13(4):e1006528.
186. Babenko V, Babenko R, Gamielien J, Markel A. FTO haplotyping underlines high obesity risk for European populations. *BMC Medical Genomics*. 2019;12(2):46.
187. Smemo S, Tena JJ, Kim K-H, Gamazon ER, Sakabe NJ, Gómez-Marín C, et al. Obesity-associated variants within FTO form long-range functional connections with IRX3. *Nature*. 2014;507(7492):371-5.
188. He D, Fu M, Miao S, Hotta K, Chandak GR, Xi B. FTO gene variant and risk of hypertension: A meta-analysis of 57,464 hypertensive cases and 41,256 controls. *Metabolism*. 2014;63(5):633-9.
189. Tao Y-X. The melanocortin-4 receptor: physiology, pharmacology, and pathophysiology. *Endocrine reviews*. 2010;31(4):506-43.
190. Ju SH, Cho G-B, Sohn J-W. Understanding melanocortin-4 receptor control of neuronal circuits: Toward novel therapeutics for obesity syndrome. *Pharmacological Research*. 2018;129:10-9.
191. Kühnen P, Krude H, Biebermann H. Melanocortin-4 Receptor Signalling: Importance for Weight Regulation and Obesity Treatment. *Trends in Molecular Medicine*. 2019;25(2):136-48.
192. Apalasy YD, Ming MF, Rampal S, Bulgiba A, Mohamed Z. Association of melanocortin-4 receptor gene polymorphisms with obesity-related parameters in Malaysian Malays. *Annals of Human Biology*. 2013;40(1):102-6.
193. Sun Y, Sun J, Wu J, Yang M. Combined effects of FTO rs9939609 and MC4R rs17782313 on elevated nocturnal blood pressure in the Chinese Han population. *Cardiovascular journal of Africa*. 2016;27(1):21-4.

194. Xi B, Zhao X, Shen Y, Wu L, Hotta K, Hou D, et al. Associations of obesity susceptibility loci with hypertension in Chinese children. *International Journal of Obesity*. 2013;37(7):926-30.
195. Bahat A, Goldman A, Zaltsman Y, Khan DH, Halperin C, Amzallag E, et al. MTCH2-mediated mitochondrial fusion drives exit from naïve pluripotency in embryonic stem cells. *Nature Communications*. 2018;9(1):5132.
196. Ruggiero A, Aloni E, Korkotian E, Zaltsman Y, Oni-Biton E, Kuperman Y, et al. Loss of forebrain MTCH2 decreases mitochondria motility and calcium handling and impairs hippocampal-dependent cognitive functions. *Scientific Reports*. 2017;7(1):44401.
197. Bar-Lev Y, Moshitch-Moshkovitz S, Tsarfaty G, Kaufman D, Horev J, Resau JH, et al. Mimp/Mtch2, an Obesity Susceptibility Gene, Induces Alteration of Fatty Acid Metabolism in Transgenic Mice. *PLoS One*. 2016;11(6):e0157850.
198. Akiyama M, Okada Y, Kanai M, Takahashi A, Momozawa Y, Ikeda M, et al. Genome-wide association study identifies 112 new loci for body mass index in the Japanese population. *Nature Genetics*. 2017;49(10):1458-67.
199. Justice AE, Winkler TW, Feitosa MF, Graff M, Fisher VA, Young K, et al. Genome-wide meta-analysis of 241,258 adults accounting for smoking behaviour identifies novel loci for obesity traits. *Nature Communications*. 2017;8(1):14977.
200. Hübel C, Gaspar HA, Coleman JRI, Finucane H, Purves KL, Hanscombe KB, et al. Genomics of body fat percentage may contribute to sex bias in anorexia nervosa. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*. 2019;180(6):428-38.
201. den Hoed M, Ekelund U, Brage S, Grontved A, Zhao JH, Sharp SJ, et al. Genetic susceptibility to obesity and related traits in childhood and adolescence: influence of loci identified by genome-wide association studies. *Diabetes*. 2010;59(11):2980-8.
202. Mei H, Chen W, Jiang F, He J, Srinivasan S, Smith EN, et al. Longitudinal replication studies of GWAS risk SNPs influencing body mass index over the course of childhood and adulthood. *PLoS One*. 2012;7(2):e31470-e.
203. Wang H-J, Hinney A, Song J-Y, Scherag A, Meng X-R, Grallert H, et al. Association of common variants identified by recent genome-wide association studies with obesity in Chinese children: a case-control study. *BMC medical genetics*. 2016;17(1):7.
204. Zhao J, Bradfield JP, Zhang H, Sleiman PM, Kim CE, Glessner JT, et al. Role of BMI-Associated Loci Identified in GWAS Meta-Analyses in the Context of Common Childhood Obesity in European Americans. *Obesity*. 2011;19(12):2436-9.
205. Jiménez-Osorio AS, Aguilar-Lucio AO, Cárdenas-Hernández H, Musalem-Younes C, Solares-Tlapechco J, Costa-Urrutia P, et al. Polymorphisms in Adipokines in Mexican Children with Obesity. *International journal of endocrinology*. 2019;2019:4764751-.
206. León-Mimila P, Villamil-Ramírez H, Villalobos-Comparán M, Villarreal-Molina T, Romero-Hidalgo S, López-Contreras B, et al. Contribution of common genetic variants to obesity and obesity-related traits in mexican children and adults. *PLoS One*. 2013;8(8):e70640-e.
207. Kulyté A, Rydén M, Mejhert N, Dungen E, Sjölin E, Arner P, et al. MTCH2 in Human White Adipose Tissue and Obesity. *The Journal of Clinical Endocrinology & Metabolism*. 2011;96(10):E1661-E5.
208. Rottiers V, Francisco A, Platov M, Zaltsman Y, Ruggiero A, Lee SS, et al. MTCH2 is a conserved regulator of lipid homeostasis. *Obesity*. 2017;25(3):616-25.
209. Kichaev G, Bhatia G, Loh P-R, Gazal S, Burch K, Freund MK, et al. Leveraging Polygenic Functional Enrichment to Improve GWAS Power. *The American Journal of Human Genetics*. 2019;104(1):65-75.

210. Kraft P, Zeggini E, Ioannidis JPA. Replication in genome-wide association studies. *Statistical science : a review journal of the Institute of Mathematical Statistics*. 2009;24(4):561-73.
211. Nakaoka H, Inoue I. Meta-analysis of genetic association studies: methodologies, between-study heterogeneity and winner's curse. *Journal of Human Genetics*. 2009;54(11):615-23.
212. Hirschhorn JN, Lohmueller K, Byrne E, Hirschhorn K. A comprehensive review of genetic association studies. *Genetics in Medicine*. 2002;4(2):45-61.
213. Border R, Johnson EC, Evans LM, Smolen A, Berley N, Sullivan PF, et al. No Support for Historical Candidate Gene or Candidate Gene-by-Interaction Hypotheses for Major Depression Across Multiple Large Samples. *American Journal of Psychiatry*. 2019;176(5):376-87.
214. Salanti G, Sanderson S, Higgins JPT. Obstacles and opportunities in meta-analysis of genetic association studies. *Genetics in Medicine*. 2005;7(1):13-20.
215. Thaler RH. Anomalies: The Winner's Curse. *Journal of Economic Perspectives*. 1988;2(1):191-202.
216. Bazerman MH, Samuelson WF. I Won the Auction But Don't Want the Prize. *Journal of Conflict Resolution*. 1983;27(4):618-34.
217. Garner C. Upward bias in odds ratio estimates from genome-wide association studies. *Genetic Epidemiology*. 2007;31(4):288-95.
218. Ioannidis JPA. Why Most Discovered True Associations Are Inflated. *Epidemiology*. 2008;19(5).
219. Zollner S, Pritchard JK. Overcoming the winner's curse: estimating penetrance parameters from case-control data. *American journal of human genetics*. 2007;80(4):605-15.
220. Bush WS, Moore JH. Chapter 11: Genome-Wide Association Studies. *PLOS Computational Biology*. 2012;8(12):e1002822.
221. Chanock SJ, Manolio T, Boehnke M, Boerwinkle E, Hunter DJ, Thomas G, et al. Replicating genotype-phenotype associations. *Nature*. 2007;447(7145):655-60.
222. Das S, Forer L, Schönherr S, Sidore C, Locke AE, Kwong A, et al. Next-generation genotype imputation service and methods. *Nature Genetics*. 2016;48(10):1284-7.
223. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2019.
224. Verbeke G, Molenberghs G. *Linear Mixed Models for Longitudinal Data*: Springer-Verlag New York; 2000.
225. Kreidler SM, Muller KE, Grunwald GK, Ringham BM, Coker-Dukowitz ZT, Sakhadeo UR, et al. GLIMPSE: Online Power Computation for Linear Models with and without a Baseline Covariate. *Journal of statistical software*. 2013;54(10):i10.
226. Center for Disease Control and Prevention. About Child & Teen BMI [updated June 29th 2020]. Available from: https://www.cdc.gov/healthyweight/assessing/bmi/childrens_bmi/about_childrens_bmi.html.
227. Solt F, Hu Y. interplot: Plot the Effects of Variables in Interaction Terms. package version 0.2.2 ed2019.
228. Fernandes AE, de Melo ME, Fujiwara CTH, Pioltine MB, Matioli SR, Santos A, et al. Associations between a common variant near the MC4R gene and serum triglyceride levels in an obese pediatric cohort. *Endocrine*. 2015;49(3):653-8.
229. Szkup M, Owczarek AJ, Schneider-Matyka D, Brodowski J, Łój B, Grochans E. Associations between the components of metabolic syndrome and the polymorphisms in the peroxisome proliferator-activated receptor gamma (PPAR- γ), the fat mass and obesity-associated (FTO), and the melanocortin-4 receptor (MC4R) genes. *Aging*. 2018;10(1):72-82.

230. Esarey J, Sumner JL. Marginal Effects in Interaction Models: Determining and Controlling the False Positive Rate. *Comparative Political Studies*. 2017;51(9):1144-76.
231. Biedrzycki RJ, Sier AE, Liu D, Dreikorn EN, Weeks DE. Spinning convincing stories for both true and false association signals. *Genetic Epidemiology*. 2019;43(4):356-64.
232. Schielzeth H, Rios Villamil A, Burri R. Success and failure in replication of genotype–phenotype associations: How does replication help in understanding the genetic basis of phenotypic variation in outbred populations? *Molecular Ecology Resources*. 2018;18(4):739-54.
233. Banerjee A, Chitnis UB, Jadhav SL, Bhawalkar JS, Chaudhury S. Hypothesis testing, type I and type II errors. *Industrial psychiatry journal*. 2009;18(2):127-31.
234. Li A, Meyre D. Challenges in reproducibility of genetic association studies: lessons learned from the obesity field. *International Journal of Obesity*. 2013;37(4):559-67.
235. Evangelou E, Ioannidis JPA. Meta-analysis methods for genome-wide association studies and beyond. *Nature Reviews Genetics*. 2013;14(6):379-89.
236. Pausova Z, Paus T, Abrahamowicz M, Bernard M, Gaudet D, Leonard G, et al. Cohort Profile: The Saguenay Youth Study (SYS). *International Journal of Epidemiology*. 2017;46(2):e19-e.
237. Engert JC, Lemire M, Faith J, Brisson D, Fujiwara TM, Roslin NM, et al. Identification of a chromosome 8p locus for early-onset coronary heart disease in a French Canadian population. *European Journal of Human Genetics*. 2008;16(1):105-14.
238. Liu Y-J, Papasian CJ, Liu J-F, Hamilton J, Deng H-W. Is replication the gold standard for validating genome-wide association findings? *PLoS One*. 2008;3(12):e4037-e.
239. Kristiansson K, Naukkarinen J, Peltonen L. Isolated populations and complex disease gene identification. *Genome biology*. 2008;9(8):109-.
240. Zondervan KT, Cardon LR. The complex interplay among factors that influence allelic association. *Nature Reviews Genetics*. 2004;5(2):89-100.
241. Service S, DeYoung J, Karayiorgou M, Roos JL, Pretorius H, Bedoya G, et al. Magnitude and distribution of linkage disequilibrium in population isolates and implications for genome-wide association studies. *Nature Genetics*. 2006;38(5):556-60.
242. Magosi LE, Goel A, Hopewell JC, Farrall M, Consortium CAD. Identifying systematic heterogeneity patterns in genetic association meta-analysis studies. *PLOS Genetics*. 2017;13(5):e1006755-e.
243. Kilpeläinen TO, Qi L, Brage S, Sharp SJ, Sonestedt E, Demerath E, et al. Physical activity attenuates the influence of FTO variants on obesity risk: a meta-analysis of 218,166 adults and 19,268 children. *PLoS medicine*. 2011;8(11):e1001116-e.
244. Young AI, Wauthier F, Donnelly P. Multiple novel gene-by-environment interactions modify the effect of FTO variants on body mass index. *Nature Communications*. 2016;7:12724-.
245. Allen M. *The SAGE Encyclopedia of Communication Research Methods*. 2017.
246. Parati G, Casadei R, Groppelli A, Di Rienzo M, Mancia G. Comparison of finger and intra-arterial blood pressure monitoring at rest and during laboratory testing. *Hypertension*. 1989;13(6_pt_1):647-55.
247. Tanaka H, Thulesius O, Yamaguchi H, Mino M, Konishi K. Continuous non-invasive finger blood pressure monitoring in children. *Acta Paediatrica*. 1994;83(6):646-52.
248. Park MK, Menard SM. Accuracy of Blood Pressure Measurement by the Dinamap Monitor in Infants and Children. *Pediatrics*. 1987;79(6):907.
249. Hernán MA, Hernández-Díaz S, Robins JM. A Structural Approach to Selection Bias. *Epidemiology*. 2004;15(5).

250. Munafò MR, Tilling K, Taylor AE, Evans DM, Davey Smith G. Collider scope: when selection bias can substantially influence observed associations. *International Journal of Epidemiology*. 2018;47(1):226-35.
251. Riedel C, Schönberger K, Yang S, Koshy G, Chen Y-C, Gopinath B, et al. Parental smoking and childhood obesity: higher effect estimates for maternal smoking in pregnancy compared with paternal smoking—a meta-analysis. *International Journal of Epidemiology*. 2014;43(5):1593-606.
252. Högberg L, Cnattingius S, Lundholm C, D'Onofrio BM, Långström N, Iliadou AN. Effects of maternal smoking during pregnancy on offspring blood pressure in late adolescence. *Journal of Hypertension*. 2012;30(4):693-9.
253. Hyde CL, Nagle MW, Tian C, Chen X, Paciga SA, Wendland JR, et al. Identification of 15 genetic loci associated with risk of major depression in individuals of European descent. *Nature Genetics*. 2016;48(9):1031-6.
254. Luciano M, Hagenaars SP, Davies G, Hill WD, Clarke T-K, Shireh M, et al. Association analysis in over 329,000 individuals identifies 116 independent variants influencing neuroticism. *Nature Genetics*. 2018;50(1):6-11.
255. Taylor AE, Jones HJ, Sallis H, Euesden J, Stergiakouli E, Davies NM, et al. Exploring the association of genetic factors with participation in the Avon Longitudinal Study of Parents and Children. *International Journal of Epidemiology*. 2018;47(4):1207-16.
256. Catalogue of bias collaboration, Bankhead CR, Spencer EA, Nunan D. Information bias. *Sackett Catalogue Of Biases* 2019.

ANNEXE A. Définitions et concepts pour études d'association génétique

Dictionnaire de termes communs

Note : À moins d'une spécification, les définitions proviennent du Dictionnaire de termes génétiques du National Cancer Institute (1).

Gène : Consiste en l'arrangement linéaire de nucléotide à une position spécifique sur un chromosome. Les gènes sont l'unité de base de l'hérédité et code pour des protéines occupant une fonction particulière.

Variant génétique : Correspond à une altération de la séquence nucléotidique en comparaison à un génome de référence. Elle peut être bénigne, pathogénique ou de conséquence inconnue.

Polymorphisme nucléotidique simple : Type de variant génétique le plus commun et correspond à l'altération d'un seul nucléotide dans la séquence d'ADN.

Allèle : Une de deux ou plus séquence d'ADN à une position spécifique sur un gène. Généralement, pour une paire d'allèles, l'un est commun et l'autre est rare.

Génotype : Le génotype peut représenter l'ensemble du bagage génétique d'un individu. Cependant, ce terme est généralement utilisé pour représenter l'allèle présent à un locus particulier.

Haplotype : Correspond à un groupe de variations génétiques présentes sur un même chromosome qui ont tendance à être héritées ensemble.

Locus : Position physique d'un gène spécifique sur un chromosome.

Recombinaison génétique : Échange de matériel génétique entre deux chromosomes ou différentes régions du même chromosome (2).

Homozygote : Génotype correspondant à la présence de deux allèles identiques à une position spécifique. Peut référer à deux allèles mineurs ou deux allèles majeurs.

Hétérozygote : Génotype correspondant à la présence de deux allèles différents à une position spécifique.

Héritabilité : La proportion de la variation d'un trait dans une population qui est attribuable à des facteurs génétiques. L'héritabilité varie de 0 à 1 et un nombre près de 1 correspond à un trait fortement héritable.

Déséquilibre de liaison : Correspond à l'observation qu'un ensemble d'allèles sont observés simultanément plus souvent que ce que l'on peut s'attendre par chance en raison d'une proximité physique.

Étude d'association pangénomique: Type d'analyse génétique examinant l'ensemble du génome pour identifier des variants génétiques associés à un certain phénotype.

Concepts de base

Les études d'association génétiques sont fréquemment utilisées afin d'identifier des facteurs de risque génétiques pour des maladies. Plus spécifiquement, on examine l'association entre un variant génétique sur un chromosome et un trait, qu'il soit sous la forme catégorique ou quantitatif. Dans la plupart des cas, les variants étudiés sont des polymorphismes nucléotidiques simples (3). On estime leur présence à 4 à 5 millions dans le génome d'une personne (3). Généralement, les variants ont alternativement deux allèles présents dans la population. Ils sont considérés comme communs lorsqu'ils sont présents dans plus de 5% des individus d'une population. Par contraste, des variants avec une fréquence allélique inférieure à 1% sont considérés comme des mutations, alors que ceux entre 1 et 5% sont des variants rares (4). L'étude présentée se focalise sur des variants ayant une fréquence allélique supérieure à 5%, ainsi les prochaines lignes se focalisent sur le fondement biologique et certains processus méthodologiques relatifs à l'étude de l'effet des variants communs sur la santé.

Dans la majorité des cas, les polymorphismes nucléotidiques simples ne provoquent pas de d'altérations importantes aux systèmes biologiques, mais dans certains cas ils peuvent affecter le risque de développer une maladie. En termes de susceptibilité génétique aux maladies, les variants communs ont généralement une pénétrance faible et un variant seul n'est donc généralement pas suffisant pour affecter un trait en grande partie. Cette observation est à la base des études d'association génétique sur des variants communs dans la population. On fait référence à ce que les maladies polygéniques comme le diabète de type II, la maladie coronarienne ou l'asthme sont affectés par plusieurs variations génétiques ayant un effet individuel relativement faible ainsi qu'une composante environnementale (5). Les GWAS représentent cette branche de

l'épidémiologie génétique qui se focalise sur l'association entre des variants présents sur l'ensemble de génome et divers traits reliés à des conditions de santé (1). Ce type d'étude est donc réalisée sans avoir d'hypothèse préalable quant aux variants spécifiques qui pourraient affecter les traits étudiés et concernant les mécanismes biologiques sous-jacents (6). Plus on effectue de tests statistiques plus il y a de chance d'identifier des associations faussement positives (inflation de l'erreur de type 1). En testant pour tous les variants dans HapMap (7), on estime que cela est équivalent à effectuer environ 1 million de tests d'association indépendants dans une population caucasienne, ce qui correspond au seuil pangénomique de 5×10^{-8} par correction de Bonferroni (8).

Lors des analyses d'associations génétiques, le génotype des individus observés peut être conceptualisé de diverses façons. Considérons une notation où le génotype des individus pour un certain variant comprend les possibilités A/A , A/a et a/a où « a » correspond à l'allèle de risque. Le type de codage le plus vastement utilisé est un codage additif où le génotype des individus est codé selon le nombre d'allèle de risque ($A/A = 0$, $A/a = 1$ et $a/a = 2$). Selon ce codage, l'allèle de risque a un effet linéaire sur le trait étudié. Le codage génotypique considère un codage catégorique (A/A , A/a et a/a) où l'effet de l'ajout d'un allèle de risque n'est pas nécessairement linéaire. Le codage récessif conceptualise une transmission récessive du trait, on y compare donc les individus avec les deux allèles de risques comparativement aux autres ($A/A = 0$, $A/a = 0$ et $a/a = 1$). Le codage dominant indique que l'on compare les individus avec au moins un allèle de risque à ceux qui n'en ont aucun ($A/A = 0$, $A/a = 1$ et $a/a = 1$). Finalement, un codage multiplicatif examine par combien de fois le risque augmente pour l'ajout de chaque allèle de risque.

Déséquilibre de liaison

Il est important de spécifier que l'identification d'une association statistiquement significative entre un variant génétique et un trait n'implique pas automatiquement un effet causal. Un variant peut ne pas avoir de relation biologique avec un trait avec lequel il est associé puisqu'ils ne se présentent pas de façon indépendante à-travers le génome, alors que certains variants sont liés entre eux. Le déséquilibre de liaison fait référence à cette présence non aléatoire de deux ou plusieurs allèles sur un chromosome, créant une association entre les deux allèles. En effet, deux allèles sont en déséquilibre de liaison lorsqu'ils ont tendance à être présents simultanément dans une population spécifique plus fréquemment que le hasard le voudrait (9). Plusieurs facteurs influencent le déséquilibre de liaison entre deux ou plusieurs allèles, dont la sélection naturelle, le

drift génétique, ou les mutations (9). Le facteur principal affectant le déséquilibre de liaison est le processus de recombinaison génétique (10). Les allèles sont légués en bloc d'allèles consécutifs dans le génome, bloc dont la location et le nombre d'allèle varie. Ce processus permet ainsi d'introduire une plus grande variabilité au matériel génétique. En général, les allèles physiquement plus près entre eux sont transmis ensemble et conservent donc un niveau de déséquilibre de liaison plus grand dans une population. Ainsi, une structure de déséquilibre de liaison en bloc peut généralement être observée par région génétique (10).

Le déséquilibre de liaison entre deux allèles se mesure par la différence (D) entre la fréquence observée de l'haplotype, ici référant à la combinaison des allèles en question, et le produit des fréquences alléliques individuelles. Ainsi, une valeur de D s'éloignant de 0 signifie un écart entre la fréquence observée de l'haplotype et la fréquence attendue en prenant en compte une distribution au hasard des allèles (9). En pratique, la valeur D' correspondant à la division de D par son maximum absolu théorique est plus communément utilisée (10). Le déséquilibre de liaison peut aussi être représenté par un coefficient de corrélation r^2 interprété comme le coefficient de Pearson. La valeur de r^2 pour deux allèles est obtenue par $D^2 / (p_1 \times p_2 \times q_1 \times q_2)$, où p_1 , p_2 , q_1 et q_2 représentent les fréquences individuelles du premier et deuxième allèle des variants étudiés (10). Une valeur de r^2 près de 1 implique que les deux allèles sont en fort déséquilibre de liaison et qu'ils sont souvent observés simultanément (5).

Le déséquilibre de liaison peut être utile à diverses fins en étude d'association génétique. Notamment, il est possible de l'utiliser afin de représenter un sous-ensemble de variants par un marqueur. La sélection d'un marqueur correspond à utiliser un SNP afin de représenter un haplotype particulier dans une région du génome. Les variants des régions en fort déséquilibre de liaison peuvent ainsi être réduits à un marqueur, duquel le génotype des autres variants de la région peut être inféré. Par exemple, les marqueurs peuvent être utilisés afin de ne présenter que des variants associés à un trait qui sont indépendants entre eux. La capacité à extrapoler la majorité des variants du génome à partir d'un sous-ensemble a aussi été essentielle à l'avancement des GWAS. Les puces de génotypage ne comprennent pas l'ensemble des polymorphismes du génome humain, mais plutôt un sous-ensemble de ceux-ci, développé afin de couvrir un large éventail de polymorphisme (9). En effet, à l'aide de données publiques sur le génome humain comme le projet

HapMap (7), il est possible d'extrapoler les variants qui n'auraient pas été génotypés directement et ainsi économiser d'importantes ressources (10).

Stratification de population

Dans les analyses épidémiologiques, certaines variables sont être considérées comme des variables de confusion potentielles lorsqu'il est possible qu'elles affectent à la fois la variable indépendante et les variables dépendantes, sans être dans le chemin causal. En étude d'association génétique, les variables dépendantes étudiées sont des variants génétiques et ne sont donc pas influencés par des facteurs de risque des traits étudiés. Cependant, il est possible que la stratification de population puisse être considérée comme une variable de confusion dans ce type d'étude. La stratification de population consiste en la différence dans la fréquence d'allèles qui peut être observée dû aux différentes ancestralités dans une population (11). Elle est induite principalement par l'isolement géographique de sous-populations sur une période de plusieurs générations. Les différentes sous-populations voient leurs fréquences alléliques évoluer aléatoirement de façon indépendante, causant des changements notables de la fréquence de plusieurs allèles après plusieurs générations (12). Souvent, les différences de fréquences alléliques observés dans les sous-populations ne causent aucun problème lors de l'analyse d'associations génétiques. Cependant, dans certains cas, la stratification de la population peut causer un biais de confusion dans une analyse entre un variant génétique et un phénotype. En effet, pour que la structure de population soit une variable de confusion, il est nécessaire que le trait d'intérêt ainsi que la fréquence allélique du variant étudié soient affectés par la présence de différentes sous-populations dans la population d'étude. Par exemple, dans une étude cas-témoins (13) portant sur le fondement génétique de la taille chez un échantillon de d'Américains d'origine européenne le variant rs4988235 du gène *LCT* était significativement associé à la taille dans l'échantillon total. Cependant, l'origine des grands-parents (Amérique, Sud de l'Europe, Nord de l'Europe) des sujets était corrélée à la fois avec la fréquence du variant et avec la taille des sujets. Ainsi, l'association observée entre le variant sur *LCT* et la taille dans l'échantillon total était dû à une sous-structure de population. En prenant en compte l'origine des grands-parents, l'association entre la fréquence du variant et la taille était inexistante ou grandement réduite (13).

Plusieurs méthodes sont utilisées afin d'évaluer la présence de structure de population dans une étude d'association génétique et pour aider à la prise de décision concernant la façon d'y

remédier au besoin. Les méthodes du contrôle génomique et d'analyse par composantes principales sont les plus fréquemment utilisées. Le contrôle génomique permet de détecter l'inflation des tests statistiques sous l'influence de biais, incluant la stratification de population. Le contrôle génomique exploite l'idée que la vaste majorité des variants testés dans un GWAS ne sera pas associée avec le phénotype d'intérêt et que la distribution empirique des statistiques de ces tests suivra la distribution de l'hypothèse nulle (14). Si présent, un biais relié à la stratification de population aurait comme effet de créer plusieurs associations de type faux-positifs dans la GWAS. Afin de détecter un tel phénomène, on peut considérer un facteur d'inflation (λ), qui est calculé en comparant la médiane empirique des statistiques de test du GWAS avec celle attendue sous l'hypothèse nulle de non-association (14). Ainsi, l'observation de valeurs de test s'éloignant de celles attendues (une valeur de λ supérieure à 1) sous l'hypothèse nulle suggère la présence possible d'une stratification de population affectant la relation entre des variants et le trait à l'étude (15). Une analyse par composante principale peut également servir à détecter la présence possible de groupes avec ancestralité distinctes dans une population d'étude. Les données génétiques individuelles y sont réduites en variables captant le maximum de variation génétique possible. Ces variables, appelées composantes principales, peuvent ensuite être représentées graphiquement et permettent de visualiser des sous-groupes au profil génétique commun et distinct des autres (12).

Advenant la détection d'une structure de population il est possible d'y remédier de diverses façons comportant chacune des avantages et inconvénients. Notamment, il est possible de limiter l'analyse des données à un sous-groupe d'ancestralité commune en y excluant les données provenant d'individus des autres sous-groupes, ce qui a toutefois pour conséquence de diminuer la taille d'échantillon (15). On peut également ajuster l'analyse statistique en y incluant les composantes principales comme covariables ou en divisant le test statistique par le facteur d'inflation (15), quoique cette méthode est parfois contestée car pour plusieurs phénotypes complexes, le facteur d'inflation peut être positif en absence de structure de population, simplement parce qu'un grand nombre de variants sont réellement associés avec le phénotype d'intérêt (16). Il est à noter que ces méthodes pour prendre en compte la stratification de population ne sont pas exhaustives et que la prise de décision varie selon la population à l'étude, les données et les ressources disponibles ou les objectifs spécifiques à l'étude.

RÉFÉRENCES ANNEXE A

1. PDQ Cancer Information Summaries. Dictionary of Genetics Terms Bethesda (MD): National Cancer Institute (US); 2002 [Available from: <https://www.cancer.gov/publications/dictionaries/genetics-dictionary/expand/R>].
2. Clancy S. Genetic recombination. *Nature Education*. 2008;1(1):40.
3. Auton A, Abecasis GR, Altshuler DM, Durbin RM, Abecasis GR, Bentley DR, et al. A global reference for human genetic variation. *Nature*. 2015;526(7571):68-74.
4. Dubé JB, Hegele RA. Genetics 100 for Cardiologists: Basics of Genome-Wide Association Studies. *Canadian Journal of Cardiology*. 2013;29(1):10-7.
5. Bush WS, Moore JH. Chapter 11: Genome-Wide Association Studies. *PLOS Computational Biology*. 2012;8(12):e1002822.
6. König IR. Validation in Genetic Association Studies. *Briefings in Bioinformatics*. 2011;12(3):253-8.
7. Gibbs RA, Belmont JW, Hardenbol P, Willis TD, Yu F, Yang H, et al. The International HapMap Project. *Nature*. 2003;426(6968):789-96.
8. Goeman JJ, Solari A. Multiple hypothesis testing in genomics. *Statistics in Medicine*. 2014;33(11):1946-78.
9. Slatkin M. Linkage disequilibrium--understanding the evolutionary past and mapping the medical future. *Nat Rev Genet*. 2008;9(6):477-85.
10. Goode EL. Linkage Disequilibrium. In: Schwab M, editor. *Encyclopedia of Cancer*. Berlin, Heidelberg: Springer Berlin Heidelberg; 2011. p. 2043-8.
11. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics*. 2006;38(8):904-9.
12. Hellwege JN, Keaton JM, Giri A, Gao X, Velez Edwards DR, Edwards TL. Population Stratification in Genetic Association Studies. *Curr Protoc Hum Genet*. 2017;95:1.22.1-1.3.
13. Campbell CD, Ogburn EL, Lunetta KL, Lyon HN, Freedman ML, Groop LC, et al. Demonstrating stratification in a European American population. *Nature Genetics*. 2005;37(8):868-72.
14. Devlin B, Roeder K, Wasserman L. Genomic Control, a New Approach to Genetic-Based Association Studies. *Theoretical Population Biology*. 2001;60(3):155-66.
15. Clarke GM, Anderson CA, Pettersson FH, Cardon LR, Morris AP, Zondervan KT. Basic statistical analysis in genetic case-control studies. *Nat Protoc*. 2011;6(2):121-33.

16. Yang J, Weedon MN, Purcell S, Lettre G, Estrada K, Willer CJ, et al. Genomic inflation factors under polygenic inheritance. *Eur J Hum Genet.* 2011;19(7):807-12.

ANNEXE B. Description des gènes reliés aux variants à l'étude

	rs16933812	rs7638110	rs9930333	rs111638368	rs7120548
Gène	PAX5	MRPS22	FTO	MC4R	MTCH2
Position	9p13.2	3q23	16q12.2	18q21.32	11p11.2
Allèle Mineur	G	T	G	T	C
Allèle majeur	T	G	T	C	T
MAF NDIT (%)	36.8	7.4	43.2	28.8	34.5
MAF HapMap CEU (%) (1)	31.8	5.1	47.5	28.8	33.8
Protéine	Paired box protein Pax-5	28S ribosomal protein S22	Alpha-Ketoglutarate Dependent Dioxygenase	Melanocortin Receptor 4	Mitochondrial carrier homolog 2
Fonctions	-Développement du système nerveux -Spermatogenèse -Différentiation des cellules lymphatiques B	-Synthèse de protéines mitochondriales -Phosphorylation oxydative	-Réparation de l'ARN alkylé par déméthylation oxydative -Métabolisme énergétique	-Métabolisme énergétique -Contrôle du poids corporel -satiété	-Apoptose des mitochondries -Métabolisme énergétique -Métabolisme lipidique
Associations répertoriées avec l'adiposité	Le variant rs3013733 sur PAX5 identifié par GWAS pour l'IMC chez les adultes (2). Aucune association répertoriée chez les jeunes.	Aucune association répertoriée chez les adultes ou les jeunes.	Le variant rs9939609 sur FTO est le premier à avoir été associé par GWAS à l'adiposité (3). Depuis, plus de 60 variants sur FTO ont été associés à l'adiposité à multiples reprises chez les adultes (4) et plusieurs ont été répliqués chez les jeunes (5).	Fait partie des premiers gènes à avoir été identifié par GWAS pour une association avec l'adiposité chez les adultes (4). Plusieurs variants sur MC4R associés à l'adiposité chez les adultes ont ensuite été répliqués chez les jeunes (5).	Fait partie des premiers gènes à avoir été identifié par GWAS pour une association avec l'adiposité chez les adultes et ont été répliqués à plusieurs reprises depuis (4). Les variants sur MTCH2 répliquent moins bien chez les jeunes. Bon nombre d'études n'observent pas

			Le variant à l'étude rs9930333 a été identifié dans une GWAS pour l'IMC (6).	Plus spécifiquement, le variant à l'étude rs17773430 était en haut LD avec le variant rs17782313 sur MC4R dans Melka et al. Ce dernier a été associé par GWAS à l'IMC à plusieurs reprises (6-8).	d'association avec l'adiposité (9-14). Le variant à l'étude rs7120548 était en haut LD avec le variant rs3817334 sur MTCH2 dans Melka et al. Ce dernier est associé à l'IMC dans plusieurs GWAS (6, 15-18)
Associations répertoriées avec la pression artérielle	Aucune association répertoriée pour la PAS chez les adultes ou les enfants.	Aucune association répertoriée chez les adultes ou les jeunes.	Seules trois GWAS pour la pression artérielle ont identifié une association avec des variants sur FTO (19-21). Les études ayant examiné spécifiquement la relation de variants sur FTO avec la pression artérielle obtiennent des résultats inconsistants (22). Quelques études ont observé une association entre variants sur FTO et la pression artérielle chez les jeunes (23-27).	Le variant rs17773430 n'était pas associé à la PAS dans l'étude de Melka et al. Un seul GWAS identifie un variant sur MC4R associé à la PAS (28). Seulement quelques études ont identifié une association entre variants sur MC4R et la pression artérielle ou l'hypertension (25, 29, 30). L'une de ces études effectuée chez des enfants (25).	La grande majorité de GWAS n'identifient pas de variants sur MTCH2 en association avec la pression artérielle. Seul le variant rs7107356 sur MTCH2 a été associé avec la PAS, la PAD et la pression pulsée par deux groupes distincts (31, 32). Aucune association répertoriée chez les jeunes.

<p>Relation possible avec l'obésité</p>	<p>Outre leur expression dans les cellules hématopoïétique, PAX5 est également exprimé au niveau du système nerveux central. Plus spécifiquement, PAX5 est exprimé dans la substance grise périaqueducale au niveau du mésencéphale, ainsi que dans l'aire postrema au niveau de la moelle allongée (33). Ces régions sont notamment impliquées dans la régulation des fonctions autonomes incluant la lipolyse médiée par activation sympathique (34).</p>	<p>MRPS22 est nécessaire à la synthèse de protéines mitochondriales composant les complexes de phosphorylation oxydative mitochondriale (35). Une altération de MRPS22 pourrait donc altérer le métabolisme énergétique et ainsi affecter l'accumulation de graisse.</p>	<p>La relation de FTO avec l'obésité est bien établie. L'association entre FTO et l'accumulation de graisse est reliée au gène IRX3. Les séquences non-codantes sur FTO associés avec l'obésité sont connectées avec les promoteurs du gène IRX3. Les variants sur FTO sont associés à un changement d'expression de IRX3 dans le cerveau. Les études sur modèles animaux ont montré que les individus privés de IRX3 montrent une altération du poids corporelle (36).</p>	<p>Plus de 150 mutations sur MC4R ont été identifiées pour leur lien avec des formes monogéniques d'obésité. Les mutations sur MC4R sont responsables de la majorité des formes monogéniques d'obésité (37).</p> <p>Physiologiquement, MC4R possède un rôle dans le contrôle du métabolisme énergétique et de la satiété en faisant partie du système mélanocortine. L'activation de MC4R provoque entre autres une augmentation de la satiété et du SNS (38).</p>	<p>MTCH2 est fortement exprimé dans le tissu adipeux. Cette expression de MTCH2 dans le tissu adipeux est notamment associée avec l'obésité (39).</p> <p>Des études effectuées sur modèles animaux suggèrent un rôle de MTCH2 dans la régulation homéostatique des lipides et l'accumulation du gras dans les cellules adipeuses (40, 41).</p>
<p>Relation possible avec la pression artérielle</p>	<p>Le système immunitaire est impliqué dans le processus d'élévation de la pression artérielle, cependant la relation est principalement médiée par les lymphocytes de type T(42). Un lien potentiel entre PAX5, impliqué dans la différenciation des lymphocytes B, et la pression artérielle reste donc difficile à établir.</p>	<p>MRPS22 possède un rôle dans la phosphorylation oxydative (35). Un lien avec le métabolisme énergétique et l'accumulation de graisse reste donc envisageable. Une relation possible entre MRPS22 et la pression artérielle pourrait donc être médiée par son effet sur le métabolisme énergétique et l'accumulation de graisse.</p>	<p>Les fonctions de FTO en dehors du métabolisme énergétique étant toujours méconnues, un lien précis avec la pression artérielle reste difficile à établir.</p> <p>Les résultats de l'étude de Pausova et al. (23) suggèrent que FTO pourrait influencer la pression artérielle par l'entremise d'une action sur l'influx sympathique vasomoteur.</p>	<p>MC4R fait partie du système mélanocortine qui participe à la régulation du poids corporel et de la satiété. Son activation provoque notamment une augmentation de l'activité du SNS qui est connu pour affecter la régulation de la pression artérielle (38).</p>	<p>Mis à part un lien indirecte à-travers la relation entre MTCH2 et l'obésité, une relation entre MTCH2 et la pression artérielle reste difficile à établir.</p>

RÉFÉRENCES ANNEXE B

1. Gibbs RA, Belmont JW, Hardenbol P, Willis TD, Yu F, Yang H, et al. The International HapMap Project. *Nature*. 2003;426(6968):789-96.
2. Zhu Z, Guo Y, Shi H, Liu C-L, Panganiban RA, Chung W, et al. Shared genetic and experimental links between obesity-related traits and asthma subtypes in UK Biobank. *Journal of Allergy and Clinical Immunology*. 2020;145(2):537-49.
3. Frayling TM, Timpson NJ, Weedon MN, Zeggini E, Freathy RM, Lindgren CM, et al. A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science*. 2007;316(5826):889-94.
4. Albuquerque D, Stice E, Rodríguez-López R, Manco L, Nóbrega C. Current review of genetics of human obesity: from molecular mechanisms to an evolutionary perspective. *Molecular Genetics and Genomics*. 2015;290(4):1191-221.
5. Felix JF, Bradfield JP, Monnereau C, van der Valk RJP, Stergiakouli E, Chesi A, et al. Genome-wide association analysis identifies three new susceptibility loci for childhood body mass index. *Hum Mol Genet*. 2016;25(2):389-403.
6. Graff M, Scott RA, Justice AE, Young KL, Feitosa MF, Barata L, et al. Genome-wide physical activity interactions in adiposity — A meta-analysis of 200,452 adults. *PLOS Genetics*. 2017;13(4):e1006528.
7. Willer CJ, Speliotes EK, Loos RJJ, Li S, Lindgren CM, Heid IM, et al. Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. *Nature genetics*. 2009;41(1):25-34.
8. Loos RJJ, Lindgren CM, Li S, Wheeler E, Zhao JH, Prokopenko I, et al. Common variants near MC4R are associated with fat mass, weight and risk of obesity. *Nature genetics*. 2008;40(6):768-75.
9. den Hoed M, Ekelund U, Brage S, Grøntved A, Zhao JH, Sharp SJ, et al. Genetic susceptibility to obesity and related traits in childhood and adolescence: influence of loci identified by genome-wide association studies. *Diabetes*. 2010;59(11):2980-8.
10. Mei H, Chen W, Jiang F, He J, Srinivasan S, Smith EN, et al. Longitudinal replication studies of GWAS risk SNPs influencing body mass index over the course of childhood and adulthood. *PloS one*. 2012;7(2):e31470-e.
11. Wang H-J, Hinney A, Song J-Y, Scherag A, Meng X-R, Grallert H, et al. Association of common variants identified by recent genome-wide association studies with obesity in Chinese children: a case-control study. *BMC Med Genet*. 2016;17(1):7.
12. Zhao J, Bradfield JP, Zhang H, Sleiman PM, Kim CE, Glessner JT, et al. Role of BMI-Associated Loci Identified in GWAS Meta-Analyses in the Context of Common Childhood Obesity in European Americans. *Obesity*. 2011;19(12):2436-9.
13. Jiménez-Orsorio AS, Aguilar-Lucio AO, Cárdenas-Hernández H, Musalem-Younes C, Solares-Tlapechco J, Costa-Urrutia P, et al. Polymorphisms in Adipokines in Mexican Children with Obesity. *Int J Endocrinol*. 2019;2019:4764751-.
14. León-Mimila P, Villamil-Ramírez H, Villalobos-Comparán M, Villarreal-Molina T, Romero-Hidalgo S, López-Contreras B, et al. Contribution of common genetic variants to obesity and obesity-related traits in Mexican children and adults. *PloS one*. 2013;8(8):e70640-e.

15. Speliotes EK, Willer CJ, Berndt SI, Monda KL, Thorleifsson G, Jackson AU, et al. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nature genetics*. 2010;42(11):937-48.
16. Justice AE, Winkler TW, Feitosa MF, Graff M, Fisher VA, Young K, et al. Genome-wide meta-analysis of 241,258 adults accounting for smoking behaviour identifies novel loci for obesity traits. *Nature Communications*. 2017;8(1):14977.
17. Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, Day FR, et al. Genetic studies of body mass index yield new insights for obesity biology. *Nature*. 2015;518(7538):197-206.
18. Akiyama M, Okada Y, Kanai M, Takahashi A, Momozawa Y, Ikeda M, et al. Genome-wide association study identifies 112 new loci for body mass index in the Japanese population. *Nature Genetics*. 2017;49(10):1458-67.
19. Feitosa MF, Kraja AT, Chasman DI, Sung YJ, Winkler TW, Ntalla I, et al. Novel genetic associations for blood pressure identified via gene-alcohol interaction in up to 570K individuals across multiple ancestries. *PLOS ONE*. 2018;13(6):e0198166.
20. Sung YJ, Winkler TW, de las Fuentes L, Bentley AR, Brown MR, Kraja AT, et al. A Large-Scale Multi-ancestry Genome-wide Study Accounting for Smoking Behavior Identifies Multiple Significant Loci for Blood Pressure. *The American Journal of Human Genetics*. 2018;102(3):375-400.
21. Takeuchi F, Akiyama M, Matoba N, Katsuya T, Nakatochi M, Tabara Y, et al. Interethnic analyses of blood pressure loci in populations of East Asian and European descent. *Nature Communications*. 2018;9(1):5052.
22. He D, Fu M, Miao S, Hotta K, Chandak GR, Xi B. FTO gene variant and risk of hypertension: A meta-analysis of 57,464 hypertensive cases and 41,256 controls. *Metabolism*. 2014;63(5):633-9.
23. Pausova Z, Syme C, Abrahamowicz M, Xiao Y, Leonard Gabriel T, Perron M, et al. A Common Variant of the FTO Gene Is Associated With Not Only Increased Adiposity but Also Elevated Blood Pressure in French Canadians. *Circulation: Cardiovascular Genetics*. 2009;2(3):260-9.
24. Luczynski W, Zalewski G, Bossowski A. The association of the FTO rs9939609 polymorphism with obesity and metabolic risk factors for cardiovascular diseases in Polish children. *J Physiol Pharmacol*. 2012;63(3):241-8.
25. Xi B, Zhao X, Shen Y, Wu L, Hotta K, Hou D, et al. Associations of obesity susceptibility loci with hypertension in Chinese children. *International Journal of Obesity*. 2013;37(7):926-30.
26. Olza J, Ruperez AI, Gil-Campos M, Leis R, Fernandez-Orth D, Tojo R, et al. Influence of FTO variants on obesity, inflammation and cardiovascular disease risk biomarkers in Spanish children: a case-control multicentre study. *BMC Med Genet*. 2013;14:123-.
27. García-Solís P, Reyes-Bastidas M, Flores K, García OP, Rosado JL, Méndez-Villa L, et al. Fat mass obesity-associated (FTO) (rs9939609) and melanocortin 4 receptor (MC4R) (rs17782313) SNP are positively associated with obesity and blood pressure in Mexican school-aged children. *British Journal of Nutrition*. 2016;116(10):1834-40.
28. Evangelou E, Warren HR, Mosen-Ansorena D, Mifsud B, Pazoki R, Gao H, et al. Genetic analysis of over 1 million people identifies 535 new loci associated with blood pressure traits. *Nature Genetics*. 2018;50(10):1412-25.
29. Apal Sammy YD, Ming MF, Rampal S, Bulgiba A, Mohamed Z. Association of melanocortin-4 receptor gene polymorphisms with obesity-related parameters in Malaysian Malays. *Annals of Human Biology*. 2013;40(1):102-6.

30. Sun Y, Sun J, Wu J, Yang M. Combined effects of FTO rs9939609 and MC4R rs17782313 on elevated nocturnal blood pressure in the Chinese Han population. *Cardiovasc J Afr.* 2016;27(1):21-4.
31. Hoffmann TJ, Ehret GB, Nandakumar P, Ranatunga D, Schaefer C, Kwok P-Y, et al. Genome-wide association analyses using electronic health records identify new loci influencing blood pressure variation. *Nature Genetics.* 2017;49(1):54-64.
32. Kichaev G, Bhatia G, Loh P-R, Gazal S, Burch K, Freund MK, et al. Leveraging Polygenic Functional Enrichment to Improve GWAS Power. *The American Journal of Human Genetics.* 2019;104(1):65-75.
33. Torlakovic E, Slipicevic A, Robinson C, DeCoteau JF, Alfsen GC, Vyberg M, et al. Pax-5 Expression in Nonhematopoietic Tissues. *American Journal of Clinical Pathology.* 2006;126(5):798-804.
34. Bartness TJ, Song CK. Thematic review series: Adipocyte Biology. Sympathetic and sensory innervation of white adipose tissue. *Journal of Lipid Research.* 2007;48(8):1655-72.
35. Smits P, Saada A, Wortmann SB, Heister AJ, Brink M, Pfundt R, et al. Mutation in mitochondrial ribosomal protein MRPS22 leads to Cornelia de Lange-like phenotype, brain abnormalities and hypertrophic cardiomyopathy. *European Journal of Human Genetics.* 2011;19(4):394-9.
36. Smemo S, Tena JJ, Kim K-H, Gamazon ER, Sakabe NJ, Gómez-Marín C, et al. Obesity-associated variants within FTO form long-range functional connections with IRX3. *Nature.* 2014;507(7492):371-5.
37. Hall JE, do Carmo JM, da Silva AA, Wang Z, Hall ME. Obesity-induced hypertension: interaction of neurohumoral and renal mechanisms. *Circ Res.* 2015;116(6):991-1006.
38. Kühnen P, Krude H, Biebermann H. Melanocortin-4 Receptor Signalling: Importance for Weight Regulation and Obesity Treatment. *Trends in Molecular Medicine.* 2019;25(2):136-48.
39. Kulyté A, Rydén M, Mejhert N, Dungner E, Sjölin E, Arner P, et al. MTCH2 in Human White Adipose Tissue and Obesity. *The Journal of Clinical Endocrinology & Metabolism.* 2011;96(10):E1661-E5.
40. Bar-Lev Y, Moshitch-Moshkovitz S, Tsarfaty G, Kaufman D, Horev J, Resau JH, et al. Mimp/Mtch2, an Obesity Susceptibility Gene, Induces Alteration of Fatty Acid Metabolism in Transgenic Mice. *PLOS ONE.* 2016;11(6):e0157850.
41. Rottiers V, Francisco A, Platov M, Zaltsman Y, Ruggiero A, Lee SS, et al. MTCH2 is a conserved regulator of lipid homeostasis. *Obesity.* 2017;25(3):616-25.
42. Case AJ, Zimmerman MC. Sympathetic-mediated activation versus suppression of the immune system: consequences for hypertension. *J Physiol.* 2016;594(3):527-36.

ANNEXE C. Description détaillée des variables du questionnaire NDI

Variable	Variation dans le temps	Cycles utilisés pour description	Cycles disponibles	Énoncé	Choix de réponse	Formatage pour analyse
Âge	Oui	1, 12, 19, 22	1-23	Date de naissance, date du questionnaire (1-20), date d'aujourd'hui (21-23)		
Sexe	Non	1	1-11	Es-tu une fille ou un garçon?	Garçon, fille	
Éducation universitaire chez la mère	Non	Questionnaire maternel	13, 17 et questionnaire maternel	Quel est le niveau d'éducation de ta mère?	N'a pas terminé l'école secondaire, terminé l'école secondaire, diplôme d'étude professionnelle, CEGEP, université, ne sais pas, non-applicable, autre	Non, oui (éducation universitaire)
Lieu de naissance	Non	1	1-11	Es-tu né...?	Au Canada (nommer la province), hors du Canada (nommer le pays)	
Famille monoparentale	Oui	1, 12, 19	1-20	La question suivante te demande avec quels adultes tu vis. Coche la case si tu vis avec cette personne... Si tu vis dans plus d'une famille (à temps partagé avec ta mère et à temps partagé avec ton père), coche toutes les cases qui s'appliquent (i) mère biologique, (ii) père biologique, (iii) belle-mère, (iv) beau-père, (v) tante(s), (vi) oncles(s), (vii) grand-mère(s), (viii) grand-père(s), (ix) autre(s) (nomme-les)	Non, oui (pour chaque personne)	Non, oui (seulement un adulte coché)
Nombre d'activités physiques modérées et vigoureuses effectuées par semaine	Oui	1, 12, 19	1-20	Pense à l'activité physique que tu as faite durant la dernière semaine, du lundi au dimanche, en dehors de tes cours réguliers d'éducation physique à l'école. Pour chaque activité que tu as effectuée pendant 5 minutes ou plus à la	Non, oui (pour chaque activité, à chaque jour)	Somme des activités catégorisées comme modérées et vigoureuses

				<p>fois, fait un X pour indiquer les jours où tu as effectué cette activité. (i) aller à l'école en vélo, faire des courses de vélo, ou faire un tour de vélo, (ii) Nager/plonger, (iii) basketball, (iv) Balle-molle, (v) football, (vi) soccer, (vii) volleyball, (viii) sports de raquettes (badminton, tennis), (ix) hockey sur glace/hockey cosum, (x) corde à danser, (xi) ski alpin, planche à neige, (xii) ski de fond, (xiii) patins à glace, (xiv) patins à roues alignées, planche à roulette, (xv) gymnastique, (xvi) exercices, conditionnement physique, (xvii) jeux de balle, (xviii) athlétisme, (xix) jeux (tag, marelle), (xx) ballet jazz/ballet classique, (xxi) danse (aérobique, folklorique, à une fête), (xxii) jeux extérieurs (grimper dans les arbres, cache-cache), (xxiii) karaté/judo/tai chi/kung fu, (xxiv) boxe, lutte, (xxv) tâches extérieures (tondre le gazon, ramasser les feuilles, jardiner), (xxvi) tâches ménagères (balayer, laver le plancher, passer la balayeuse), (xxvii) mélange de marche rapide, course, jogging, (xxviii) marche, (xxix) course, jogging, (xxx) autres</p>		effectuées par semaine
Consommation d'alcool des trois derniers mois	Oui	1, 12, 19	1-20	<p>Au cours des 3 derniers mois, à quelle fréquence as-tu...? (i) fume un cigare ou un cigarillo, (ii) utilisé du tabac à chiquer ou du tabac à priser, (iii) bu de l'alcool (bière, vin, autre alcool)</p>	<p>Jamais, un peu pour essayer, une fois ou quelques fois par mois, une fois ou quelques fois par semaine, habituellement chaque jour</p>	<p>Non, oui (un peu pour essayer à plus)</p>
Déjà fumé la cigarette	Oui	1, 12, 19, 22	1-23	<p>Cycles 1-20: Coche la case qui te décrit le mieux.</p>	<p>Cycles 1-20: Je n'ai jamais fumé la cigarette, j'ai déjà fumé la cigarette</p>	<p>Cycles 1-20: Non, oui (j'ai déjà fumé la cigarette (même une seule bouffée)</p>

				<p>Cycles 21-23: Avez-vous déjà au cours de votre vie fume la cigarette, ne serait-ce qu'une bouffée (poffe)?</p>	<p>(même une seule bouffée) mais pas depuis les douze derniers mois, j'ai fumé la cigarette une ou plusieurs fois durant les douze derniers mois, je fume la cigarette une ou plusieurs fois par mois, je fume la cigarette une ou plusieurs fois par semaine, je fume la cigarette tous les jours</p> <p>Cycle 21-23: Non, oui, 1 ou 2 fois, oui, 3 ou 4 fois, oui, 5 à 10 fois, Yes, plus de 10 fois</p>	<p>mais pas depuis les douze derniers mois à plus)</p> <p>Cycles 21-23: Non, oui (oui, 1 ou 2 fois à plus)</p>
<p>Nombre moyen de cigarettes fumées au cours des trois derniers mois</p>	Oui	1, 12, 19, 22	1-23	<p>Pour chacun des trois derniers mois : Pendant le mois de _____, combien de jours as-tu fumé la cigarette, même si c'est juste une bouffée?</p> <p>Les jours où tu as fumé au cours du dernier mois, combien de cigarettes as-tu fumées par jour habituellement?</p>	<p>0, 1, 2-3, 4-5, 6-10, 11-15, 16-20, 21-30, chaque jour, je ne sais pas</p> <p><1, 1, 2-3,4-5, 6-10, 11-15, 16-20, 21-25, >25, je ne sais pas</p>	<p>Nombre de jours multiplié par le nombre de cigarettes fumées par jour, moyenne des trois derniers mois</p>

ANNEXE D. Tableaux additionnels

Tableau A1: Calcul de puissance PAS

		Moyenne attendue (mm Hg) par groupe									
SNP	n	12 ans	15 ans	17 ans	24 ans	Écart-type	Erreur de type 1	Test statistique	Groupe (écoles)	Corrélation intra groupe	Puissance
15%											
0	550	105	108	111	111	11,0	0,05	Wilks Likelihood Ratio	n=70 par école	0,1	0,68
1	110	105,5	108,5	111,5	111,5						
2	55	106	109	112	112						
20%											
0	490	105	108	111	111	11,0	0,05	Wilks Likelihood Ratio	n=70 par école	0,1	0,75
1	140	105,5	108,5	111,5	111,5						
2	70	106	109	112	112						
31%											
0	360	105	108	111	111	11,0	0,05	Wilks Likelihood Ratio	n=70 par école	0,1	0,80
1	270	105,5	108,5	111,5	111,5						
2	90	106	109	112	112						
43%											
0	240	105	108	111	111	11,0	0,05	Wilks Likelihood Ratio	n=70 par école	0,1	0,79
1	360	105,5	108,5	111,5	111,5						
2	120	106	109	112	112						

Tableau A2: Calcul de puissance IMC

		Moyenne attendue (kg/m ²) par groupe									
SNP	n	12 ans	15 ans	17 ans	24 ans	Écart-type	Erreur de type 1	Test statistique	Groupe (écoles)	Corrélation intra groupe	Puissance
15%											
0	550	20,0	21,25	22,5	24,5	11,0	0,05	Wilks Likelihood Ratio	n=70 par école	0,1	0,83
1	110	20,25	21,50	22,75	24,75						
2	55	20,5	21,75	23,0	25,0						
20%											
0	490	20,0	21,25	22,5	24,5	11,0	0,05	Wilks Likelihood Ratio	n=70 par école	0,1	0,89
1	140	20,25	21,50	22,75	24,75						
2	70	20,5	21,75	23,0	25,0						
31%											
0	360	20,0	21,25	22,5	24,5	11,0	0,05	Wilks Likelihood Ratio	n=70 par école	0,1	0,93
1	270	20,25	21,50	22,75	24,75						
2	90	20,5	21,75	23,0	25,0						
43%											
0	240	20,0	21,25	22,5	24,5	11,0	0,05	Wilks Likelihood Ratio	n=70 par école	0,1	0,92
1	360	20,25	21,50	22,75	24,75						
2	120	20,5	21,75	23,0	25,0						